
UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

Dissertação de Mestrado

Dinâmica de torneios e símbolos

Haroldo Valentin Ribeiro

Orientador: Prof. Dr. Renio dos Santos Mendes

Maringá, Março de 2010.

UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

Dissertação de Mestrado

Dinâmica de torneios e símbolos

Haroldo Valentin Ribeiro

Dissertação de Mestrado
submetida ao Departamento de Física
da Universidade Estadual de Maringá.

Orientador: Prof. Dr. Renio dos Santos Mendes

Maringá, Março de 2010.

*Em memória de
Aparecida da Silva Valentin.*

Agradecimentos

Sou especialmente grato ao professor Renio, orientador e amigo, com quem tenho aprendido muitas coisas relacionadas à pesquisa científica e outra tantas relacionadas à vida. Suas muitas ideias e entusiasmo contagiante foram essenciais para o bom andamento deste trabalho.

Agradeço igualmente ao professor Perseu pela orientação nos primeiros projetos de iniciação científica e pela confiança em mim depositada. Não posso deixar de agradecer também aos professores Sergio, Ervin e Malacarne pelo grande apoio e participação ativa nos projetos de pesquisa.

Agradeço a meus pais, Vera e Gilberto, pelo apoio nas minhas escolhas e a minha namorada Mônica, pelo companheirismo e apoio emocional.

Agradeço aos meus colegas e amigos de curso pela ajuda, paciência e pelos mais de dez mil cafés: Rodolfo, Marcos Paulo, Roberto, Angel, Douglas e Diogo. Nessa contagem dos cafés, não posso esquecer o professor Renato.

Finalmente, agradeço à CAPES e ao CNPq pelo apoio financeiro e, pelo suporte computacional, ao CENAPAD-SP.

Resumo

Esta dissertação está dividida em duas partes principais, ambas relacionadas às caminhadas aleatórias. Na primeira delas, um modelo baseado em caminhada aleatória foi utilizado para discutir aspectos estatísticos de torneios. Este modelo foi aplicado a ligas de futebol com ênfase nas pontuações das equipes. Este sistema foi computacionalmente simulado e os resultados mostraram um bom acordo com os dados empíricos provenientes das ligas de futebol inglesa, alemã e espanhola. O presente ponto de vista permitiu caracterizar um processo difusivo em que as pontuações não estão distribuídas gaussianamente e exibem um regime superdifusivo. Nós argumentamos que o comportamento não-gaussiano está relacionado à diferença entre as equipes e com a assimetria do sistema de pontuação. E por fim, como uma aplicação do nosso modelo, comparamos dois sistemas de torneios: os de pontos corridos e os eliminatórios. Na segunda parte, investigamos o comportamento de sequências simbólicas com correlações de longo alcance via simulação computacional. Analisamos sequências com dois, três e quatro símbolos que poderiam se repetir l vezes, em que l é um número aleatório com distribuição de probabilidade $p(l) \propto l^{-\mu}$. Para essas sequências, verificamos que a entropia usual cresce mais lentamente quando as sequências estão correlacionadas e que a entropia S_q de Tsallis exibe um comportamento linear para uma escolha particular do parâmetro q . Adicionalmente, estudamos as sequências como se fossem uma caminhada aleatória e observamos um comportamento difusivo não-usual para alguns valores de μ . Especificamente, verificamos que o regime de difusão é superdifusivo quando há correlação de longo alcance e, além disso, a distribuição de probabilidade mostrou-se em bom acordo com uma q -gaussiana.

Abstract

This dissertation is divided into two main sections, both related to random walks. In the first one, a random walk-like model is considered to discuss statistical aspects of tournaments. The model is applied to soccer leagues with emphasis on the scores. This competitive system was computationally simulated and the results are compared with empirical data from the English, the German and the Spanish leagues and showed a good agreement with them. The present approach enabled us to characterize a diffusion where the scores are not normally distributed and exhibit a superdiffusive regime. We argue that this non-Gaussian behavior is related with the difference between the teams and with the asymmetry of the score systems. In addition, we compared two tournament systems: the all-play-all and the elimination tournaments. In the second section, we investigated symbolic sequences with long-range correlations by using computational simulation. We analyze sequences with two, three and four symbols that could be repeated l times, with the probability distribution $p(l) \propto l^{-\mu}$. For these sequences, we verified that the usual entropy increases more slowly when the symbols are correlated and the Tsallis entropy S_q exhibits, for a suitable choice of q , a linear behavior. We also study the sequences as a random walk-like process and observed a nonusual diffusive behavior depending on the values of the parameter μ . Specifically, we found that the diffusion regime is superdiffusive when there are long-range correlations and that the probability distribution function is in good agreement with a q -Gaussian.

Sumário

Resumo	3
Abstract	4
Introdução	7
1 Dinâmica de Torneios	10
1.1 As tabelas de pontuação	10
1.2 Apresentação dos dados observacionais	12
1.3 Análise dos dados	13
1.4 Um modelo minimalista de equipes idênticas	21
1.5 Um modelo de equipes não-idênticas	26
1.6 Número de vitórias, empates e derrotas	35
1.7 Comparando dois sistemas de torneio	37
1.8 Equipes não-idênticas e o comportamento não-gaussiano	39
2 Sequências Simbólicas	41
2.1 Interações de longo alcance	41
2.2 Gerando sequências correlacionadas	43
2.3 Entropia de sequências simbólicas	45
2.4 Difusão em sequências simbólicas	51

2.5	Leis de escala e distribuição de probabilidade	53
	Conclusões	58
	Referências Bibliográficas	61

Introdução

Esta dissertação de mestrado está focada no estudo de fenômenos e modelos de natureza aleatória. Antes de mais nada, há de se notar que muitos fenômenos naturais mostram-se imprevisíveis ou aleatórios e, além disso, a maioria das teorias da Física apresentam alguma relação com teoria de probabilidades como, por exemplo, a Mecânica Estatística e a Mecânica Quântica.

O motivo pelo qual fenômenos aleatórios são tão presentes está relacionado ao fato de que, para uma ampla classe de sistemas físicos, as informações sobre o fenômeno estão disponíveis até um certo nível. Mesmo se estivermos no limite clássico e admitirmos conhecer a forma exata das interações de um sistema com muitos graus de liberdade, necessitaríamos de um conjunto enorme de condições iniciais e teríamos que resolver um sistema acoplado de equações dinâmicas igualmente grande. Esta tarefa seguramente assumimos ser impossível e uma nova abordagem estatística ou probabilística se faz necessária.

Além disso, nas últimas décadas, tem havido uma grande tendência multidisciplinar nos vários ramos das ciências e a Física não é exceção, especialmente a Mecânica Estatística. Muitos desses estudos recentes constituem um campo emergente e bastante ativo denominado de *Sistemas Complexos*. Grosseiramente, podemos dizer que um sistema é um conjunto de elementos interagentes que geram um comportamento global[1]. Se por um lado alguns sistemas podem ser muito complicados, por outro eles não são necessariamente considerados complexos. Não

há uma definição precisa de sistemas complexos, mas existe um consenso entre a maioria dos autores no que se refere às propriedades essenciais que um sistema deve possuir para ser considerado complexo. São elas:

- os sistemas consistem de um grande número de agentes interagentes;
- as interações são, em geral, não-locais e/ou não-lineares;
- existe um comportamento coletivo, auto-organizado, o qual é difícil de antecipar a partir do conhecimento da dinâmica individual dos agentes;
- esse comportamento coletivo não resulta da existência de um controle central.

Um bom exemplo de sistema complexo são as colônias de formigas[2]. A colônia opera sem qualquer controle central, isto é, a rainha não decide o trabalho que um indivíduo deve fazer e nenhuma formiga possui uma visão geral da colônia para dar instruções aos demais membros. Cada indivíduo tem disponível apenas informações locais e usando essa informação ele decide qual, das muitas funções existentes, deve fazer para a colônia.

Formigueiros não são os únicos sistemas de múltiplos agentes que possuem um comportamento coordenado sem a existência de um controle central. Outros grupos de animais também apresentam tal característica, por exemplo, cardumes[3] e pássaros em revoada[4, 5]. Além disso, outros sistemas distintos como tráfego de veículos[6], artigos científicos[7, 8], redes sociais como a Word Wide Web[9], mercado financeiro[10] e até mesmo religião[11] apresentam um comportamento emergente similar.

Em contraste com a Mecânica Estatística de equilíbrio, que pode ser resumida de uma maneira precisa e simples, a física estatística dos sistemas complexos [12, 13, 14, 15, 16] não possui um formalismo geral, mas sim alguns formalismos individuais aplicados a casos específicos. Uma dessas tentativas constitui a chamada Mecânica Estatística Não-Extensiva[17], proposta em 1988 por Tsallis, a

qual pode ser considerada como uma generalização da teoria de Boltzmann-Gibbs, baseada em uma entropia não-aditiva S_q ($q \in \mathbb{R}$). Esta última tem se mostrado bastante útil em situações em que as hipóteses básicas da mecânica estatística de Boltzmann-Gibbs (caos molecular e ergodicidade, por exemplo) não podem ser todas verificadas.

No primeiro capítulo desta dissertação, estudaremos uma classe de sistemas com algumas dessas características: os torneios[18]. Apresentaremos uma tentativa de modelagem livre, baseada em caminhada aleatória, para discutir propriedades estatísticas de campeonatos, em particular os de futebol. Faremos uma comparação com dados empíricos provenientes de quatro ligas de futebol e veremos que o modelo apresenta um bom acordo com todas elas. Usando ideias de caminhada aleatória, veremos que a distribuição de pontos não é gaussiana e que apresenta um comportamento superdifusivo. Argumentaremos que a origem do comportamento não-gaussiano está nas diferenças entre as equipes e, por fim, apresentaremos uma comparação quantitativa entre dois sistemas de torneios.

O segundo capítulo consiste de um modelo simbólico que permite explorar regiões em que não se verificam todas as hipóteses da mecânica estatística usual. Este modelo foi proposto por M. Buiatti *et al.*[19] e revisitado recentemente[20]. Trata-se de um experimento numérico que gera sequências correlacionadas e essa correlação faz com que a hipótese de aleatoriedade seja violada, tornando a entropia de Boltzmann-Gibbs não-extensiva. Este fato nos conduz ao uso da entropia S_q de Tsallis, a qual veremos ser extensiva para uma escolha particular do parâmetro q . Adicionalmente, interpretaremos essas sequências como partículas em movimento errático, o que nos permitirá caracterizar o processo de difusão como sendo balístico, superdifusivo ou usual, a depender dos parâmetros do modelo. Na última seção, encerramos esta dissertação apresentando nossas conclusões.

Capítulo 1

Dinâmica de Torneios

Neste capítulo, aplicaremos conceitos de caminhada aleatória para investigar atividades humanas bastante populares: os torneios[18], em particular os de futebol. Veremos que é possível interpretar a evolução das tabelas de pontuação como um processo difusivo. Além disso, construiremos um modelo estatístico que representa muito bem diversos aspectos das ligas de futebol.

1.1 As tabelas de pontuação

Para a maioria dos brasileiros, as tabelas de pontuação de campeonatos são bastante comuns. Elas são atualizadas a cada rodada dos campeonatos e contém diversas informações sobre as equipes, como, por exemplo, número de pontos, vitórias, derrotas, empates, saldo de gols, etc. Para o caso de torneios não-eliminatórios (os chamados torneios de pontos corridos), ao final de todas as rodadas, as tabelas indicam ainda a equipe campeã.

Visando empregar conceitos da mecânica estatística para descrever a dinâmica dos torneios, podemos imaginar que cada uma das equipes executa um passeio aleatório no espaço das pontuações. O próprio esquema de pontuação induz essa

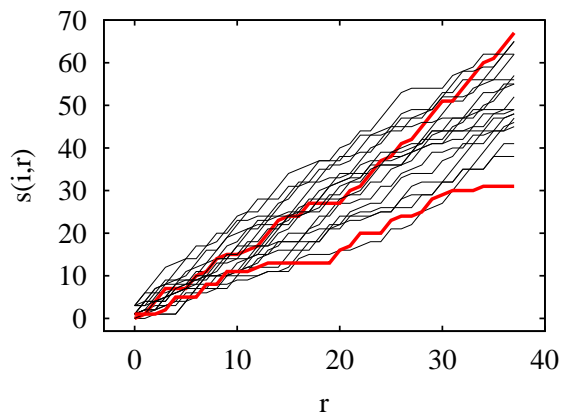


Figura 1.1: Interpretação da tabela de pontos como um processo difusivo. Os dados são da primeira divisão do campeonato brasileiro de 2009 e as duas trajetórias destacadas representam o primeiro colocado (Flamengo) e o último (Sport). Aqui, $s(i, r)$ representa a quantidade de pontos da i -ésima equipe na rodada r .

interpretação. Podemos imaginar que as equipes fazem o papel das partículas de um movimento browniano. O “movimento” das equipes, ou seja, a evolução temporal do número de pontos, é governado pelo resultado de cada partida que a equipe realiza. Em cada rodada (tempo) uma equipe (partícula) pode dar três passos para a direita (no caso de vitória), um passo para a direita (no caso de empate) ou manter-se na mesma posição (no caso de derrota). A figura (1.1) ilustra esta interpretação.

Podemos resumir o processo dinâmico descrito no parágrafo anterior de uma maneira matemática. Para isso, vamos usar $s(i, r)$ para representar o número de pontos da i -ésima equipe na rodada r . Suponha, agora, que haja uma partida entre duas equipes i e j . Podemos, então, escrever

$$\begin{cases} s(i, r + 1) = s(i, r) + \xi_i \\ s(j, r + 1) = s(j, r) + \xi_j \end{cases}, \quad (1.1)$$

em que ξ_i é um número aleatório que pode assumir três valores: no caso em que a

equipe i sai vitoriosa da partida $\xi_i = 3$, no caso de sair derrotada $\xi_i = 0$ e em caso de empate $\xi_i = 1$. Note que, devido ao vínculo que existe entre as duas equipes no momento da partida, uma vez sorteado ξ_i , ξ_j fica bem definido. Isto é, $\xi_j = 0$ em caso de vitória de i , $\xi_j = 3$ em caso de derrota de i e $\xi_j = 1$ em caso de empate. Naturalmente, se considerássemos outro tipo de torneio que não fosse o futebol, ξ_i e ξ_j seguiriam a regra pertinente a esse outro tipo de competição.

1.2 Apresentação dos dados observacionais

Como base de dados observacionais, usaremos os resultados provenientes de quatro ligas de futebol:

- 41 temporadas da liga de futebol alemã[21] (*German Bundesliga*) no período de 1965 a 2007, excluindo-se a temporada de 1991 devido esta ter um número de equipes diferente das demais¹;
- 12 temporadas da liga de futebol inglesa[22] (*English Premier League*) no período de 1995 a 2007;
- 11 temporadas da primeira divisão da liga de futebol espanhola[23] (*La Liga de Fútbol Profesional*) no período de 1996 a 2007;
- 9 temporadas da segunda divisão da liga de futebol espanhola[23] (*La Liga de Fútbol Profesional*) no período de 1996 a 2007.

Em todos esses torneios, as equipes disputam duas partidas com todas as outras, uma vez em seu próprio campo e uma outra no campo do adversário. Dessa maneira, cada equipe disputa um número total de partidas igual a M , em que M depende do número de equipes T envolvidas em cada uma das ligas, $M = 2(T - 1)$.

¹Não consideraremos dados dos campeonatos brasileiros pois nestes têm havido sucessivas mudanças na quantidade de equipes.

O sistema de pontuação mais comum, usado em muitas ligas esportivas, especialmente em torneios de futebol, estabelece 3 pontos para cada vitória, 1 ponto para cada empate e nenhum ponto para derrotas. Contudo, este esquema de pontuação já foi diferente e no passado usava-se 2 pontos para cada vitória ao invés de 3. O ano em que ocorreu essa mudança varia de liga para liga e em nossos dados somente a liga alemã mistura os dois esquemas de pontuação. Esta liga passou a usar o esquema atual de pontuação em 1995, ano no qual a FIFA (*Fédération Internationale de Football Association*) adotou oficialmente o novo sistema, tornando-o padrão em torneios internacionais e em ligas nacionais. Assim, para garantir que os nossos dados tenham um único esquema de pontuação, os pontos das temporadas anteriores a 1995 foram recalculados usando o sistema de pontuação atual.

