UNIVERSIDADE ESTADUAL DE MARINGÁ

CENTRO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE FÍSICA

PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

DIEGO DOMINGUES LOPES

VOLEIBOL: DINÂMICA NÃO-MARKOVIANA E VANTAGEM DE JOGAR EM CASA

Orientador: Prof. Dr. Renio dos Santos Mendes

Maringá
13 de junho de 2017

DIEGO DOMINGUES LOPES

VOLEIBOL: DINÂMICA NÃO-MARKOVIANA E VANTAGEM DE JOGAR EM CASA

Dissertação apresentada como requisito parcial para a obtenção do título de mestre em Física do programa de Pós-Graduação em Física da Universidade Estadual de Maringá.

Orientador: Prof. Dr. Renio dos Santos Mendes

Maringá
13 de junho de 2017

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

L864v

Lopes, Diego Domingues

Voleibol: Dinâmica Não-Markoviana e vantagem de jogar em casa / Diego Domingues Lopes. -- Maringá, 2017.

50 f. : il., color., figs., tabs.

Orientador: Prof. Dr. Renio dos Santos Mendes.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Departamento de Física, Programa de Pós-Graduação em Fisíca, 2017.

1. Vantagem de jogar em casa. 2. Voleibol. 3. Probabilidade. 4. Probabilidade condicional. I. Mendes, Renio dos Santos, orient. II. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Física. Programa de Pós-Graduação em Física. III. Título.

CDD 21.ed. 519.2

AHS

Sumário

Lista de Figuras

Lista de Tabelas

Agradecimentos

Meus sinceros agradecimentos ao prof. Dr. Renio dos Santos Mendes pela excelente orientação.

Resumo

Usando conceitos e ferramentas da teoria de probabilidades, são analisados empiricamente vários aspectos de uma possível vantagem de jogar em casa em jogos de voleibol. Primeiramente, são apresentados alguns conceitos da teoria de probabilidades. Em seguida, são exibidas as investigações desta dissertação. Iniciamos com uma exposição dos dados relativos aos jogos de voleibol e uma primeira análise da vantagem de jogar em casa nesses jogos. Esta análise é estendida para cada *set* dos jogos e então é investigada a probabilidade de um time vencer um *set* dependendo dos *sets* anteriores. Os resultados indicam uma vantagem de jogar em casa nos jogos de voleibol, e que essa vantagem é mais evidente nos três primeiros *sets* do jogo. Além disso, verificamos que a probabilidade de um time vencer cada *set* depende do resultado dos *sets* anteriores. Concluímos também que a dinâmica de jogos de voleibol tem características não-markovianas.

Palavras chave: vantagem de jogar em casa, voleibol, probabilidade, probabilidade condicional.

Abstract

By using concepts and tools of probability theory, several aspects of home advantage in volleyball games were analyzed. Firstly, an introduction to set theory and its basic operations is presented. Based on set theory, some concepts of probability theory are exhibited and discussed. Secondly, the investigations of this dissertation are presented. We start with a presentation of the data related to volleyball games and a first analysis of the advantage of playing at home in these games. This analysis is extended for each set of games and the dependence of the probability of a team winning a set depending on the previous sets is investigated. The results indicate a home advantage in volleyball games and that this advantage is most evident in the first three sets of the game. In addition, the probability of a team winning each set depends on the outcome of the previous sets. We also conclude that the volleyball dynamics presents non-Markovian features.

Keywords: home advantage, volleyball, probability, conditional probability.

Introdução

Em geral, as pessoas possuem grande dificuldade para lidar com fenômenos aleatórios, pois costuma-se criar associações, crenças, superstições e correlações que muitas vezes não passam de acontecimentos de natureza aleatória. Se faz necessário, então, para o melhor entendimento de tais fenômenos, que em muitas áreas de conhecimento sejam utilizados modelos probabilísticos. A Física não é exceção, e muitas de suas teorias apresentam alguma relação com a teoria de probabilidades como a Física Estatística e a Mecânica Quântica.

A Física Estatística é considerada uma área muito bem consolidada e, recentemente, tem provado ser frutífera para descrever fenômenos fora de seu domínio tradicional [1], o que tem atraído pesquisadores não só da Física, mas de muitas outras áreas. Muitos desses estudos recentes constituem um campo emergente e bastante ativo denominado *Sistemas Complexos*. Esse caráter abrangente dos Sistemas Complexos dá margem à possibilidade de análise dos mais diversos sistemas possíveis, como aqueles relacionados à economia [2, 3], eleições [4–6] e esportes [7–9].

Em particular, os esportes oferecem ótima oportunidade para diversos tipos de estudos, pois têm a vantagem de proporcionar bases de dados cada vez maiores, além de apresentarem aspectos sociológicos relacionados a competições e serem assunto de conhecimento comum. Alguns exemplos desses estudos incluem caminhadas aleatórias ou interpretações difusivas do processo de pontuação [10–14], discussões relativas à eficiência de competições esportivas [10, 12, 15, 16], jogo cooperativo [17], vantagem de jogar em casa [7, 9, 18–21], etc. Estas e outras características já foram estudadas em vários esportes diferentes. Nesta dissertação, tratamos, particularmente, da possível vantagem de jogar em casa e como esta se manifesta no jogo de voleibol.

Nos últimos anos, o jogo de voleibol e jogos "similares" têm atraído a atenção de pesquisadores por causa de sua estrutura dividida em *sets*. Esse tipo de jogo facilita o estudo individual de cada etapa das partidas e permite diversos tipos de investigações, tais como simulações e cálculos estatísticos sobre a sequência de pontuação e de *sets* [22–32].

A vantagem de jogar em casa é um fenômeno bastante observado em esportes coletivos e tem interessado pesquisadores de diversas áreas [7, 18–21]. É representada pela consistência com que as equipes vencem mais de 50% dos jogos disputados em casa, desde que realizem o mesmo número de jogos em casa e fora, defrontando os mesmos adversários [7]. Esse fenômeno provavelmente é causado pelos efeitos da torcida, familiaridade com o local, efeitos de viagem, parcialidade dos árbitros, estados psicológicos, as próprias regras do jogo [33, 34] e possivelmente alguns outros motivos.

O entendimento teórico da vantagem de jogar em casa foi tentado por várias abordagens diferentes, como teorias de territorialidade baseadas em modelos biológicos, teorias psicológicas de pulsão[1] e teorias de cognição social [7, 18, 19]. No entanto, não existem evidências fortes o suficiente para preferenciar uma teoria sobre as outras, provavelmente porque as possíveis causas da vantagem de jogar em casa devem operar em conjunto, cada uma interagindo com a outra de maneiras difíceis de se investigar, isolar e quantificar [36, 37].

Para tentar verificar a existência da vantagem de jogar em casa, foram realizados estudos [7] que a comparavam em muitas modalidades coletivas. Os valores obtidos mostram que o fator casa varia de modalidade para modalidade, e foi encontrada maior vantagem para o futebol (69%). O beisebol foi a modalidade que menos usufruiu desse fator (53,5 %).

A vantagem de jogar em casa também foi analisada em jogos olímpicos [8], em que foram analisados cinco grupos de esporte dos Jogos Olímpicos de Verão entre os anos de 1896 e 1996. Os cinco grupos foram atletismo e levantamento de peso (julgamento predominantemente objetivo), boxe e ginástica (julgamento predominantemente subjetivo) e esportes coletivos (envolve decisões subjetivas). Os resultados mostraram que houve uma significativa vantagem em casa para os grupos de julgamento subjetivo ou que envolvem decisões subjetivas, entretanto não houve vantagem em casa nos dois grupos em que o julgamento é objetivo [8, 38].

Curiosamente, apesar da importância e popularidade do jogo de Voleibol [39], não existem muitos estudos referentes à vantagem de jogar em casa nesse esporte coletivo. Os estudos que foram feitos centram-se principalmente na influência do fator casa nas ações do jogo que podem levar uma equipe à vitória [40–43]. Apenas uma das referências encontradas trata, além dos fatores de performance, de como a vantagem de jogar em casa varia de acordo com os *sets* [20].

Em certo sentido, modelos sobre aproveitamento rápido assinalam que uma performance inicial favorável melhora o momento psicológico e pode levar os times a uma vi-

[1]Pulsão designa em psicologia um impulso interno que direciona o comportamento do indivíduo. O comportamento gerado pelas pulsões é diferente daquele gerado por decisões, por ser gerado por forças internas, inconscientes e alheias ao processo de decisão [35].

tória [44, 45]. Por outro lado, existem estudos que investigam a importância de uma boa performance nos últimos momentos das partidas [46–48]. Não existe nenhum consenso acerca da importância hierárquica desses períodos, mas existem evidências o suficiente para suportar a ideia de que algum período é mais importante que os outros [20]. Adicionalmente, a natureza dos jogos coletivos implica um processo de interação dinâmica que provavelmente influencia a vantagem de jogar em casa [49]. Todavia, poucos estudos foram realizados sobre como a vantagem de jogar em casa varia ao longo dos períodos do jogo. Em [50], a vantagem de jogar em casa foi comparada a cada quarto dos jogos de basquete, e das análises de 1189 jogos da NBA foi concluído que os times da casa tinham vantagem em todos os quartos, mas com mais relevância no primeiro. Esse tipo de informação poderia ser um dos pontos de partida para ajudar a isolar e quantificar as causas do fenômeno da vantagem de jogar em casa. Por exemplo, é provável que a familiaridade com o local do jogo afete os jogadores de forma mais pronunciada no primeiro *set* do jogo de voleibol [20].

Além de como a vantagem de jogar em casa varia ao longo do jogo, notamos a ausência de estudos que investigam como um período do jogo pode influenciar os outros, diminuindo ou aumentando as chances de vitória, dependendo da performance no período anterior.

Neste trabalho, investigamos a vantagem de jogar em casa nos jogos de voleibol, bem como analisamos o modo pelo qual essa vantagem muda de acordo com o *set* do jogo. Visamos também estudar a chance de vitória em cada *set* condicionada ao *set* anterior para entendermos como ganhar ou perder um *set* influencia o próximo *set* do jogo. Conectada com esse estudo, discutimos ainda a possibilidade do sistema em questão apresentar características não-markovianas. Apresentamos essas análises no Capítulo 2, após uma breve introdução à teoria de probabilidades, no Capítulo 1, na qual discorremos sobre os conceitos básicos utilizados para os cálculos das probabilidades nos jogos de voleibol.

Capítulo 1

Introdução à teoria de probabilidades

Na natureza, existem muitos exemplos de situações para as quais modelos determinísticos são apropriados. Por exemplo, as leis da gravitação explicam com bastante precisão o que acontece a um corpo que cai sob determinadas condições [51]. Contudo, existem também muitos fenômenos que requerem modelos matemáticos não-determinísticos para sua investigação.

Experimentos que ao serem repetidos nas mesmas condições não produzem o mesmo resultado são denominados experimentos não determinísticos ou probabilísticos. Por exemplo, em um jogo de azar, se lançarmos um dado sobre uma superfície plana e observarmos o número que aparece na face superior, não podemos determinar a *priori* qual será esse número.

Ainda que os jogos de azar normalmente estejam associados a histórias de ruína pessoal, há vários casos em que a relação entre matemáticos e jogadores foi bastante profícua. O grande legado dessa relação foi a teoria das probabilidades [52].

As primeiras tentativas de se atribuir probabilidades a eventos aleatórios surgiram na Idade Média. Os jogos de dados eram praticados desde vários milênios antes da era cristã, mas não há menção sobre cálculos associados a chances de ocorrência de resultados dos lançamentos [53]. Jakob Bernoulli escreveu um livro influente, *Ars Conjectandi* (Arte da conjectura), em que apresentava importantes contribuições, como a lei dos grandes números. Para esse autor, a teoria de probabilidades oferece um método geral de inferência [54, 55].

A teoria de probabilidades é de grande importância em áreas diversas. Existe grande número de fenômenos observáveis modelados de maneira probabilística. Por esse motivo, técnicas estatísticas estão entre as mais importantes ferramentas dos cientistas e engenheiros [51].

De início, é muito importante distinguir o próprio fenômeno do modelo matemático que o descreve. Naturalmente, não exercemos influência sobre aquilo que observamos. No entanto, ao escolher um modelo, podemos lançar mão de nosso julgamento.

Em geral, nos esportes, os modelos necessários para descrever qualquer característica dos jogos são probabilísticos, já que os resultados não podem ser, usualmente, previstos de maneira completa. Isso não é diferente para o voleibol e, para tentar entender aspectos sobre a dinâmica dos resultados dos jogos e dos *sets*, se faz necessário o uso da teoria de probabilidades.

1.1 Introdução à teoria dos conjuntos

A fim de expor os conceitos básicos do modelo probabilístico que desejamos utilizar, é conveniente conhecer algumas ideias e conceitos da teoria matemática dos conjuntos. Este é um assunto bastante extenso e muito se tem escrito sobre ele. Contudo, necessitamos apenas de algumas noções fundamentais.

Um conjunto é uma coleção de objetos que podem ou não ser de natureza matemática[1] [56]. A coleção de troféus conquistados por um atleta é um conjunto tanto quanto a coleção dos números naturais. Geralmente, os conjuntos são representados por letras maiúsculas, A, B etc. Essa notação será mantida aqui.

Os objetos que individualmente formam um conjunto são denominados *membros* ou *elementos* do conjunto. Quando a for um elemento do conjunto A, escrevemos $a \in A$, e quando a não for um elemento de A, escrevemos $a \notin A$.

Existem três maneiras de descrever quais objetos estão contidos em um conjunto A:

1. Podemos fazer uma lista dos elementos de A. Por exemplo, $A = \{1, 2, 3, 4\}$ descreve o conjunto formado pelos inteiros positivos 1, 2, 3 e 4.

2. Podemos descrever o conjunto A por meio de palavras. Por exemplo, podemos dizer que A é formado por todos os números reais entre 0 e 1.

3. Para descrevermos o conjunto A, podemos simplesmente escrever $A = \{x \in \mathbb{R} | \ 0 \leq x \leq 1\}$.

Um conjunto especial, que merece ser destacado, pode surgir da seguinte maneira: suponhamos que o conjunto A seja descrito como o conjunto de todos os números reais x que satisfaçam à equação $x^2 + 1 = 0$. É evidente que não existem tais números, isto é, o conjunto A não contém qualquer elemento. Essa situação ocorre tão frequentemente que se justifica a introdução de um nome especial para esse conjunto. Por isso, definimos

[1] Essa é uma definição conhecida como definição *ingênua* dos conjuntos, que é muito simples mas leva a alguns paradoxos. Para que se evitem tais problemas, se faz necessária a definição *axiomática* dos conjuntos. No entanto, essa definição vai muito além dos propósitos deste trabalho, e portanto, a deixamos de lado.

o *conjunto vazio* como o conjunto que não contenha elemento algum. Geralmente, se representa esse conjunto por $\emptyset$.

