Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia

Caminhadas Deterministas Parcialmente Auto-repulsivas: Resultados Analíticos para o Efeito da Memória do Turista na Exploração de Meios Desordenados

César Augusto Sangaletti Terçariol

Tese submetida ao Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo como parte dos requisitos para a obtenção do título de Doutor em Física.

Orientador: Prof. Dr. Alexandre Souto Martinez

Ribeirão Preto, novembro de 2008.

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE DOCUMENTO, POR MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Terçariol, C. A. S.
Caminhadas Deterministas Parcialmente Auto-repulsivas: Resultados Analíticos para o Efeito da Memória do Turista na Exploração de Meios Desordenados/ César Augusto Sangaletti Terçariol; orientador Prof. Dr. Alexandre Souto Martinez.
– Riberão Preto/SP, 2008.
133 f.

Tese (Doutorado – Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo.

Palavras Chaves: caminhada determinista, caminhada do turista, meios aleatórios, caminhada com memória, distribuição conjunta, distribuição de tempos de transiente, distribuição de período de atratores, memória crítica, modelo de distâncias aleatórias, modelo de mapeamento aleatório.

Aos meus pais.

Agradecimentos

Quero agradecer ao meu orientador Alexandre Souto Martinez pelo incentivo e entusiasmo demonstrados durante nossos estudos, por ter-se mantido fiel ao nosso projeto de pesquisa mesmo diante de tantas mudanças ocorridas em nossas vidas.

Agradeço aos professores doutores

- Carlos Frederico de Oliveira Graeff
- José Augusto Baranauskas
- Osame Kinouchi Filho
- Renato Tinós

pelo empenho e dedicação dispensados durante as disciplinas que cursei.

Gostaria de agradecer aos componentes da minha banca de qualificação, professores doutores

- Domingos Alves
- Nelson Augusto Alves
- Ubiraci Pereira da Costa Neves

pelas críticas sinceras que fizeram à minha monografia.

Agradeço aos componentes da minha banca de defesa, professores doutores

- Marco Antonio Alves da Silva
- Rafael Andres Rosales Mitrowsky
- Reinaldo Roberto Rosa
- Tereza Cristina da Rocha Mendes

pelas sugestões e contribuições que fizeram a este trabalho.

Agradeço aos colegas de equipe pela oportunidade de intercâmbio de conhecimento. Sou grato ao colega doutorando Rodrigo Silva González pelas correções feitas no manuscrito desta Tese.

"Do zero ao um,
do um ao infinito..."

César Augusto Sangaletti Terçariol

Resumo

Considere um meio desordenado constituído por N pontos cujas coordenadas são geradas aleatoriamente de maneira uniforme e independente nas arestas unitárias de um hipercubo d-dimensional. As probabilidades de vizinhança entre os pares de pontos deste meio são expressas através da fórmula de Cox. Um caminhante parte de um dado ponto deste meio desordenado e se movimenta obedecendo à regra determinista de ir para o ponto mais próximo que não tenha sido visitado nos últimos μ passos. Este processo foi denominado de caminhada determinista do turista. Cada trajetória gerada por esta dinâmica possui uma parte inicial não-periódica de t passos (transiente) e uma parte final periódica de p passos (atrator). Neste trabalho, obtemos analiticamente algumas distribuições estatísticas para a caminhada determinista do turista com memória μ arbitrária em sistemas unidimensionais e com memória $\mu = 2$ no modelo Random Link (que corresponde ao limite $d \to \infty$). Estes resultados nos permitiram compreender o papel da memória no comportamento exploratório do turista e explicar a equivalência não-trivial entre o modelo Random Link e o modelo Random Map (que é um caso limite das redes de Kauffman). Enfatizamos que o número de pontos explorados pelo turista é a grandeza fundamental nos problemas considerados. As distribuições analíticas obtidas foram validadas através de experimentos numéricos. Também obtivemos uma dedução alternativa para a fórmula de Cox, apresentando os resultados finais em termos de distribuições estatísticas elementares.

Palavras Chaves: caminhada determinista, caminhada do turista, meios aleatórios, caminhada com memória, distribuição conjunta, distribuição de tempos de transiente, distribuição de período de atratores, memória crítica, modelo de distâncias aleatórias, modelo de mapeamento aleatório.

Abstract

Consider a medium characterized by N points whose coordinates are randomly and independently generated by a uniform distribution along the unitary edges of a d-dimensional hypercube. The neighborhood probabilities between any pair of points in this medium are given by the Cox formula. A walker leaves from each point of this disordered medium and moves according to the deterministic rule to go the nearest point which has not been visited in the preceding μ steps. This process has been called the deterministic tourist walk. Each trajectory generated by this dynamics has an initial non-periodic part of t steps (transient) and a final periodic part of p steps (attractor). In this work, we obtain analytically some statistical distributions for the deterministic tourist walk with arbitrary memory μ in one-dimensional systems and with memory $\mu = 2$ in the random link model (which corresponds to $d \to \infty$ limit). These results enable us to understand the main role played by the memory on the tourist's exploratory behavior and explain the non-trivial equivalence between the random link model and the random map model (which is a limiting case of the Kauffman model). We stress that the number of explored points is the fundamental quantity in the considered problems. The obtained distributions have been validated by numerical experiments. We also obtain an alternative derivation for the Cox formula, writing the final results in terms of known statistical distributions.

Keywords: deterministic walk, tourist walk, random media, walk with memory, joint distribution, transient time distribution, attractor period distribution, critical memory, random distance model, random map model.

Sumário

1 Introdução p. 19

2 Modelo da Caminhada Determinista do Turista p. 23

2.1 Meio Desordenado: Aspecto Estático . . . . . p. 23

2.1.1 Efeito de Borda . . . . . p. 23

2.1.2 Aproximação de Campo Médio . . . . . p. 24

2.1.2.1 Modelo de Distâncias Aleatórias (Random Link) . . . p. 25

2.1.2.2 Modelo de Mapeamento Aleatório (Random Map) . . . p. 25

2.2 Caminhada Determinista do Turista: Aspecto Dinâmico . . . . . p. 26

2.2.1 Turista Preguiçoso ($\mu = 0$) . . . . . p. 26

2.2.2 Turista sem Memória ($\mu = 1$) . . . . . p. 27

2.2.3 Turista com Memória ($\mu \geq 2$) . . . . . p. 27

2.3 Conclusão . . . . . p. 28

3 Alguns Aspectos Computacionais da Caminhada do Turista p. 29

3.1 Problema de Pontos Aleatórios . . . . . p. 29

3.2 Modelo Random Link . . . . . p. 31

3.2.1 Implementação Convencional . . . . . p. 31

3.2.2 Método One-Seed . . . . . p. 32

3.2.3 Método Multiple-Seed . . . . . p. 33

3.2.4 Comentários sobre Geradores de Números Aleatórios . . . . . p. 34

3.3 Conclusão . . . . . p. 36

4 Modelo Random Link p. 37

4.1 Distribuição do Número de Sítios Explorados Antes da Primeira Revisita p. 38

4.1.1 Analogia à Distribuição Geométrica . . . . . p. 42

4.1.2 Dedução Alternativa . . . . . p. 43

4.1.2.1 Probabilidades de Revisita e de Exploração . . . . . p. 43

4.1.2.2 Forma Exponencial (Acumulada Meia-Gaussiana) . . . p. 44

4.2 Distribuição do Número Total de Sítios Explorados . . . . . p. 45

4.2.1 Probabilidades de Revisita e de Aprisionamento em Função do Posto . . . . . p. 45

4.2.1.1 Probabilidade de Revisita . . . . . p. 46

4.2.1.2 Probabilidade de Aprisionamento . . . . . p. 47

4.2.2 Probabilidades de Captura e de Subsistência . . . . . p. 48

4.2.2.1 Comparação com o modelo Random Map . . . . . p. 48

4.2.2.2 Forma Exponencial . . . . . p. 49

4.3 Distribuição Conjunta de Transiente e Período . . . . . p. 50

4.3.1 Distribuição Marginal de Tempos de Transiente . . . . . p. 50

4.3.2 Distribuição Marginal de Períodos dos Ciclos . . . . . p. 51

4.4 Conclusão . . . . . p. 52

5 Sistemas Unidimensionais p. 53

5.1 Meio Semi-infinito . . . . . p. 53

5.1.1 Distribuição do Número de Sítios Explorados . . . . . p. 53

5.1.1.1 Resultado Exato . . . . . p. 54

5.1.1.2 Aproximação de Campo Médio . . . . . p. 58

5.1.2 Probabilidades de Exploração e de Recuo . . . . . p. 59

5.1.3 Dedução Alternativa . . . . . p. 62

5.2 Meio Finito . . . . . p. 62

5.2.1 Probabilidade de Percolação . . . . . p. 63

5.2.2 Memória Crítica . . . . . p. 64

5.3 Conclusão . . . . . p. 66

6 Probabilidade de Vizinhança em Ambientes Poissônicos: Fórmula de Cox p. 67

6.1 Dedução Alternativa da Fórmula de Cox . . . . . p. 68

6.2 Probabilidades de Vizinhança no Modelo Random Link e em Alta Dimensionalidade . . . . . p. 70

6.2.1 Limite Termodinâmico . . . . . p. 70

6.2.2 Sistema com Tamanho Finito . . . . . p. 72

6.3 Probabilidades de Vizinhança no Modelo Random Map . . . . . p. 75

6.4 Conclusão . . . . . p. 75

7 Conclusão p. 77

Apêndice A – Integrais Gaussianas p. 81

A.0.1 Ordem Zero . . . . . p. 81

A.0.2 Ordem Arbitrária . . . . . p. 82

Apêndice B – Algumas Funções Especiais p. 83

B.1 Funções do Tipo Erro . . . . . p. 83

B.2 Funções do Tipo Gama . . . . . p. 84

B.3 Funções do Tipo Beta . . . . . p. 85

B.4 Funções do Tipo Hipergeométrica . . . . . p. 86

Apêndice C – Áreas de Hipersuperfícies, Volume de Hiperesferas, de Hipercalotas e de Outras Hiperformas p. 89

C.1 Coordenadas Retangulares . . . . . p. 90

C.1.1 Volume de uma Hipercalota Esférica . . . . . p. 93

C.1.2 Volume Relativo de um Crescente . . . . . p. 94

C.2 Coordenadas Esféricas . . . . . p. 95

C.3 Integrais Gaussianas . . . . . p. 98

Apêndice D – Distribuições Estatísticas em Meios Desordenados p. 101

D.1 Distribuições-base . . . . . p. 102

D.1.1 Distribuição de Bernoulli . . . . . p. 102

D.1.2 Distribuição Uniforme (ou Retangular) . . . . . p. 103

D.2 Meios Ilimitados . . . . . p. 103

D.2.1 Meios Discretos . . . . . p. 104

D.2.1.1 Número de Pontos num Intervalo – Distribuição Binomial . . . . . p. 104

D.2.1.2 Distribuição Multinomial (ou Polinomial) . . . . . p. 105

D.2.1.3 Distância entre Pontos Consecutivos – Distribuição Geométrica . . . . . p. 106