1.3 Análise dos dados

Em geral, as equipes que participam de um dado torneio variam com o tempo. Nas ligas de futebol, normalmente, existem pelo menos duas divisões: a primeira e a segunda divisão. Cada divisão tem o seu próprio torneio, no qual participam unicamente as equipes que estão naquela divisão. Ocorre que, ao final dos torneios, as equipes melhores classificadas da segunda divisão são promovidas à primeira e as equipes da primeira divisão com pior desempenho são rebaixadas para a segunda. Esta variação de equipes dificulta seguir uma dada equipe durante muitas temporadas e ainda que fizéssemos isso, o número de dados contínuos para as equipes seria muito reduzido.

Para contornar o problema descrito anteriormente, não vamos seguir os dados de uma dada equipe ao longo das temporadas, mas sim de uma dada classificação m . Como veremos mais adiante, nossa abordagem visa caracterizar os torneios como sendo um sistema “físico” passível de um comportamento global e neste

sentido não há perda de generalidade se desprezarmos as identidades das equipes. Assim, permita uma pequena mudança de notação: representaremos a pontuação da classificação m (aquela do final do campeonato), em uma temporada i , na rodada número r , por $s_i(m, r)$.

A pontuação $s_i(m, r)$ faz o papel de uma medida microscópica e por isso está sujeita a flutuações. Visando minimizar o efeito dessas flutuações, vamos iniciar nossas análises investigando o valor médio dessa quantidade ao final de todas as temporadas ($r = M$), isto é,

$$\bar{s}(m) = \frac{1}{N} \sum_{i=1}^N s_i(m, M), \quad (1.2)$$

em que N é o número de temporadas. Na figura (1.2), mostramos $\bar{s}(m)$ para o nosso conjunto de dados empíricos. Observe que $\bar{s}(m)$ apresenta uma forma similar para todas as ligas, mesmo sendo o número de equipes diferente entre as ligas. Este fato sugere uma tendência geral ou um comportamento aproximadamente universal. Note também, que parece haver independência temporal nesta forma, como sugere a figura (1.2f), em que mostramos $\bar{s}(m)$ para quatro períodos distintos da liga alemã.

Outra possibilidade é investigar as flutuações em torno do valor médio $\bar{s}(m)$. Para tal, consideraremos o desvio padrão de $s_i(m, M)$ como função da classificação:

$$\sigma_s^2(m) = \frac{1}{N-1} \sum_{i=1}^N [s_i(m, M) - \bar{s}(m)]^2. \quad (1.3)$$

Mostramos na figura (1.3) esta quantidade para os nossos dados observacionais. Como no caso dos valores médios, $\sigma_s(m)$ também apresenta uma forma similar entre as ligas, caracterizada por flutuações maiores nas extremidades, isto é, nas primeiras e nas últimas classificações. Uma forma muito parecida também é observada nas flutuações da fração de vitórias no beisebol, como reportado por Sire e Redner[24] e mostrado na figura (1.3f), reforçando o possível caráter universal dessas formas.

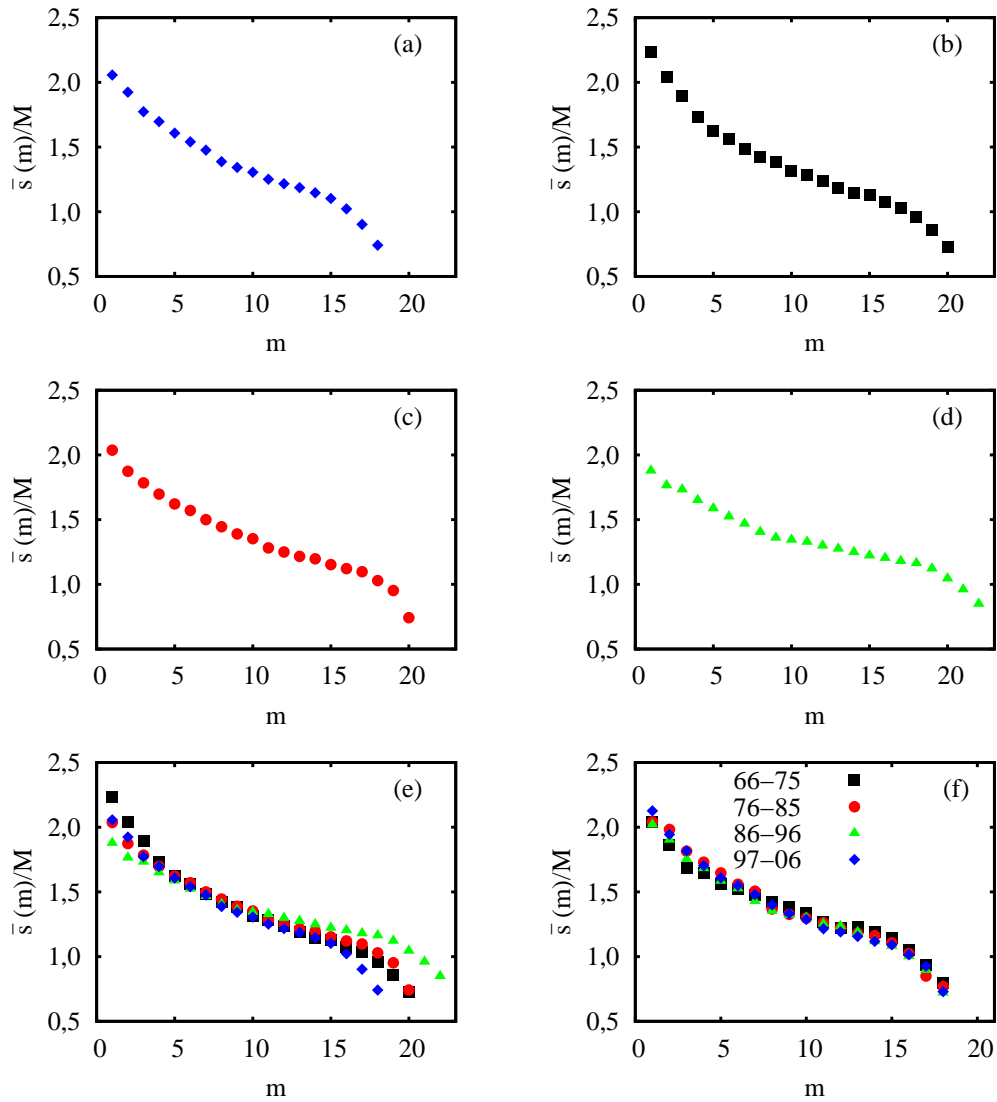


Figura 1.2: Valor médio de $s_i(m, r)$ ao final do torneio ($r = M$) dividido por M , em função da classificação m , para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B. Na figura (e), mostramos todas as ligas superpostas em um único gráfico, preservando os símbolos das figuras anteriores e na figura (f) mostramos a quantidade $\bar{s}(m)$ para quatro períodos da liga alemã.

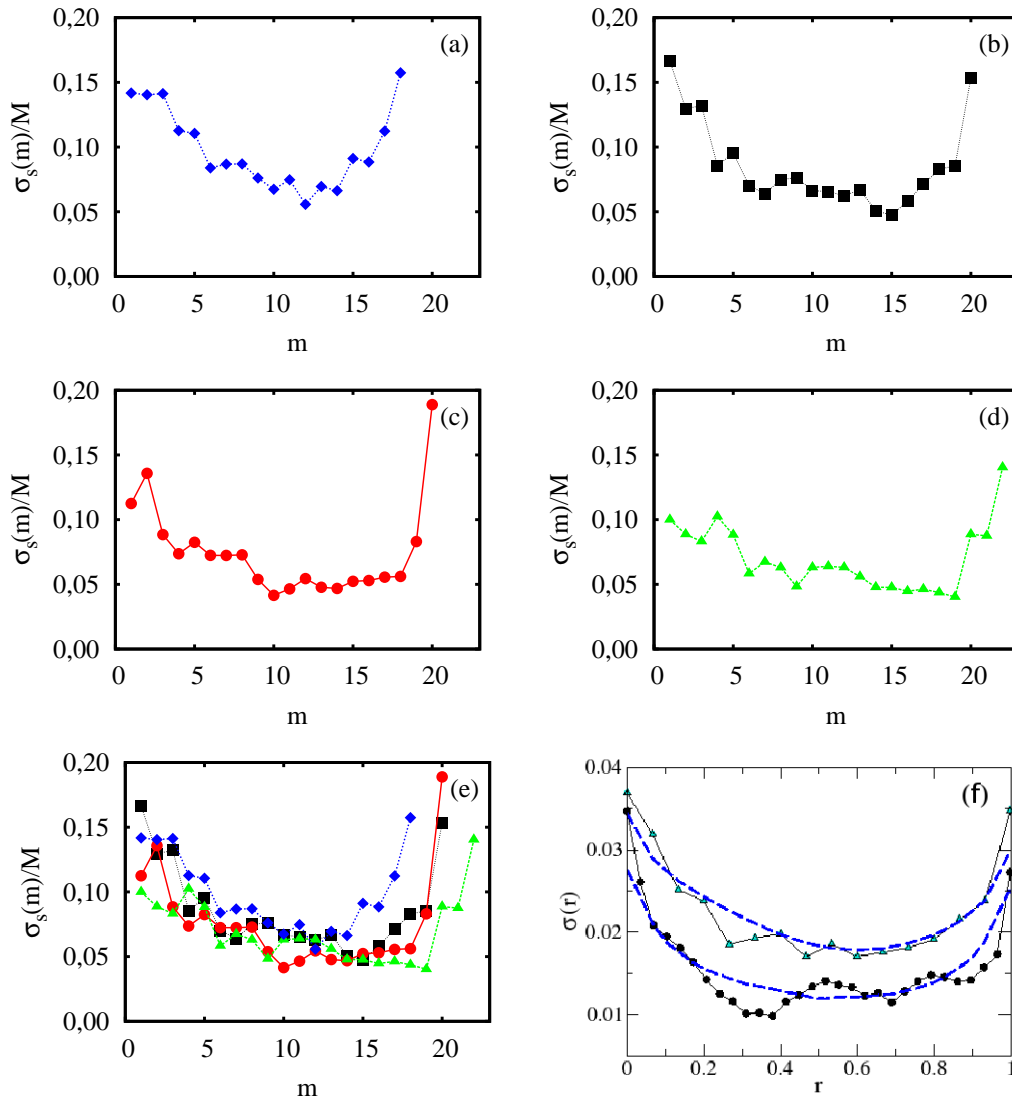


Figura 1.3: Desvio padrão de $s_i(m, M)$, $\sigma_s(m)$ dividido por M , em função da classificação m , para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B. Na figura (e), mostramos todas as ligas superpostas em um único gráfico, preservando os símbolos das figuras anteriores. (f) Flutuação da fração de vitórias da *Major-League baseball*, referência [24].

Nossa análise pode ser estendida para o ponto de vista microscópico usando a interpretação de caminhada aleatória que foi apresentada na seção 1.1. Naquela seção, estabelecemos uma correspondência entre a evolução das equipes e trajetórias erráticas de partículas executando um movimento browniano. Supondo que as equipes façam o papel das partículas e que o movimento seja governado pelos resultados das partidas, teremos que uma equipe pode saltar 3 passos (caso vença) ou 1 passo (caso empate) ou permanecer na mesma posição (caso perca). Dessa maneira, estamos lidando com um processo difusivo, em que na primeira rodada temos apenas 3 posições permitidas (0, 1, 3), na segunda rodada temos 6 (0, 1, 2, 3, 4, 6), e assim sucessivamente.

Neste tipo de análise, é mais conveniente uma quantidade maior de dados e por isso usaremos somente os resultados provenientes da liga alemã, visto que esta liga possui o maior número de temporadas. Começaremos investigando o desvio padrão sobre todas as temporadas em função do número da rodada r (tempo). Matematicamente, temos:

$$\sigma^2(r) = \frac{1}{NT - 1} \sum_{i=1}^N \sum_{m=1}^T (s_i(m, r) - \mu(r))^2, \quad (1.4)$$

em que

$$\mu(r) = \frac{1}{NT} \sum_{i=1}^N \sum_{m=1}^T s_i(m, r),$$

N é número de temporadas e T é o número de equipes. O gráfico da figura (1.4a) mostra $\sigma^2(r)$ versus r em escala logarítmica dupla, no qual a linha contínua representa um ajuste linear dos dados. Neste caso, temos $\sigma^2(r) \propto r^\gamma$, com $\gamma \simeq 1,5$ para $r > 10$, indicando que o processo de difusão não é usual, mas sim um processo superdifusivo, segundo a classificação abaixo.

Sabemos que para o movimento browniano usual $\sigma^2(r) \propto r$, que é uma consequência direta do teorema do limite central e da natureza markoviana do processo estocástico subjacente. Contudo, muitos sistemas apresentam um crescimento não

linear de $\sigma^2(r)$, o qual é frequentemente descrito por uma lei de potência

$$\sigma^2(r) \propto r^\gamma. \quad (1.5)$$

O comportamento difusivo anômalo é normalmente distinguido pelo valor do expoente γ [43]: quando $0 < \gamma < 1$, temos subdifusão e para $\gamma > 1$, superdifusão. O limiar entre subdifusão e superdifusão corresponde à difusão usual e o caso $\gamma = 2$ é conhecido como regime de difusão balístico.

Em adição a $\sigma^2(r)$, calculamos também a distribuição de probabilidade dos pontos para o conjunto de dados de todas as rodadas juntas (25704 pontos) e comparamos com uma distribuição gaussiana, como mostrado na figura (1.4b). Para tal, construímos um histograma. O procedimento usual para obtenção de um histograma consiste em agrupar as N realizações de uma variável aleatória x em subconjuntos mutuamente excludentes de tamanho Δx e representar a distribuição de probabilidade graficamente. Em tal situação, teremos Δn elementos correspondendo ao intervalo entre $x - \frac{\Delta x}{2}$ e $x + \frac{\Delta x}{2}$ e, neste caso, a função densidade de probabilidade pode ser aproximada pela relação

$$p(x) \approx \frac{\Delta n}{N \Delta x}. \quad (1.6)$$

O quão boa é esta aproximação depende de dois fatores: (i)

$$\int_{x - \frac{\Delta x}{2}}^{x + \frac{\Delta x}{2}} p(x) dx \approx p(x) \Delta x,$$

que é melhor à medida em que Δx torna-se pequeno e (ii)

$$\int_{x - \frac{\Delta x}{2}}^{x + \frac{\Delta x}{2}} p(x) dx \approx \frac{\Delta n}{N},$$

que é melhor à medida em que Δn e N tornam-se maiores.

Observe que a pontuação não está normalmente distribuída, visto que a mesma possui caudas mais curtas que as de uma gaussiana e assimétricas, estendendo-se

mais para os valores positivos. Características como esta tornam-se mais evidentes quando analisamos os coeficientes de curtose κ e de assimetria ν , como mostrado respectivamente nas figuras (1.4c) e (1.4d).

O coeficiente de curtose κ , definido como

$$\kappa = \frac{\langle (x - \langle x \rangle)^4 \rangle}{\langle (x - \langle x \rangle)^2 \rangle^2}, \quad (1.7)$$

é adimensional, invariante perante trocas de escala e de origem, servindo para medir se a forma de uma distribuição é muito achatada ou não, quando comparada com uma gaussiana. Em geral, se $\kappa > 3$, temos uma distribuição de cauda mais longa e com pico mais agudo; se $\kappa < 3$ trata-se de uma distribuição de cauda mais curta e com pico mais achatado e quando $\kappa = 3$ a distribuição é tão achatada quanto uma distribuição gaussiana.

Já o coeficiente de assimetria ν é definido por

$$\nu = \frac{\langle (x - \langle x \rangle)^3 \rangle}{\langle (x - \langle x \rangle)^2 \rangle^{\frac{3}{2}}}. \quad (1.8)$$

Ele é uma medida de assimetria da distribuição de probabilidade. De uma maneira geral, se $\nu > 0$ a cauda à direita do valor médio da distribuição é maior, se $\nu < 0$ a cauda à esquerda é a maior e no caso de $\nu = 0$, temos uma distribuição simétrica em relação a seu valor médio.