Tendo em mente o que são os conjuntos, uma ideia muito útil é a de combiná-los a fim de formar um novo conjunto. Nesse sentido, existem duas operações fundamentais apresentadas a seguir.

- Consideremos dois conjuntos A e B. Definimos C como a *união* de A e B, como segue: $C = \{x|x \in A \text{ ou } x \in B \text{ (ou ambos)}\}$. Escrevemos a união de A e B como $C = A \cup B$. Desse modo, C será formado por todos os elementos que estejam em A, ou em B, ou em ambos.

- Definiremos D como a *interseção* de A e B, da seguinte maneira: $C = \{x|x \in A \text{ e } x \in B \text{ (em ambos)}\}$. Escreveremos a interseção de A e B como $D = A \cap B$. Portanto, D será formado por todos os elementos que estão em A e em B.

Além dessas duas operações básicas, denominamos *produto cartesiano* de A e B, denotado por $A \times B$, o conjunto $\{(a, b), a \in A,\ b \in B\}$, isto é, o conjunto de todos os pares ordenados nos quais o primeiro é tirado de A e o segundo de B. Por exemplo, suponhamos $A = \{1, 2\}$ e $B = \{3, 4\}$, então $A \times B = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$.

Por fim, introduzimos a noção de *complemento* de um conjunto A em relação a um conjunto fundamental U que contém A. O conjunto denotado por A^c, constituído por todos os elementos que não estejam em A (mas que estejam no conjunto fundamental U), é denominado complemento de A. Isto é, $A^c = \{x|x \notin A\}$.

Também é possível visualizarmos os conjuntos e as operações entre eles por meio de um recurso gráfico, conhecido como *Diagrama de Venn*, como mostra a Figura 1.1.

Algumas identidades são muito empregadas na teoria dos conjuntos, assim como na maioria das áreas da matemática. São elas:

- $A \cup B = B \cup A$;

- $A \cap B = B \cap A$;

- $A \cup (B \cup C) = (A \cup B) \cup C$;

- $A \cap (B \cap C) = (A \cap B) \cap C$;

As duas primeiras são denominadas leis comutativas e as duas últimas leis associativas. Existem outras identidades de conjuntos que englobam união, interseção e complemento. Citamos na sequência as mais importantes.

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;

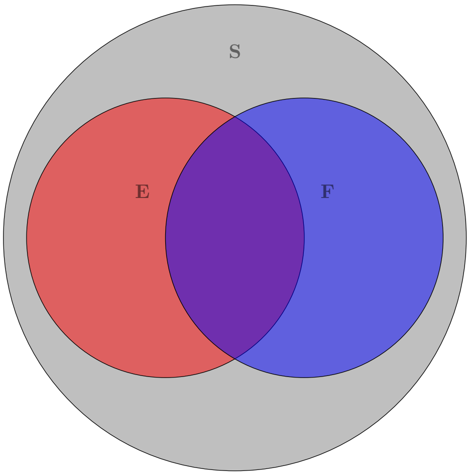


Figura 1.1: **Diagrama de Venn.**

Podemos ver claramente as combinações básicas dos conjuntos. Se tomarmos toda a região vermelha e azul, temos $E \cup F$. Na região onde as duas cores se sobrepõem, temos $E \cap F$, e na região cinza temos $(E \cup F)^c$.

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
- $A \cup \emptyset = A$;
- $A \cap \emptyset = \emptyset$;
- $(A \cap B)^c = A^c \cup B^c$;
- $(A \cup B)^c = A^c \cap B^c$;
- $(A^c)^c = A$.

A validade de cada uma das identidades pode ser demonstrada sem grandes dificuldades, utilizando-se as definições de cada operação.

O número de elementos de um conjunto tem grande importância para nós. Se existir um número finito de elementos em um conjunto A, dizemos que A é finito. Se existir um número infinito de elementos em A, os quais possam ser postos em correspondência biunívoca com os inteiros positivos, dizemos que A é enumerável ou infinito enumerável. Finalmente, devemos considerar o caso de um conjunto infinito não enumerável. Esse tipo de conjunto possui um número infinito de elementos que não podem ser enumerados. O conjunto dos números reais é um bom exemplo desse tipo de conjunto. Nesta dissertação, apenas lidamos com conjuntos finitos.

Os conceitos apresentados até aqui, muito embora representem apenas um rápido exame da teoria dos conjuntos, são suficientes para nossos objetivos, que são expor, com razoável precisão, as ideias fundamentais da teoria da probabilidade.

1.2 Espaço amostral e eventos

Vamos considerar um experimento cujo resultado não é previsível a *priori*. No entanto, embora o resultado não possa ser previsto de antemão, suponhamos que o conjunto de todos os possíveis resultados seja conhecido. Esse conjunto é conhecido como *espaço amostral* do experimento e geralmente é denotado por S, notação que também adotamos neste trabalho (nesse contexto, S representa o conjunto fundamental, que mencionamos anteriormente) [51, 57].

Como exemplo, podemos considerar o lançamento de um dado no qual observamos o número mostrado na face de cima. Nesse caso, o espaço amostral seria representado pelo conjunto $S_1 = \{1, 2, 3, 4, 5, 6\}$, que contém todos os possíveis resultados do lançamento do dado. Outro exemplo pode ocorrer ao lançarmos uma moeda quatro vezes e observarmos o número de caras obtido. Nesse caso, o espaço amostral seria $S_2 = \{0, 1, 2, 3, 4\}$, com todos os resultados possíveis após os quatro lançamentos da moeda.

A fim de descrever um espaço amostral associado a um experimento, devemos ter uma ideia bastante clara daquilo que estamos mensurando ou observando. Por isso, devemos falar de "um" espaço amostral associado a um experimento, e não de "o" espaço amostral. Salientamos também que o resultado de um experimento não é necessariamente um número.

Outra noção fundamental é o conceito de *evento*. Qualquer subconjunto E de um espaço amostral S é um evento de S. Nesse sentido, o próprio conjunto S constitui um evento, assim como o conjunto vazio $\emptyset$. Qualquer resultado individual também pode ser tomado como um evento.

Para ilustrarmos a noção de evento, consideramos os resultados dos *sets* em jogos de voleibol. Se representarmos a vitória do time da casa por 0 e a sua derrota por 1, o espaço amostral para cada *set* seria dado por $S = \{0, 1\}$ e teríamos dois eventos possíveis, $E_1 = \{0\}$ e $E_2 = \{1\}$.

Agora, podemos empregar aos eventos as várias técnicas de combinar conjuntos e obter novos eventos, da forma como apresentamos anteriormente. Segue então, da teoria dos conjuntos, que

1. Se E e F são eventos, então $E \cup F$ será o evento que ocorrerá se, e somente se, E ou F (ou ambos) ocorrerem.

2. Se E e F são eventos, então $E \cap F$ será o evento que ocorrerá se, e somente se, E e F ocorrerem.

3. Se E é um evento, então E^c será o evento que ocorrerá se, e somente se, não ocorrer E.

4. Se $E_1, \ldots, E_n$, for qualquer coleção finita de eventos, então, $\bigcup_{i=1}^{n} E_i$ será o evento que ocorrerá se, e somente se, ao menos um dos eventos E_i ocorrer.

5. Se $E_1, \ldots, E_n$, for qualquer coleção finita de eventos, então, $\bigcap_{i=1}^{n} E_i$ será o evento que ocorrerá se, e somente se, todos os eventos E_i ocorrer.

6. Suponhamos que S represente o espaço amostral de algum experimento ε, e que executemos ε duas vezes. Então, $S \times S$ poderá ser empregado para representar todos os resultados dessas duas repetições. Portanto, $(S_1, S_2) \in S \times S$ significa que s_1 resultou quando ε foi executado pela primeira vez e que s_2 resultou quando ε foi executado pela segunda vez.

7. Esse exemplo pode ser generalizado. Consideramos n repetições de um experimento ε cujo espaço amostral é S. Desse modo,

$$S \times S \times \ldots \times S = \{(s_1, s_2, \ldots, s_n),\ s_i \in S,\ i = 1, \ldots, n\}$$

   representa o conjunto de todos os possíveis resultados, quando ε for executado n vezes. De certo modo, $S \times S \times \ldots \times S$ é ele próprio um espaço amostral, a saber, o espaço amostral associado a n repetições de ε.

Uma última definição sobre eventos, usada com frequência, é que dois eventos, E e F, são ditos *mutuamente excludentes* se não puderem ocorrer juntos. Exprimimos isso escrevendo $E \cap F = \emptyset$, isto é, a interseção de E e F é o conjunto vazio. Se levarmos em conta os dois primeiros *sets* de um jogo de voleibol, temos o espaço amostral $S = \{(0,0), (0,1), (1,0), (1,1)\}$, em que usamos novamente 0 e 1 para vitória e derrota do time da casa respectivamente. Nesse caso, podemos ver claramente que os eventos de S são mutuamente excludentes, já que um time não pode ganhar e perder o mesmo *set*, o que implica que quando um dos resultados de S ocorre, os outros não podem ocorrer.

1.3 Frequência relativa e axiomas da probabilidade

Uma das características fundamentais do conceito de experimento probabilístico, como apresentamos anteriormente, é que não sabemos qual resultado particular ocorrerá quando o experimento for realizado. Em outras palavras, se E for um evento associado a um experimento, não podemos afirmar com certeza se E irá ocorrer ou não. Por isso, torna-se muito importante tentar associar um número ao evento E, o qual medirá de alguma maneira o quão verossímil é que o evento E venha a ocorrer. Essa tarefa nos leva à teoria da probabilidade.

Uma maneira de definir a probabilidade de um evento é em termos da sua frequência relativa. Tal definição geralmente é dada como segue: suponhamos que um experimento, cujo espaço amostral é S, é realizado repetidamente sobre as mesmas condições. Para cada evento E do espaço amostral S, definimos $n(E)$ como sendo o número de vezes que o evento E ocorre nas primeiras n repetições. Então, a probabilidade $P(E)$ do evento E ocorrer é definida por

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}. \tag{1.1}$$

Ou seja, $P(E)$ é definida como a proporção de vezes que E ocorre, ou a frequência relativa de E. Essa definição pode ser usada para calcular a probabilidade do time da casa ganhar o primeiro *set* em jogos de voleibol. Para isto, basta observarmos uma quantidade n de jogos, contarmos em quantos deles o time da casa venceu o primeiro *set* e dividir por n. Quanto mais jogos forem observados, mais precisa será a probabilidade calculada.

Embora essa definição seja intuitivamente aceitável, devemos ter em mente que ela carrega, constantemente, um inconveniente. Como sabemos se $n(E)/n$ converge para algum limite constante que será o mesmo para cada possível sequência de repetições do experimento? Por exemplo, suponhamos que o experimento a ser repetido constantemente seja o lançamento de uma moeda. Como podemos saber se a proporção de caras obtidas nos n primeiros lançamentos converge para algum valor constante quando n se torna grande? E mesmo se convergir, como sabermos se esse valor será o mesmo quando realizarmos o experimento uma segunda vez?

Normalmente, estamos intuitivamente a par desse fenômeno de estabilização, muito embora nunca o tenhamos realmente verificado. Fazê-lo exige considerável porção de tempo e de paciência, porque inclui um grande número de repetições de um experimento. Uma maneira de responder a essas objeções é supor que $\frac{n(E)}{n}$ converge para uma constante por meio de um axioma. Entretanto, assumir que $\frac{n(E)}{n}$ irá, necessariamente, convergir para um valor constante parece ser uma suposição muito complexa. Embora esperemos que tal limite exista, não é nem um pouco evidente que deva ser esse o caso. De fato, seria mais razoável supor um conjunto de axiomas mais simples e autoconsistentes sobre probabilidade e então tentar provar que tal limite constante, de certa forma, existe [2] [57]. Essa é a abordagem axiomática moderna para a teoria da probabilidade que adotamos neste trabalho. Em particular, devemos assumir que para cada evento E no espaço amostral S existe um valor $P(E)$ referente à probabilidade de E. Vamos então assumir que as probabilidades satisfazem um certo conjunto de axiomas que estejam de acordo com nossa noção intuitiva de probabilidade.

Consideremos um experimento cujo espaço amostral é S. Para cada evento E do

[2] A existência desse limite é conhecida como a lei dos grandes números, e pode, em certo sentido, ser demonstrada (Apêndice A).

espaço amostral S, assumimos que um número $P(E)$ é definido e satisfaz os três axiomas:

Axioma 1. $0 \leq P(E) \leq 1$;

Axioma 2. $P(S) = 1$;

Axioma 3. Para qualquer sequência de eventos mutuamente excludentes $E_1, E_2, \ldots$ (isto é, $E_i \cap E_j = \emptyset$ quando $i \neq j$),

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

Nos referimos a $P(E)$ como a probabilidade (não necessariamente a probabilidade frequencista) do evento E.

O **axioma 1** apenas revela que a probabilidade de certo evento ocorrer em um experimento é um número entre 0 e 1. O **axioma 2** afirma que, com probabilidade 1, o resultado de um experimento tem de estar em seu espaço amostral. Por fim, o **axioma 3** dá a entender que, para qualquer sequência de eventos mutuamente excludentes (conjuntos disjuntos [3]), a probabilidade de pelo menos um desses eventos ocorrer é simplesmente a soma de suas respectivas probabilidades.

Uma decorrência importante dos axiomas da probabilidade decorre do seguinte raciocínio: para qualquer evento E, podemos escrever $E = E \cup \emptyset$. Uma vez que A e $\emptyset$ são disjuntos, decorre do **axioma 3** que $P(E) = P(E) + P(\emptyset)$ logo, vemos de imediato que $P(\emptyset) = 0$[4].

1.3.1 Algumas proposições simples

Demonstramos algumas proposições simples relacionadas aos axiomas da probabilidade.

Proposição 1. $P(E^c) = 1 - P(E)$

Demonstração. Primeiro, notamos que E e E^c são sempre mutuamente excludentes, ou $E \cup E^c = S$, e pelos **axiomas 2** e **3** temos que

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c).$$

□

[3] Conjuntos disjuntos são conjuntos que não possuem elementos em comum.

[4] A recíproca desse resultado não é verdadeira. Isto é, se $P(E) = 0$, não podemos, em geral, concluir que $E = \emptyset$, porque existem situações nas quais atribuímos probabilidade zero a um evento que pode ocorrer.