D.2.1.4 Distância entre Pontos Não-consecutivos – Distribuição de Pascal (ou Binomial Negativa) . . . . . p. 107

D.2.2 Meios Contínuos . . . . . p. 108

D.2.2.1 Número de Pontos num Intervalo – Distribuição de Poisson . . . . . p. 108

D.2.2.2 Distância entre Pontos Consecutivos – Distribuição Exponencial . . . . . p. 110

D.2.2.3 Distância entre Pontos Não-consecutivos – Distribuição Gama . . . . . p. 112

D.3 Meios Finitos . . . . . p. 113

D.3.1 Meios Discretos . . . . . p. 114

D.3.1.1 Número de Pontos num Intervalo – Distribuição Hipergeométrica . . . . . p. 114

D.3.1.2 Abscissa do k-ésimo Posto – Distribuição Hipergeométrica Multivariada . . . . . p. 116

D.3.2 Meios Contínuos . . . . . p. 118

D.3.2.1 Número de Pontos num Intervalo – Distribuição Binomial . . . . . p. 118

D.3.2.2 Abscissa do Primeiro Posto – Distribuição Potência . . p. 119

D.3.2.3 Abscissa do k-ésimo Posto – Distribuição Beta . . . . . p. 121

Generalização da Distribuição Beta – Estatística de Ordem . . . . . p. 124

D.4 Distribuição do Produto de Variáveis Uniformes . . . . . p. 124

D.5 Distribuição Gaussiana (ou Normal) . . . . . p. 126

D.5.1 Distribuição Meia-Gaussiana . . . . . p. 128

D.5.2 Distribuição Log-Normal . . . . . p. 128

Referências p. 131

1 Introdução

As caminhadas aleatórias em meios regulares e em meios desordenados [1] são um assunto bastante explorado, capaz de modelar uma grande variedade de fenômenos, como por exemplo, problemas de transporte (difusão) [2, 3]. Apesar de não serem estudadas tão minunciosamente quanto as caminhadas aleatórias, as caminhadas deterministas em meios regulares [4, 5] e em meios desordenados [6, 7, 8] apresentam resultados muito interessantes, tendo, por exemplo, aplicação em forrageamento [9, 10, 11].

Estamos propondo o modelo de uma caminhada determinista sobre um meio desordenado. Este meio é caracterizado por N pontos cujas coordenadas são geradas aleatoriamente, de maneira uniforme, nas arestas unitárias de um hipercubo de d dimensões. De cada ponto deste meio, parte um caminhante que se movimenta obedecendo à regra determinista de ir para o ponto mais próximo que não tenha sido visitado nos últimos μ passos. A trajetória gerada por esta dinâmica possui uma parte inicial não-periódica de t passos, denominada transiente, e uma parte final periódica de p passos, denominada de atrator, na qual os mesmos sítios serão visitados sempre na mesma seqüência. Apesar de poder ser enunciada de forma simples, esta dinâmica possui um comportamento complexo, com resultados não-triviais para $\mu \geq 2$. Esta caminhada foi denominada de *Caminhada Determinista do Turista* [12, 13] e tem sido usada para caracterizar tesauro [14], como um algoritmo de reconhecimento de padrões [15] e em análise de imagem [16, 17].

Observamos que a regra determinista acima pode ser relaxada, permitindo que o viajante visite com probabilidades maiores os sítios mais próximos, sendo esta variante denominada *Caminhada Estocástica do Turista* [18, 19]. O objetivo desta Tese é analisar o efeito da memória na Caminhada Determinista do Turista, através da dedução analítica das distribuições estatísticas de interesse.

No Cap. 2 introduzimos o meio desordenado onde se realiza a dinâmica da caminhada e dois modelos que apresentam a característica da aproximação de campo médio e são casos limites dos espaços euclidianos (modelo Random Link) e das redes de Kauffman [20]

(modelo Random Map). A seguir descrevemos o modelo da caminhada do turista, enfatizando seus aspectos mais importantes. Apresentamos os resultados já obtidos para os casos do Turista Preguiçoso ($\mu = 0$) e do Turista Sem Memória ($\mu = 1$).

No Cap. 3 discutimos alguns aspectos relativos à implementação computacional do meio desordenado. Apresentamos uma técnica (máscara) aplicável aos espaços euclidianos que reduz o consumo de memória RAM da ordem de $O(N^2)$ para $O(N)$. Salientamos que esta técnica não pode ser aplicada ao modelo Random Link e apresentamos dois algoritmos alternativos para tratar numericamente este modelo.

No Cap. 4 determinamos analiticamente a distribuição do número de sítios explorados por um turista com memória $\mu = 2$ caminhando no modelo Random Link e, através de uma analogia à distribuição geométrica, calculamos as probabilidades de captura e de subsistência. Este resultado nos permite explicar a mudança drástica no comportamento exploratório do turista, quando comparado ao caso $\mu = 1$. Em seguida, obtemos a distribuição conjunta de tempos de transiente e período dos ciclos e mostramos a equivalência não-trivial entre o modelo Random Link com $\mu = 2$ e o modelo Random Map com $\mu = 0$.

No Cap. 5 focalizamos os sistemas unidimensionais. Iniciamos com o meio semi-infinito, calculando a distribuição exata do número de sítios explorados por um turista com memória μ arbitrária que parte da origem deste meio. Através das probabilidades de exploração e de recuo, obtemos uma forma aproximada para a distribuição do número de sítios explorados. Em seguida, mostramos a equivalência entre o meio semi-infinito e o meio finito para a caminhada do turista. Esta equivalência explica observações anteriores (obtidas através de simulações numéricas) que apontavam um efeito de tamanho finito pouco pronunciado para o meio unidimensional. Explorando esta equivalência, obtemos a probabilidade de percolação para o meio finito e destacamos a existência de uma memória crítica.

No Cap. 6 apresentamos uma dedução alternativa e mais simples para a fórmula de Cox [21], que fornece as probabilidades de vizinhança em processos poissônicos. A seguir fazemos uma aproximação para altas dimensionalidades e calculamos estas probabilidades de vizinhança para os modelos de campo médio. Todos os resultados obtidos são convenientemente expressos em termos de funções especiais e de distribuições estatísticas elementares.

As conclusões, considerações finais e perspectivas de trabalhos futuros são apresentados no Cap. 7.

No Apêndice A calculamos as integrais Gaussianas. No Apêndice B reunimos as funções especiais que foram usadas ao longo do texto. No Apêndice C calculamos o volume de uma hiperesfera (e de outras hiperformas correlatas) de três maneiras distintas. Finalmente, no Apêndice D discutimos as distribuições estatísticas elementares usadas nesta Tese, dando-lhes significados intuitivos no contexto de meios desordenados.

2 Modelo da Caminhada Determinista do Turista

Neste Capítulo apresentamos o modelo da Caminhada Determinista do Turista. Iniciamos caracterizando o meio desordenado e discutimos os efeitos de borda. Apresentamos dois modelos com característica de aproximação de campo médio. Em seguida, descrevemos a regra determinista da caminhada e destacamos as principais diferenças entre os casos *sem* e *com* memória.

2.1 Meio Desordenado: Aspecto Estático

O meio desordenado (mapa) é composto por N pontos (sítios) de um espaço euclidiano de d dimensões, cujas coordenadas $x_i^{(k)}$ (com $i = 1, 2, \ldots, N$ e $k = 1, 2, \ldots, d$) são geradas aleatoriamente de maneira independente seguindo uma função de densidade de probabilidade (pdf) uniforme (Eq. D.2) no intervalo $[0, 1]$. Neste meio, define-se a distância $D_{i,j}$ entre dois sítios s_i e s_j quaisquer através da métrica euclidiana:

$$D_{i,j} = \sqrt{\sum_{k=1}^{d} \left[x_i^{(k)} - x_j^{(k)}\right]^2} . \tag{2.1}$$

Matematicamente, este meio representa N pontos dispostos aleatoriamente de maneira uniforme no interior de um hipercubo de d dimensões com arestas unitárias.

2.1.1 Efeito de Borda

Nem todos os sítios do meio desordenado se relacionam igualmente com seus vizinhos. Os sítios situados próximos às faces do hipercubo sofrem um certo isolamento em relação aos demais, pois têm uma região de vizinhança menor. Este isolamento, denominado efeito de borda, aumenta acentuadamente à medida que a dimensionalidade do meio cresce. Diz-se que este meio tem condições de contorno abertas.

Uma maneira de minimizar os efeitos de borda é conectar a extremidade 0 de cada dimensão à sua extremidade 1, tornando-as circulares. Em vez de um hipercubo, obtém-se dessa forma uma hipersuperfície toroidal de d dimensões. Diz-se, então, que o meio tem condições periódicas de contorno. As Figs. 1 e 2 ilustram este processo para $d = 1$ e $d = 2$, respectivamente.

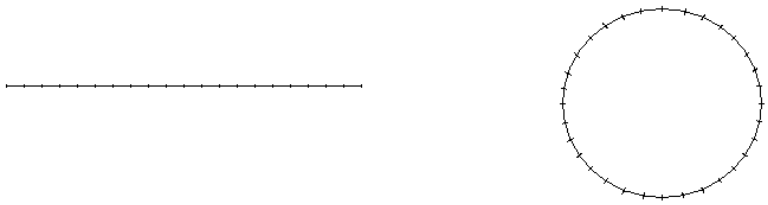


Figura 1: Condições abertas e periódicas de contorno em 1 dimensão.

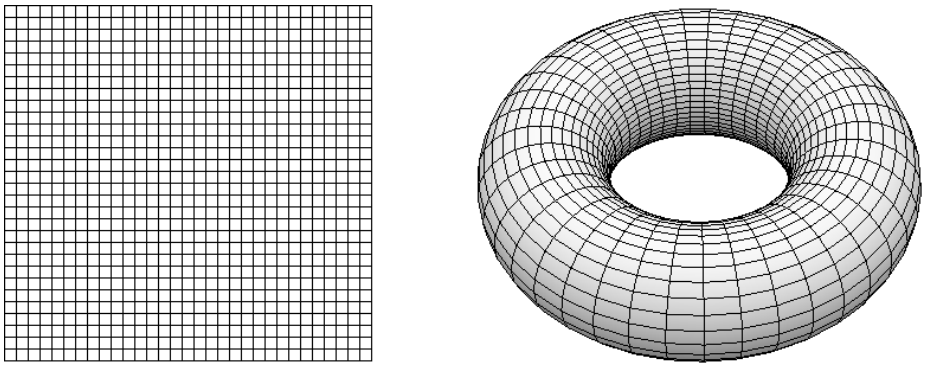


Figura 2: Condições abertas e periódicas de contorno em 2 dimensões.