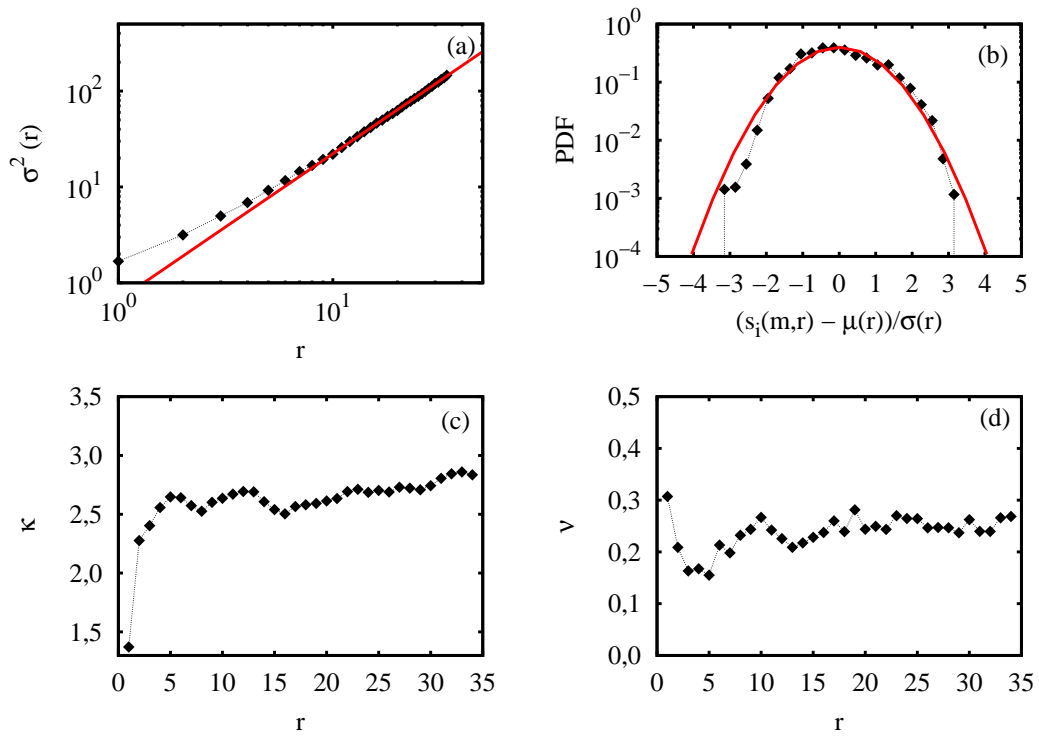


Figura 1.4: Análises estatísticas da liga alemã: (a) variância $\sigma^2(r)$, (b) distribuição de probabilidade dos pontos comparada com uma gaussiana (linha contínua), (c) coeficiente de curtose κ e (d) coeficiente de assimetria ν .

1.4 Um modelo minimalista de equipes idênticas

Na seção anterior, apresentamos algumas características e análises dos dados empíricos. Agora, iremos construir um modelo minimalista fazendo uma aproximação do tipo campo médio. Nesta aproximação, consideraremos que as equipes são idênticas, isto é, em uma partida a probabilidade de cada uma vencer é a mesma. Além disso, vamos supor que os resultados das partidas sejam obtidos por meio de um algoritmo numérico, o qual passaremos a descrever.

Para simular uma partida entre a equipe i e a equipe j , sorteamos dois números aleatórios uniformemente distribuídos no intervalo $[0, 1]$, x_i e x_j . Em seguida, usamos os dois números no seguinte algoritmo:

$$\begin{aligned} &\text{SE } |x_i - x_j| \leq \delta \\ &\quad \text{o jogo termina empatado } (\xi_i = \xi_j = 1); \\ &\text{SE NÃO} \\ &\quad \text{SE } x_i > x_j \tag{1.9} \\ &\quad \quad \text{vence a equipe } i \text{ } (\xi_i = 3 \text{ e } \xi_j = 0); \\ &\quad \text{SE NÃO} \\ &\quad \quad \text{vence a equipe } j \text{ } (\xi_i = 0 \text{ e } \xi_j = 3); \end{aligned}$$

do qual obtemos o resultado da partida. No caso de um torneio em que não ocorram empates, como o basquete, a primeira etapa deste algoritmo deve ser suprimida. Podemos também usar a interpretação de caminhada aleatória da equação (1.1), considerando agora que as variáveis aleatórias ξ_i e ξ_j sejam determinadas pelo algoritmo anterior.

Se estivermos considerando algum tipo de torneio em que não há empates, o modelo acima não possui nenhum parâmetro livre. No caso típico do futebol, sob a hipótese anterior de equipes iguais, nosso modelo contém apenas um parâmetro δ associado com as ocorrências de empate. De fato, se definirmos uma variável

aleatória $\eta = x_i - x_j$, pode-se mostrar que a distribuição de probabilidade dessa variável é dada por

$$p(\eta) = \begin{cases} \eta + 1 & \text{se } -1 \leq \eta < 0 \\ -\eta + 1 & \text{se } 0 \leq \eta \leq 1. \end{cases} \quad (1.10)$$

Sendo assim, a probabilidade de ocorrer empate fica dada por $\int_{-\delta}^{\delta} p(y)dy = \delta(2-\delta)$.

Empregando este procedimento para simular um campeonato inteiro muitas vezes (10^5), podemos obter todas as quantidades analisadas para comparação com os dados empíricos. Para obter os valores ótimos de δ , usamos o método dos mínimos quadrados para minimizar a diferença entre os valores simulados de $\bar{s}(m)$ e os valores observacionais da figura (1.2), com valores incrementais de δ . Os valores obtidos são próximos entre si: $\delta = 0.11$ para a liga alemã e para a liga espanhola A, $\delta = 0.15$ para a liga espanhola B e $\delta = 0.10$ para a liga inglesa.

Nas figuras (1.5) e (1.6), apresentamos uma comparação entre os valores simulados de $\bar{s}(m)$ e $\sigma_s(m)$ e os valores observacionais. Como podemos ver, este modelo minimalista é apenas capaz de descrever qualitativamente o comportamento dos dados empíricos. Observe que a discrepância entre o modelo e os dados é menor para liga espanhola B e maior para a liga inglesa, veja as figuras (1.5b) e (1.5d). Veremos mais adiante que essa discrepância está relacionada com a diferença entre as equipes, a qual é maior na liga inglesa e menor na liga espanhola B.

Por completeza, também comparamos a variância $\sigma^2(r)$, a distribuição de probabilidade e os coeficientes de curtose κ e de assimetria ν , como mostrado na figura (1.7). Contudo, uma vez que este modelo minimalista é como uma aproximação de campo médio, não esperamos que ele seja capaz de descrever o comportamento não-gaussiano presente nos dados. De fato, neste caso, estamos lidando com uma soma de muitas variáveis aleatórias distribuídas identicamente e, sendo assim, o teorema do limite central estabelece que a distribuição deve tender para uma gaus-

siana. Este fato torna-se evidente observando que os coeficientes κ e ν tendem para os valores normais.

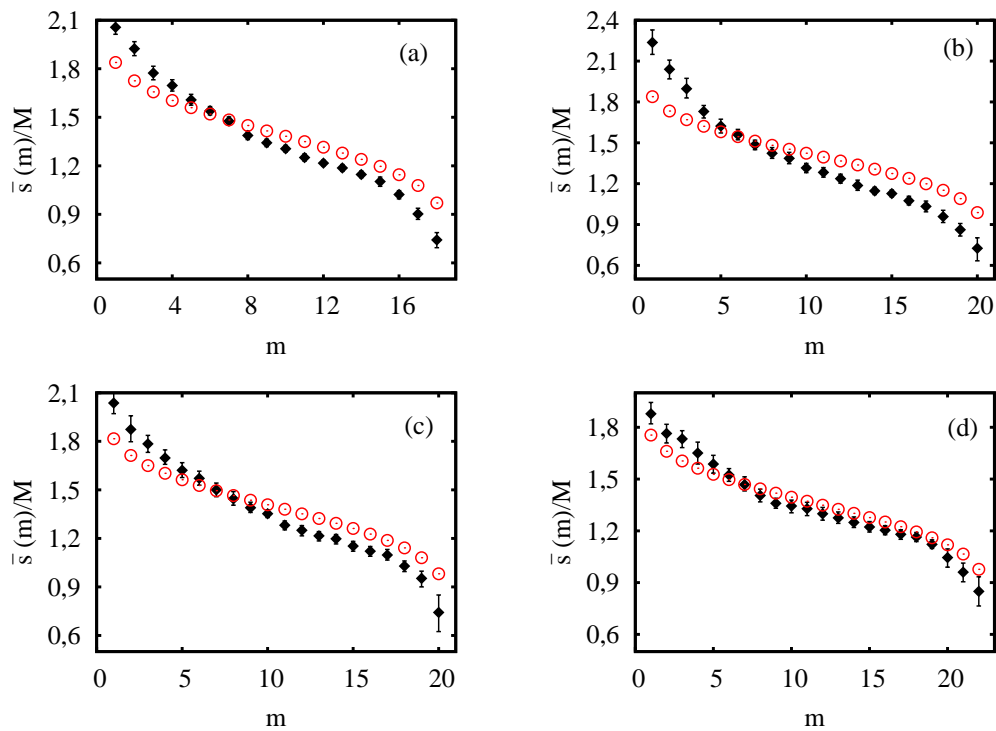


Figura 1.5: Valor médio da pontuação por partida $\bar{s}(m)/M$: comparação entre os dados empíricos (losangos) e o modelo minimalista (círculos) para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

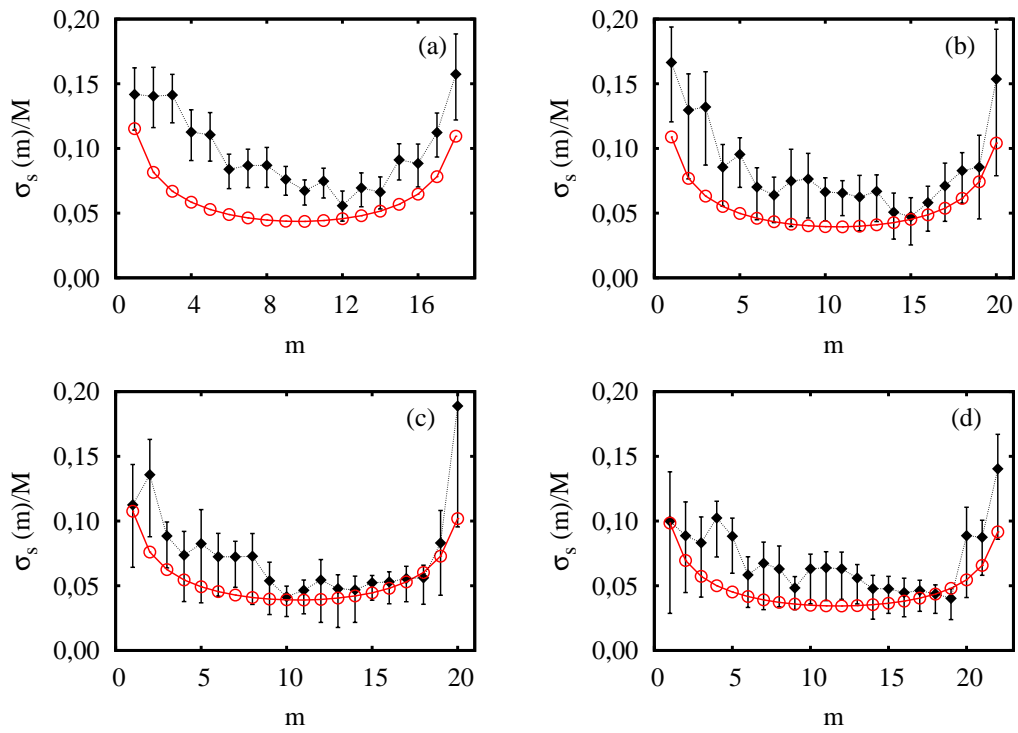


Figura 1.6: Desvio padrão da pontuação por partida $\sigma_s(m)/M$: comparação entre os dados empíricos (losangos) e o modelo minimalista (círculos) para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

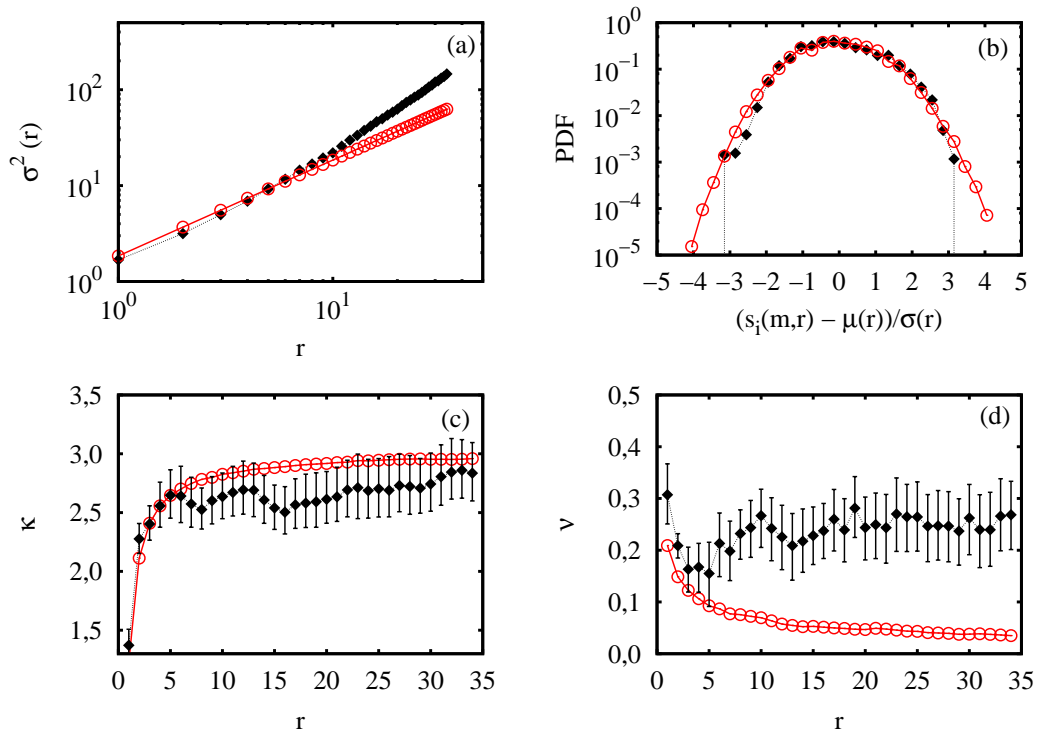


Figura 1.7: Análise estatística da liga alemã: comparação entre dados empíricos (losangos) e o modelo minimalista (círculos) para (a) a variância $\sigma^2(r)$, (b) a distribuição de probabilidade dos pontos, (c) o coeficiente de curtose κ e (d) o coeficiente de assimetria ν .

1.5 Um modelo de equipes não-idênticas

Uma vez que o modelo minimalista (aproximação de campo médio) não foi suficiente para descrever os dados observacionais, precisamos de um modelo com mais ingredientes para tentar reduzir as discrepâncias indicadas na seção anterior. É evidente que uma partida entre duas equipes de futebol (ou outra modalidade) é governada por muitos aspectos, previsíveis e imprevisíveis. Esses aspectos, geralmente, são muito difíceis de serem levados em conta. Dessa maneira, focaremos nossa atenção no que acreditamos ser mais importante: as diferenças entre as equipes. Parece ser de senso comum que as equipes não são idênticas, visto que há um comportamento humano subjacente o qual pode se refletir, por exemplo, em diferentes qualidades defensivas e ofensivas.