A segunda proposição afirma que se um evento E está contido em um evento F, então a probabilidade de E não pode ser maior do que a probabilidade de F.

Proposição 2. *Se $E \subset F$, então $P(E) \leq P(F)$.*

Demonstração. Desde que $E \subset F$, podemos expressar F como

$$F = E \cup (E^c \cap F).$$

Então, como E e $E^c \cap F$ são mutuamente excludentes, obtemos pelo **axioma 3** que

$$P(F) = P(E) + P(E^c \cap F).$$

Como $P(E^c \cap F) \geq 0$, concluímos que

$$P(E) \leq P(F).$$

□

A proposição acima revela, por exemplo, que a probabilidade de ocorrer o resultado 1 no lançamento de um dado é menor ou igual à probabilidade de um resultado ímpar ocorrer. Esse resultado é bem intuitivo, pois afirma que E deve ocorrer sempre que F ocorra; consequentemente, E não pode ser mais provável que F.

A próxima proposição nos dá uma relação entre a probabilidade da união de dois eventos em termos das probabilidades individuais e da probabilidade da interseção.

Proposição 3. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Demonstração. Para derivar a equação para $P(E \cup F)$, primeiro precisamos escrever $E \cup F$ como a união dos dois eventos disjuntos E e $E^c \cap F$. Assim, pelo **axioma 3** obtemos que

$$\begin{aligned} P(E \cup F) &= P[E \cup (E^c \cap F)] \\ &= P(E) + P(E^c \cap F). \end{aligned}$$

Além disso, como $F = (E \cap F) \cup (E^c \cap F)$ (ver a Figura 1.2), podemos obter, novamente pelo **axioma 3**, que

$$P(F) = P(E \cap F) + P(E^c \cap F)$$

ou, equivalentemente,

$$P(E^c \cap F) = P(F) - P(E \cap F)$$

que demonstra a proposição 3. $\square$

Essa proposição representa uma extensão imediata do **axioma 3**, porque se $E \cap F = \emptyset$, obtemos exatamente seu enunciado.

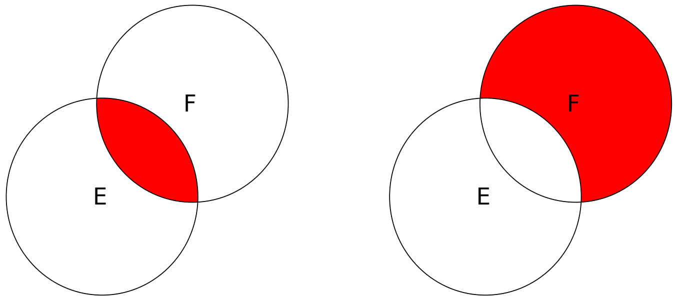


Figura 1.2: **Diagrama de Venn para $E \cap F$ e $E^c \cap F$.**

Na primeira figura, temos $E \cap F$ e na segunda temos $E^c \cap F$. Podemos ver claramente que $F = (E \cap F) \cup (E \cap F)$.

A proposição 3 pode ser estendida para o caso em que haja três eventos, o que nos daria

$$P(E \cup F \cup G) = P(E) + P(F)P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G).$$

Nesse caso, a demonstração consiste em escrever $E \cup F \cup G$ na forma $(E \cup F) \cup G$ e aplicar o resultado da proposição 3.

Por fim, podemos generalizar esse resultado para n eventos como segue

$$\begin{aligned} P(E_1 \cup E_2 \cup \ldots \cup E_n) &= \sum_{i=1}^{n} P(E_i) - \sum_{i<j=2}^{n} P(E_i \cap E_j) + \sum_{i<j<k}^{n} P(E_i \cap E_j \cap E_k) \\ &\quad + \ldots + (-1)^{k-1} P(A_1 \cap A_2 \cap \ldots \cap A_n). \end{aligned}$$

Esse resultado pode ser demonstrado utilizando-se a técnica de indução matemática. No entanto omitimos essa demonstração.

1.3.2 Resultados igualmente prováveis

Em muitos experimentos, é natural assumir que todos os resultados no espaço amostral são igualmente prováveis de ocorrer. Nesse caso, segue que cada probabilidade será $p_i = \frac{1}{n}$. Consequentemente, a condição $\sum_{i=1}^{n} p_i = 1$ torna-se $np_i = 1$ para todo i. Pelo **axioma 3** decorre que, para qualquer evento E

$$P(E) = \frac{\text{número de pontos em E}}{\text{número de pontos em S}}.$$

Em outras palavras, se assumirmos que todos os resultados de um experimento são igualmente prováveis, então a probabilidade de qualquer evento E ocorrer é a proporção de pontos no espaço amostral que estão contidos em E.

Seja $P(E) = \frac{r}{n}$, em que r é o número de pontos do evento E e n é o número de pontos do espaço amostral S. Para calcularmos a probabilidade de um evento qualquer, precisamos apenas contar o número de pontos desse evento [53].

Um procedimento muito elementar de contagem tem sido apresentado sob o título de regra da multiplicação. Suponhamos que uma tarefa pode ser executada em duas etapas. A primeira etapa pode ser realizada de n maneiras e a segunda etapa de m maneiras. Suponhamos também que cada maneira de executar a primeira etapa possa ser seguida por qualquer uma das maneiras de se executar a segunda etapa, então a tarefa completa pode ser executada de nm maneiras [5].

Temos também a regra da adição. Suponhamos que um procedimento, designado por 1, possa ser realizado de n maneiras. Admitamos que um segundo procedimento, designado por 2, possa ser realizado de m maneiras. Além disso, suponhamos que não seja possível que ambos os procedimentos 1 e 2 sejam realizados em conjunto. Então, o número de maneiras pelas quais podemos realizar ou 1 ou 2 será $n + m$ [6].

1.4 Probabilidade condicional

Nesta seção, introduzimos um dos conceitos centrais da teoria de probabilidades, a probabilidade condicional. A importância desse conceito é dupla. Em primeiro lugar, estamos frequentemente interessados em calcular probabilidades quando alguma informação parcial, que diz respeito ao resultado do experimento, está disponível. Em tal situação, as probabilidades desejadas são condicionais. Em segundo, mesmo quando nenhuma informação parcial está disponível, as probabilidades condicionais podem ser usadas para

[5]Obviamente, essa regra pode ser estendida a qualquer número de etapas. Se existirem k etapas e a i-ésima etapa puder ser realizada de n_i maneiras, $i = 1, 2, \ldots, k$, então a etapa formada por 1, seguida de 2,..., seguida pela etapa k, poderá ser executada de $n_1 n_2 \ldots n_k$ maneiras.

[6]Essa regra também pode ser generalizada de maneira análoga à anterior.

calcularmos as probabilidades desejadas mais facilmente [57]. Introduzimos o conceito de probabilidade condicional considerando uma situação especial em que o espaço amostral tem eventos equiprováveis.

Consideramos um experimento que consiste em lançar um dado duas vezes em uma superfície plana e observar o número de pontos na face superior do dado em cada um dos lançamentos. Suponhamos que não presenciemos os lançamentos do dado, mas recebamos a seguinte informação: "em cada um dos lançamentos, o número de pontos observados é menor ou igual a dois". Denotamos por F esse evento. Nessas condições, perguntamos: qual é a probabilidade de que a soma dos pontos nos dois lançamentos seja igual a quatro? ou seja, designando por E o evento "soma dos pontos nos dois lançamentos igual a quatro", queremos saber qual é a probabilidade de ocorrer o evento E, sabendo que o evento F ocorreu. Para o espaço amostral associado aos dois lançamentos e para os eventos F e E, temos:

$$S = \{(i, j) : \ i, j \text{ são inteiros } 1 \leq i \leq 6, \, 1 \leq j \leq 6\}$$

$$F = \{(1, 3), (2, 2), (3, 1)\}, \, E = \{(1, 1), (1, 2), (2, 1), (2, 2)\}.$$

Afirmar que o evento F ocorreu é equivalente a assinalar que não é necessário se levar em conta qualquer ponto do espaço amostral que não pertença a F, ou seja, podemos considerar o evento F como novo espaço amostral para o experimento. Desta maneira, a probabilidade de E ocorrer dado F é igual a $\frac{1}{4}$, pois dos quatro pontos de E apenas o ponto $(2, 2) \in F$ e os quatro pontos são equiprováveis.

Para espaços amostrais com eventos equiprováveis, podemos adotar esse procedimento como definição de probabilidade condicional do evento E dado o evento F $P(E|F)$. Uma fórmula geral para $P(E|F)$ que seja válida para todos os eventos E e F pode ser derivada da mesma maneira: se o evento F ocorre, então para que E ocorra é necessário que esse resultado seja um ponto de E e de F, isto é, deve estar em $E \cap F$. Agora, como sabemos que F ocorreu, segue que o evento F se torna nosso novo, ou reduzido, espaço amostral. Logo, a probabilidade de que o evento $E \cap F$ ocorra será igual à probabilidade de $E \cap F$ relativa à probabilidade de F. Assim, temos a seguinte definição:

Definição 1. Se $P(F) > 0$, então

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \tag{1.2}$$

Retornamos ao nosso exemplo dos dados. Temos $E \cap F = \{(2, 2)\}$ e portanto $P(E \cap F) = \frac{1}{36}$, $P(F) = \frac{4}{36}$. Aplicando a equação 1.2, obtemos o valor $\frac{1}{4}$, já encontrado anteriormente.

Da equação 1.2, que define a probabilidade condicional do evento E dado o evento F, obtemos a seguinte expressão:

$$P(E \cap F) = P(F)P(E|F). \tag{1.3}$$

Essa expressão e sua generalização para uma interseção de n eventos permitem construir probabilidades em espaços amostrais que representam experimentos realizados em sequência, em que a ocorrência de um evento na k-ésima etapa depende das ocorrências na $k-1$ etapas anteriores. Essa generalização é, às vezes, referida como a regra da multiplicação:

$$P(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)\ldots P(E_n|E_1 \cap \ldots \cap E_{n-1}).$$

Para provarmos a regra da multiplicação, basta aplicarmos a probabilidade condicional em seu lado direito. Isto nos dá

$$P(E_1)\frac{P(E_1 \cap E_2)}{P(E_1)}\frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1 \cap E_2)}\cdots\frac{P(E_1 \cap E_2 \cap \ldots E_n)}{P(E_1 \cap E_2 \cap \ldots E_{n-1})} = P(E_1 \cap E_2 \cap \ldots \cap E_n).$$

Nossa definição de $P(E|F)$ é consistente com a interpretação de probabilidade como sendo uma frequência relativa a longo prazo. Para tal, suponhamos que foram realizadas n repetições de um experimento, em que n é grande. Afirmamos que se considerarmos somente aqueles experimentos em que F ocorre, então $P(E|F)$ será igual à proporção de longo prazo daqueles em que E também ocorre [57]. Para verificar isso, notamos que como $P(F)$ é a proporção à longo prazo dos experimentos em que F ocorre, segue que, nas n repetições do experimento, F ocorrerá aproximadamente $nP(F)$ vezes. Similarmente, em aproximadamente $nP(E \cap F)$ desses experimentos, ambos, E e F, irão ocorrer. Então, fora dos (aproximadamente) $nP(F)$ experimentos em que F ocorre, a proporção daqueles em que E também ocorre é aproximadamente igual a

$$\frac{nP(E \cap F)}{nP(F)} = \frac{P(E \cap F)}{P(F)}.$$

Como essa aproximação se torna exata à medida que n se torna maior, temos a apropriada definição de $P(E|F)$. No contexto dos jogos de voleibol, essa definição pode ser empregada no cálculo da probabilidade de um time vencer um *set* levando-se em conta que venceu (ou perdeu) o *set* anterior.

1.4.1 P(E|F) é uma probabilidade

As probabilidades condicionais satisfazem todas as propriedades que as probabilidades ordinárias satisfazem. Demonstramos esse fato na proposição a seguir, que mostra que $P(E|F)$ satisfaz os três axiomas da probabilidade.

Proposição 4.

1. *$0 \leq P(E|F) \leq 1$.*

2. *$P(S|F) = 1$.*

3. *Se E_i, $i = 1, 2, \ldots$ são mutuamente excludentes, então*

$$P\left(\bigcup_{i=1}^{\infty} E_i | F\right) = \sum_{i=1}^{\infty} P(E_i|F).$$

Demonstração. Para demonstrar a parte (1), precisamos mostrar apenas que $0 \leq \frac{P(E\cap F)}{P(F)} \leq 1$. O lado esquerdo da inequação é óbvio, enquanto que no lado direito segue que $(E \cap F) \subset F$, que implica que $P(E \cap F) \leq P(F)$.

A parte (2) é simples,

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1.$$

Na parte (3), segue que

$$\begin{aligned}
P\left(\bigcup_{i=1}^{\infty} E_i | F\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} E_i\right) \cap F\right)}{P(F)} \\
&= \frac{P\left(\bigcup_{i=1}^{\infty} E_i \cap F\right)}{P(F)} \text{ já que } \left(\bigcup_{i=1}^{\infty} E_i\right) \cap F = \bigcup_{i=1}^{\infty} E_i \cap F \\
&= \frac{\sum_{i=1}^{\infty} P(E_i \cap F)}{P(F)} \\
&= \sum_{i=1}^{\infty} P(E_i|F).
\end{aligned}$$

A antepenúltima igualdade é válida, porque $E_i \cap E_j = \emptyset$, implica que $E_i \cap F \cap E_j \cap F = \emptyset$.

$\square$

Se definirmos $Q(E) = P(E|F)$, então segue da proposição 1.4.1 que $Q(E)$ pode ser tratado como uma função de probabilidade nos eventos de S. Então, todas as proposições provadas anteriormente para probabilidades se aplicam a $Q(E)$.

1.4.2 Independência de eventos

Geralmente, a probabilidade condicional de E dado F $P(E|F)$ não é igual à probabilidade incondicional de E $P(E)$. Em outras palavras, saber que F ocorreu geralmente muda a probabilidade de E ocorrer. Nos casos especiais em que $P(E|F)$ é igual à $P(E)$, dizemos que E é independente de F. Isto é, E é independente de F se o conhecimento de F não muda em nada a probabilidade de E ocorrer.

Como $P(E|F) = \frac{P(E \cap F)}{P(F)}$, podemos ver que E é independente de F se

$$P(E \cap F) = P(E)P(F). \tag{1.4}$$

Como a equação 1.4 é simétrica em E e F, ela mostra que sempre que E é independente de F, F também é independente de E.

Um resultado bastante interessante é dado na seguinte proposição.