Deve-se observar que a circunferência da Fig. 1 é geometricamente bidimensional, mas topologicamente unidimensional. Do mesmo modo, o toróide da Fig. 2 tem três dimensões geométricas, mas é topologicamente bidimensional, pois os pontos do meio desordenado estão distribuídos apenas sobre sua superfície. Não é possível construir figuras que ilustrem meios com condições periódicas de contorno para $d \geq 3$, mas o processo continua sendo algebricamente válido.

2.1.2 Aproximação de Campo Médio

A aproximação de campo médio consiste em fazer com que cada nodo de um grafo, ou sítio de uma rede, ou cidade de um mapa, interaja com todos os outros nodos (sítios do sistema). Vamos considerar aqui dois sistemas que têm esta característica: o modelo de distâncias aleatórias e o modelo de mapeamento aleatório. O primeiro é o limite natural

de alta dimensionalidade para a caminhada do turista e o segundo foi introduzido no contexto das redes de Kauffman [20].

2.1.2.1 Modelo de Distâncias Aleatórias (Random Link)

À medida que a dimensionalidade d do espaço euclidiano aumenta, as correlações entre as distâncias (como a desigualdade triangular) se enfraquecem. Por isso, no limite de alta dimensionalidade ($d \to \infty$), estas correlações podem ser desprezadas e as distâncias entre os pontos passam a ser consideradas como variáveis aleatórias independentes e identicalmente distribuídas (i.i.d.) uniformemente no intervalo $[0, 1]$ [22, 23, 24, 25, 26, 27], permanecendo apenas duas restrições euclidianas:

1. a distância de um ponto a si mesmo é sempre nula ($D_{i,i} = 0, \forall\, i$) e
2. a distâncias de ida é igual à distância de volta ($D_{i,j} = D_{j,i}, \forall\, i, j$).

Este é o modelo de campo médio denominado Random Link (RL), proposto originalmente por Mézard e Parisi [22] e posteriormente explorado por Percus e Martin [26].

2.1.2.2 Modelo de Mapeamento Aleatório (Random Map)

No modelo RL, descrito na seção anterior, as distâncias entre os sítios são consideradas variáveis aleatórias independentes, porém a simetria $D_{i,j} = D_{j,i}$ entre as distâncias de ida e de volta é preservada. Pode-se, entretanto, construir um modelo de distâncias aleatórias *assimétricas*, no qual as distâncias de ida e volta não sejam necessariamente iguais. Assim, em geral $D_{i,j} \neq D_{j,i}$. Esta desigualdade se manifesta, por exemplo, numa cidade com avenidas de mão única, de forma que o caminho de ida geralmente difere do caminho de volta. Observamos que neste caso a única característica remanescente do modelo anterior é que a distância de um sítio a ele mesmo é zero ($D_{i,i} = 0, \forall i$).

Uma das variantes do modelo de distâncias aleatórias assimétricas estudada por Derrida e Flyvbjerg [28], denominada modelo Random Map (RM) [29], consiste em eliminar a restrição $D_{i,i} = 0$, permitindo que a distância de um sítio a ele mesmo seja diferente de zero. Esta distância pode representar, por exemplo, um custo para um caminhante permanecer num dado sítio. Esta variante é o modelo mais descorrelacionado possível, pois para um mapa de N pontos as distâncias são consideradas N^2 variáveis aleatórias independentes. Este modelo corresponde a uma aproximação de campo médio para as redes de Kauffman [20].

2.2 Caminhada Determinista do Turista: Aspecto Dinâmico

Considere um caminhante que parte de um dado ponto do meio desordenado e se move, a cada passo discreto de tempo, obedecendo à regra determinista de ir para o ponto mais próximo (menor distância euclidiana), que não tenha sido visitado nos μ passos precedentes. Esta dinâmica é conhecida como *Caminhada Determinista do Turista* [12]. A grandeza μ, denominada memória, representa o tempo (em número de passos) necessário para a regeneração dos sítios visitados (tempo refratário).

Partindo de cada sítio do mapa (ou de vários mapas), o viajante percorre trajetórias diferentes que possuem uma parte inicial não-periódica (trasiente de tamanho t) e terminam em ciclos (atratores de período p). Uma vez preso num ciclo, o caminhante não mais visita novos pontos.

2.2.1 Turista Preguiçoso ($\mu = 0$)

O caso mais simples a ser considerado é o do turista preguiçoso, no qual $\mu = 0$. Todos os sítios são permitidos e a caminhada do turista é trivial: o caminhante não se move a cada passo de tempo, ficando preso no sítio de partida, pois ele é, a cada passo, o sítio mais próximo permitido. Dessa forma, todos os sítios são atratores de período unitário do sistema. A distribuição de transientes $S^{(N)}_{\mu,d}(t)$ vale simplesmente

$$S^{(N)}_{0,d}(t) \;=\; \delta_{t,0} \;,$$

e a distribuição de períodos $S^{(N)}_{\mu,d}(p)$ é dada por

$$S^{(N)}_{0,d}(p) \;=\; \delta_{p,1} \;.$$

Neste caso, a distribuição conjunta de tempos de transiente e período de ciclos $S^{(N)}_{\mu,d}(t,p)$ é simplesmente

$$S^{(N)}_{0,d}(t,p) \;=\; S^{(N)}_{0,d}(t) \cdot S^{(N)}_{0,d}(p) = \delta_{t,0} \cdot \delta_{p,1} \;,$$

onde $\delta_{i,j}$ é o delta de Kronecker.

Embora esta situação seja trivial, sua extensão para a caminhada estocástica do turista (e para o modelo RM [30]) produz resultados analíticos onde observa-se uma mudança abrupta de comportamento [19].

2.2.2 Turista sem Memória ($\mu = 1$)

No caso $\mu = 1$, apenas o sítio atual é proibido. O turista é obrigado, a cada passo, a abandonar o sítio em que se encontra e mover-se até o sítio mais próximo dentre os demais. A denominação *turista sem memória* se deve ao fato de que, a cada passo, o turista tem conhecimento apenas do sítio em que está, mas não se "lembra" de nenhum dos sítios visitados nos passos precedentes. As trajetórias sempre terminam em dois sítios que são vizinhos mutuamente mais próximos. Diz-se que estes pontos formam um casal. Os casais são os atratores do sistema. Portanto, a distribuição de períodos vale

$$S_{1,d}^{(N)}(p) \;=\; \delta_{p,2} \;,$$

ou seja, todos os ciclos têm período 2. No entanto, a distribuição de transientes não é concentrada num único valor.

Para $N \gg 1$, a distribuição de tempos de transiente foi obtida analiticamente para dimensionalidade arbitrária em trabalhos anteriores a esta Tese [30]:

$$S_{1,d}^{(\infty)}(t) = \frac{(t + p_d^{-1})\Gamma(1 + p_d^{-1})}{\Gamma(t + 2 + p_d^{-1})} \;,$$

onde $\Gamma(z)$ é a função gama (dada pela Eq. B.3) e p_d é dado pela Eq. 6.1. Este caso não leva a uma exploração eficiente do meio aleatório, pois após um tempo de transiente muito curto, o turista fica preso num casal.

2.2.3 Turista com Memória ($\mu \geq 2$)

Fenômenos interessantes ocorrem quando se consideram valores de memória $\mu \geq 2$. Neste caso, a distribuição de períodos não é mais concentrada em $p_{min} = \mu + 1$, mas apresenta um espectro completo de ciclos com períodos $p \geq p_{min}$, com possível decaimento lei de potência [12, 14, 31], favorecendo a exploração do meio pelo caminhante. Até mesmo com $\mu = 2$ é possível obter trajetórias com transientes e períodos grandes. Nos Capítulos 4 e 5 desta Tese, obtemos resultados analíticos para a caminhada do turista com memória $\mu = 2$ no modelo Random Link [32] e com memória μ arbitrária em sistemas uni-dimensionais [33], respectivamente.

2.3 Conclusão

Neste Capítulo caracterizamos o meio desordenado, destacamos o efeito de borda e apresentamos uma alternativa para minimizar este efeito através das condições periódicas de contorno. Também apresentamos dois modelos com característica de aproximação de campo médio: o modelo Random Link, que representa o limite de alta dimensionalidade para espaços euclidianos; e o modelo Random Map, que representa um caso limite para as redes de Kauffman. Descrevemos o modelo da caminhada do turista, uma caminhada determinista sobre o meio desordenado que, apesar de poder ser enunciado de forma simples, possui um comportamento complexo quando se introduz a memória.

3 *Alguns Aspectos Computacionais da Caminhada do Turista*

Neste Capítulo discutimos alguns aspectos computacionais relativos à implementação do meio desordenado. Para o problema de pontos aleatórios, apresentamos o recurso da *máscara*, que reduz o consumo de memória RAM da ordem de $O(N^2)$ (na implementação direta) para $O(N)$. Salientamos que tal recurso não pode ser aplicado diretamente ao modelo RL e apresentamos dois algoritmos alternativos (one- e multiple-seed) que exploram a reproducibilidade dos geradores de números pseudo-aleatórios para gerar as distâncias descorrelacionadas deste modelo. Concluímos que o algoritmo multiple-seed é o mais eficiente, pois leva a um consumo de memória da ordem de $O(N)$, em vez de $O(N^2)$ obtido pela implementação ingênua, mas com a mesma dependência temporal $O(N^2)$.

3.1 Problema de Pontos Aleatórios

O problema de pontos aleatórios (PPA) é uma abordagem clássica para construir meios desordenados. As coordenadas $x_i^{(k)}$ de N pontos são independente e aleatoriamente geradas seguindo uma função de densidade de probabilidade (pdf) uniforme (Eq. D.2) nas arestas unitárias de um hipercubo d-dimensional. As distâncias $D_{i,j}$ entre os pares de pontos i e j são obtidas pela métrica euclidiana (Eq. 2.1). Uma possível implementação computacional consiste basicamente dos seguintes passos:

1. gerar aleatoriamente as coordenadas $x_i^{(k)}$ e armazená-las numa matriz $N \times d$, denominada *matriz de coordenadas*,

2. usar a métrica euclidiana para calcular a distância $D_{i,j}$ entre todos os pares de pontos e armazená-las numa matriz $N \times N$, denominada *matriz de distâncias*,

3. ordenar as distâncias a fim de determinar os μ vizinhos mais próximos de cada ponto

e armazená-los numa matriz $N \times \mu$, denominada *tabela de vizinhança*.