Supor que as equipes não sejam idênticas faz com que surja uma questão prática: como modelar essa diferença? Comentaristas de futebol dirão, certamente, que existem dois grupos: as “equipes grandes” e as “equipes pequenas”. Contudo, uma definição descontínua como essa não parece muito razoável e talvez seja mais natural imaginar que as equipes se diferenciem de uma maneira mais contínua. Seguindo por essa direção, vamos supor que as equipes sejam completamente caracterizadas por um único parâmetro Q , um fator de qualidade. Dessa maneira, temos um modelo com muitos parâmetros: um para cada equipe e o parâmetro δ associado aos empates. Um modelo com tantos parâmetros causaria espanto e grandes dificuldades relacionadas com o ajuste de dados. Torna-se, então, mais conveniente empregar uma forma funcional para Q . Nosso palpite, baseado nas formas dos gráficos da figura (1.2), que se assemelham muito ao gráfico da função $f(x) \sim -x|x|^{\alpha-1}$ transladado, é considerar Q , em função da classificação m , com a seguinte forma bastante ajustável

$$Q(m) = 2 + \frac{T - 2m - \epsilon - \beta/2}{T} \left| \left(\frac{T - 2m - \epsilon - \beta/2}{T} \right)^{\alpha-1} \right|, \quad (1.11)$$

em que ϵ é um número pequeno ($\epsilon \ll 1$), $\alpha \geq 0$ e β são parâmetros que ditam a forma de $Q(m)$. Se $\alpha = 0$ as equipes se distribuem em dois grupos $Q_1 = 1$ e $Q_2 = 3$; à medida que α cresce, as equipes se distinguem de uma maneira mais contínua; a função é translada para direita quando $\beta < 0$ e para a esquerda quando $\beta > 0$. Além disso, quando $\beta = 0$, a função $Q(m) - 2$ é ímpar com respeito ao ponto $m = T/2$. Na figura (1.8) ilustramos o comportamento de $Q(m)$ para alguns valores de α e β .

Para introduzir este fator de qualidade na simulação dos resultados das partidas, vamos fazer uma pequena mudança no algoritmo (1.9). Os dois números aleatórios x_i e x_j serão distribuídos uniformemente em intervalos diferentes, $[0, Q(i)]$ e $[0, Q(j)]$, em que $Q(i)$ é o fator de qualidade da equipe i e $Q(j)$ é o análogo para a equipe j . Além disso, visando tornar o parâmetro δ relativo, nós o substituiremos por $\delta' = \delta \frac{Q(i)}{Q(j)}$, com $i > j$. Para resumir as modificações no algoritmo (1.9), podemos reescrevê-lo como

$$\begin{aligned}
x_i &\rightsquigarrow \text{Unif}[0, Q(i)] & x_j &\rightsquigarrow \text{Unif}[0, Q(j)] \\
\text{SE } |x_i - x_j| &\leq \delta \frac{Q(i)}{Q(j)} \\
&\text{o jogo termina empatado } (\xi_i = \xi_j = 1); \\
\text{SE N\~{A}O} \\
&\text{SE } x_i > x_j & & (1.12) \\
&\text{vence a equipe } i \text{ } (\xi_i = 3 \text{ e } \xi_j = 0); \\
\text{SE N\~{A}O} \\
&\text{vence a equipe } j \text{ } (\xi_i = 0 \text{ e } \xi_j = 3).
\end{aligned}$$

Usando o novo algoritmo, simulamos por muitas vezes (10^5) os torneios e variamos os parâmetros do modelo (α , β e δ) visando minimizar, por meio do método dos mínimos quadrados, as diferenças entre os valores simulados de $\bar{s}(m)$ e os valores empíricos. Na tabela (1.1) mostramos os valores dos parâmetros ótimos para todas

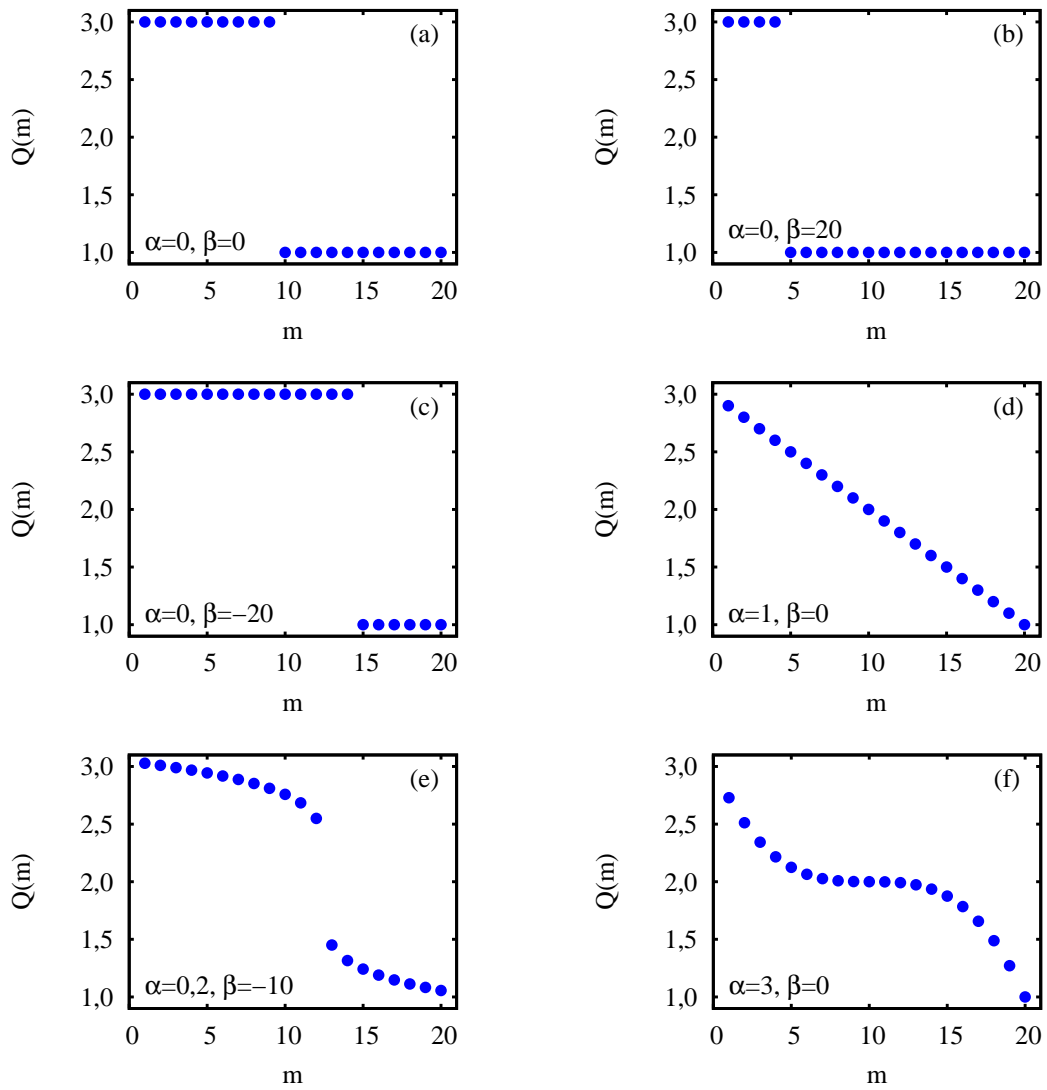


Figura 1.8: Distribuições de qualidade $Q(m)$ para alguns valores de α e β indicados nas figuras.

Tabela 1.1: Valores ótimos dos parâmetros do modelo de equipes não-idênticas.

Torneio	Período	α	β	δ
Liga alemã	1965-2007	2,10	-8,29	0,40
Liga alemã	1997-2007	2,08	-12,00	0,42
Liga inglesa	1995-2007	2,99	-17,29	0,45
Liga espanhola A	1996-2007	2,33	-6,94	0,41
Liga espanhola B	1998-2007	5,18	-3,30	0,40

as ligas investigadas. Observe que, apesar de possíveis mudanças nas propriedades estatísticas durante os 41 anos analisados da liga alemã, quando consideramos somente os dados correspondentes à última década desta liga, os parâmetros α e δ não variam muito.

Com este novo ingrediente no modelo (diferença entre as equipes), somos capazes de descrever corretamente o comportamento de $\bar{s}(m)$, como mostrado na figura (1.9) e também o comportamento da variância $\sigma_s^2(m)$, como podemos ver na figura (1.10).

Além do comportamento estático dos dados, este novo modelo explica muito bem os aspectos do processo difusivo não-usual. De fato, como podemos ver na figura (1.11), o modelo de equipes não-idênticas descreve corretamente o comportamento da variância $\sigma^2(r)$, o coeficiente de curtose κ , o coeficiente de assimetria ν e a distribuição de probabilidade.

Do ajuste dos dados obtivemos os parâmetros α , β e δ ótimos e, conseqüentemente, a forma funcional de $Q(m)$ para cada liga de futebol, como mostrado na figura (1.12). Esta função nos dá informação a respeito da competitividade em cada campeonato. Vimos na seção anterior que o modelo de equipes idênticas descrevia melhor a liga espanhola B, veja figura (1.5d). Agora, notamos que a forma de $Q(m)$ para este campeonato possui o maior número de equipes com aproxima-

damente o mesmo fator de qualidade. Dessa maneira, podemos dizer que a liga espanhola B é o campeonato mais equilibrado que analisamos. Por outro lado, a liga inglesa apresenta a distribuição de qualidade menos equilibrada. Podemos observar que o fator de qualidade das primeiras classificações é substancialmente maior do que os demais, indicando que nesta liga existem algumas equipes que são muito melhores do que as outras. Realmente, no período de 1995 a 2007 da liga inglesa apenas 3 equipes² diferentes foram campeãs; em contraste tivemos 11 campeões³ diferentes no mesmo período da liga espanhola B.

²Manchester United, Arsenal and Chelsea.

³Las Palmas, Sporting B, Cacereño, Getafe, Universidad LPGC, Atlético B, Barakaldo, Universidad LPGC, Pontevedra, Real Madrid B and Pontevedra.

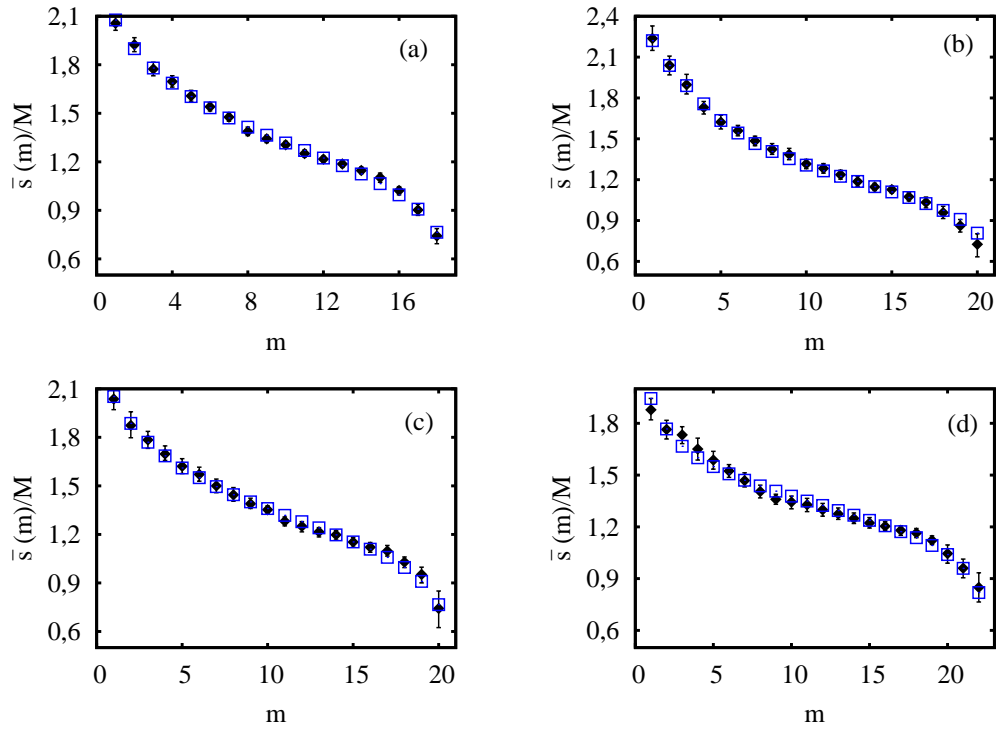


Figura 1.9: Valor médio da pontuação por partida $\bar{s}(m)/M$: comparação entre os dados empíricos (losangos) e o modelo de equipes não-idênticas (quadrados) para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

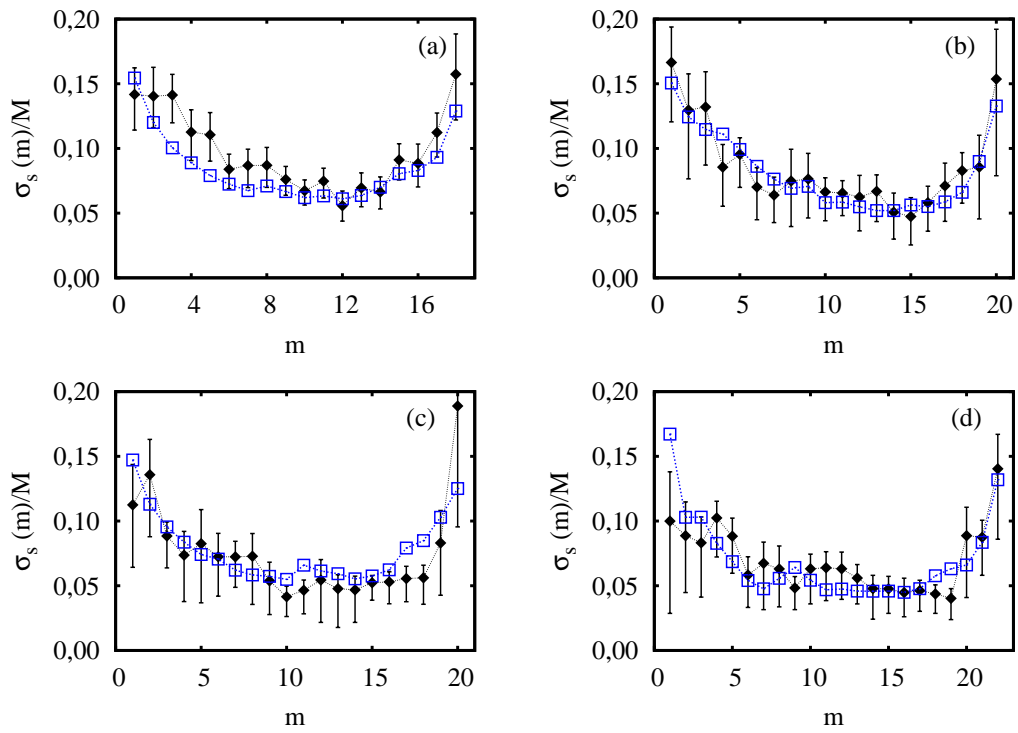


Figura 1.10: Desvio padrão da pontuação por partida $\sigma_s(m)/M$: comparação entre os dados empíricos (losangos) e o modelo de equipes não-identicas (quadrados) para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

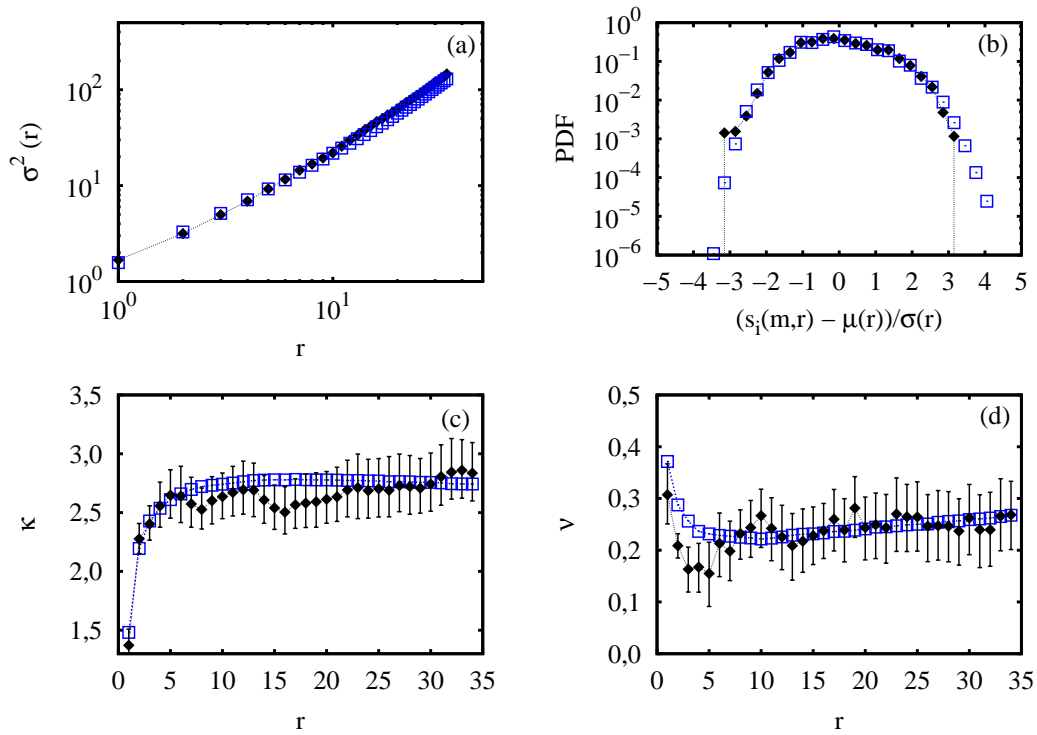


Figura 1.11: Análise estatística da liga alemã: comparação entre dados empíricos (losangos) e o modelo de equipes não-idênticas (quadrados) para (a) a variância $\sigma^2(r)$, (b) a distribuição de probabilidade dos pontos, (c) o coeficiente de curtose κ e (d) o coeficiente de assimetria ν .