Proposição 5. *Se E e F são independentes, então E e F^c também são independentes.*

Demonstração. Suponhamos que E e F são independentes. Como $E = (E \cap F) \cup (E \cap F^c)$, e $(E \cap F)$ e $(E \cap F^c)$ são mutuamente excludentes, temos que

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap F^c) \\ &= P(E)P(F) + P(E \cap F^c) \end{aligned}$$

ou equivalentemente,

$$\begin{aligned} P(E \cap F^c) &= P(E)(1 - P(F)) \\ &= P(E)P(F^c) \end{aligned}$$

e o resultado está demonstrado. □

Assim, se E é independente de F, então a probabilidade de E ocorrer não é modificada pela informação da ocorrência de F nem da não ocorrência de F.

Para o caso de três eventos independentes, temos a seguinte definição:

Definição 2. Os três eventos E, F e G, são ditos independentes se

$$\begin{aligned} P(E \cap F \cap G) &= P(E)P(F)P(G) \\ P(E \cap F) &= P(E)P(F) \\ P(E \cap G) &= P(E)P(G) \\ P(F \cap G) &= P(F)P(G) \end{aligned}$$

Devemos notar que se E, F e G são independentes, então E será independente de qualquer evento formado por F e G. Por exemplo, E é independente de $F \cup G$, já que

$$\begin{aligned}
P(E \cap (F \cup G)) &= P((E \cap F) \cup (E \cap G)) \\
&= P(E \cap F) + P(E \cap G) - P(E \cap F \cap G) \\
&= P(E)P(F) + P(E)P(G) - P(E)P(F \cap G) \\
&= P(E)(P(F) + P(G) - P(F \cap G)) \\
&= P(E)P(F \cup G)
\end{aligned}$$

É claro que também devemos estender a definição de independência para mais de três eventos. Os eventos $E_1, E_2, \ldots, E_n$ são ditos ser independentes se para cada subconjunto $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leq n$, desses eventos

$$P(E_{1'}, E_{2'}, \ldots, E_{r'}) = P(E_{1'})P(E_{2'}) \ldots P(E_{r'}). \tag{1.5}$$

Em alguns casos, o experimento probabilístico considerado consiste de uma sequência de sub-experimentos. Por exemplo, se o experimento consiste no lançamento contínuo de uma moeda, podemos pensar em cada lançamento como sendo um sub-experimento. Em muitos casos, é razoável assumir que os resultados de qualquer grupo de sub-experimentos não tem efeito nas probabilidades de outros sub-experimentos. Se esse for o caso, dizemos que os sub-experimentos são independentes. Mais formalmente, dizemos que os sub-experimentos são independentes se $E_1, E_2, \ldots, E_n, \ldots$ é necessariamente uma sequência independente de eventos, sempre que E_i é um evento cuja ocorrência é completamente determinada pelo resultado do i-ésimo sub-experimento.

Se cada sub-experimento é idêntico (isto é, cada sub-experimento tem o mesmo espaço amostral e a mesma probabilidade em seus eventos), então os sub-experimentos são chamados de *ensaios*.

1.5 Fórmula de Bayes

Sejam E e F dois eventos. Assim como já fizemos antes, podemos expressar E como $E = (E \cap F) \cup (E \cap F^c)$. Para que um ponto esteja em E, precisa estar tanto em E quanto em F ou em E mas não em F. Como $E \cap F$ e $E \cap F^c$ são mutuamente excludentes, temos pelo axioma 3 que

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap F^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \end{aligned} \tag{1.6}$$

A equação 1.6 revela que a probabilidade do evento E é um peso médio da probabilidade condicional de E dado que F ocorreu e a probabilidade condicional de E, dado que F não ocorreu, com cada probabilidade condicional tendo o peso equivalente à probabilidade do evento a que está condicionada ocorrer.

Essa é uma fórmula muito útil, porque seu uso geralmente nos permite determinar a probabilidade de um evento, primeiramente condicionando sobre a ocorrência, ou não, de um segundo evento. Isto é, existem muitas situações em que é difícil calcular a probabilidade de um evento diretamente, mas é direto calculá-la, uma vez que sabemos se um segundo evento ocorreu ou não.

A equação 1.6 pode ser generalizada da seguinte maneira: suponhamos que $F_1, F_2, \ldots, F_n$ são eventos mutuamente excludentes de tal forma que

$$\bigcup_{i=1}^{n} F_i = S.$$

Em outros termos, exatamente um dos eventos $F_1, F_2, \ldots, F_n$ deve ocorrer. Escrevendo

$$E = \bigcup_{i=1}^{n} E \cap F_i$$

e usando o fato que os eventos $E \cap F_i$, $i = 1, \ldots, n$ são mutuamente excludentes, obtemos que

$$\begin{aligned} P(E) &= \sum_{i=1}^{n} P(E \cap F_i) \\ &= \sum_{i=1}^{n} P(E|F_i)P(F_i) \end{aligned} \tag{1.7}$$

Dados os eventos $F_1, F_2, \ldots, F_n$, em que apenas um deles deve ocorrer, a equação 1.7 mostra como podemos calcular $P(E)$ primeiro condicionando sobre qual dos eventos F_i ocorreu. Isto é, a equação 1.6 revela que $P(E)$ é igual à média ponderada de $P(E|F_i)$, cada termo tendo como peso a probabilidade do evento ao qual está condicionado. Agora suponhamos que E tenha ocorrido e que estamos interessados em determinar qual dos F_j

também ocorreu. Pela equação 1.7, temos a seguinte equação:

$$\begin{aligned} P(F_j|E) &= \frac{P(E \cap F_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}. \end{aligned} \tag{1.8}$$

A equação 1.8 é conhecida como a fórmula de Bayes. Se pensarmos nos eventos F_j como sendo possíveis hipóteses sobre algum assunto, então a fórmula de Bayes pode ser interpretada de maneira a nos mostrar como as opiniões sobre essas hipóteses realizadas antes do experimento seriam modificadas pela evidência do experimento.

1.6 A probabilidade como uma medida de crença

Até aqui, interpretamos a probabilidade de um dado evento acorrer como sendo a medida do quão frequentemente esse evento ocorre quando o experimento é continuamente repetido. Há, no entanto, situações em que a repetição do experimento não pode ser realizada e outras em que não pode ser realizada em idênticas condições. Por exemplo: um paciente é submetido a um novo tipo de cirurgia e desejamos saber se ele ficará bom, ou desejamos saber se haverá um tremor de terra no Rio Grande do Norte no próximo ano. Podemos ainda querer saber quem vencerá o próximo jogo entre São Paulo e Palmeiras.

No primeiro exemplo, não podemos falar em repetição do experimento, pois se trata de uma nova técnica cirúrgica que estará sendo empregada. No segundo, temos notícia de raras ocorrências de tremores de terra no Rio Grande do Norte. No caso do jogo entre São Paulo e Palmeiras, sabemos que há estatísticas de um grande número de jogos entre os dois times, mas que as condições entre um jogo e outro variam bastante.

As probabilidades nesses tipos de casos pode ser interpretada como a crença que o observador possui na ocorrência do evento. Desse modo, a probabilidade será, em geral, diferente para distintas pessoas em decorrência das diferentes opiniões que elas têm sobre a ocorrência do evento.

Para lidar com essa questão, são estabelecidas regras de comportamento racional para o observador. Dessa forma, parece lógico que essas regras sejam, justamente, os axiomas da probabilidade. Assim, se interpretarmos a probabilidade como uma medida de crença ou como uma frequência de ocorrência, suas propriedades matemáticas permanecem inalteradas.

1.7 Processos markovianos

Dentro da teoria de probabilidades, um *processo estocástico* é uma família de variáveis aleatórias [7]representando a evolução de um sistema de valores com o tempo. Ao contrário de um processo que possui um único modo de evoluir, como nas soluções de equações diferenciais ordinárias, por exemplo, em um processo estocástico há uma indeterminação. Mesmo que se conheça a condição inicial, existem várias (por vezes infinitas) maneiras nas quais o processo pode evoluir.

Em casos de tempo discreto, o processo estocástico é uma sequência de variáveis aleatórias. Suponhamos que uma variável aleatória x_t assuma valores inteiros e t os valores 0, 1, 2, 3, ... Neste caso um processo estocástico fica completamente definido até o instante l pela distribuição de probabilidade [7]

$$P_l(n_0, n_1, \ldots, n_l)$$

de que x_t tome o valor n_0 no instante $t = 0$, o valor n_1 no instante $t = 1$, ..., e o valor n_l no instante $t = l$.

Em seguida, consideremos a probabilidade condicional

$$P_{l+1}(n_{l+1}|n_0, n_1, \ldots, n_l)$$

de que a variável aleatória x_t tome o valor n_{l+1} no instante $t = l + 1$, dado que ela tenha tomado o valor n_o no instante $t = 0$, o valor n_1 no instante $t = 1$, ..., e o valor n_l no instante $t = l$. Se ela for igual à probabilidade condicional

$$P_{l+1}(n_{l+1}|n_l)$$

de que x_t tome o valor n_{l+1} no instante $t = l+1$, dado que ela tenha tomado o valor n_l no instante $t = l$, então o processo estocástico é um *processo markoviano*. Em outros termos, um processo markoviano é aquele em que a probabilidade condicional de x_t tomar um determinado valor em um determinado instante depende somente do valor que ela tenha tomado no instante anterior [58]. Devemos ressaltar que existe uma pequena sutileza nessa dependência com o valor tomado no instante anterior. Se examinarmos uma caminhada aleatória unidimensional, por exemplo, a posição que será ocupada no passo seguinte obviamente depende da posição atual; no entanto, a probabilidade de que o passo seja para frente ou para trás não é alterada pela posição atual. Nesse sentido, em um jogo de voleibol, se a probabilidade de vencer um *set* depende do resultado do *set* anterior ou de outros, este não será um processo markoviano. Por exemplo, inúmeras aplicações

[7]ver Apêndice A

em Física empregam a equação mestra e a equação de Fokker-Planck que, por sua vez, descrevem somente processos markovianos.

Capítulo 2

A vantagem de jogar em casa em jogos de voleibol

Neste capítulo, discutimos a ocorrência da vantagem de jogar em casa em jogos de voleibol. Além disso, analisamos como essa vantagem se manifesta em cada set do jogo e como uma vitória ou derrota em algum set influencia a probabilidade de vencer o próximo set.

2.1 Apresentação dos Dados

A base de dados do presente estudo foi retirada de um *site* que tem como enfoque divulgar os resultados de uma grande variedade de esportes de vários países [59]. Esse site foi escolhido devido à maior organização dos resultados e maior facilidade para a obtenção dos dados. Os resultados referentes ao voleibol estavam expostos com a pontuação com que cada uma das equipes havia terminado em cada set da partida. Assim, foi possível obter o resultado do jogo e também o resultado final de cada set.

Para a realização deste trabalho, utilizamos os resultados de 23575 jogos, oriundos de 39 países. A Tabela 2.1 mostra o total de jogos disponíveis para cada país. Originalmente, a base de dados obtida foi um pouco maior do que a apresentada aqui, no entanto alguns jogos apresentavam resultados impossíveis de ocorrer, de forma que tivemos que descartá-los.

Os dados são referentes ao principal campeonato de voleibol masculino de cada país, que são compostos pelas equipes de nível mais alto e englobam o período de 2011 a 2015.

Depois da obtenção dos dados, o próximo passo foi organizá-los. Sendo assim, dividimos as equipes em função de jogar em casa ou fora de casa e observamos qual time venceu cada set, para que pudéssemos analisar a estrutura interna do jogo. Seguindo essa organização, cada jogo foi escrito como uma sequência binária, ou seja, a cada set atribuímos o

País	Nº de jogos
Alemanha	629
Argentina	560
Austrália	159
Áustria	624
Bélgica	513
Brasil	730
Bulgária	466
Catar	428
China	282
Chipre	311
Coreia do Sul	593
Croácia	818
Dinamarca	470
Eslováquia	700
Eslovênia	737
Espanha	548
Estônia	354
Finlândia	1097
França	982
Grécia	721
Holanda	700
Inglaterra	302
Iran	376
Israel	282
Itália	875
Japão	535
Marrocos	188
Noruega	308
Polônia	871
Portugal	913
República Tcheca	573
Romênia	789
Rússia	1089
Sérvia	560
Suécia	665
Suíça	654
Turquia	776
Ucrânia	626
USA	771
Total	**23575**

Tabela 2.1: **Países utilizados na base de dados e número de jogos em cada país.**

resultado 0 ou 1 caso o time da casa ou o time visitante tenha vencido, respectivamente, esse set. Por exemplo, se o time da casa ganhou os dois primeiros sets, perdeu o terceiro e quarto set e ganhou o último, o jogo é escrito como (0 0 1 1 0). Essa notação facilita muito a visualização rápida do resultado de cada jogo, bem como o manuseio dos resultados.

Os cálculos foram realizados de acordo com a probabilidade frequencista (frequência relativa), em que utilizamos os conceitos de probabilidade simples e condicional. Além disso, os intervalos de confiança[1] foram calculados por meio da técnica de *bootstrapping* (Apêndice B). O intervalo adotado neste trabalho foi de 99%. Todos os cálculos foram realizados por meio do *software* Mathematica [60].

2.2 Resultados

De acordo com a notação introduzida na seção anterior, o jogo de voleibol apresenta vinte resultados possíveis, e metade desses resultados é relativa à vitória do time da casa e a outra metade à vitória do time visitante. A Tabela 2.2 mostra cada um dos resultados possíveis, a probabilidade com que cada um deles acontece e a probabilidade de cada resultado acontecer caso as partidas fossem totalmente aleatórias[2]. Nessa tabela, fica claro que os resultados em que o jogo acaba mais rápido (3-0) são razoavelmente mais prováveis, já que compreendem cerca de 45,9% dos dados[3]. Os jogos que terminam em quatro sets (3-1) são um pouco menos prováveis, mas ainda assim englobam aproximadamente 32,29% dos dados. Os últimos 21,81% dos dados são referentes aos jogos que duram cinco sets (3-2). É possível notar que geralmente os jogos não se estendem até o quinto set.

Podemos ver também que o resultado (0 0 0) acontece mais vezes (27,8%) do que o resultado (1 1 1) (18,01%), ou seja, o time da casa vence 60,57% dos jogos com resultado 3-0, o que já nos indica a existência da vantagem de jogar em casa. Essa diferença também ocorre nos jogos que duram quatro sets, porém é menos pronunciada. A saber, o time da casa vence cerca de 56,15% dos jogos com resultado 3-1, e aproximadamente metade dos jogos com resultado 3-2, em que aparentemente não existe vantagem para nenhum dos times.