Estes procedimentos estão esquematizados no diagrama da Fig. 3.

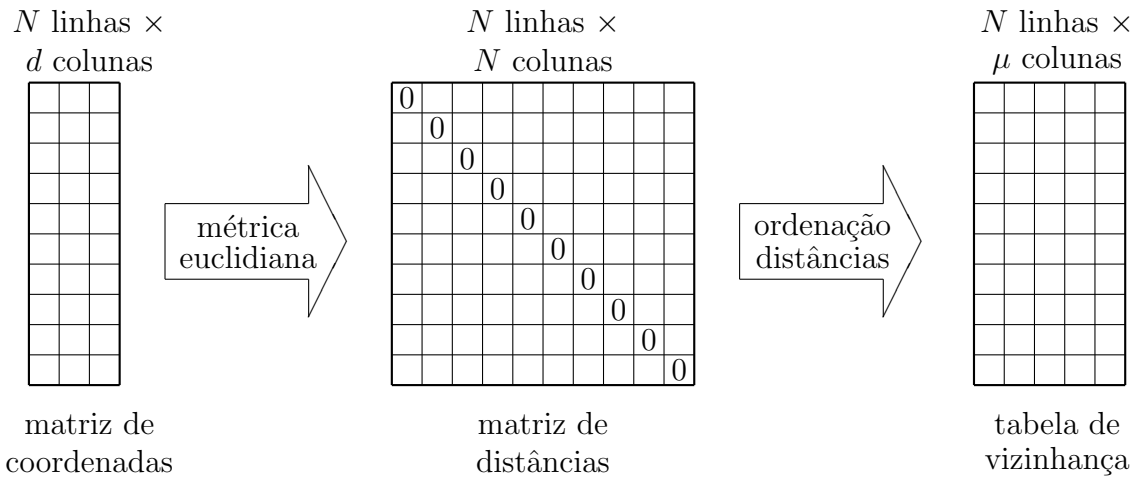


Figura 3: Implementação convencional do PPA.

Neste algoritmo ingênuo, os consumos de tempo e de memória são ambos proporcionais a N^2. Em aplicações numéricas, isto limita o tamanho do sistema tipicamente a $N = 15\,000$ pontos quando se usam técnicas avançadas de alocação da memória de acesso aleatório (RAM). Observe que o consumo de memória da ordem de $O(N^2)$ deve-se à declaração da matriz de distâncias.

Um procedimento alternativo e melhor para implementar o PPA consiste em substituir a matriz de distância de tamanho $N \times N$ por um vetor de tamanho N, denominado *máscara*, e calcular apenas as distâncias relativas a um dado ponto a cada passo de tempo, como mostra a Fig. 4. O único desperdício computacional é calcular a mesma distância ($D_{i,j} = D_{j,i}$) duas vezes (ela poderia ser calculada apenas uma única vez se a matriz de distâncias estivesse à disposição). Todavia, a dependência para com o tempo é mantida proporcional a N^2 e o consumo de memória se torna proporcional a N.

A caminhada determinista do turista é realizada sobre a tabela de vizinhança. Por isso estamos interessados somente nos postos de vizinhança dos pontos aleatórios. As distâncias euclidianas são apenas uma maneira de obter estes postos.

Para minimizar os efeitos de borda, é importante considerar condições periódicas de contorno e manter fixa a separação média entre pontos ($\ell = L/N^{1/d} = \rho^{-1/d}$, onde L é o tamanho típico do sistema e ρ é a densidade de pontos). Por esta razão tem-se que aumentar N à medida que a dimensionalidade do sistema aumenta. Uma vez que sistemas de alta

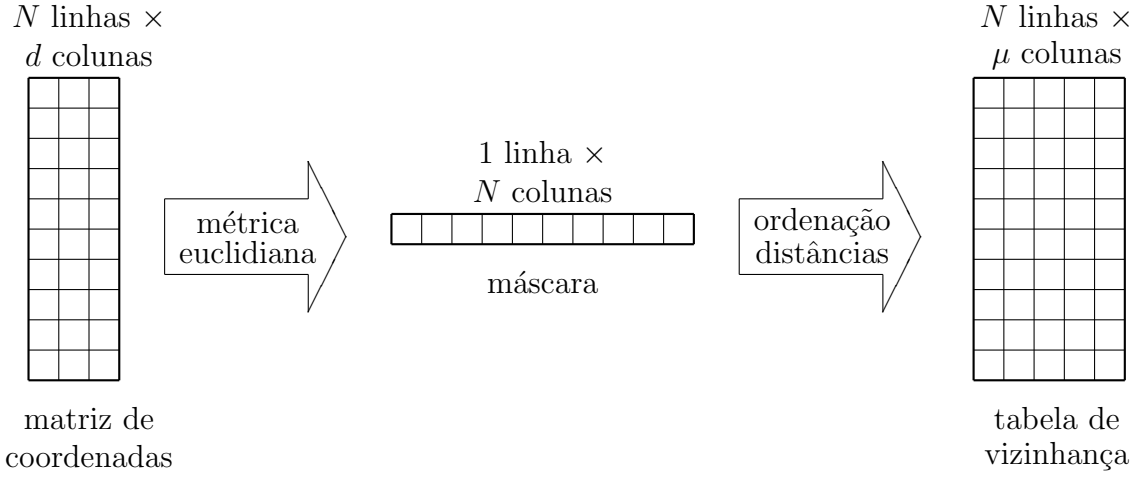


Figura 4: Implementação do PPA com o uso da máscara.

dimensionalidade serão considerados, até mesmo declarar a matriz de coordenadas pode consumir muita memória RAM. Isto introduz dificuldades computacionais adicionais, já que o tamanho do sistema encontra uma barreia imposta pela dimensionalidade.

Entretanto, ao se usar condições periódicas de contorno, quanto maior é a dimensionalidade do sistema, mais fracas são as correlações entre as distâncias (desigualdade triangular, por exemplo). No limite de alta dimensionalidade ($d \to \infty$), as distâncias entre todos os pares de pontos distintos podem ser consideradas como $N(N-1)/2$ variáveis aleatórias i.i.d. e o PPA converge para o modelo RL.

3.2 Modelo Random Link

Para trabalhar numericamente com o modelo RL, devem-se gerar diretamente as $N(N-1)/2$ distâncias aleatórias i.i.d., em vez de gerar aleatoriamente as coordenadas dos pontos e calcular as distâncias aplicando a métrica euclidiana à matriz de coordenadas. Isto resolve o problema de se alocar muita memória RAM no computador devido à alta dimensionalidade do sistema, já que a matriz de coordenadas não existe mais.

3.2.1 Implementação Convencional

Numa implementação direta para gerar as distâncias aleatórias e i.i.d. do modelo RL, declara-se uma matriz de distâncias de tamanho $N \times N$ e simplesmente aplicam-se as duas restrições euclidianas mencionadas na introdução:

1. atribuir 0 a toda sua diagonal principal e

2. gerar aleatoriamente cada distância $D_{i,j}$, com $i = 1, 2, \ldots, N-1$ e $j = i+1, i+2, \ldots, N$ (lado direito da diagonal principal), e atribuir seu valor a $D_{j,i}$ (lado esquerdo da diagonal principal) para obedecer à simetria da métrica.

A dependência para com o tempo e o consumo de memória deste algoritmo convencional são ambos proporcionais a N^2 (veja Tab. 1). Nesta implementação, a restrição imposta pela simetria $D_{i,j} = D_{j,i}$ traz sérias dificuldades numéricas, já que uma matriz de distâncias de tamanho $N \times N$ deve ser então declarada.

Método	Convencional	Single-Seed	Multiple-Seed
Tempo	N^2	N^3	N^2
Memória	N^2	N	N

Tabela 1: Consumos de memória e de tempo para os métodos convencional, one-seed e multiple-seed descritos. O método multiple-seed é o melhor pois combina o baixo consumo de memória $O(N)$ do método one-seed e baixo tempo de processamento $O(N^2)$ gasto pelo mótodo de implementação convencional.

Note que no PPA as coordenadas de pontos estão à disposição e pode-se calcular todas as distâncias relativas a apenas um ponto de cada vez, levando a um consumo de memória da ordem de $O(N)$ e que no caso do modelo RL não é possível utilzar o recurso da máscara, pois a matriz de coordenadas não está à disposição para reobtermos as distâncias já geradas e obedecermos à simetria da métrica euclidiana. Apresentamos a seguir dois métodos que substituem a matriz de distâncias por uma máscara, levando a consumo de memória da ordem de $O(N)$, como no caso do PPA. Para re-obter a distância $D_{i,j}$ (e obedecer à restricão da simetria), a característica determinista do gerador de números pseudo-aleatórios é extensivamente explorada.

3.2.2 Método One-Seed

O método one-seed usa duas variáveis inteiras alternadamente como parâmetro do gerador de números aleatórios, a saber *seed* (usada para gerar novas distâncias no lado direito da diagonal principal da matriz de distâncias) e *dummy* (usada para reobter as distâncias já geradas e obedecer à simetria da métrica). Como primeiro passo, inicialize *seed* com um valor arbitrário S_1 e salve este valor numa outra variável. Na i-ésima linha, com i variando de 1 a N:

1. atribua $D_{i,i} = 0$,

2. use a variável *seed* para gerar seqüencialmente as novas distâncias $D_{i,i+1}$, $D_{i,i+2}$, $\ldots$, $D_{i,N}$, e

3. para reobter cada distância $D_{i,j}$, com j variando de 1 a $i-1$, use a variável *dummy* nos passos seguintes:

   (a) inicialize-a com S_1,

   (b) use-a para fazer $i-j-1$ chamadas ao gerador (estas chamadas são feitas para que *dummy* possua o mesmo valor de *seed* usado para gerar a distância $D_{j,i}$, os valores de retorno dessas chamadas não serão usados),

   (c) faça uma única chamada para reobter a distância desejada $D_{i,j} = D_{j,i}$, e

   (d) faça mais $N-i$ chamadas (apenas para preparar *dummy* para a próxima iteração j – novamente, os valores de retorno não serão usados, veja Fig. 5).

Este método nos possibilita construir numericamente sistemas muito maiores do que aqueles que o método convencional pode construir, mas com um gasto de tempo computacional maior. Para gerar todas as distâncias de um mapa com N pontos, o consumo de memória é da ordem de $O(N)$ (veja Tab. 1) enquanto que o tempo requerido é proporcional a N^3, devido às dispendiosas chamadas ao gerador (veja Fig. 6).