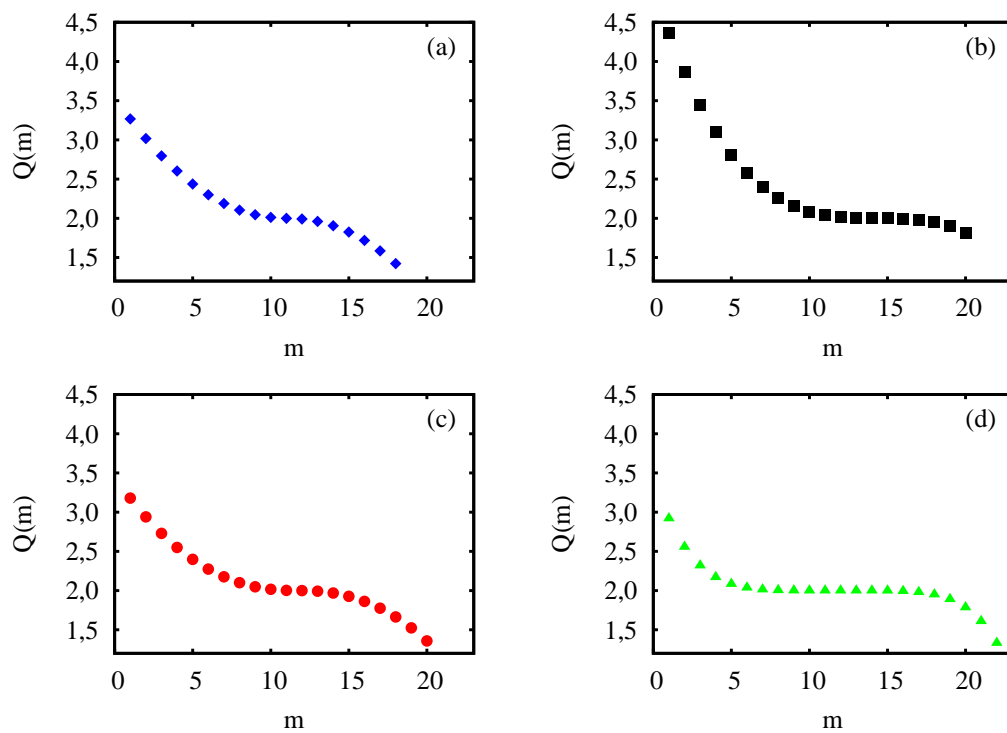


Figura 1.12: Fatores de qualidade ajustados para a liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

1.6 Número de vitórias, empates e derrotas

Até agora, analisamos somente as pontuações das equipes. Uma outra perspectiva seria investigar o número de vitórias, empates e derrotas. Para fazer isto, calculamos essas quantidades usando nosso conjunto de dados empíricos e também os resultados provenientes do modelo de equipes não-idênticas. Estes últimos, obtivemos por meio da simulação de uma temporada completa usando os parâmetros ótimos da tabela (1.1). Para cada simulação, ao final do torneio ($r = M$), contamos o número de vitórias, empates e derrotas de cada classificação m e posteriormente calculamos seus valores médios usando 10^5 realizações do torneio. Na figura (1.13), mostramos uma comparação entre o modelo e os dados empíricos.

Ainda que nosso modelo seja baseado e ajustado usando os valores das pontuações, ele também está em bom acordo com esses novos dados observacionais. Podemos notar que as flutuações são maiores nessas novas variáveis, o que é plausível, visto que as pontuações são construídas como combinações lineares dessas três variáveis. Da figura (1.13), notamos que o número de vitórias e empates possuem uma forma hierárquica bem definida, sendo o número de vitórias maior para as primeiras classificações e menor para as últimas. Contudo, não é claro que exista uma forma hierárquica bem definida para o caso do número de empates, pelo contrário, os dados empíricos sugerem que o número de empates é aproximadamente constante, não dependendo das classificações m .

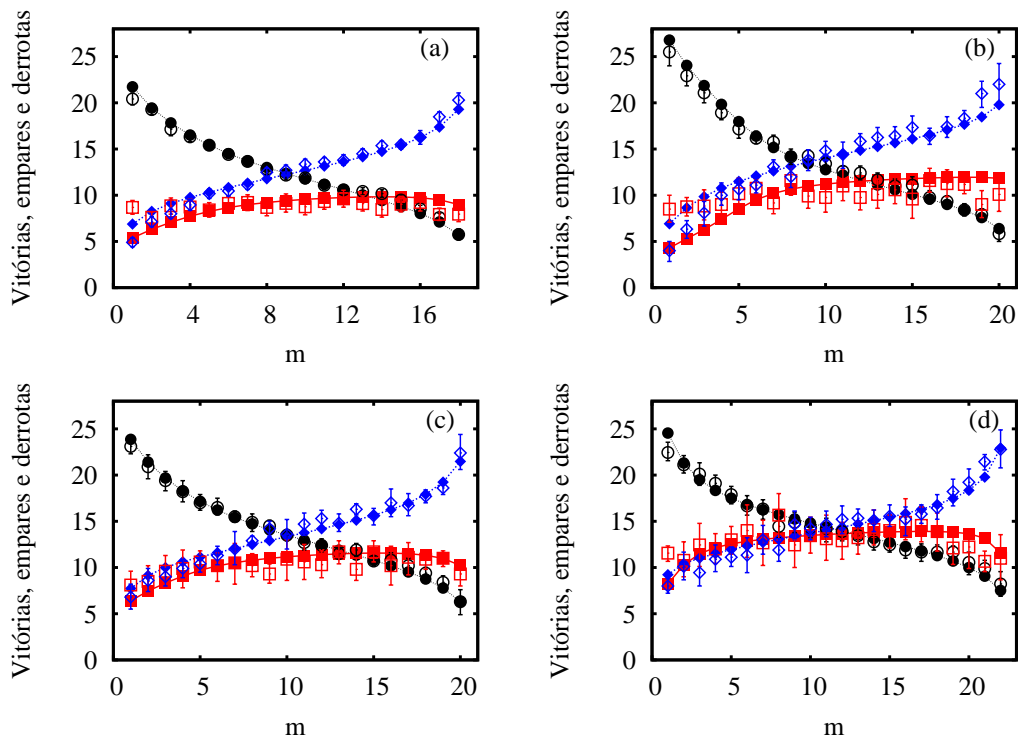


Figura 1.13: Número de vitórias (círculos), empates (quadrados) e derrotas (losangos), em função da classificação, para os dados observacionais (pontos não preenchidos) e para o modelo de equipes não-identicas (pontos preenchidos), no caso da liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

1.7 Comparando dois sistemas de torneio

Entre os fãs de futebol são comuns as discussões sobre qual sistema de torneio é melhor: aqueles em que todas as equipes disputam duas partidas com todas as outras ou aqueles em que existe um emparelhamento inicial e a partir daí as equipes se enfrentam, sendo que somente as vitoriosas seguem na competição. O primeiro deles é conhecido como torneio de pontos corridos, o qual temos como exemplo o Campeonato Brasileiro; o segundo são os torneios eliminatórios ou “mata-mata”, como é o caso das etapas finais da Copa do Mundo de Futebol e da copa Libertadores da América. Nesta seção, vamos dar uma resposta objetiva para esta questão usando nosso modelo de equipes não-idênticas. Para fazer a simulação, tomamos as equipes com os 16 melhores fatores de qualidade, obtidos após os ajustes dos dados experimentais, e realizamos o campeonato 10^6 vezes, contando o número de vezes que a equipe com qualidade $Q(m)$ foi campeã. Obtemos, assim, a probabilidade dessa equipe vencer a competição, seja ela do tipo pontos corridos ou eliminatória. No caso das competições eliminatórias, usamos um emparelhamento inicial no qual as primeiras partidas eram sorteadas. Os gráficos da figura (1.14) mostram os resultados obtidos.

Podemos notar que a probabilidade da equipe com maior fator de qualidade vencer o campeonato é substancialmente maior quando o campeonato é de pontos corridos. Observamos, ainda, que as equipes com fator de qualidade menor possuem maiores probabilidades de serem campeãs quando o campeonato é eliminatório. Estes resultados indicam que os campeonatos eliminatórios apresentam uma maior aleatoriedade, possibilitando que equipes menos qualificadas possam ser campeãs. Por outro lado, os campeonatos de pontos corridos são implacáveis com as equipes de menor qualidade, reduzindo praticamente a zero a probabilidade dessas equipes vencerem a competição.

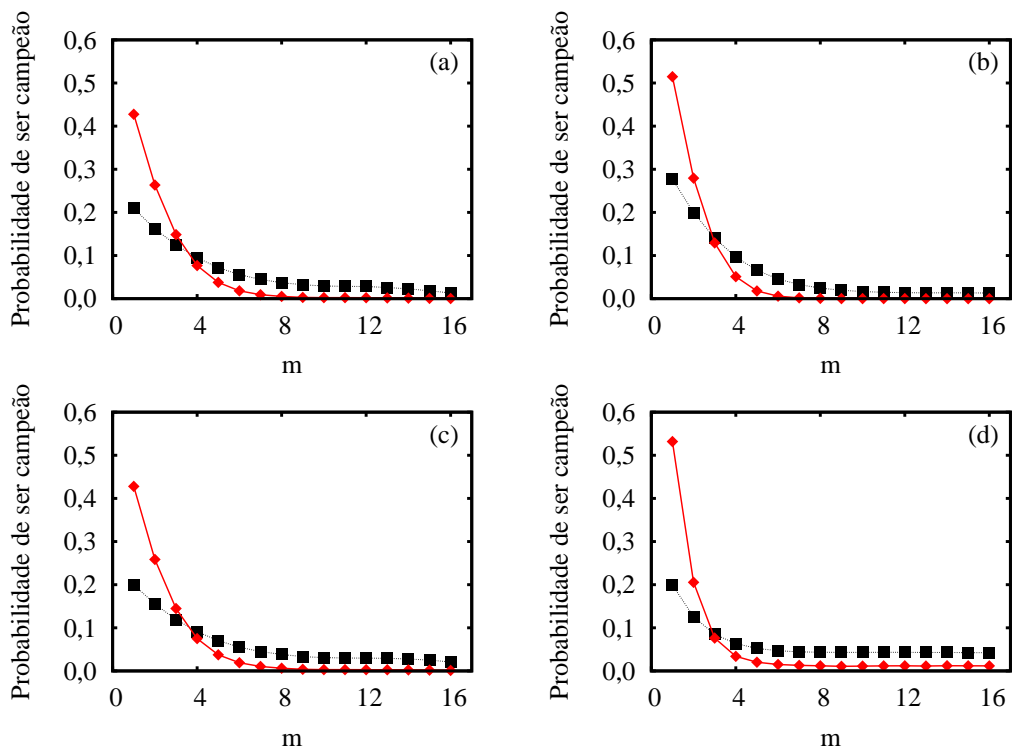


Figura 1.14: Probabilidade da classificação m ser campeã em um torneio de pontos corridos (losangos) e em um torneio eliminatório (quadrados), considerando os parâmetros do modelo de equipes não idênticas no caso da liga (a) alemã, (b) inglesa, (c) espanhola A e (d) espanhola B.

1.8 Equipes não-idênticas e o comportamento não-gaussiano

Vimos na seção 1.5 que o comportamento não-gaussiano presente nos dados observacionais só é corretamente explicado se usarmos o modelo de equipes não-idênticas. Este fato sugere que existe uma relação entre o comportamento não-gaussiano e as diferenças entre as equipes. Para testar essa hipótese, calculamos a variância $\sigma^2(r)$ para um conjunto de parâmetros dos fatores de qualidade, como mostrado na figura (1.15a). Fixando os valores de $\beta = -2$ e $\delta = 0, 4$, executamos nossas simulações variando α no intervalo de 0 a 100. Desta maneira, quando α aumenta, a diferença entre as equipes diminui, como ilustra a figura (1.15b).

Nós já vimos que o processo de difusão usual é caracterizado por $\sigma^2(r) \propto r^\gamma$, com $\gamma = 1$. Para $\gamma < 1$ temos um subdifusão e para $\gamma > 1$ superdifusão. Neste contexto, investigamos a relação γ versus α , como mostrado na figura (1.15c). Estes resultados indicam que, quando as equipes tendem a ser idênticas ($\alpha \gg 1$), o regime de difusão tende para o usual ($\gamma \rightarrow 1$). Por outro lado, quando as equipes se tornam mais diferentes, o regime de difusão se torna mais superdifusivo ($\gamma > 1$).

Investigamos também o papel do número de pontos p , relacionado com a vitória, na cauda assimétrica da distribuição de probabilidade das pontuações, caracterizada pelos valores positivos do coeficiente de assimetria ν . Especificamente, calculamos o valor médio de ν usando 10^5 temporadas, simuladas em função do número de pontos p , o qual fizemos variar no intervalo de 1 a 10. Nestas simulações, usamos o modelo de equipes idênticas e alguns valores de δ , como indicado na figura (1.15d). Essa figura indica que diferentes intervalos entre os pontos relacionados aos possíveis resultados causam a assimetria nas caudas. No caso mais simétrico, isto é, nenhum ponto para derrota, 1 ponto para empate e 2 pontos para vitória, o coeficiente ν é aproximadamente zero para todos valores de δ . Observe que

quando p aumenta, ν também aumenta. Contudo, para p grande comparado com 1 (ponto relacionado ao empate) temos um comportamento assintótico constante. Isto ocorre, porque valores grandes de p dominam as pontuações e, consequentemente, a assimetria satura.