[1] Em estatística, o *intervalo de confiança* é o intervalo estimado na qual uma medida estatística de uma amostra tem uma dada probabilidade de ocorrer. Comumente define-se esse intervalo no qual há 95% (ou 99%) de probabilidade da medida verdadeira do parâmetro ocorrer (Apêndice B).

[2] O cálculo para o caso aleatório é razoavelmente simples. Considere que cada time tem probabilidade $\frac{1}{2}$ de vencer cada set, neste caso cada tripla tem probabilidade $\frac{1}{2^3}$, cada quádrupla tem probabilidade $\frac{1}{2^4}$ e cada quíntupla tem probabilidade $\frac{1}{2}$, levando em conta que os resultados de cada set são independentes e utilizando a equação 1.5, obtemos os resultados da Tabela 2.2

[3] É importante lembrar que todos os resultados possíveis são mutuamente excludentes, ou seja, não é possível que dois deles ocorram na mesma partida. O resultado 3-0 acontece quando ocorre (0 0 0) ou (1 1 1) de forma que a probabilidade do jogo terminar em 3-0 é simplesmente a soma das probabilidades de cada um destes dois resultados ocorrerem. O mesmo raciocínio vale para os jogos que terminam em 3-1 e 3-2.

É interessante notar que para cada resultado possível existe um resultado exatamente simétrico que pode ser obtido apenas trocando 0 por 1 e 1 por 0, por exemplo: o resultado simétrico ao resultado (0 1 0 0) é (1 0 1 1). Apesar de simétricos, esses resultados apresentam probabilidades diferentes de acontecer, de acordo com a Tabela 2.2. Essa simetria é útil para investigarmos os resultados possíveis do ponto de vista do time da casa e do time de fora, como fazemos na sequência.

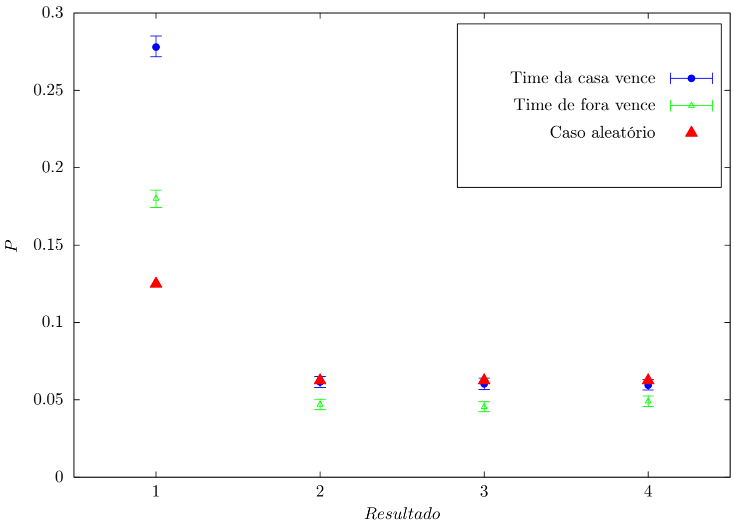
	Resultado do jogo	**Caso aleatório (%)**	**Dados (%)**	**Intervalo de confiança (%)**
1	(0 0 0)	12,500	27,812	27,173 - 28,513
	(1 1 1)	12,500	18,015	17,425 - 18,558
2	(0 1 0 0)	6,250	6,155	5,799 - 6,515
	(1 0 1 1)	6,250	4,700	4,365 - 5,039
3	(0 0 1 0)	6,250	6,032	5,667 - 6,405
	(1 1 0 1)	6,250	4,547	4,233 - 4,878
4	(1 0 0 0)	6,250	5,951	5,633 - 6,303
	(0 1 1 1)	6,250	4,908	4,577 - 5,247
5	(0 0 1 1 0)	3,125	1,743	1,553 - 1,943
	(1 1 0 0 1)	3,125	1,603	1,413 - 1,790
6	(0 1 0 1 0)	3,125	1,837	1,616 - 2,049
	(1 0 1 0 1)	3,125	1,837	1,620 - 2,053
7	(1 0 0 1 0)	3,125	1,921	1,731 - 2,134
	(0 1 1 0 1)	3,125	1,578	1,379 - 1,756
8	(0 1 1 0 0)	3,125	1,756	1,565 - 1,951
	(1 0 0 1 1)	3,125	1,684	1,489 - 1,896
9	(1 0 1 0 0)	3,125	1,743	1,548 - 1,943
	(0 1 0 1 1)	3,125	1,879	1,680 - 2,070
10	(1 1 0 0 0)	3,125	1,769	1,565 - 1,968
	(0 0 1 1 1)	3,125	1,964	1,743 - 2,163

Tabela 2.2: **Probabilidade com que cada resultado acontece no caso aleatório e baseado nos dados.**

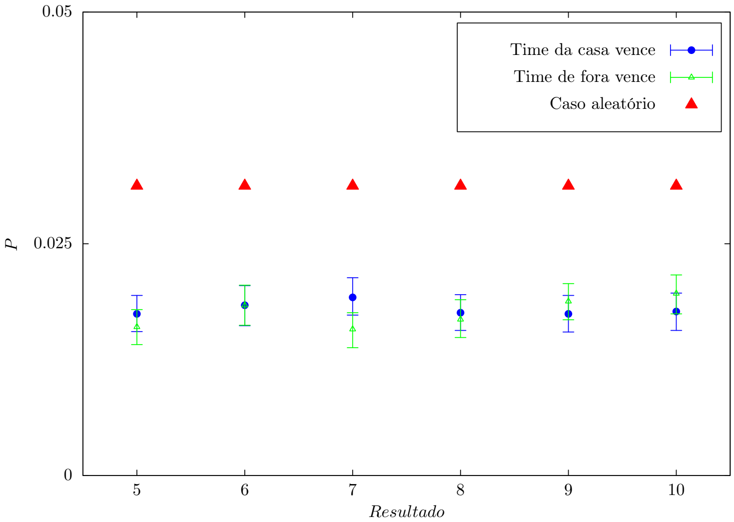
A comparação entre a probabilidade de cada resultado acontecer e o caso aleatório pode ser melhor vista na Figura 2.1, que apresenta a probabilidade do resultado acontecer no eixo y e a numeração do resultado no eixo x (de acordo com a primeira coluna da tabela 2.2). Nessa figura dividimos os resultados em dois gráficos. No Gráfico 2.1(a), estão os resultados 1-4 e na Figura 2.1(b) estão os resultados 5-10.

Na Figura 2.1(a), fica evidente que os resultados 3-0 são muito mais prováveis e, além disso, fica clara a vantagem de jogar em casa nesse resultado. É notável também que os resultados 3-0 acontecem com uma frequência muito maior do que era de se esperar pelo valor do caso aleatório. Vemos ainda que a probabilidade de vitória nos casos em que a partida termina em 3-1 é ligeiramente menor do que no caso aleatório, e que ainda assim aparece uma ligeira vantagem para o time da casa em todos os resultados.

A Figura 2.1(b) mostra as probabilidades dos resultados em que o jogo termina em 3-2. Talvez por esse caso englobar partidas entre times mais equilibrados, não aparece vantagem significativa para o time da casa. Vale notar que a probabilidade de ocorrer esses resultados está consideravelmente abaixo do caso aleatório, evidenciando assim que geralmente a partida não dura até esse estágio.



(a) Resultados 1-4 de acordo com a tabela 2.2



(b) Resultados 5-10 acordo com a tabela 2.2

Figura 2.1: **Comparação dos dados com o caso aleatório.**
Na Figura (a), apresentamos a frequência da ocorrência dos jogos que terminaram em 3-0 e as diferentes possibilidades para os jogos com resultado 3-1. Em (b), elencamos todos os possíveis resultados para os jogos que terminaram em 3-2.

2.2.1 A vantagem de jogar em casa

De acordo com o que vimos até aqui, é possível esperarmos que exista a vantagem de jogar em casa nos jogos de voleibol. De fato, verificamos nos dados que o time da casa apresenta uma probabilidade significativamente maior de vencer a partida. Os valores que obtivemos foram similares aos valores observados em [38], como mostra a Tabela 2.3.

	Prob. de vencer (P) (%)	**Intervalo de confiança (IC) (%)**
Time da casa	56,92	56,17 - 57,72
Time visitante	43,08	42,27 - 43,82

Tabela 2.3: **Probabilidade de ganhar o jogo em casa e fora de casa.**

Vimos que a vantagem de jogar em casa aparece de maneira mais expressiva em jogos com resultado 3-0 e 3-1; além disso, notamos que essa vantagem não aparece significantemente nos jogos mais equilibrados, aqueles com resultado 3-2. Isto nos sugere que uma análise mais detalhada da estrutura interna do jogo de voleibol é necessária para que possamos saber como a vantagem de jogar em casa varia de acordo com o *set* do jogo.

2.2.2 A vantagem de jogar em casa *set* a *set*

O estudo das diferentes fases de uma partida é de grande importância em muitos esportes coletivos, pois permite saber quais são os momentos cruciais de uma partida. Não existe um consenso sobre qual fase é mais importante, mas existem vários estudos que investigam essa questão [20, 44–48].

Nossa intenção é analisar como a vantagem de jogar em casa se modifica de acordo com o *set* do jogo. Dessa forma, podemos apurar em qual fase do jogo essa vantagem é mais pronunciada.

A Tabela 2.4 apresenta os resultados para a probabilidade de vencer o *set* tanto para o time da casa quanto para o time de fora. Podemos notar, primeiramente, que as probabilidades do time da casa vencer os três primeiros *sets* são maiores e aproximadamente as mesmas. No entanto, essa vantagem diminui no quarto *set* e deixa de ser significativa no quinto *set*. Como as probabilidades para o time de fora são complementares às do time da casa, estas começam baixas nos três primeiros *sets* e aumentam no dois *sets* subsequentes. Esta mudança nas probabilidades em cada *set* fica bastante clara na Figura 2.2.

Esse resultado concorda em partes com o resultado obtido em [20], em que eles observam, utilizando uma base de dados de 800 jogos, maior vantagem para o time da casa no início do jogo porém, diferente de nós, eles também encontram essa vantagem no final do jogo.

	1º *Set*	2º *Set*	3º *Set*	4º *Set*	5º *Set*
Probabilidade do time da casa vencer (%)	55,88	55,58	55,36	52,86	50,30
IC (%)	54,78 - 56,38	54,86 - 56,39	54,46 - 56,14	51,79 - 53,84	49,18 - 52,33
Probabilidade do time visitante vencer (%)	44,12	44,42	44,64	47,14	49,68
IC (%)	43,28 - 44,87	43,66 - 45,11	43,91 - 45,35	46,11 - 48,15	48,07 - 51,22

Tabela 2.4: **Probabilidade de ganhar o jogo em casa e fora de casa.**

Vários fatores podem ter um papel importante na vantagem de jogar em casa e na maneira como ela se manifesta em cada *set*. Dessa forma, um estudo probabilístico é certamente insuficiente para concluirmos qual fator é mais relevante se tivermos apenas informações sobre os resultados dos *sets* dos jogos. No entanto, os dados parecem sugerir que quando os times estão equilibrados, o time da casa não tem vantagem alguma, mas se os times mantêm diferença de força, jogar em casa parece evidenciar essa diferença.

Os fatores emocionais também podem desempenhar grande papel em uma partida de voleibol [43, 61]. Infelizmente, não temos como saber de que maneira esses fatores se manifestam em um jogo, mas podemos analisar, por exemplo, como ganhar ou perder um *set* influencia na probabilidade de vencer o próximo *set*, o que pode estar relacionado, mesmo que levemente, às condições emocionais dos jogadores.

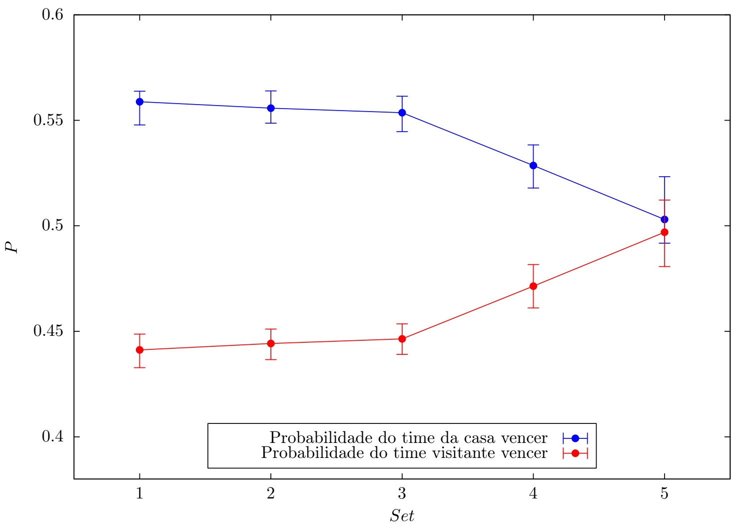


Figura 2.2: **Probabilidade do time da casa ou de fora vencer cada *set*.**

2.2.3 Vantagem de jogar em casa condicionada ao *set* anterior

Vimos que a vantagem de jogar em casa varia de acordo com o *set*, mas intencionamos saber se as chances de ganhar um *set* aumentam ou diminuem de acordo com o resultado do *set* anterior.

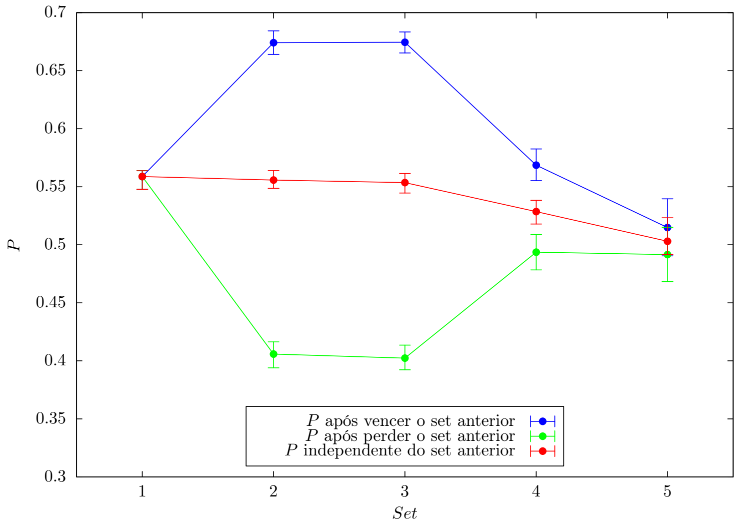
A Figura 2.3 mostra como as probabilidades de vencer o *set* se alteram, para o time da casa e de fora, a partir das probabilidades simples (em vermelho). Em azul e em verde estão representadas as probabilidades de vencer o *set* quando o time ganha ou perde o *set* respectivamente. Podemos notar que quando um time ganha um *set*, a probabilidade de vencer o próximo aumenta, no entanto, quando um time perde um *set*, as chances de ganhar o próximo diminuem. Essas mudanças ocorrem tanto para o time da casa quanto para o time de fora. Uma explicação possível para esse fato é que quando um time ganha um *set*, geralmente ele está melhor do que seu adversário, o que justifica a probabilidade maior de ganhar o próximo *set*. Outro fator que pode influenciar esse resultado é o aumento da confiança dos jogadores, pois quando estes estão motivados, apresentam melhor performance no jogo [43].