3.2.3 Método Multiple-Seed

Obtém-se uma melhora no método one-seed substituindo a variável inteira *dummy* por um vetor de inteiros de tamanho N, denominado *s_vec*. A semente inicial de cada linha (S_1, S_N, S_{2N-2} $\ldots$) é armazenada na posição correspondente de *s_vec*. Vale mencionar que apenas $N-1$ entradas serão realmente usada, porém declarar *s_vec* com tamanho N evita indesejáveis testes de limites. Neste método, a variável inteira *seed* também é usada. Inicialize-a com algum valor arbitrário S_1. Para enterder a implementação, vamos supor que a i-ésima linha está sendo gerada, com i variando de 1 a N como antes:

1. atribua $D_{i,i} = 0$,

2. armazene o valor atual de *seed* em *s_vec*(i) (i-ésima posição de *s_vec*) para ser usada mais tarde,

3. use *seed* para gerar seqüencialmente as distâncias $D_{i,i+1}$, $D_{i,i+2}$, $\ldots$, $D_{i,N}$, e

4. para reobter as distâncias $D_{i,j} = D_{j,i}$, com j variando de 1 a $i-1$, simplesmente use $s_vec(j)$ para fazer uma única chamada ao gerador, e então o valor atual de $s_vec(j)$ será modificado para o cálculo da próxima iteração (veja Fig. 5).

O armazenamento e manipulação das sementes pelo método multiple-seed durante a construção de todos os pares de distâncias reduz drasticamente o tempo de computação à ordem de N^2 como na implementação convencional, porém com um consumo de memória da ordem de N, como no método one-seed.

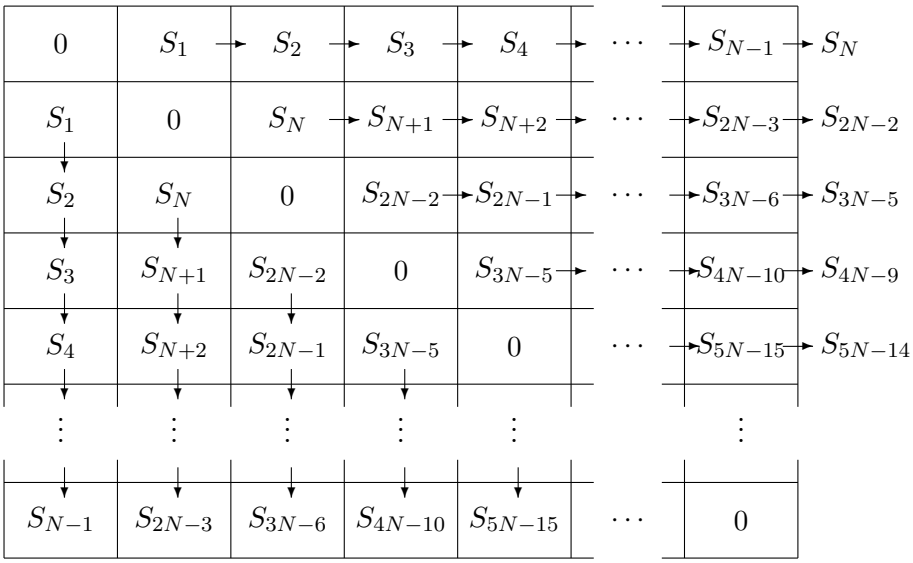


Figura 5: Ilustração da evolução da semente a cada chamada ao gerador de número aleatório ao longo da matriz de distâncias nos algoritmos convencional, one-seed e multiple-seed. Cada semente é usada para gerar a distância localizada em sua posição correspondente.

3.2.4 Comentários sobre Geradores de Números Aleatórios

Os algoritmos apresentados aqui têm em comum o uso de geradores de números pseudo-aleatórios. Em sua mais simples implementação, trata-se de um algoritmo determinista, que requer um único parâmetro inteiro, denominado *semente*. O algoritmo funciona da seguite maneira: antes da primeira chamada ao gerador, a semente é inicializada com um valor (usualmente) arbitrário S_1. Nas chamadas subseqüentes ao gerador, o valor da semente é deterministicamente modificado e dá-se início à longa seqüência descorrelacionada de inteiros S_1, S_2, ..., S_{2^m}, uniformemente distribuídos no intervalo $[0, 2^m - 1]$, para algum inteiro $m > 0$. Depois de percorrer toda a seqüência, a semente reassume seu valor inicial S_1 e o ciclo de período 2^m reinicia. Cada valor inicial S_1 tipica-

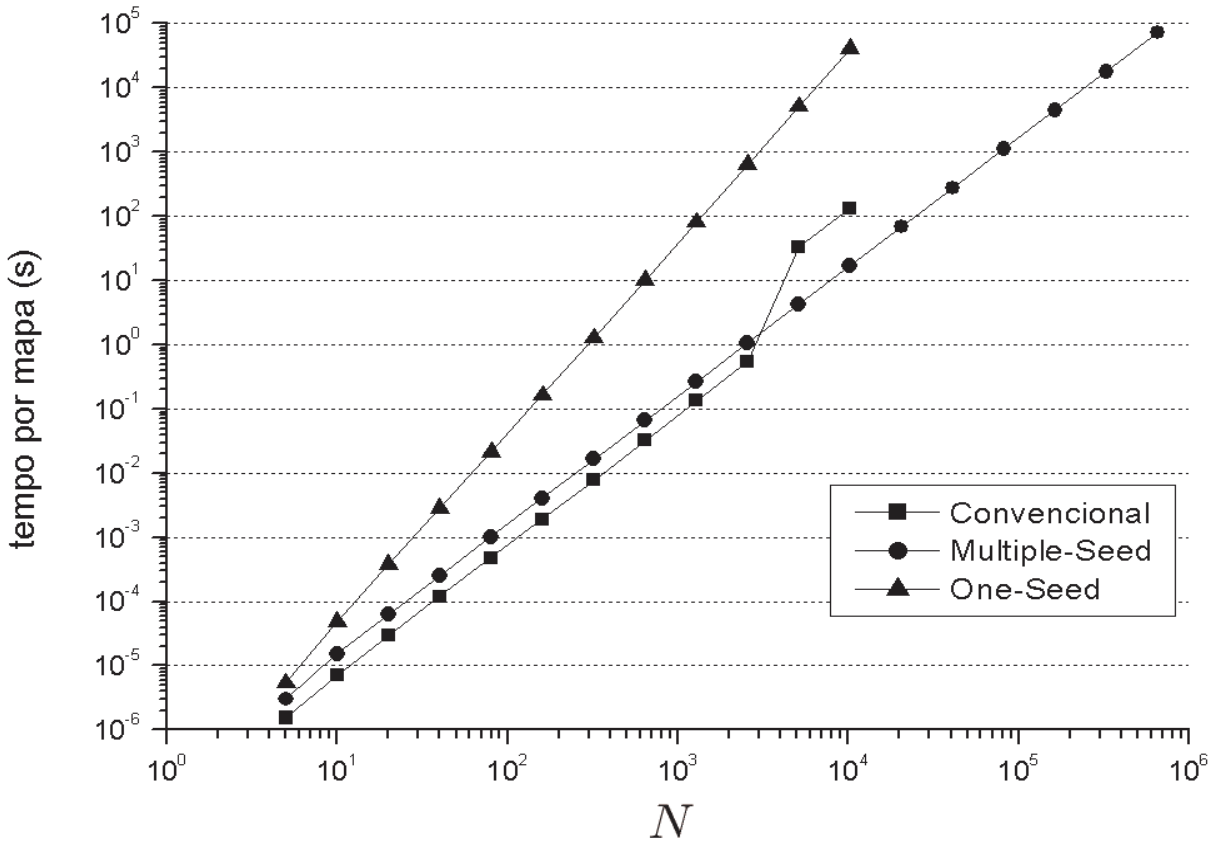


Figura 6: Tempo de processamento em função do tamanho do sistema. Esta dependência é bem descrita por uma lei de potência com expoente 2 para os métodos convencional e multiple-seed e exponent 3 para o método one-seed. Enquanto o método convencional pode lidar com sistemas de tamanho em torno de 10^4, o método multiple-seed pode lidar com sistemas mais de 100 vezes maiores. A descontinuidade na curva do método convencional corresponde ao momento quando se esgota a memória volátil e o sistema operacional passa a fazer uso do disco rígido.

mente retornará uma seqüência aleatória diferente. Entretanto, um *mesmo* valor inicial de S_1 sempre levará à *mesma* seqüência S_2, S_3, ..., S_{2^m}. A base do nosso algoritmo é justamente esta característica determinista do gerador de números aleatórios. Para cada valor inteiro S_i da semente, o gerador usualmente retorna um número real $S_i/2^m$, que está uniformemente distribuído no intervalo $[0, 1[$, enquanto segue a seqüência de sementes. Em simulações numéricas comumente utiliza-se este número real em vez da semente em si.

Deve-se tomar cuidado com um aspecto particular em programação numérica. Os algoritmos propostos aqui assumem que todas as variáveis locais do gerador de números aleatórios são dinâmicas. Para entender este aspecto, vamos brevemente descrever como variáveis são geralmente tratadas pelas subrotinas. Quando um código numérico entra numa rotina, um endereço na memória RAM é alocado para cada variável dinâmica local. Quando o código deixa a rotina, esses endereços são desalocados. Assim, os valores das variáveis internas à rotina não são preservados. Os geradores de números aleatórios mais sofisticados fazem uso de variáveis internas estáticas para manter seus valores armazenados para uso futuro. Ao contrário das variáveis dinâmicas, as variáveis estáticas têm seus endereços e valores atuais alocados na RAM mesmo quando a rotina termina. Neste tipo de gerador, duas chamadas com a *mesma* semente não produz o *mesmo* resultado, pois os valores das variáveis estáticas não são os mesmos nestas duas chamadas. Para obter a reprodutibilidade desejada neste tipo de gerador de números aleatórios, estas variáveis locais estáticas devem ser controladas da mesma forma que a semente: um vetor de tamanho N deve ser declarado para cada uma delas para recuperar seus valores. Mesmo neste caso, a ordem de magnitude dos consumos de memória e de tempo pelo algoritmo proposto não são alterados.

3.3 Conclusão

Neste Capítulo tratamos da implementação computacional do meio desordenado. Enfatizamos que a caminhada determinista do turista é realizada sobre a tabela de vizinhança e que as distâncias são apenas um meio para se obter as ordens de vizinhança. Como a declaração da matriz de distâncias é desnecessária, apresentamos recurso da máscara, que reduz o consumo de memória da ordem de $O(N^2)$ para $O(N)$. Argumentamos que a máscara não pode ser usada para gerar o modelo Random Link e apresentamos dois algoritmos alternativos que também reduzem o consumo de memória da ordem de $O(N^2)$ para $O(N)$. Através de resultados numéricos, mostramos que o algoritmo multiple-seed é o melhor compromisso entre os consumos de tempo e de memória.