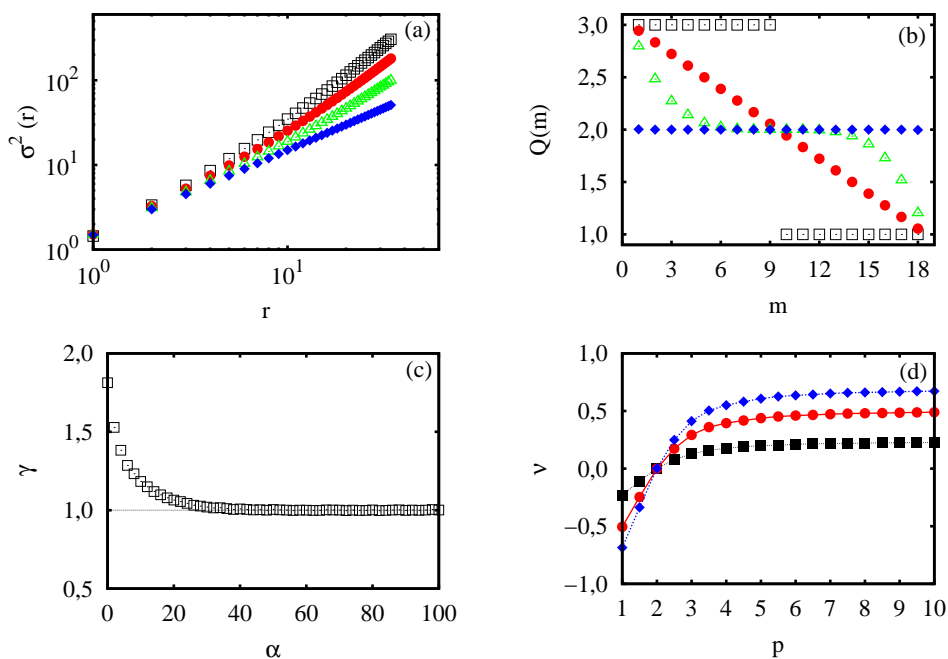


Figura 1.15: Análise do modelo: (a) variância $\sigma^2(r)$ para valores crescentes de α (de cima para baixo: 0, 1, 4 e 100) com $\beta = -2$ e $\delta = 0, 4$; (b) formas de $Q(m)$ para $\alpha = 0$ (quadrados), $\alpha = 1$ (círculos), $\alpha = 4$ (triângulos) e $\alpha = 100$ (losangos), com $\beta = 0$; (c) γ versus α com $\beta = -2$ e $\delta = 0, 4$; (d) valor médio do coeficiente de assimetria ν versus número de pontos p , para $\delta = 0, 2$ (quadrados), $\delta = 0, 4$ (círculos) e $\delta = 0, 6$ (losangos).

Capítulo 2

Sequências Simbólicas

A exemplo do capítulo anterior, empregaremos conceitos de caminhadas aleatórias para discutir algumas propriedades de sequências simbólicas geradas por meio de um experimento computacional. Veremos que uma entropia não-extensiva pode ser relacionada a essas sequências e que elas apresentam um cenário difusivo rico, indo de usual a anômalo[20].

2.1 Interações de longo alcance

É bastante comum a utilização de potenciais truncados em simulações computacionais, ou seja, potenciais que a partir de uma certa distância b são considerados nulos. Em geral, um potencial interatômico truncado satisfaz três condições:

$$u(r) = 0 \quad \text{se } r > b, \quad (2.1)$$

$$\lim_{r \rightarrow a} u(r) = \infty, \quad (2.2)$$

$$u(r) \geq -\varepsilon. \quad (2.3)$$

A primeira delas é a condição que de fato limita o alcance da interação, a segunda condição define uma distância máxima a de aproximação entre as partículas e a

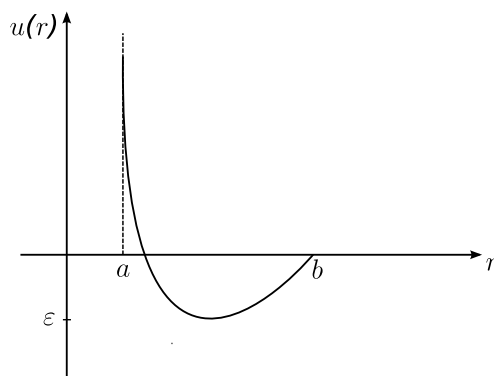


Figura 2.1: Potencial truncado típico de interação entre moléculas.

última condição assegura que a atração entre as partículas é finita. Uma possível forma de potencial que satisfaz estas condições é mostrada na figura (2.1).

Para esses potenciais, o teorema de Yang-Lee [25] garante a existência do limite termodinâmico e, para sistemas que possuem interações desse tipo, a Mecânica Estatística e a Termodinâmica usuais são válidas.

Sabe-se que a condição (2.1) pode ser menos restritiva, desde que o potencial vá a zero suficientemente rápido para $r > b$, mais especificamente,

$$u(r) \leq Ar^{-\lambda} \quad (\lambda > d), \quad (2.4)$$

em que d é a dimensão espacial do sistema (no caso usual $d = 3$). Um potencial muito conhecido e utilizado em simulações, e que satisfaz essa condição, é o potencial de Lennard-Jones. Este potencial possui, para grandes distâncias, uma parte atrativa devido às forças de Van der Waals (a qual vai a zero com r^{-6}) e uma parte repulsiva para pequenas distâncias devido ao princípio de exclusão de Pauli (a qual varia com r^{-12}).

Contudo, interações fundamentais da natureza como a eletromagnética e a gravitacional não satisfazem todas essas condições. Para estas duas interações, temos

$$u(r) \propto \frac{1}{r}, \quad (2.5)$$

ou seja, $\lambda = 1$, não satisfazendo a condição (2.4). Existem para interações desse tipo grandes dificuldades relacionadas à formulação padrão da Termodinâmica, em particular, a respeito da extensividade da entropia e da convergência do limite termodinâmico. No caso do limite termodinâmico, quando temos interações de curto alcance, a energia livre cresce linearmente com o número de partículas, isto é, $F \propto N$ e desta maneira o limite termodinâmico $\lim_{N \rightarrow \infty} \frac{F}{N}$ tende a um valor finito não-nulo. Porém, quando as interações são de longo alcance, a energia livre pode crescer não linearmente com o número de partículas, tal como, $F \propto N^\eta$ (com $\eta > 1$) e assim o limite termodinâmico $\lim_{N \rightarrow \infty} \frac{F}{N}$ diverge. Demonstrou-se que estes sistemas são não-extensivos[26], não-ergódicos[27] e que apresentam características muito particulares como, por exemplo, calor específico negativo[28], metaestabilidades[29] e distribuição de velocidades não-gaussiana no equilíbrio[30].

Existem também outros tipos de sistemas não relacionados com partículas e movimentos, no sentido mais clássico dessas palavras, que apresentam interações ou, mais apropriadamente, correlações de longo alcance. Como exemplos de sistemas que apresentam tal característica, dos muitos existentes, podemos citar sequências de DNA[31] e textos literários[32]. Visando investigar sistemas que podem estar relacionados a esses, na próxima seção apresentaremos um modelo capaz de gerar sequências de símbolos com correlações de longo alcance.

2.2 Gerando sequências correlacionadas

Apresentaremos um modelo, proposto originalmente por M. Buiatti *et al.*[19], o qual permite a construção de cadeias de símbolos correlacionados. Veremos que trata-se de um experimento numérico simples, mas que é equivalente a processos de geração de sequências com propriedades semelhantes as de DNA[33, 34] e que também pode ser interpretado como um gás de rede[35].

O experimento numérico consiste, essencialmente, de uma rede unidimensional com N sítios. Há uma variável associada a cada sítio que assume unicamente valores discretos ou símbolos como, por exemplo, 0 ou 1. Em outras palavras, esta rede é a sequência simbólica. O procedimento considera dois números aleatórios: um deles, x , possui distribuição de probabilidade uniforme no intervalo $[0,1]$ e o outro emerge da expressão

$$y = A \left[\frac{1}{(1-x)^{1/(\mu-1)}} - 1 \right], \quad (2.6)$$

sendo $A > 0$ e $\mu > 1$. Percorremos toda a rede (sequência) preenchendo $N_y = [y]+1$ sítios com o mesmo símbolo z , em que $[y]$ denota a parte inteira de y e z é um gerador de sinais (valores) discretos e equiprováveis. Por exemplo,

$$Q = \left\{ \underbrace{0, 0, 0}_{N_y=3}, \underbrace{1, 1, 1, 1, 1}_{N_y=5}, \underbrace{0, 0}_{N_y=2}, \underbrace{1, 1, 1, 1}_{N_y=4}, \underbrace{0}_{N_y=1} \right\}.$$

pode representar uma cadeia simbólica de dois símbolos (0,1). É interessante escrever a distribuição de probabilidade da variável y ,

$$p(y) = (\mu - 1) \frac{A^{\mu-1}}{(A+y)^\mu} \quad (A > 0, \mu > 1) \quad (2.7)$$

e calcular o primeiro e o segundo momento, isto é,

$$\langle y \rangle = \frac{A}{(\mu - 2)} \quad \text{e} \quad \langle y^2 \rangle = \frac{2A^2}{(\mu - 2)(\mu - 3)}. \quad (2.8)$$

Com isso, podemos notar que o primeiro momento é divergente quando $\mu \rightarrow 2$, consistentemente com o fato de que para valores de $\mu \leq 2$ a integral $\int_0^\infty yp(y)dy$ não converge, de modo que para $\mu \leq 2$ o primeiro momento não está definido. Assim, quando μ apresentar valores próximos a 2, N_y poderá assumir valores grandes, preenchendo uma parte significativa da sequência com um mesmo símbolo. Por outro lado, valores de μ maiores e distantes de 2 tornam raros valores grandes de N_y , fazendo com que exista uma alternância maior de símbolos. Note ainda, que somente para $\mu > 3$ o segundo existe e é finito.

2.3 Entropia de sequências simbólicas

Uma vez apresentado o procedimento de construção das sequências, podemos investigar as possíveis correlações existentes entre os símbolos. Para isso, usaremos um outro procedimento que permite calcular a entropia dessas sequências simbólicas.

O procedimento inicia-se com o cálculo das probabilidades de todas as possíveis configurações de símbolos presentes em uma sequência de tamanho N , quando a percorremos usando janelas de comprimento L . Numericamente, fixamos uma janela de comprimento L para ser movida ao longo da cadeia. Ao movê-la, contamos quantas vezes observamos uma dada configuração i , definindo a sua probabilidade p_i de ocorrer na cadeia. Por exemplo, suponha que tenhamos a seguinte cadeia para analisar:

$$Q = \{0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1\} .$$

Escolhemos uma janela de tamanho 2 e a movemos através da cadeia. Esquemáticamente temos:

$$Q = \left\{ \underbrace{0, 0}_1, \underbrace{1, 0}_2, \underbrace{0, 1}_3, \underbrace{1, 1}_4, \underbrace{0, 0}_5, \underbrace{0, 1}_6, \underbrace{0, 0}_7, \underbrace{0, 0}_8, \underbrace{1, 0}_9, \underbrace{1, 1}_{10}, \underbrace{1, 1}_{11}, \underbrace{0, 0}_{12}, \underbrace{1, 0}_{13}, \underbrace{0, 0}_{14}, \underbrace{1, 1}_{15} \right\} ,$$

em que o número abaixo das chaves indica como a janela está se movendo. O próximo passo é contar o número de vezes que uma dada configuração ocorreu. Por exemplo, a configuração $\{0, 0\}$ ocorreu nos “instantes” 1, 5, 7, 8, 12 e 14; assim ela ocorreu 6 vezes o que lhe dá uma probabilidade igual a $6/15$. Procedendo de forma análoga calculamos as probabilidades das demais configurações e também para outros tamanhos de janela.

De posse das probabilidades, podemos calcular a entropia da sequência. Para tal, consideraremos a forma entrópica proposta por Tsallis, que é definida para um

sistema com W microestados e probabilidade de ocupação p_i como

$$S_q = \frac{1 - \sum_{i=1}^W p_i^q}{q - 1}, \quad (2.9)$$

sendo q um parâmetro real. Note que no limite de $q \rightarrow 1$ recuperamos a consagrada forma entrópica de Boltzmann-Gibbs, isto é,

$$S_1 = S_{BG} = - \sum_{i=1}^W p_i \ln p_i = \ln W, \quad (2.10)$$

em que usamos na última passagem a equiprobabilidade dos microestados, ou seja, $p_i = 1/W$.

É de importância fundamental, do ponto de vista termodinâmico, que S_{BG} satisfaça algumas propriedades matemáticas. Entre elas, a aditividade ($S_{BG}(A + B) = S_{BG}(A) + S_{BG}(B)$) e a extensividade (S_{BG} cresce linearmente com o número de subsistemas). A entropia de Tsallis é, para sistemas não-correlacionados, não-aditiva e não-extensiva, surgindo deste fato o nome *Mecânica Estatística Não-Extensiva*. Porém, veremos mais adiante que quando existem correlações, S_q pode se tornar extensiva para um valor particular do parâmetro q , com $q \neq 1$. A forma entrópica de Tsallis e o formalismo não-extensivo como um todo têm sido empregados com sucesso na descrição de fenômenos complexos. Em particular, eles têm se mostrado bastante úteis no tratamento de sistemas que apresentam comportamentos do tipo leis de potência[36], sistemas no limiar do caos[37], turbulência[38] e sistemas biológicos[39].

Para testar a extensividade, calcularemos a entropia S_q para cada tamanho L das janelas. Assim, podemos reescrever a expressão (2.9) como função de L :

$$S_q(L) = \frac{1 - \sum_{i=1}^W p_i(L)^q}{q - 1}. \quad (2.11)$$

Observe que existem s^L configurações possíveis para uma cadeia de s símbolos.

Para ganhar intuição com a forma entrópica de Tsallis, consideremos um caso particular mais simples, no qual as configurações para um dado tamanho L tenham

igual probabilidade de ocorrer, isto é,

$$p_i(L) = \frac{1}{s^L}. \quad (2.12)$$

Neste caso, podemos calcular $S_q(L)$ analiticamente substituindo este resultado na equação (2.11), de onde obtemos

$$S_q(L) = \begin{cases} \frac{1-s^{-L(q-1)}}{q-1} & \text{se } q \neq 1 \\ L \ln(s) & \text{se } q = 1 \end{cases}. \quad (2.13)$$

Note que, neste caso, S_q cresce linearmente com L (é extensiva) somente para $q = 1$. Os gráficos da figura (2.2) mostram o comportamento de S_q versus L para alguns valores de q e de s .

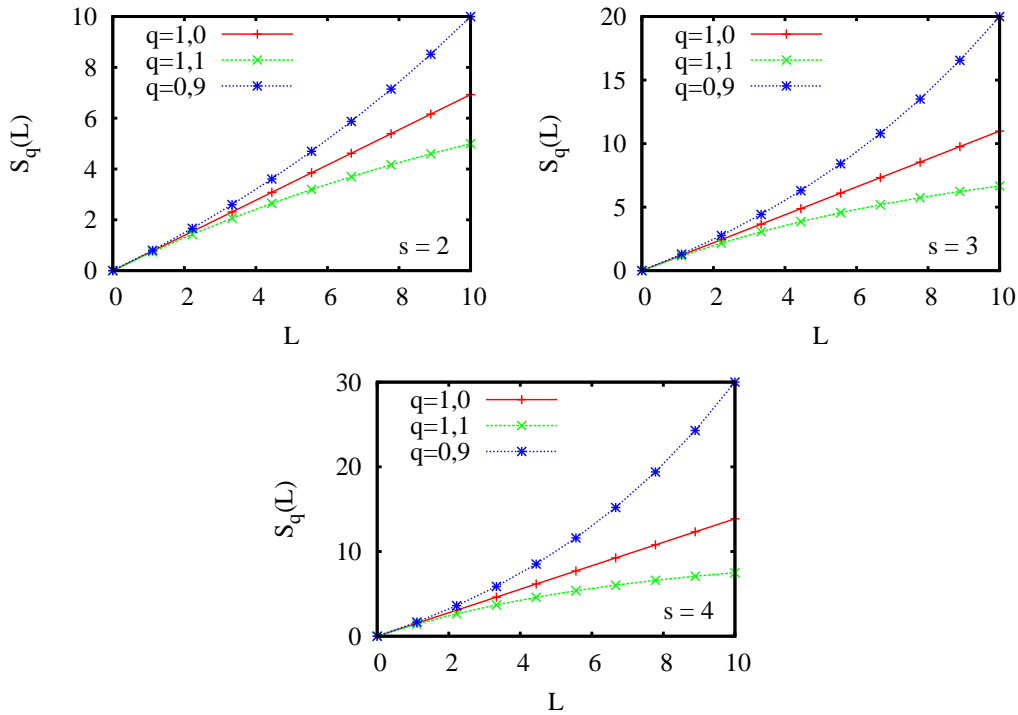


Figura 2.2: Entropia S_q versus L para alguns valores de q e s indicados nos gráficos, assumindo que as todas as configurações têm igual probabilidade de ocorrer.