As probabilidades para cada resultado possível são apresentadas nas Figuras 2.4 e 2.5. Nessas figuras, P_0 significa a probabilidade de que o time da casa vença o *set* e P_1 é a probabilidade de que o time visitante vença o *set*. A figura segue do primeiro para o quinto *set*, e mostra toda a evolução dos *sets* de uma partida de voleibol, além da probabilidade de cada time (da casa ou de fora) vencer o próximo *set* de acordo com a história do jogo até o momento. A Figura 2.4 mostra a história do jogo quando o time da casa ganha o primeiro *set* e a Figura 2.5 mostra a história do jogo quando o time de fora ganha o primeiro *set*.

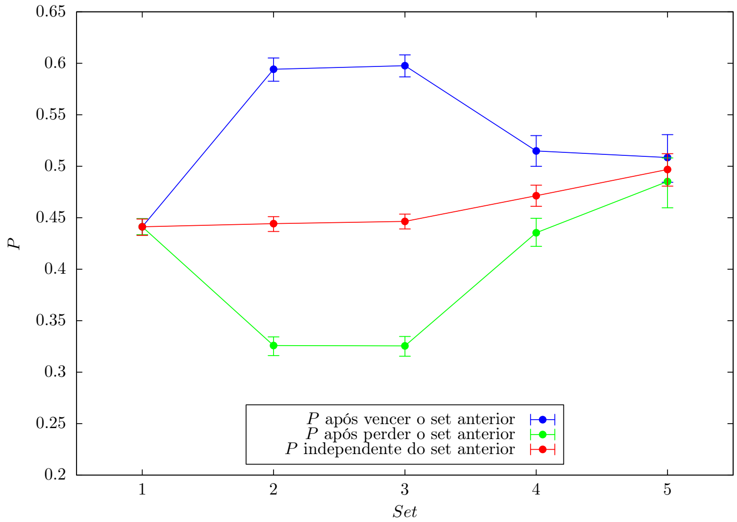
É interessante notar que resultados opostos ((0 0 0) e (1 1 1) por exemplo) têm o mesmo padrão de probabilidades, ou seja, a vitória de um *set* seguida pela derrota do próximo tem um probabilidade menor de ocorrer do que a vitória de um *set* seguida pela vitória do próximo. Esse resultado está mostrado na Figura 2.6 e reforça o que foi mostrado na Figura 2.3.

Existem alguns estudos que propõem modelos para a predição de resultados de jogos de voleibol e para a reprodução das sequências de pontuação e de *sets* geralmente sugerindo que a probabilidade de vencer um *set* se mantém constante ao longo do jogo [23, 24, 27, 31, 32]. No entanto, as Figuras 2.4 e 2.5 indicam que a situação não é tão simples e deveria ser encarada com certa cautela, já que as probabilidades são diferentes para cada sequência possível de *sets*. Esse tipo de estudo também ocorre com bastante frequência em jogos de tênis [22, 25, 26, 28–30]. Como a estrutura do jogo de tênis é parecida com a do jogo de voleibol, é possível que os mesmos cuidados tenham de ser tomados.

Na seção 1.7 foi dito que em um processo markoviano a probabilidade condicional de uma variável assumir um valor depende apenas do seu valor no instante anterior. No entanto, os resultados que acabaram de ser expostos apontam na direção de um processo não-markoviano, pois a probabilidade de um *set* ser vencido mostrou-se depender, em geral, do resultado de cada um dos *sets* anteriores, isto é, da história do jogo.



(a) Probabilidade do time da casa vencer cada *set*



(b) Probabilidade do time visitante vencer cada *set*

Figura 2.3: **Probabilidade de vencer cada *set* condicionada ao *set* anterior.** Em vermelho temos a probabilidade simples de vencer o *set*. Em azul e em verde estão as probabilidades de vencer o *set* quando o time vence e perde o *set* anterior respectivamente. Na figura (a) são mostrados os resultados para o time da casa e na figura (b) os resultados para o time de fora.

Por fim, a investigação desenvolvida nesta dissertação poderia ser mais aprofundada no sentido de entender com mais detalhes a possibilidade de tendências não apresentadas nas análises. Em particular, caberia investigar a provável influência dos placares dos *sets* anteriores nas probabilidades condicionais. E, com isso, aumentar informações que poderiam ajudar em possíveis estratégias, por parte dos participantes, ao longo dos jogos.

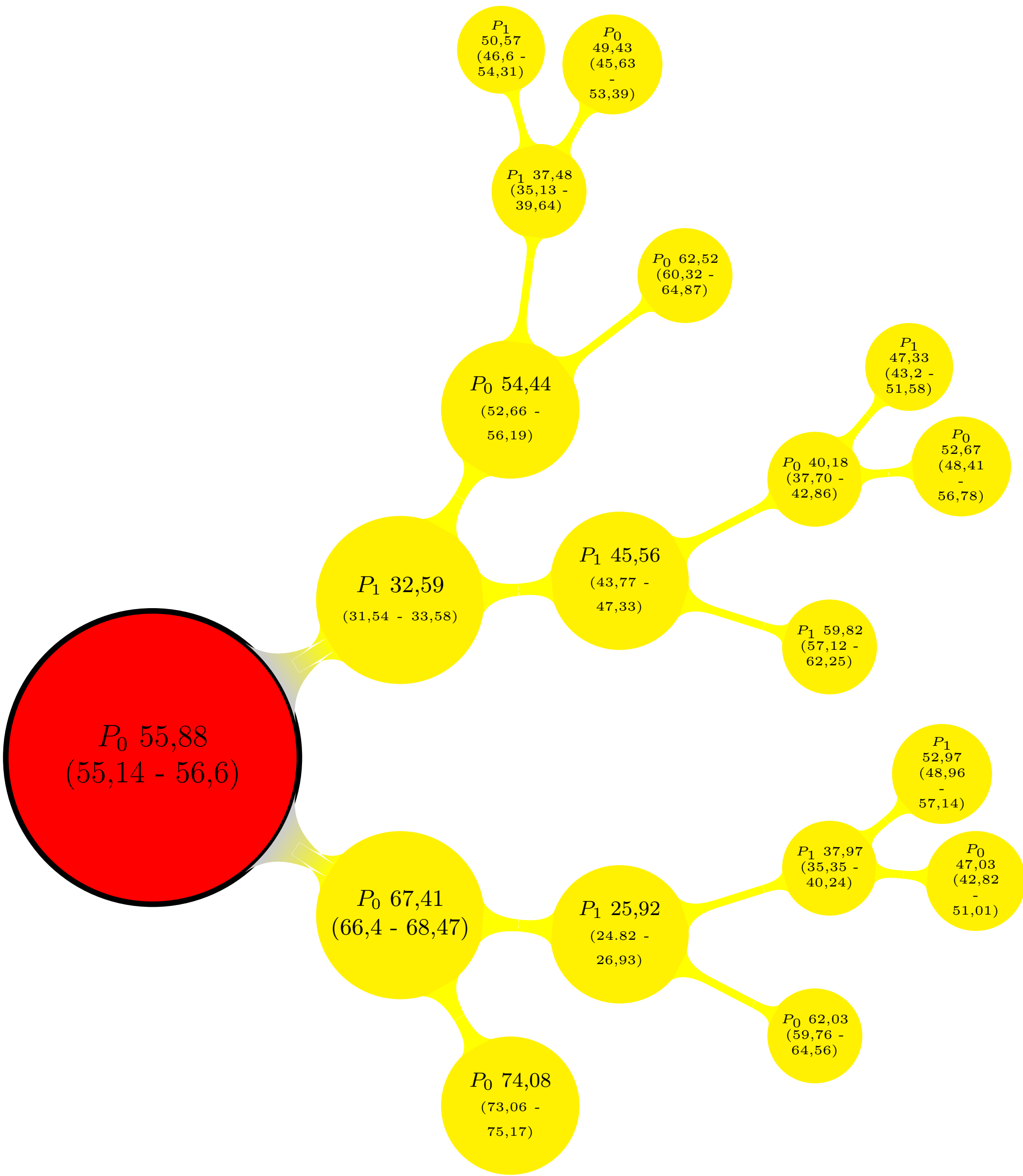


Figura 2.4: **Probabilidade condicional de vencer cada *set* (em %) quando o time da casa começa ganhando.** Nesta figura, P_0 significa a probabilidade de que o time da casa vença o *set* e P_1 é a probabilidade de que o time visitante vença o *set*. A figura segue do primeiro *set* (círculo maior) para o quinto (círculos menores).

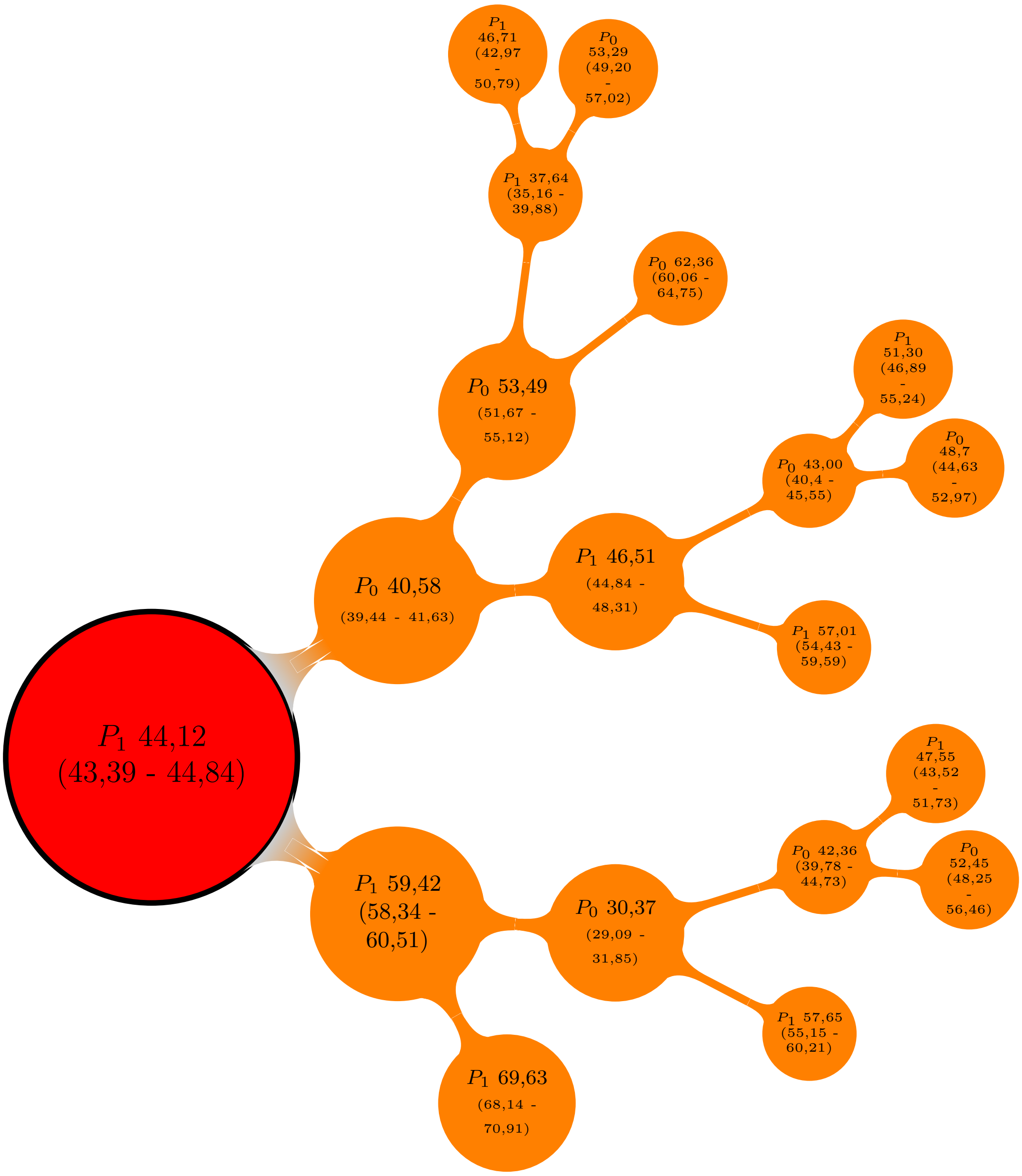


Figura 2.5: **Probabilidade condicional de vencer cada *set* (em %) quando o time de fora começa ganhando.** Assim como na figura anterior, P_0 significa a probabilidade de que o time da casa vença o *set* e P_1 é a probabilidade de que o time visitante vença o *set*. A figura segue do primeiro *set* (círculo maior) para o quinto (círculos menores).