4 Modelo Random Link

Neste Capítulo, obtemos resultados analíticos para a caminhada do turista, com memória $\mu = 2$ no modelo RL. Estes resultados possibilitam explicar o principal mecanismo que torna as situações $\mu = 1$ e $\mu \geq 2$ tão distintas e permitem estabelecer uma relação entre os modelos de campo médio RL e RM. A caminhada com memória $\mu = 2$ no modelo de distâncias aleatórias simétricas (RL) é equivalente à caminhada sem memória ($\mu = 0$) no modelo de distâncias aleatórias assimétricas (RM), que já foi resolvido na Ref. [30]. Também conseguimos explicar a razão da mudança abrupta (observada numericamente) na distribuição conjunta de transientes e períodos entre o transiente nulo ($t = 0$) e os outros subsequentes ($t \geq 1$).

A apresentação desses resultados será feita como segue. Na Sec. 4.1, calculamos a probabilidade $\tilde{S}^{(N)}_{2,rl}(\tilde{n})$ do turista, com memória $\mu = 2$, visitar $\tilde{n}$ sítios distintos antes da primeira revisita a algum sítio, caminhando sobre o modelo RL com N sítios. Começamos calculando a distribuição complementar acumulada $\tilde{F}^{(N)}_{2,rl}(\tilde{n})$ (upper-tail distribution). Em seguida, através de uma analogia à distribuição geométrica, obtemos as probabilidades de revisitação $\tilde{p}^{(N)}_{2,rl}(j)$ (primeira passagem) e de exploração $\tilde{q}^{(N)}_{2,rl}(j)$. Usando uma dedução alternativa, obtemos expressões mais simples para estas probabilidades, que nos leva a uma expressão analítica fechada para $\tilde{F}^{(N)}_{2,rl}(\tilde{n})$. Na Sec. 4.2, mostramos que a probabilidade do turista ser aprisionado num ciclo ao revisitar um sítio da trajetória é $p_t = 2/3$ – um resultado contra-intuitivo. Este resultado (combinado com os anteriores) nos permite obter a distribuição acumulada complementar $F^{(N)}_{2,rl}(n)$ para o número total n de sítios visitados (até que o turista entre num atrator) e explicar a equivalência entre os modelos RL e RM. Na Sec. 4.3, obtemos a distribuição conjunta $S^{(N)}_{2,rl}(t,p)$ do tempo de transiente t e do período de ciclo p e mostramos a drástica diferença entre os casos $t = 0$ e $t \neq 0$. Distribuições universais de probabilidade são obtidas re-escalando variáveis. Uma discussão final é apresentada na Sec. 4.4 e estudos futuros são propostos.

4.1 Distribuição do Número de Sítios Explorados Antes da Primeira Revisita

Considere que o turista, caminhando com memória $\mu = 2$ sobre o modelo Random Link com N pontos, tenha visitado $\tilde{n} \geq 3 = \mu + 1 = \tilde{n}_{min}$ sítios distintos e então revisite algum destes sítios. Com o intuito de obter a distribuição $\tilde{S}_{2,rl}^{(N)}(\tilde{n})$ do número $\tilde{n}$ de sítios visitados antes da primeira revisita, calcularemos inicialmente a distribuição acumulada complementar

$$\tilde{F}_{2,rl}^{(N)}(\tilde{n}) = \sum_{k=\tilde{n}}^{N} \tilde{S}_{2,rl}^{(N)}(k)$$

ou seja, a probabilidade do turista explorar no mínimo $\tilde{n}$ sítios antes da primeira revisita.

No esquema ilustrado na Fig. 7, o turista parte de um dado sítio s_1 (primeiro passo, $j = 1$) e percorre a trajetória $s_1, s_2, \ldots, s_{\tilde{n}}$, explorando $\tilde{n} = 9$ sítios distintos, sem revisita alguma. Para $1 \leq i \leq \tilde{n} - 1$, denotemos por

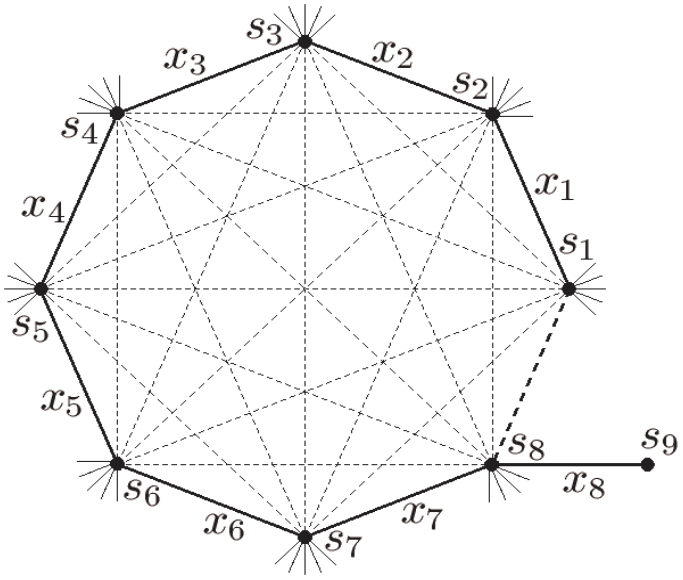


Figura 7: Representação esquemática de uma caminhada do turista com no mínimo $\tilde{n} = 9$ sítios visitados antes da primeira revisita. O turista parte do sítio s_1 e percorre a trajetória $s_1, s_2, \ldots, s_9$.

- x_i a distância entre os sítios consecutivos s_i e s_{i+1} da trajetória (linhas contínuas espessas na ilustração da Fig. 7),
- $y_{i,k}$ as distâncias entre o sítio s_i da trajetória e os sítios não-pertencentes à trajetória (linhas contínuas delgadas parcialmente representadas na Fig. 7),

- $z_{i,k}$ a distância entre os sítios não-consecutivos s_i e s_k da trajetória (linhas tracejadas da Fig. 7).

Pela definição do modelo RL, todas estas distâncias x_i, $y_{i,k}$ e $z_{i,k}$ têm distribuição uniforme (Eq. D.2) no intervalo $[0, 1]$.

As condições para que o turista percorra a trajetória s_1, s_2, ..., $s_{\tilde{n}}$ nos primeiros $\tilde{n}$ passos são:

1. No caso $\mu = 1$ (já resolvido na Ref. [30]), as distâncias x_i devem satisfazer a desigualdade $x_{\tilde{n}-1} < x_{\tilde{n}-2} < \cdots < x_1$, pois o turista não mais explora novos sítios quando $x_{i+1} > x_i$, dando origem a um ciclo de período $p = 2$. Mas para o caso $\mu = 2$, cada a distância x_i pode variar irrestritamente no intervalo $[0, 1]$, pois a memória $\mu = 2$ proíbe que o turista retroceda do sítio s_{i+1} para o sítio s_i (mesmo se $x_{i+1} > x_i$).

2. Quando o turista está prestes a percorrer a distância x_i (e mover-se de s_i para s_{i+1}), há $N - i$ sítios ainda não-explorados à sua disposição.

3. Para cada sítio s_i, todas as $N - (\tilde{n} - 1)$ distâncias $y_{i,k}$ devem ser maiores que x_i. A probabilidade disto ocorrer é $\left[\int_{x_i}^{1} \mathrm{d}y_{i,k}\right]^{N-(\tilde{n}-1)} = (1 - x_i)^{N-\tilde{n}+1}$. A única exceção é o sítio $s_{\tilde{n}-1}$, para o qual a probabilidade é $\left[\int_{x_{\tilde{n}-1}}^{1} \mathrm{d}y_{\tilde{n}-1,k}\right]^{N-\tilde{n}} = (1 - x_{\tilde{n}-1})^{N-\tilde{n}}$ (veja Fig. 7, onde $s_{\tilde{n}-1}$ corresponde a s_8).

4. Para evitar atalhos e revisitas, cada uma das distâncias $z_{i,k}$ deve ser simultaneamente maior que as distâncias x_i e x_k.

Estas condições nos levam a:

$$\begin{aligned}\tilde{F}_{2,rl}^{(N)}(\tilde{n}) &= \prod_{i=1}^{\tilde{n}-2} \int_0^1 \mathrm{d}x_i (N-i)(1-x_i)^{N-\tilde{n}+1} \int_0^1 \mathrm{d}x_{\tilde{n}-1} (N-\tilde{n}+1)(1-x_{\tilde{n}-1})^{N-\tilde{n}} \times \\ &\quad \prod_{i=1}^{\tilde{n}-3} \prod_{k=i+2}^{\tilde{n}-1} \int_{\text{máx}(x_i,x_k)}^{1} \mathrm{d}z_{i,k} \,. \end{aligned} \tag{4.1}$$

Vale observar que ainda não fizemos aproximação alguma e, por isso, a Eq. 4.1 produz resultados exatos mesmo para pequenos valores de N, como mostra a Tab. 2.

Apesar da Eq. 4.1 ser exata, a função $\text{máx}(x_i, x_k)$ nos limites inferiores das integrais em $z_{i,k}$ torna difícil sua resolução, pois teríamos que desmembrar esta expressão e considerar todas as $(\tilde{n} - 1)!$ ordenações possíveis entre as distâncias x_i. A seguir serão feitas aproximações na Eq. 4.1 para o caso $N \gg 1$ com o intuito de facilitar o cálculo.

$\tilde{n}$	$\tilde{F}^{(6)}_{2,rl}(\tilde{n})$	$\tilde{S}^{(6)}_{2,rl}(\tilde{n})$	média	erro-padrão	diferença	diferença (em erros-padrão)
3	1	0,15625	0,15624	$1 \cdot 10^{-5}$	$7 \cdot 10^{-6}$	0,62
4	27/32	0,29533	0,29535	$1 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	1,13
5	9 459/17 248	0,33785	0,33784	$1 \cdot 10^{-5}$	$1 \cdot 10^{-5}$	0,82
6	107 301/509 600	0,21055	0,21056	$1 \cdot 10^{-5}$	$3 \cdot 10^{-6}$	0,22

Tabela 2: Validação numérica da Eq. 4.1. As colunas $\tilde{F}^{(6)}_{2,rl}(\tilde{n})$ e $\tilde{S}^{(6)}_{2,rl}(\tilde{n})$ referem-se a valores analíticos e as colunas *média* e *erro padrão* representam simulações numéricas. As caminhadas foram realizadas sobre 300 000 000 mapas, sendo que cada mapa tem $N = 6$ pontos.