Passaremos agora a apresentar os resultados obtidos via execução do experimento computacional descrito anteriormente. Os gráficos à esquerda na figura (2.3) mostram o comportamento de S_q versus L para alguns valores de μ e q no caso de uma sequência com dois símbolos. Observe que para cada valor de μ existe somente um valor que $q = q^*$ para o qual a relação S_q versus L é linear. Este fato torna-se evidente quando calculamos o coeficiente de correlação linear de Pearson[40], o qual é definido para um conjunto de n variáveis discretas x_i e y_i , como

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \quad (2.14)$$

sendo

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad \sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{e} \quad \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2,$$

em que \bar{x} e \bar{y} representam os valores médios de x_i e y_i . Esse coeficiente indica o grau da dependência linear entre as variáveis x_i e y_i , quando mais próximo de 1 ou -1 ele for, mais acentuada será a correlação linear entre as variáveis.

Os gráficos contidos à direita na figura (2.3) mostram o coeficiente de Pearson em função do índice entrópico q . Como dissemos anteriormente, existe um valor particular de q , $q = q^*$, que maximiza esse coeficiente e, portanto, faz a relação S_{q^*} versus N o mais linear possível, tornando a entropia S_{q^*} de Tsallis extensiva.

Usando o coeficiente de Pearson para determinar q^* , investigamos a relação entre q^* e μ para sequências com dois, três e quatro símbolos, como mostra a figura (2.4a). Observamos que quanto mais símbolos existem na sequência, mais rapidamente q^* tende a unidade. A razão deste comportamento está relacionada ao fato de que os símbolos ocorrem com igual probabilidade dentro de cada janela N_y . Dessa maneira, um número maior de símbolos implica numa maior alternância de símbolos na sequência, o que a torna mais aleatória, conduzindo à região de aplicabilidade do formalismo usual de Boltzmann-Gibbs .

Por completeza, investigamos também o papel do parâmetro A na relação q^* versus μ , como mostrado na figura (2.4b). Podemos ver que para valores pequenos de A , q^* tende mais rápido à unidade. Se expandirmos a expressão (2.7) para grande valores de y , veremos que $p(y) \propto A^{\mu-1}y^{-\mu}$. Assim, a cauda da distribuição $p(y)$ será menos significativa quanto menor for o valor de A . Com a cauda menos significativa, N_y assume valores menores, o que gera uma maior alternância de símbolos e, conseqüentemente, uma sequência mais aleatória.

Uma vez exposto este comportamento, devemos enfatizar que a entropia Boltzmann-Gibbs não é extensiva para valores pequenos de μ ($\mu < 3$). Nesta região de valores de μ , temos grandes correlações entre os símbolos o que impede a aplicação do formalismo usual. Por outro lado, quando introduzimos correlações, a entropia de Tsallis, a qual não é extensiva sob a hipótese de aleatoriedade completa, torna-se extensiva para uma escolha particular de q , $q = q^*$.

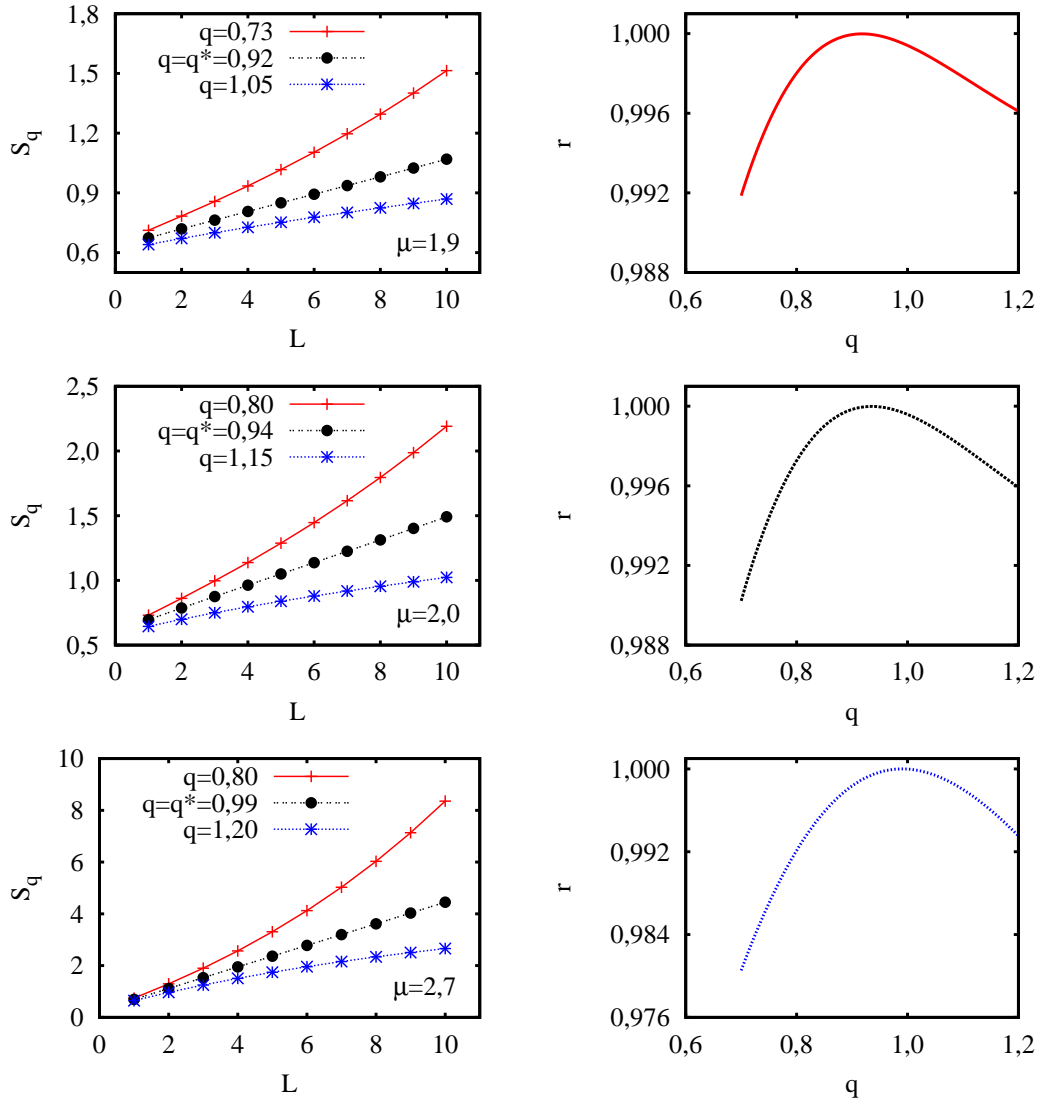


Figura 2.3: À esquerda: entropia S_q versus L para alguns valores de q e s indicados nos gráficos. Os parâmetro da simulação foram: $A = 2$ e $N = 10^8$. À direita: coeficiente de correlação linear de Pearson r versus q .

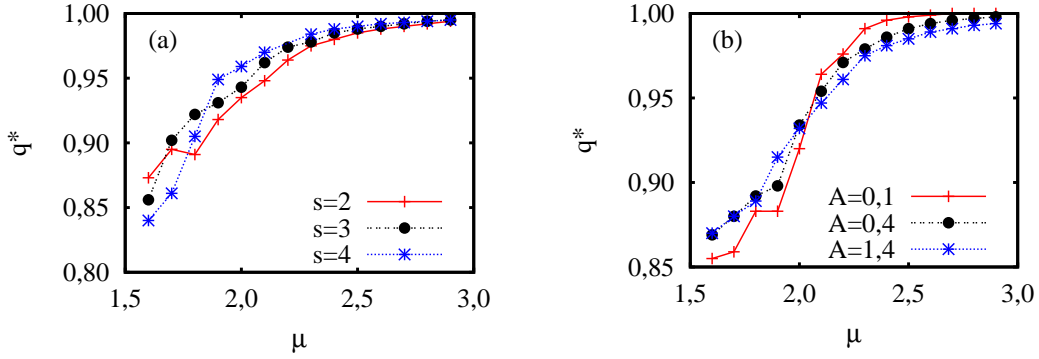


Figura 2.4: Relação entre o índice entrópico q^* e o parâmetro μ : (a) para sequências com 2, 3 e 4 símbolos, fixando-se $A = 2$ e $N = 10^8$; e (b) para sequências com 2 símbolos e $N = 10^8$, variando-se A como indicado da figura.

2.4 Difusão em sequências simbólicas

Além das análises entrópicas apresentadas na seção anterior, podemos fazer uma outra baseada em difusão se considerarmos as sequências simbólicas como trajetórias erráticas. Com esse propósito, consideremos a variável

$$\xi(n) = \sum_{i=1}^n Q_i, \quad (2.15)$$

sendo Q_i o símbolo presente na i -ésima posição da sequência. Por simplicidade, vamos assumir que a sequência tenha apenas dois símbolos: $Q_i = \pm 1$. Esta nova variável representa uma caminhada aleatória, na qual n faz o papel do tempo. Na figura (2.5), mostramos várias realizações de $\xi(n)$ para três valores de μ com $A = 1$. Observe que as trajetórias são notoriamente diferentes, dependendo do valor de μ . Já vimos que valores pequenos de μ conduzem a valores grandes de N_y , o que faz surgir os grandes saltos nas trajetórias da figura (2.5) para $\mu = 2,2$ e $\mu = 2,4$. Por outro lado, para $\mu = 3$ esses saltos grandes não ocorrem. Em busca de caracterizar o processo difusivo, iremos analisar a dependência da variância

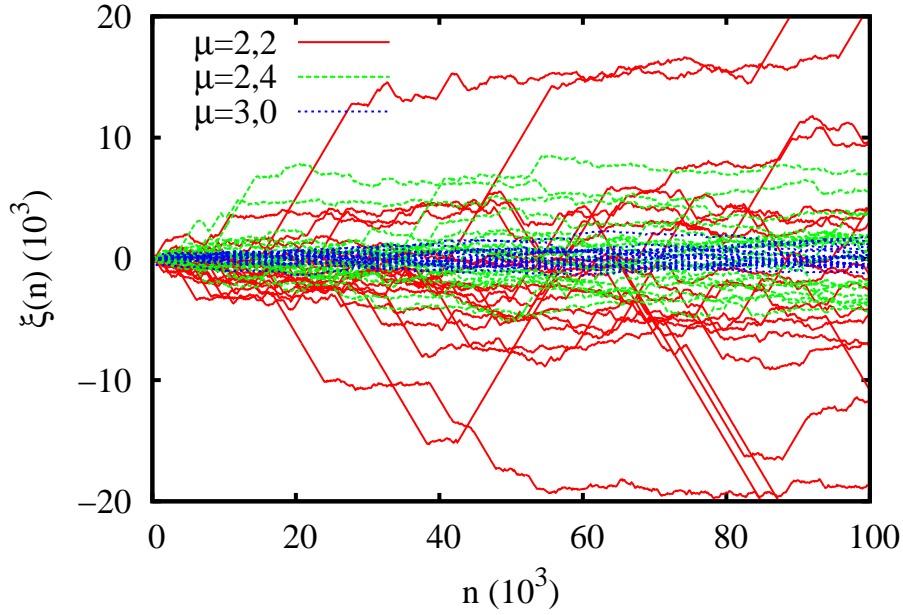


Figura 2.5: $\xi(n)$ para três valores de μ mostrados na figura.

$\sigma^2(n) = \langle (\xi(n) - \langle \xi(n) \rangle)^2 \rangle$ com o tempo n , em que $\langle \dots \rangle$ denota o valor médio sob um conjunto de realizações. Assim, calculamos $\sigma^2(n)$ para alguns valores de μ como mostrado na figura (2.6a). Relembrando que $\sigma^2(n) \propto n^\gamma$ com $\gamma = 1$ indica difusão usual, $\gamma < 1$ subdifusão e $\gamma > 1$ superdifusão.

Note que a inclinação da reta na figura (2.6a) é numericamente igual ao expoente γ , observe também que essa inclinação varia com μ . Por esse motivo, investigamos a relação γ versus μ , como mostrado na figura (2.6b). Os resultados sugerem três regimes de difusão: (i) para $\mu < 2$ temos difusão balística; (ii) para $2 < \mu < 3$ temos superdifusão e (iii) para $\mu > 3$ recuperamos o movimento browniano.

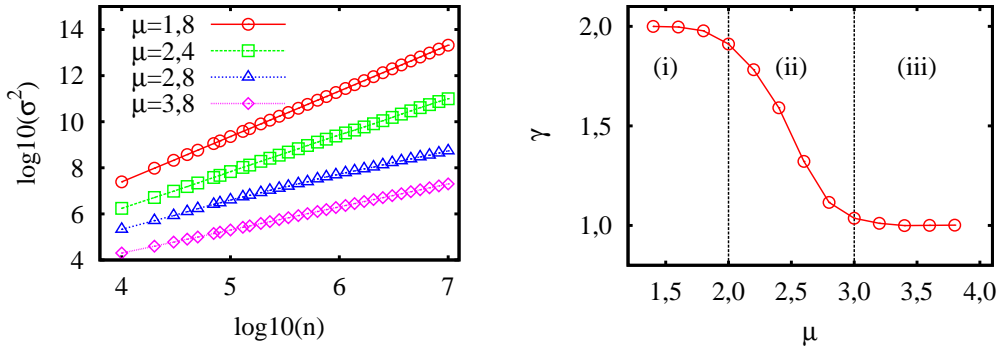


Figura 2.6: (a) Logaritmo da variância $\sigma^2(r)$ versus logaritmo de n para alguns valores de μ indicados na figura. (b) O expoente γ que caracteriza a difusão versus μ . Nesses resultados, usamos um conjunto de 2×10^8 realizações e $A = 1$.

2.5 Leis de escala e distribuição de probabilidade

Quando trabalhamos com processos estocásticos, uma questão natural é sabermos para qual distribuição de probabilidade o mesmo converge. Já vimos que para $\mu > 3$ o processo estocástico descrito pela equação (2.15) converge para um processo de Wiener. Nesse limite, nossa análise entrópica indica que não existem correlações na sequência. Dessa maneira, a equação (2.15) é uma soma de n variáveis aleatórias suficientemente independentes e, para este caso, o teorema do limite central assegura que a distribuição de probabilidade será uma gaussiana. Contudo, para $\mu < 3$, a análise entrópica também revelou que a sequência apresenta correlações de longo alcance caracterizada pelo lento crescimento da entropia de Boltzmann-Gibbs (mais lento do que o de uma função linear). Nesta região, o teorema do limite central não se aplica e surge uma questão: qual a forma da distribuição de probabilidade neste limite?

Para tentar responder esta questão, consideremos novamente a soma na equação

(2.15), mas agora escalada por um fator diferente:

$$\xi(n) = \frac{1}{n^\lambda} \sum_{i=1}^n Q_i, \quad (2.16)$$

em que λ é um fator de escala escolhido para revelar uma possível invariância de escala. Para determinar λ , calculamos a distribuição de probabilidade $p(\xi)$ para diferentes valores de n e variamos λ no intervalo $[0, 2]$. O valor ótimo de λ é aquele que faz o colapso das distribuições de probabilidades, isto é, minimiza a diferença entre as distribuições. Na figura (2.7), mostramos $p(\xi)$ para diversos valores de μ e três valores de n . Observe que há um bom colapso das distribuições, indicando a existência de invariância de escala nos dados. Investigamos também a dependência de λ com o parâmetro μ , como mostrado na figura (2.8a). Note que λ é aproximadamente $1/2$ para $\mu > 3$ e que quando μ diminui o fator de escala aumenta.

Analisando as distribuições da figura (2.7), podemos observar que para $\mu < 3$ um perfil não-gaussiano surge, caracterizado por uma distribuição de cauda mais longa e consistente com a não existência do segundo momento $\langle y^2 \rangle$ (equação 2.8). Visando promover um ajuste a esses dados, uma forma bastante robusta, capaz de descrever o comportamento leptocúrtico e que também é compatível com a existência de segundo momento é a q -gaussiana[17], dada por

$$p(x) = a \left(1 + (q-1) \frac{x^2}{b} \right)^{-\frac{1}{q-1}}, \quad (2.17)$$

em que a é um parâmetro de normalização, b está relacionado com a variância e q é o parâmetro que dita o comportamento de cauda da distribuição. Essa distribuição surge no contexto da Mecânica Estatística Não-Extensiva por meio do processo de maximização da entropia de Tsallis e tem sido empregada no estudo de muitos sistemas fortemente interagentes, como por exemplo, vidros de spin[41] e plasma[42]. Além disso, note que quando $q \rightarrow 1$ recuperamos a distribuição gaussiana.