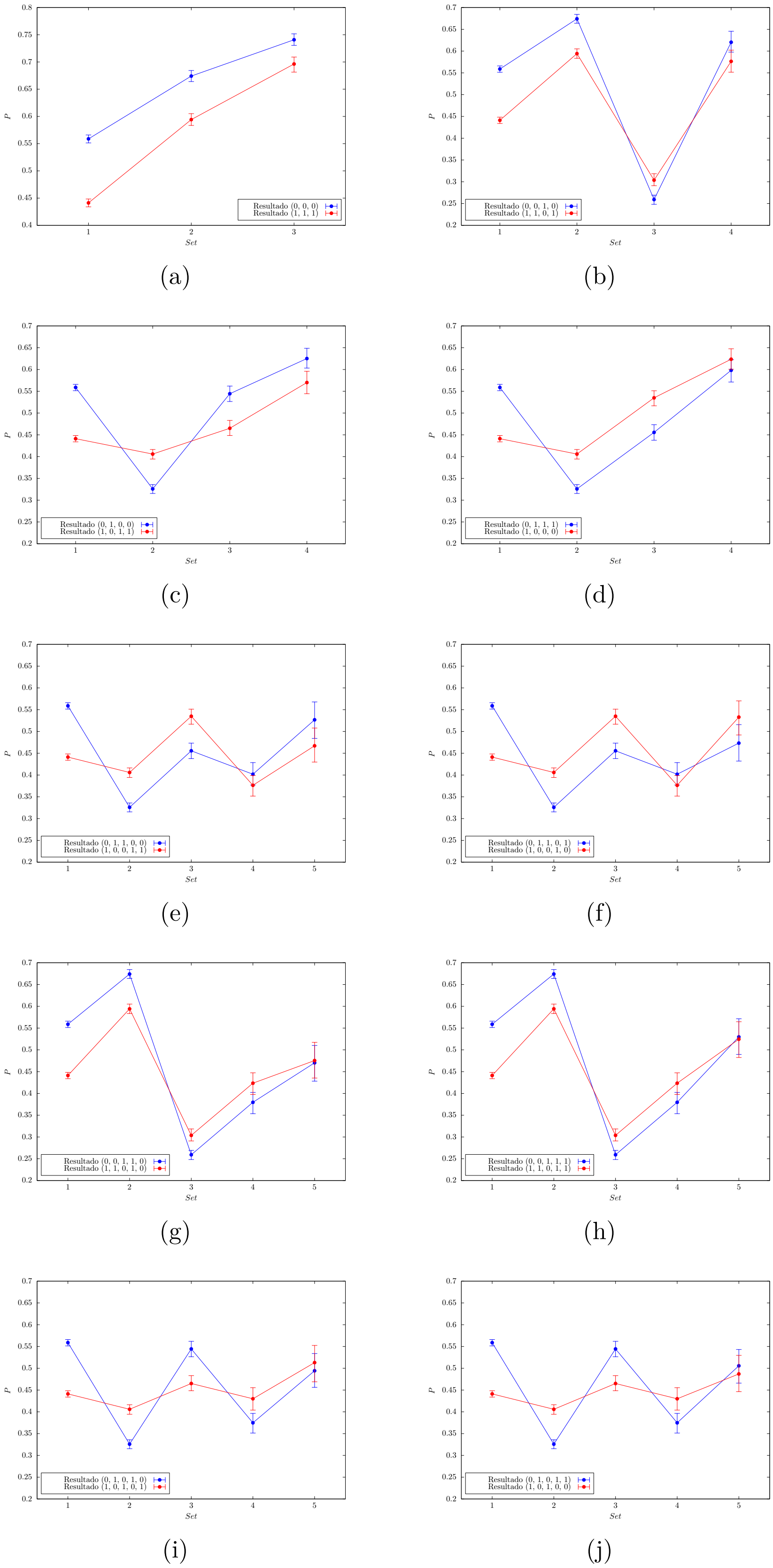


Figura 2.6: **Probabilidade de vencer cada *set* dependendo do *set* anterior.** Os gráficos mostram as probabilidades para os resultados específicos do jogo, em que é possível observar que a probabilidade de vencer um *set* aumenta ou diminui após a vitória ou derrota do *set* anterior, respectivamente

Conclusões

Neste trabalho, abordamos a vantagem de jogar em casa em jogos de voleibol. Para esse fim, utilizamos os conceitos de probabilidade frequencista simples e condicional, sobre os quais apresentamos uma breve introdução no Capítulo 1.

A partir de uma base de dados com tamanho considerável, pudemos calcular com que frequência o time da casa vence os jogos. Dessa forma, foi possível estimar que, ao jogar em casa, o time tem uma probabilidade relativamente maior de vencer a partida (56,92%). Esse resultado está de acordo com o que já foi encontrado antes na literatura.

Analisamos como essa vantagem se manifesta em cada *set* e vimos que o time da casa tem certa vantagem que se mantém aproximadamente constante nos três primeiros *sets*, mas que diminui no quarto *set* e se extingue no último *set* do jogo 2.2, o que pode estar relacionada de alguma forma com o fato da vantagem de jogar em casa aparecer mais quando os times são desequilibrados em força (jogos terminados em 3-0 e 3-1) e não existir quando os times são equilibrados (jogos terminados em 3-2).

Calculamos também a probabilidade de um time vencer um *set* quando perdeu ou ganhou o *set* anterior. Foi possível ver claramente que a chance de vencer um *set* aumenta consideravelmente após vencer o *set* anterior e diminui bastante após perder o *set* anterior. Esse fato depende de vários fatores, mas pode estar relacionado de alguma maneira a variáveis emocionais.

Por fim, foi possível calcular as probabilidades de vencer cada *set* da partida dependendo de toda a história do jogo, de tal forma que, a partir desses resultados, é possível construir as probabilidades de qualquer resultado final ou parcial. No melhor do nosso conhecimento, esse tipo de resultado que depende da história do jogo ainda não tinha sido investigado. Esses resultados podem servir de base para simulações, isto é, ao invés de usar probabilidades iguais de se vencer cada *set*, usa-se as probabilidades condicionais dos tipos das obtidas neste trabalho. Além disso caberia um estudo da influência do placar dos *sets* anteriores na probabilidade do seguinte. Talvez isso possibilitaria um ajuste mais fino das probabilidades condicionais.

De maneira geral, os resultados mostrados aqui indicam que a dinâmica do voleibol não é consistente com a de processos markovianos, pois a probabilidade de um *set* ser vencido mostrou-se depender de toda a história do jogo.

Ainda existe muito o que se estudar em relação a jogos de voleibol, como, por exemplo a vantagem de jogar em casa em campeonatos femininos, já que em jogos de tênis foi verificado que a vantagem de jogar em casa é grande em campeonatos masculinos, mas não é significativa em campeonatos femininos [21] e seria interessante investigar se isto também acontece no voleibol. Também seria muito interessante estudar estruturas mais internas do jogo, como a dinâmica das pontuações em cada *set* e também a influência dos pedidos de tempo no resultados dos *sets* e dos jogos.

Os resultados aqui apresentados podem servir também de base para comissões técnicas tomarem decisões. Por exemplo, em um jogo em andamento, sabendo da história do jogo (e, portanto das probabilidades condicionais aqui obtidas) pode-se empregar mais ou menos esforços na motivação dos jogadores.

Apêndice A

A lei dos grandes números

Desejamos tornar mais precisa uma ideia usada em todo este trabalho. À medida que o número de repetições de um experimento cresce, a frequência relativa f_E de algum evento E converge (em um sentido probabilístico que será explicado mais adiante) para a probabilidade teórica $P(E)$. É esse fato que nos permite identificar a frequência relativa de um evento, baseada em um grande número de repetições, com a probabilidade do evento [51].

Para que sejamos bem sucedidos nessa tarefa, inicialmente definimos alguns conceitos importantes na teoria de probabilidade e na estatística, de acordo com [51, 53, 57].

Definição 3. *Variável Aleatória.* Sejam ε um experimento e S um espaço amostral associado ao experimento aleatório. Uma função X, que associe a cada elemento $s \in S$ um número real, $X(s)$, é denominada *variável aleatória.* Se o número de valores possíveis de X for finito ou infinito enumerável, denominaremos X uma *variável aleatória discreta.*

Em outras palavras, a variável aleatória possui um valor único (definido aleatoriamente) para cada resultado possível do experimento.

Definição 4. *Distribuição de Probabilidade.* A distribuição de probabilidades $f(x)$, associa uma probabilidade a cada resultado numérico de um experimento, ou seja, dá a probabilidade de cada valor de uma variável aleatória $f(x) = P(X = x)$. Temos duas regras que se aplicam a qualquer distribuição de probabilidade:

1. $0 \leq f(x) \leq 1$ para todo i,

2. $\sum_{x=-\infty}^{\infty} f(x) = 1$.

Uma distribuição de grande importância no caso discreto pode ser formulada da seguinte maneira: suponhamos um experimento cujos resultados possíveis possam ser classificados como sucesso ou falha. Se associarmos $X = 1$ ao resultado sucesso e $X = 0$ ao resultado falha, então a distribuição de probabilidade da função X é dada por

$$f(0) = P(X = 0) = 1 - p$$

$$f(1) = P(X = 1) = p$$

em que p, $0 \leq p \leq 1$, é a probabilidade do resultado sucesso ser obtido.

Essa distribuição é conhecida como distribuição de Bernoulli. Suponha gora que n diferentes tentativas são realizadas, cada uma com probabilidade de sucesso p e probabilidade de falha $1-p$. Se X representa o número de sucessos que ocorrem em n tentativas, então X é dita uma variável aleatória binomial.

Definição 5. *Valor Esperado.* Seja X uma variável aleatória discreta tendo uma função densidade de probabilidade $f(x)$, o valor esperado de X, denotado por $E(X)$, é dado por

$$E(X) = \sum_x x f(x).$$

Definição 6. *Variância.* Se X for uma variável aleatória com média μ, então a variância de X, denotada por $Var(X)$, é dada por

$$Var(X) = E[(X - \mu)^2].$$

Para a distribuição de binomial temos

$$E(X) = np,$$

$$Var(X) = np(1 - p).$$

Existe uma conhecida desigualdade, devida ao matemático russo Tchebychev, que desempenha um importante papel em nosso trabalho subsequente. Se conhecermos a distribuição de probabilidade de uma variável aleatória X, podemos calcular $E(X)$ e $Var(X)$, se existirem. Contudo, a recíproca não é verdadeira. Isto é, do conhecimento de $E(X)$ e $Var(X)$ não podemos reconstruir a distribuição de probabilidade de X e, consequentemente, calcular quantidades como $P(|X - E(X)| \leq c)$. Não obstante, verificamos que, muito embora não possamos calcular tais probabilidades, podemos estabelecer um limite superior (ou inferior) muito útil para essas probabilidades. Esse resultado está contido na desigualdade de Tchebychev [51, 53].

A.1 Desigualdade de Tchebychev

Seja X uma variável aleatória, com $E(X) = \mu$, e seja c um número real qualquer. Então, se $E[(X-c)^2]$ for finita e ϵ for qualquer número positivo, teremos

$$P(|X-c| \geq \epsilon) \leq \frac{1}{\epsilon^2} E[(X-c)^2].$$

Demonstração. Consideremos

$$P(|X-c| \geq \epsilon) = \sum_{x:|x-c|\geq\epsilon} f(x).$$

Mas, $|x-c| \geq \epsilon$ é equivalente a $\frac{(x-c)^2}{\epsilon^2} \leq 1$. Consequentemente, a soma acima é

$$\leq \sum_{x:|x-c|\geq\epsilon} \frac{(x-c)^2}{\epsilon^2} f(x).$$

Essa soma é, por sua vez,

$$\leq \sum_{-\infty}^{\infty} \frac{(x-c)^2}{\epsilon^2} f(x),$$

que é igual a

$$\frac{1}{\epsilon^2} E[(X-c)^2],$$

como queríamos demonstrar.[1] □

Se escolhemos $c = \mu$ e $\epsilon = k\sigma$, em que $\sigma^2 = Var(X) > 0$, obtemos

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

A.2 Demonstração da lei dos grande números

Lei fraca dos grandes números. Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2. Seja $S_n = \sum_{i=1}^{n} X_i$, temos que

$$\frac{S_n}{n} \to \mu.$$

Demonstração. Aplicamos a desigualdade de Tchebichev à variável aleatória $\frac{S_n}{n}$. Como

[1] Este resultado é bastante notável, já que muito pouco é suposto a respeito do comportamento probabilístico da variável aleatória X.

$E\left(\frac{S_n}{n}\right) = \mu$ e $Var\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$, a desigualdade se transforma em

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Daí segue, tomando o limite para $n \to \infty$, que

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0.$$

□

Um caso particular importante é aquele em que $X_i = 1$ (sucesso) com probabilidade p e $X_i = 0$ (falha) com probabilidade $(1-p)$, para $i = 1, 2, \ldots, n$, que corresponde a uma sequência de ensaios de Bernoulli. $\frac{S_n}{n}$ representa a frequência relativa de sucessos para n ensaios. Nesse caso, na desigualdade de Tchebichev temos $\mu = p$ e $\sigma^2 = p(1-p)$. Assim,

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

A expressão do lado direito tende a zero para $n \to \infty$.

A lei fraca diz que $\frac{S_n}{n}$, a frequência relativa de sucessos, converge em probabilidade para p.

É importante salientar a diferença entre a convergência referida (convergência em probabilidade) e o tipo de convergência frequentemente mencionada em cálculo infinitesimal. Quando dizemos que 2^{-n} converge para zero quado $n \to \infty$, queremos dizer que para n suficientemente grande, 2^{-n} se torna e permanece arbitrariamente próximo de zero. Quando dizemos que $\frac{S_n}{n}$ converge para p, queremos dizer que a probabilidade do evento $P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right)$ pode se tornar arbitrariamente próxima de zero, ao se tomar n suficientemente grande.

Apêndice B

Bootstrapping

A ideia do método *bootstrapping* é simples. Em vez de fazermos suposições sobre a distribuição de probabilidades da amostra para usarmos argumentos teóricos, se pudéssemos ter não apenas uma, mas várias amostras de dados, a medida estatística de interesse (a média por exemplo) poderia ser calculada para todas as amostras e poderíamos ter uma estimativa dos valores entre os quais esta medida estatística se encontra [62].

Podemos, então, criar amostras adicionais, com o mesmo tamanho da amostra original, fazendo uma re-amostragem aleatória dos dados com reposição, ou seja, um elemento de um conjunto de dados é sorteado, adicionado a um novo conjunto e é mantido no conjunto original, podendo ser sorteado novamente. Esse método foi proposto por Bradley Efron em 1979 [63] e pode ser usado para a construção de intervalos de confiança [1] [64, 65].

Vamos supor uma amostra $E = \{e_1, e_2, \ldots, e_n\}$, da qual desejamos saber alguma medida estatística $\hat{\theta}$. Podemos proceder como segue.

Primeiramente, construímos N subconjuntos da amostra E, em que cada elemento a_j destas subamostras é sorteado do conjunto original.

$$A_i = \{a_1, a_2, \ldots, a_n\},\ i = 1, 2, \ldots, N.$$

Para cada subconjunto A_i calculamos a medida $\hat{\theta}$, ou seja,

$$\Theta = \{\hat{\theta}[A_1], \hat{\theta}[A_2], \ldots, \hat{\theta}[A_N]\}.$$

O intervalo de confiança para a medida $\hat{\theta}$, com nível de confiança α, é dado por

$$\begin{aligned}(\hat{\theta}[A]_{inferior}) &= \hat{Q}_{\frac{\alpha}{2}}[\Theta],\\ (\hat{\theta}[A]_{superior}) &= \hat{Q}_{1-\frac{\alpha}{2}}[\Theta],\end{aligned}$$

[1] É importante ressaltar que *bootstrapping* é um método para se obter intervalos de confiança, e **não** para se obter melhores estimativas de alguma medida.

em que $\hat{Q}_\beta$ é o β-quantil do conjunto Θ. Ou seja, $\hat{Q}_\beta$ retorna o valor de x tal que a probabilidade de encontrarmos um evento menor do que x seja β.