Observe que as integrais em $z_{i,k}$ referem-se às linhas tracejadas da Fig. 7. De cada sítio partem exatamente $\tilde{n} - 4$ linhas tracejadas, com exceção dos sítios s_1 e $s_{\tilde{n}-1}$, dos quais partem $\tilde{n} - 3$ linhas tracejadas, devido à distância adicional $z_{1,\tilde{n}-1}$ (linha tracejada espessa da Fig 7). Algebricamente, isto significa que cada uma das variáveis x_2, x_3, ..., $x_{\tilde{n}-2}$ aparece exatamente $\tilde{n}-4$ vezes na Eq. 4.1 como argumento da função máx() e apenas as variáveis x_1 e $x_{\tilde{n}-1}$ aparecem $\tilde{n}-3$ vezes, devido à integral em $z_{1,\tilde{n}-1}$. Visando obter uma expressão mais regular, podemos eliminar esta integral para que cada variável x_i apareça exatamente $\tilde{n}-4$ vezes na função máx(). Esta eliminação é plausível, pois, devido à regra determinista da caminhada, cada distância x_i é o mínimo de $N-2$ variáveis aleatórias uniformemente distribuídas no intervalo $[0, 1]$ e segue uma pdf potência, dada pela Eq. D.23, com média $\langle x_i \rangle = 1/(N-1) \approx 1/N$ e desvio-padrão $\sigma_{x_i} = \sqrt{(N-2)/[N(N-1)^2]} \approx 1/N$, de forma que no limite $N \gg 1$, x_i assume (em média) valores próximos de 0 e o valor da integral $\int_{\text{máx}(x_1, x_{\tilde{n}-1})}^{1} \mathrm{d}z_{1,\tilde{n}-1}$ tende a 1.

Alterando o expoente do integrando em $x_{\tilde{n}-1}$ de $N - \tilde{n}$ para $N - \tilde{n} + 1$, todos os integrandos passam a ter o mesmo expoente. A expressão resultante é algebricamente simétrica em relação às variáveis x_1, x_2, ..., $x_{\tilde{n}-1}$. Portanto todas as $(\tilde{n} - 1)!$ ordenações possíveis entre as variáveis x_i ocorrem com a mesma probabilidade. Assim, podemos considerar a ordenação específica $x_1 < x_2 < \cdots < x_{\tilde{n}-1}$ e reescrever a Eq. 4.1 sem usar a inconveniente função máx().

$$\begin{aligned} \frac{\tilde{F}^{(N)}_{2,rl}(\tilde{n})}{(\tilde{n}-1)!} &= \prod_{i=1}^{\tilde{n}-1} (N-i) \int_0^1 \mathrm{d}x_1 (1-x_1)^{N-\tilde{n}+1} \times \\ & \prod_{i=2}^{\tilde{n}-2} \int_{x_{i-1}}^1 \mathrm{d}x_i (1-x_i)^{N-\tilde{n}+i-1} \int_{x_{\tilde{n}-2}}^1 \mathrm{d}x_{\tilde{n}-1} (1-x_{\tilde{n}-1})^{N-3} \,, \end{aligned} \tag{4.2}$$

onde o fator extra $(\tilde{n}-1)!$ leva em consideração todas as possíveis ordenações das variáveis x_i. Alterando o expoente do integrando em x_1 de $N - \tilde{n} + 1$ para $N - \tilde{n}$, os expoentes de

x_1, x_2, ..., $x_{\tilde{n}-2}$ formam uma progressão aritmética. Resolvendo as integrais da Eq. 4.2, temos

$$\begin{aligned}\tilde{F}^{(N)}_{2,rl}(\tilde{n}) &= \frac{(\tilde{n}-1)!(N-1)(N-2)(N-3)\cdots(N-\tilde{n}+1)}{(N-2)(2N-4)(3N-7)(4N-11)\cdots\left\{(\tilde{n}-1)N-\left[\frac{(\tilde{n}-1)\tilde{n}}{2}+1\right]\right\}} \\ &= \prod_{k=1}^{\tilde{n}-1}\frac{k(N-k)}{kN-\frac{k(k+1)}{2}-1} = \prod_{j=4}^{\tilde{n}}\frac{N-j+1}{N-\frac{j}{2}-\frac{1}{j-1}}\,, \end{aligned} \tag{4.3}$$

onde aplicamos a substituição $j = k+1$ e o limite inferior do produtório foi alterado de $j = 2$ para $j = 4$ porque os fatores para $j = 2$ e $j = 3$ não têm significado físico, como mostraremos na Sec. 4.1.1. A distribuição de $\tilde{n}$ é calculada pela diferença

$$\begin{aligned}\tilde{S}^{(N)}_{2,rl}(\tilde{n}) &= \tilde{F}^{(N)}_{2,rl}(\tilde{n}) - \tilde{F}^{(N)}_{2,rl}(\tilde{n}+1) \\ &= \left[1-\frac{N-\tilde{n}}{N-\frac{\tilde{n}+1}{2}-\frac{1}{\tilde{n}}}\right]\prod_{j=4}^{\tilde{n}}\frac{N-j+1}{N-\frac{j}{2}-\frac{1}{j-1}}\,. \end{aligned} \tag{4.4}$$

A expressão da Eq. 4.3 é similar àquela obtida para o caso $\mu = 1$ (usando as Eqs. 9 e 10 da Ref. [30] e substituindo $\tilde{n} = t+2$): $\tilde{F}^{(N)}_{1,rl}(\tilde{n}) = [\prod_{j=3}^{\tilde{n}}\frac{N-j+1}{N-j/2}]/(\tilde{n}-1)!$. A principal diferença é a presença do fator $1/(\tilde{n}-1)!$. Este fatorial surge da necessidade de se impor a condição de que a distância a ser percorrida num dado passo seja menor que a distância percorrida no passo imediatamente anterior. Assim, no caso $\mu = 1$ deve-se considerar somente a ordenação espefícica $x_{\tilde{n}-1} < x_{\tilde{n}-2} < \cdots < x_2 < x_1$, e a probabilidade disto ocorrer é dada por

$$\begin{aligned}\int_{0<x_{\tilde{n}-1}<x_{\tilde{n}-2}<\cdots<x_2<x_1<1}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{\tilde{n}-2}\mathrm{d}x_{\tilde{n}-1} &= \\ \int_0^1\mathrm{d}x_1\int_0^{x_1}\mathrm{d}x_2\int_0^{x_2}\mathrm{d}x_3\cdots\int_0^{x_{\tilde{n}-3}}\mathrm{d}x_{\tilde{n}-2}\int_0^{x_{\tilde{n}-2}}\mathrm{d}x_{\tilde{n}-1} &= \frac{1}{(\tilde{n}-1)!}\end{aligned}$$

Através destes resultados, somos capazes de entender o principal papel da memória na caminhada do turista. Para $\mu = 1$, o turista deve ir imperativamente para o vizinho mais próximo. A estatística de extremos é intrínseca a esta dinâmica. Entretanto, proibindo o turista de retornar ao último sítio visitado tem-se a possibilidade de ir para o primeiro ou para o segundo vizinho mais próximo, transformando a estatística de extremos para a estatística combinatorial. Matematicamente, isto é expresso pela ausência do fator $1/(\tilde{n}-1)!$ na Eq. 4.3. Isto explica porque há uma mudança radical no comportamento exploratório do turista entre os casos $\mu = 1$ (exploração altamente localizada) e $\mu = 2$ (exploração estendida) mas entre os casos $\mu \geq 2$ a mudança é suave e gradativa, como mostra a Fig. 8.

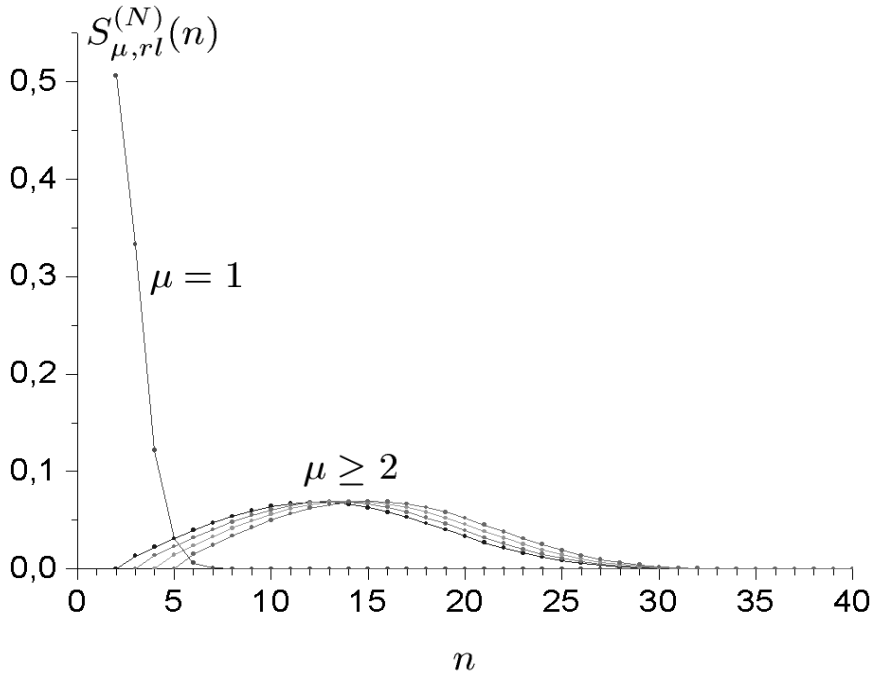


Figura 8: Distribuição do número n de sítios explorados pelo turista num mapa com $N = 40$ pontos. Da esquerda para a direita, as 4 distribuições em forma de sino se referem aos casos $\mu = 2$, $\mu = 3$, $\mu = 4$ e $\mu = 5$, respectivamente.

4.1.1 Analogia à Distribuição Geométrica

Em analogia à distribuição geométrica, podemos escrever a Eq. 4.4 como

$$\tilde{S}^{(N)}_{2,rl}(\tilde{n}) = \tilde{p}^{(N)}_{2,rl}(\tilde{n}+1) \prod_{j=4}^{\tilde{n}} \tilde{q}^{(N)}_{2,rl}(j)$$

onde

$$\tilde{q}^{(N)}_{2,rl}(j) = \frac{N-j+1}{N-\frac{j}{2}-\frac{1}{j-1}} \tag{4.5}$$

é a probabilidade de exploração no j-ésimo passo (tomada como fracasso) e

$$\tilde{p}^{(N)}_{2,rl}(j) = 1 - \tilde{q}^{(N)}_{2,rl}(j)$$

é a probabilidade de revisita no j-ésimo passo (tomada como sucesso).