Assim, ajustamos a q -gaussiana da equação anterior aos histogramas e os resultados são mostrados na figura (2.7), de onde podemos ver um bom acordo com os dados. Analisamos também a relação entre q e μ , como mostrado na figura (2.8b). Podemos observar que q tende a unidade quando μ aumenta, em acordo com o fato de que o processo estocástico descrito pela equação (2.16) converge para um processo de Wiener. Por outro lado, quando μ diminui, q aumenta, evidenciando os fatos de que as sequências apresentam correlações de longo alcance e que a distribuição de probabilidade não é gaussiana nesta região.

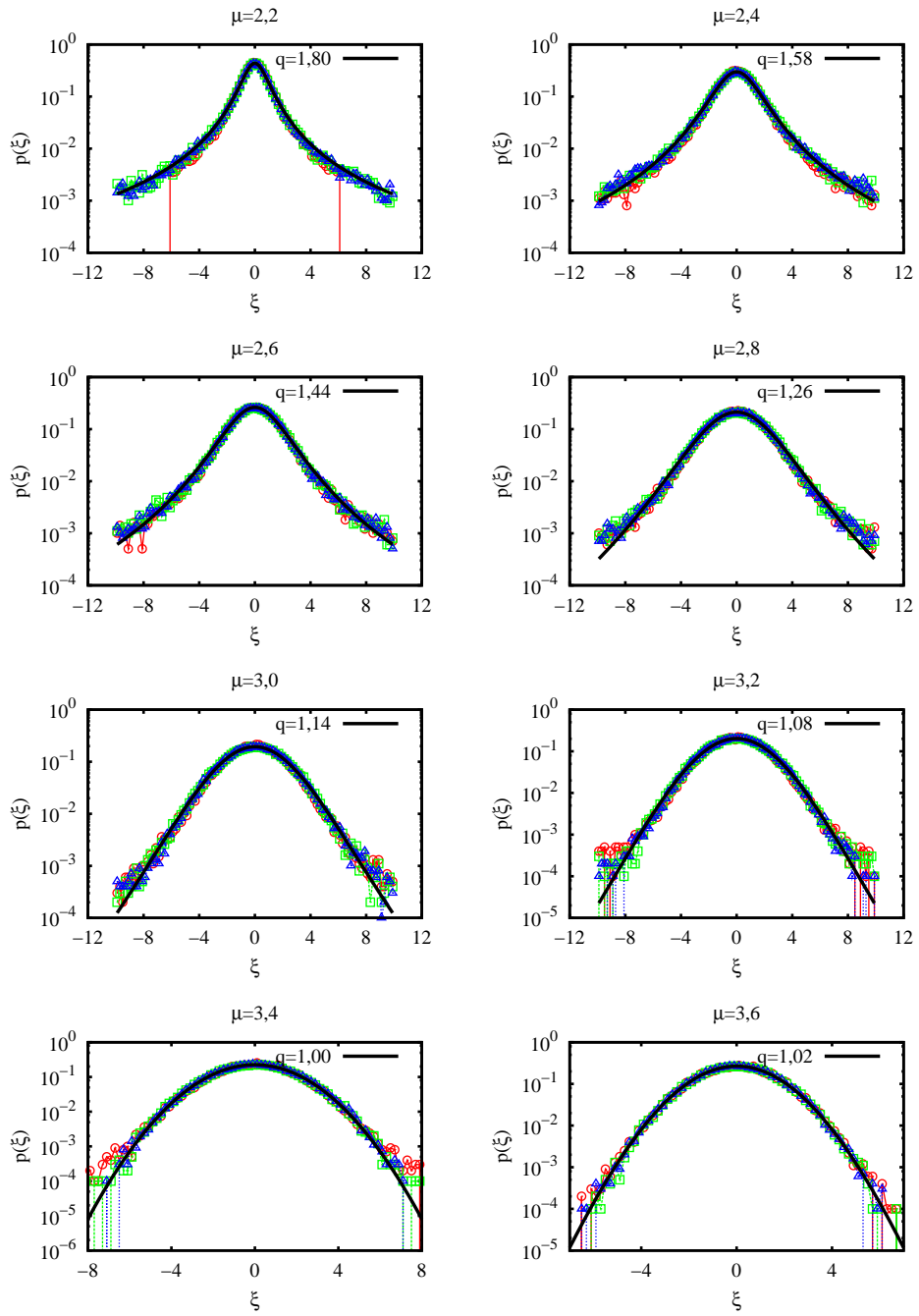


Figura 2.7: Distribuição de probabilidade $p(\xi)$ para $n = 10^4$ (círculos), 5×10^6 (quadrados) e 10^7 (triângulos) para alguns valores de μ . A linha contínua representa uma q -gaussiana, com os valores de q indicados nas figuras.

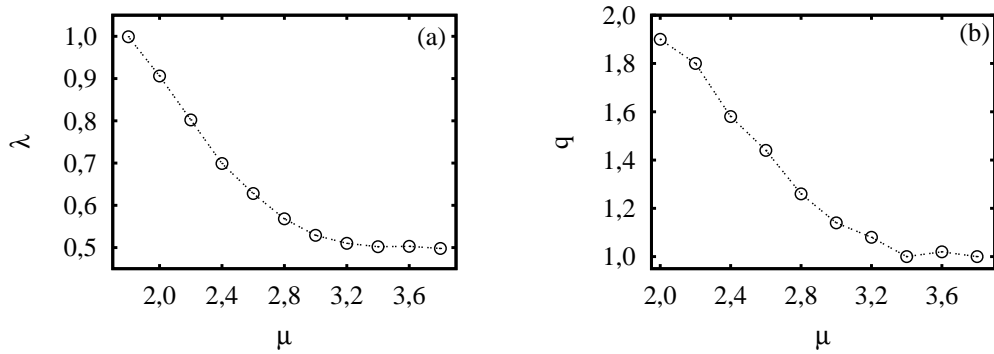


Figura 2.8: (a) O fator de escala λ versus μ . (b) O parâmetro q da q -gaussiana versus μ .

Conclusões

Estudamos nesta dissertação dois sistemas distintos, mas que apresentam características comuns no que diz respeito à aplicabilidade da Mecânica Estatística usual. Nossas investigações indicaram que em ambos os casos as hipóteses da mecânica estatística de equilíbrio não são todas válidas.

No primeiro capítulo, estudamos alguns aspectos estatísticos de torneios de futebol. A dinâmica desse sistema competitivo foi simulada por meio de um modelo probabilístico, o qual reproduziu bem algumas características relevantes dessas ligas, tais como, valor médio e variância das pontuações. A comparação foi realizada usando dados provenientes das ligas alemã, inglesa e espanhola e mostrou-se em bom acordo com todas elas. Investigamos a evolução temporal da liga alemã como uma caminhada aleatória, o que nos permitiu caracterizar o processo de difusão como sendo não-gaussiano e superdifusivo. Quando comparados com o modelo, esses resultados indicaram que o comportamento não-gaussiano está relacionado com a diferença entre as equipes e também com a assimetria do sistema de pontuação. Como uma aplicação do modelo, comparamos os dois sistemas de torneios mais populares: pontos corridos e eliminatórios. Esta comparação indicou que os torneios eliminatórios apresentam uma maior aleatoriedade, o que facilita equipes com menor preparo serem campeãs. Por outro lado, nos torneios de pontos corridos essa aleatoriedade é menor, tornando mais provável que a melhor equipe seja a campeã. Neste contexto, devido ao bom acordo do modelo com os dados empíricos,

acreditamos que o mesmo possa ser empregado em outras modalidades esportivas e também em outros sistemas em que exista competição, como dinâmica de firmas e mercado financeiro. Além disso, temos a intenção de investigar outros aspectos relacionados à difusão em sistemas caracterizados por múltiplos agentes que interagem entre si de maneira diferente, mediados por uma função peso análoga a distribuição de qualidades. Os resultados obtidos para o futebol parecem indicar que em um sistema complexo mais geral, comportamentos difusivos anômalos possam estar relacionados com o fato das interações não serem todas iguais entre os agentes.

No segundo capítulo, apresentamos um modelo simbólico que permite gerar sequências de símbolos com correlações de longo alcance. Estas correlações tornaram-se evidentes quando calculamos a entropia de Boltzmann-Gibbs dessas sequências e observamos um crescimento não-linear da mesma com o tamanho L da janela movida ao longo da sequência. Este crescimento não-linear, ou seja, a não-extensividade da entropia usual, nos conduziu ao uso da entropia S_q de Tsallis, a qual preservou a extensividade para uma escolha particular do parâmetro q , mesmo na região onde existiam correlações de longo alcance. Além disso, interpretamos esse modelo simbólico como uma caminhada aleatória, o que permitiu descobrir que o modelo exhibe um comportamento difusivo anômalo para valores específicos do parâmetro μ . Especificamente, encontramos três regimes de difusão, a depender do parâmetro μ : balístico ($\mu < 3$), superdifusivo ($2 < \mu < 3$) e usual ($\mu > 3$). Verificamos que a distribuição de probabilidade pode ser colapsada para todo n se a soma $\xi(n)$ for escalada por $n^{-\lambda}$, sugerindo uma invariância de escala nos dados, e que a mesma encontra-se em bom acordo com distribuições q -gaussianas. Contudo, ressaltamos que esse ajuste é apenas uma tentativa baseada nas formas das distribuições de probabilidade e que outros modelos mais fundamentais podem conduzir a resultados igualmente bons. Em particular, como uma continuidade

desses estudos, poderiam ser investigados modelos do tipo caminhada aleatória contínua (*continuous time random walk*). Mas, de uma maneira geral, este modelo simbólico mostrou-se bastante robusto e pode ser um bom candidato para a modelagem de dados empíricos que apresentem algumas destas características.

Referências Bibliográficas

- [1] B. Hannon, M. Ruth, 2nd ed. *Dynamic modeling* (Springer, New York, 2001).
- [2] D. M. Gordon, *The organization of work in social insects colonies*. Nature **380**, 121-124 (1996).
- [3] T. H. Clutton-Brock *et al.*, *Selfish sentinels in cooperative mammals*. Science **284**, 1640-1644 (1999).
- [4] D. B. Bahr, M. Bekoff, *Predicting flock vigilance from simple passerine interactions: modelling with cellular automata*. Animal Behavior **58**, 831-839 (1999).
- [5] J. Toner, Y. Tu, *Long-range order in a two-dimensional dynamical XY model: how birds fly together*. Phys. Rev. Lett. **75**, 4326-4329 (1995).
- [6] D. Chowdhury, L. Santen, A. Schadschneider, *Statistical physics of vehicular traffic and some related systems*. Phys. Rep. **329**, 199-329 (2000).
- [7] S. Redner, *How popular is your paper? An empirical study of the citation distribution*. Eur. Phys. J. B **4**, 131-134 (1998).
- [8] S. Picoli Jr., R. S. Mendes, L. C. Malacarne, E. K. Lenzi, *Scaling behavior in the dynamics of citations to scientific journals*. Europhys. Lett. **75**, 673-679 (2006).

- [9] M. Barthelemy, L. A. N. Amaral, *Small-world networks: evidence for a crossover picture*. Phys. Rev. Lett. **82**, 3180-3183 (1999).
- [10] R. N. Mantegna, H. E. Stanley, *An introduction to econophysics* (Cambridge University Press, Cambridge, 1999).
- [11] S. Picoli Jr., R. S. Mendes, *Universal features in the growth dynamics of religious activities*. Phys. Rev. E **77**, 036105 (2008).
- [12] S. Y. Auyang, *Foundations of complex-systems* (Cambridge University Press, Cambridge, 1998).
- [13] H. J. Jensen, *Self-organized criticality* (Cambridge University Press, Cambridge, 1998).
- [14] R. Albert, A.-L. Barabási, *Statistical mechanics of complex networks*. Rev. Mod. Phys. **74**, 47-74 (2002).
- [15] N. Boccaro, *Modeling complex systems* (Springer-Verlag, New York, 2004)
- [16] H. Haken, *Information and self-organization* (Springer, Berlin, 2006).
- [17] C. Tsallis, *Nonextensive statistics: theoretical, experimental and computational evidences and connections*. Braz. J. Phys. **29**, 1-35 (1999).
- [18] H. V. Ribeiro, R. S. Mendes, L. C. Malacarne, S. Picoli Jr., P. A. Santoro, *Dynamics of tournaments: the soccer case*. arXiv:1001.2854 (2010).
- [19] M. Buiatti, P. Grigolini, L. Palatella, *A non extensive approach to the entropy of symbolic sequences*. Physica A **268**, 214-224 (1999).
- [20] H. V. Ribeiro, E. K. Lenzi, R. S. Mendes, G. A. Mendes, L. R. da Silva, *Symbolic sequences and Tsallis entropy*. Braz. J. Phys. **39**, 444-447 (2009).

- [21] <http://www.dfb.de/>
- [22] <http://www.premierleague.com/>
- [23] <http://www.lfp.es/>
- [24] C. Sire, S. Redner, *Understanding baseball team standings and streak*. Eur. Phys. J. B **67**, 473-481 (2009).
- [25] S.-K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985).
- [26] T. Rohlf, C. Tsallis, *Long-range memory elementary 1D cellular automata: dynamics and nonextensivity*. Physica A **379**, 465-470 (2007).
- [27] A. Pluchino, A. Rapisarda, C. Tsallis, *Nonergodicity and central-limit behavior for long-range Hamiltonians*. Europhys. Lett. **80**, 26002 (2007).
- [28] E. P. Borges, C. Tsallis, *Negative specific heat in a Lennard-Jones-like gas with long-range interactions*. Physica A **305**, 148-151 (2002).
- [29] L. G. Moyano, A. P. Majtey, C. Tsallis, *Weak chaos and metastability in a symplectic system of many long-range-coupled standard map*. Eur. Phys. J. B **52**, 493-500 (2006).
- [30] V. Latora, A. Rapisarda, C. Tsallis, *Non-Gaussian equilibrium in a long-range Hamiltonian system*. Phys. Rev. E **64**, 056134 (2001).
- [31] C.-K. Peng *et al.*, *Long-range correlations in nucleotide-sequences*. Nature **356**, 168-170 (1992).
- [32] W. Ebeling, T. Poschel, *Entropy and long-range correlations in literary english*. Europhys. Lett. **26**, 241-246 (1994).

- [33] P. Allegrini, P. Grigolini, B. J. West, *A dynamical approach to DNA sequences*. Phys. Lett. A **211**, 217-222 (1996).
- [34] P. Allegrini, M. Buiatti, P. Grigolini, B. J. West, *Non-Gaussian statistics of anomalous diffusion: the DNA sequences of prokaryotes*. Phys. Rev. E **58**, 3640 (1998).
- [35] P. D. Siders, *Driven on- and off-lattice gas*. Phys. Rev. E **79**, 061101 (2009).
- [36] M. Gell-Mann, C. Tsallis, *Nonextensive entropy - interdisciplinary applications* (Oxford University Press, New York, 2004).
- [37] F. Baldovin, A. Robledo, *Sensitivity to initial conditions at bifurcations in one-dimensional nonlinear maps: rigorous nonextensive solutions*. Europhys. Lett. **60**, 518-524 (2002).
- [38] T. Arimitsu, N. Arimitsu, *PDF of velocity fluctuation in turbulence by a statistics based on generalized entropy*. Physica A **305**, 218-226 (2002).
- [39] C. Tsallis, G. Bemsiki, R. S. Mendes, *Is re-association in folded proteins a case of nonextensivity?*. Phys. Lett. A **257**, 93-98 (1999).
- [40] R.V. Hogg, A. Craig, 5th ed. *Introduction to mathematical statistics* (Prentice Hall, New York, 1995).
- [41] R. M. Pickup, R. Cywinski, C. Pappas, B. Farago, P. Fouquet, *Generalized spin-glass relaxation*. Phys. Rev. Lett. **102**, 097202 (2009).
- [42] B. Liu, J. Goree, *Superdiffusion and non-Gaussian statistics in a driven-dissipative 2D dusty plasma*. Phys. Rev. Lett. **100**, 055003 (2008).
- [43] R. Metzler, J. Klafter, *The random walk's guide to anomalous diffusion: a fractional dynamics approach*. Phys. Rep. **339**, 1-77 (2000).