Assim, podemos interpretar α da seguinte maneira: existe uma chance de $\alpha.100\%$ da medida $\hat{\theta}$ estar fora do intervalo de confiança, ou ainda, garante-se, com uma chance de (1 - α).100% que α está dentro do intervalo de confiança. Os valores mais comuns de α são $\alpha = 0,05$ e $\alpha = 0,01$.

Como vimos, o *bootstrapping* é um método simples e prático para obtermos intervalos de confiança. Mas, para que esse método seja confiável, duas condições dever ser obedecidas

- A amostra original deve ser suficientemente grande e relativamente limpa, isto é, não apresentar grandes oscilações em seus valores ou outras anomalias.

- A quantidade estimada deve depender suavemente dos dados, ou seja, não deve depender criticamente de alguns poucos pontos dos dados.

O número de amostras a serem consideradas vai depender do tamanho da amostra original. Se os pontos da amostra original são poucos, criar muitas amostras do *bootstrapping* gerará a mesma subamostra várias vezes. No entanto, se a amostra original for grande o suficiente, isso não será um problema, uma vez que o número de subamostras possíveis cresce muito rápido com o número de dados da amostra original. Assim, será muito improvável que a mesma subamostra se repita.

Referências Bibliográficas

[1] E. Bertin, *Statistical Physics of Complex Systems*. New York: Springer, 2016.

[2] R. N. Mantegna and H. E. Stanley, *Introduction to econophysics: correlations and complexity in finance*. Cambridge: Cambridge university press, 1999.

[3] T. K. D. Peron, L. da Fontoura Costa, and F. A. Rodrigues, "The structure and resilience of financial market networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 1, p. 013117, 2012.

[4] L. E. Araripe, R. N. Costa Filho, H. J. Herrmann, and J. S. Andrade Jr, "Plurality voting: the statistical laws of democracy in Brazil," *International Journal of Modern Physics C*, vol. 17, no. 12, pp. 1809–1813, 2006.

[5] M. C. Mantovani, H. V. Ribeiro, M. V. Moro, S. Picoli Jr, and R. S. Mendes, "Scaling laws and universality in the choice of election candidates," *EPL (Europhysics Letters)*, vol. 96, no. 4, p. 48001, 2011.

[6] A. Chatterjee, M. Mitrović, and S. Fortunato, "Universality in voting behavior: an empirical analysis," *Scientific Reports*, vol. 3, no. 1049, 2013.

[7] K. S. Courneya and A. V. Carron, "The home advantage in sport competitions: a literature review.," *Journal of Sport & Exercise Psychology*, vol. 14, no. 1, 1992.

[8] N. J. Balmer, A. M. Nevill, and A. M. Williams, "Modelling home advantage in the summer olympic games," *Journal of Sports Sciences*, vol. 21, no. 6, pp. 469–478, 2003.

[9] H. V. Ribeiro, S. Mukherjee, and X. H. T. Zeng, "The advantage of playing home in NBA: Microscopic, team-specific and evolving features," *PLOS ONE*, vol. 11, p. e0152440, mar 2016.

[10] E. Ben-Naim, S. Redner, and F. Vazquez, “Scaling in tournaments,” *EPL (Europhysics Letters)*, vol. 77, no. 3, p. 30005, 2007.

[11] E. Ben-Naim and N. Hengartner, “Efficiency of competitions,” *Physical Review E*, vol. 76, no. 2, p. 026106, 2007.

[12] H. V. Ribeiro, R. S. Mendes, L. C. Malacarne, S. Picoli, and P. A. Santoro, “Dynamics of tournaments: the soccer case,” *The European Physical Journal B*, vol. 75, no. 3, pp. 327–334, 2010.

[13] A. Gabel and S. Redner, “Random walk picture of basketball scoring,” *Journal of Quantitative Analysis in Sports*, vol. 8, no. 1, 2012.

[14] H. V. Ribeiro, S. Mukherjee, and X. H. T. Zeng, “Anomalous diffusion and long-range correlations in the score evolution of the game of cricket,” *Physical Review E*, vol. 86, no. 2, 2012.

[15] E. Ben-Naim, F. Vazquez, and S. Redner, “What is the most competitive sport?,” *Journal- Korean Physical Society*, vol. 5, no. 1, pp. 124–126, 2006.

[16] Y. de Saá Guerra, J. M. González, S. S. Montesdeoca, D. R. Ruiz, A. García-Rodríguez, and J. García-Manso, “A model for competitiveness level analysis in sports competitions: Application to basketball,” *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 10, pp. 2997–3004, 2012.

[17] E. L. Uhlmann and C. M. Barnes, “Selfish play increases during high-stakes NBA games and is rewarded with more lucrative contracts,” *PLoS ONE*, vol. 9, no. 4, p. e95745, 2014.

[18] A. M. Nevill and R. L. Holder, “Home advantage in sport,” *Sports Medicine*, vol. 28, no. 4, pp. 221–236, 1999.

[19] A. V. Carron, T. M. Loughhead, and S. R. Bray, “The home advantage in sport competitions: Courneya and Carron’s (1992) conceptual framework a decade later,” *Journal of Sports Sciences*, vol. 23, no. 4, pp. 395–407, 2005.

[20] R. Marcelino, I. Mesquita, J. Palao, and J. Sampaio, “Home advantage in high-level volleyball varies according to set number,” *Journal of Sports Science and Medicine*, vol. 8, no. 3, 2009.

[21] R. H. Koning, “Home advantage in professional tennis,” *Journal of Sports Sciences*, vol. 29, no. 1, pp. 19–27, 2011.

[22] W. H. Carter Jr and S. L. Crews, "An analysis of the game of tennis," *The American Statistician*, vol. 28, no. 4, pp. 130–134, 1974.

[23] P. E. Pfeifer and S. J. Deutsch, "A probabilistic model for evaluation of volleyball scoring systems," *Research Quarterly for Exercise and Sport*, vol. 52, no. 3, pp. 330–338, 1981.

[24] G. W. Fellingham, B. J. Collings, and C. M. McGown, "Developing an optimal scoring system with a special emphasis on volleyball," *Research Quarterly for Exercise and Sport*, vol. 65, no. 3, pp. 237–243, 1994.

[25] F. J. Klaassen and J. R. Magnus, "Forecasting the winner of a tennis match," *European Journal of Operational Research*, vol. 148, no. 2, pp. 257–267, 2003.

[26] T. Barnett, A. Brown, and S. Clarke, "Developing a model that reflects outcomes of tennis matches," in *Proceedings of the 8th Australasian Conference on Mathematics and Computers in Sport, Coolangatta, Queensland*, pp. 3–5, 2006.

[27] B. Kovacs, "The effect of the scoring system changes in volleyball: a model and an empirical test," *Journal of Quantitative Analysis in Sports*, vol. 5, no. 3, pp. 1559–0410, 2009.

[28] P. K. Newton and K. Aslam, "Monte Carlo tennis: a stochastic Markov chain model," *Journal of Quantitative Analysis in Sports*, vol. 5, no. 3, pp. 1–42, 2009.

[29] I. McHale and A. Morton, "A bradley-terry type model for forecasting tennis match results," *International Journal of Forecasting*, vol. 27, no. 2, pp. 619–630, 2011.

[30] D. Paindaveine and Y. Swan, "A stochastic analysis of some two-person sports," *Studies in Applied Mathematics*, vol. 127, no. 3, pp. 221–249, 2011.

[31] D. L. González, "Un modelo exactamente soluble para los marcadores en partidos de voleibol," *Revista Brasileira de Ensino de Fısica*, vol. 35, no. 2, p. 2302, 2013.

[32] T. K. Shimizu, F. Louzada, and A. K. Suzuki, "Analyzing volleyball data on a compositional regression model approach: An application to the Brazilian men's volleyball super league 2011/2012 data," *arXiv preprint arXiv:1412.5848*, 2014.

[33] T. M. Loughead, A. V. Carron, S. R. Bray, and A. J. Kim, "Facility familiarity and the home advantage in professional sports," *International Journal of Sport and Exercise Psychology*, vol. 1, no. 3, pp. 264–274, 2003.

[34] H. M. Wallace, R. F. Baumeister, and K. D. Vohs, "Audience support and choking under pressure: A home disadvantage," *Journal of Sports Sciences*, vol. 23, no. 4, pp. 429–438, 2005.

[35] "Pulsão." `https://pt.wikipedia.org/wiki/Puls~ao`. Accessed: 19-01-2017.

[36] R. Pollard and G. Pollard, "Home advantage in soccer: A review of its existence and causes," *International Journal of Soccer and Science Journal*, vol. 3, no. 1, pp. 28–38, 2005.

[37] J. Sampaio, S. J. Ibanez Godoy, M. A. Gomez Ruano, A. Lorenzo Calvo, and E. Ortega Toro, "Game location influences basketball players performance across playing positions.," *International Journal of Sport Psychology*, vol. 39, no. 3, pp. 43–50, 2008.

[38] N. O. de Melo, "Vantagem de jogar em casa no voleibol de elevado rendimento," Dissertação de Mestrado, Depertamento de Educação Física, Universidade Federal do Rio Grande do Sul, 2011.

[39] M. D. Tillman, C. J. Hass, D. Brunt, and G. R. Bennett, "Jumping and landing techniques in elite women's volleyball," *J Sports Sci Med*, vol. 3, no. 1, pp. 30–36, 2004.

[40] F. Angioluci Diniz Campos, B. Natale Pasquarelli, L. C. Benetti Campos, E. Heidi Ozaki, R. Stanganélli, and L. Cláudio, "Análise da vantagem de jogar em casa no voleibol masculino brasileiro," *Brazilian Journal of Biomotricity*, vol. 7, no. 3, pp. 144–149, 2013.

[41] F. A. Campos, L. C. Stanganélli, B. N. Pasquarelli, L. C. Campos, and M.-Á. Gómez, "Performance indicators analysis at brazilian and italian women's volleyball leagues according to game location, game outcome, and set number," *Perceptual and motor skills*, vol. 118, no. 2, pp. 347–361, 2014.

[42] F. A. D. Campos, I. L. Pellegrinotti, B. N. Pasquarelli, E. H. Ozaki, and L. C. R. Stanganéllli, "Análise da vantagem de jogar em casa no voleibol feminino brasileiro," *Revista Brasileira de Ciência e Movimento*, vol. 23, no. 1, pp. 40–47, 2015.

[43] F. C. Rabaz, R. J. Castuera, M. C. Suárez, L. G. González, and M. P. M. Arroyo, "Cognitive and motivational variables as predictors of performance in game actions in young volleyball players," *European Journal of Human Movement*, vol. 35, pp. 68–84, 2015.

[44] P. A. Richardson, W. Adler, and D. Hankes, "Game, set, match: Psychological momentum in tennis," *The Sport Psychologist*, vol. 2, no. 1, pp. 69–76, 1988.

[45] K. L. Burke and S. Houseworth, “Structural charting and perceptions of momentum in intercollegiate volleyball,” *Journal of Sport Behavior*, vol. 18, no. 3, p. 167, 1995.

[46] M. Bar-Eli and G. Tenenbaum, “Observations of behavioral violations as crisis indicators in competition,” *The Sport Psychologist*, vol. 3, no. 3, pp. 237–44, 1989.

[47] J. Sampaio, A. Ferreira, S. Ibañez, and C. Ribeiro, “Success in the last 5 minutes of basketball close games: investigating final outcome of ball possession, duration of ball possession, number of players’ involves, defensive opposition and court location,” in *Book of abstracts of the World Congress of Performance Analysis of Sport (13)*, pp. 56–58, 2004.

[48] R. M. Navarro, M. A. Gómez, J. Lorenzo, and S. Jiménez, “Qualitative analysis of critical moments in basketball,” *Revista de Psicología del Deporte*, vol. 22, no. 1, pp. 249–251, 2013.

[49] M. Lames, “Modelling the interaction in game sports–relative phase and moving correlations,” *Journal of Sports Science and Medicine*, vol. 5, no. 4, pp. 556–560, 2006.

[50] M. Jones, “Home advantage in the NBA as a game-long process,” *Journal of Quantitative Analysis in Sports*, vol. 3, no. 4, p. 2, 2007.

[51] P. L. Meyer, “Probabilidade: aplicações à estatística,” in *Probabilidade: aplicações à estatística*, Rio de Janeiro: Livro Técnico, 2011.

[52] F. M. Mendes, “Processos estocásticos em física: teoria e fundamentos,” Tese de Doutorado, Universidade de Brasília, Instituto de Física, 2009.

[53] C. A. B. Dantas, *Probabilidade: Um Curso Introdutório.* Edusp, 2013.

[54] G. Shafer, “The significance of Jacob Bernoulli’s Ars Conjectandi for the philosophy of probability today,” *Journal of Econometrics*, vol. 75, no. 1, pp. 15–32, 1996.

[55] M. Mattmüller, “The difficult birth of stochastics: Jacob Bernoulli’s Ars Conjectandi (1713),” *Historia Mathematica*, vol. 41, no. 3, pp. 277–290, 2014.

[56] N. A. Lemos, *Convite à Física Matemática.* São Paulo: Editora Livraria da Física, 2013.

[57] R. Sheldon, *A first course in probability.* University of California, Berkeley, 1998.

[58] T. Tomé and M. J. de Oliveira, *Dinâmica estocástica e irreversibilidade.* Edusp, 2001.

[59] “Resultados dos jogos de vôlei..” `https://http://www.scorespro.com/volleyball/`. Accessed: 19-01-2017.

[60] “Mathematica.” `https://www.wolfram.com/mathematica/`.

[61] M. M. F. da Silva, M. B. de Paiva Vidual, R. O. Afonso, H. M. Yoshida, J. P. Borin, and P. T. Fernandes, “Ansiedade e desempenho de jogadoras de voleibol em partidas realizadas dentro e fora de casa,” *Journal of Physical Education*, vol. 25, no. 4, pp. 585–596, 2014.

[62] P. K. Janert, *Data analysis with open source tools*. Beijing: "O’Reilly Media, Inc.", 2010.

[63] B. Efron, “Bootstrap methods: Another look at the jackknife,” *The Annals of Statistics*, vol. 7, pp. 1–26, jan 1979.

[64] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.

[65] B. Efron *et al.*, “Second thoughts on the bootstrap,” *Statistical Science*, vol. 18, no. 2, pp. 135–140, 2003.