Nota-se que a expressão da Eq. 4.5 é similar à expressão correspondente para o caso $\mu = 1$ (adaptando as Eqs. 9 e 10 da Ref. [30] de seus conceitos originais de probabilidade de subsistência para o conceito de probabilidade de exploração tratado aqui): $(j-1)\tilde{q}^{(N)}_{1,rl}(j) = (N-j+1)/(N-j/2)$. A principal diferença é o fator extra $j-1$, que é uma conseqüência da restrição $x_{\tilde{n}-1} < x_{\tilde{n}-2} < \cdots < x_1$. Este fator extra explica a mudança abrupta no

comportamento exploratório entre os casos $\mu = 1$ e $\mu = 2$: por um lado, para $\mu = 1$ a probabilidade de exploração (no limite termodinâmico) diminui com $1/(j-1)$ ao longo da trajetória; por outro lado, para $\mu = 2$ esta probabilidade tende a 1, quando $N \to \infty$.

Observe que a memória $\mu = 2$ garante a exploração nos três primeiros passos. Desta forma, só faz sentido falar em probabilidade de exploração a partir do quarto passo. De fato, para o primeiro passo ($j = 1$) a Eq. 4.5 não tem valor definido, para o segundo passo produz $\tilde{q}_{2,rl}^{(N)}(2) = (N-1)/(N-2) > 1$, que é um absurdo, e para o terceiro $\tilde{q}_{2,rl}^{(N)}(3) = 1$. Foi por este motivo que alteramos o limite inferior do produtório da Eq. 4.3 de $j = 2$ para $j = 4$. É interessante mencionar que para o passo $j = N+1$ (após o turista ter percorrido todos os N sítios do mapa), a Eq. 4.5 coerentemente produz $\tilde{q}_{2,rl}^{(N)}(N+1) = 0$.

Como no j-ésimo passo há $j-3$ sítios igualmente prováveis a serem revisitados e $\tilde{p}_{2,rl}^{(N)}(j)$ é a probabilidade do turista revisitar **qualquer um** destes sítios, no limite $N \gg j \gg 1$ a probabilidade $\tilde{p}_{rl}$ do turista revisitar **um sítio específico** s_k é

$$\tilde{p}_{rl} = \frac{1}{j-3}\,\tilde{p}_{2,rl}^{(N)}(j) = \frac{1}{j-3}\,\frac{\frac{j}{2} - 1 - \frac{1}{j-1}}{N - \frac{j}{2} - \frac{1}{j-1}} \approx \frac{1}{2N}\,, \tag{4.6}$$

que é metade da probabilidade do turista explorar um novo sítio específico [a saber $\tilde{q}_{rl} = 1/(N-j) \approx 1/N$].

4.1.2 Dedução Alternativa

No que segue, obteremos expressões mais simples para as probabilidades de revisita e de exploração para $N \gg 1$, através de um raciocínio alternativo. A partir destas probabilidades, obteremos expressões analíticas fechadas para $\tilde{F}_{2,rl}^{(N)}(\tilde{n})$.

4.1.2.1 Probabilidades de Revisita e de Exploração

Suponha que o turista tenha percorrido a trajetória s_1, s_2, ..., $s_{\tilde{n}}$ ($\tilde{n} > 3$) sem ter revisitado sítio algum. Inicialmente, vamos reobter a probabilidade $\tilde{p}_{rl}$ do turista revisitar um sítio específico s_k (fora da janela de exclusão, isto é, $k \leq \tilde{n} - 2$) no próximo passo. Para tanto, temos as seguintes condições (veja Fig. 7):

1. a distância $z_{\tilde{n},k}$ deve ser menor que $x_{\tilde{n}}$.

2. como no $(k+1)$-ésimo passo o turista migrou do sítio s_k para o sítio s_{k+1}, a distância $z_{\tilde{n},k}$ é, por conseqüência, maior que a distância x_k.

Resumidamente, $z_{\tilde{n},k}$ deve variar entre x_k e $x_{\tilde{n}}$, ou seja, $0 < x_k < z_{\tilde{n},k} < x_{\tilde{n}} < 1$.

Como cada distância x_i segue uma pdf potência $g(x_i) = (N-2)(1-x_i)^{N-3}$, dada pela Eq. D.23, e $z_{\tilde{n},k}$ tem distribuição uniforme (pela definição do modelo RL), para $N \gg 1$ a probabilidade $\tilde{p}_{rl}$ é dada por

$$\begin{aligned}
\tilde{p}_{rl} &= P(x_k < z_{\tilde{n},k} < x_{\tilde{n}}) \\
&= \int_0^1 \mathrm{d}x_k (N-2)(1-x_k)^{N-3} \int_{x_k}^1 \mathrm{d}x_{\tilde{n}} (N-2)(1-x_{\tilde{n}})^{N-3} \int_{x_k}^{x_{\tilde{n}}} \mathrm{d}z_{\tilde{n},k} \\
&= \frac{N-2}{(N-1)(2N-3)} \approx \frac{1}{2N} \ ,
\end{aligned}$$

que concorda com a Eq. 4.6.

Para um passo j genérico há $j-3$ sítios suscetíveis à revisita, e então as probabilidades de revisita e de exploração para este passo são:

$$\tilde{p}_{2,rl}^{(N)}(j) = \frac{j-3}{2N} = 1 - \tilde{q}_{2,rl}^{(N)}(j) \ ,$$

que é uma aproximação para a Eq. 4.5, e de onde resulta

$$\begin{aligned}
\tilde{F}_{2,rl}^{(N)}(\tilde{n}) &= \prod_{j=4}^{\tilde{n}} \tilde{q}_{2,rl}^{(N)}(j) = \prod_{j=4}^{\tilde{n}} \left[1 - \frac{j-3}{2N}\right] \\
&= \frac{\Gamma(2N)}{\Gamma(2N-\tilde{n}+3)(2N)^{\tilde{n}-3}} \ ,
\end{aligned} \tag{4.7}$$

que é uma forma analítica fechada para a Eq. 4.3.

4.1.2.2 Forma Exponencial (Acumulada Meia-Gaussiana)

No limite $N \gg 1$, a Eq. 4.5 pode ser escrita como

$$\tilde{q}_{2,rl}^{(N)}(j) = \left(1 - \frac{1}{2N}\right)^{j-3}$$

de forma que a Eq. 4.7 assume sua forma exponencial:

$$\begin{aligned}
\tilde{F}_{2,rl}^{(N)}(\tilde{n}) &= \prod_{j=4}^{\tilde{n}} \tilde{q}_{2,rl}^{(N)}(j) = \left(1 - \frac{1}{2N}\right)^{\tilde{\omega}} \\
&\approx e^{-\tilde{\omega}/(2N)} = \exp\left[-\frac{(\tilde{n}-3)^2}{4N}\left(1 + \frac{1}{\tilde{n}-3}\right)\right] \ ,
\end{aligned} \tag{4.8}$$

onde

$$\tilde{\omega} = \sum_{j=4}^{\tilde{n}} (j-3) = \frac{(\tilde{n}-2)(\tilde{n}-3)}{2} \tag{4.9}$$

admite uma interpretação física simples: é justamente o número de distâncias $z_{i,k}$ entre sítios não-consecutivos da trajetória. Observe que a trajetória exibida na Fig. 7 é topologicamente equivalente a um polígono de $\tilde{n}-1$ lados, que tem $(\tilde{n}-1)(\tilde{n}-4)/2$ diagonais. Todas estas diagonais juntamente com o lado $s_1 s_{\tilde{n}-1}$ totalizam $\tilde{\omega} = (\tilde{n}-2)(\tilde{n}-3)/2$ caminhos (linhas tracejadas da Fig. 7), que possibilitam a revisita.

Para $\tilde{n}-3 \gg 1$, pode-se ignorar $1/(\tilde{n}-3)$ na Eq. 4.8, o que leva a uma distribuição Meia-Gaussiana (Eq D.30): $y = \tilde{F}^{(N)}_{2,rl}(\tilde{n}) = e^{-[(\tilde{n}-3)/\sqrt{2N}]^2/2}$, indicando que a variável escalada é $x = (\tilde{n}-\tilde{n}_{min})/\sqrt{2N}$ com $\tilde{n}_{min} = \mu+1 = 3$, levando à curva universal $y = e^{-x^2/2}$, com $x \geq 0$. Observe que nós apenas mantivemos $\tilde{n}_{min}$ para eventualmente num futuro próximo podermos comparar com uma possível generalização destes cálculos para o caso de memória pequena $\mu \ll N$.

4.2 Distribuição do Número Total de Sítios Explorados

Até este ponto estivemos focados no número $\tilde{n}$ de sítios explorados antes da primeira revisita a algum sítio. Na caminhada do turista com $\mu = 1$, a revisita a um sítio implica que o turista entrou num atrator de período $p = 2$ e encerrou a exploração de novos sítios [30]. Porém, com $\mu = 2$ a revisita não implica em captura. No que segue, calcularemos a probabilidade p_t do turista ser aprisionado durante uma revisita a um sítio e então obteremos as probabilidades de captura $p^{(N)}_{2,rl}(j)$ e subsistência $q^{(N)}_{2,rl}(j)$, assim como a distribuição acumulada complementar $F^{(N)}_{2,rl}(n)$ para o número n de sítios visitados em toda a trajetória.

4.2.1 Probabilidades de Revisita e de Aprisionamento em Função do Posto

Considere que o turista tenha percorrido a trajetória $s_1, s_2, \ldots, s_{\tilde{n}}$ ilustrada na Fig. 7 sem revisita alguma. Suponha que no próximo passo ele revisite o sítio s_k (fora da janela de exclusão, $k \leq \tilde{n}-2$). Devido à regra determinista, duas situações podem ocorrer:

- se $x_k < x_{k-1}$, o turista avança para o sítio s_{k+1} e é aprisionado por um atrator de período $p = \tilde{n}-k+1$.
- se $x_{k-1} < x_k$, o turista retrocede para o sítio s_{k-1} e foge do atrator.

Portanto, o aprisionamento ou fuga depende de qual das distâncias x_{k-1} ou x_k é a menor. A única exceção é uma revisita ao sítio de partida s_1, quando o turista é incondicionalmente aprisionado, dando origem a uma trajetória com tempo de transiente nulo ($t = 0$) e um ciclo de período $p = \tilde{n}$.

Levando em conta que todas as $(\tilde{n} - 1)!$ ordenações possíveis entre as distâncias x_1, x_2, ..., $x_{\tilde{n}-1}$ são igualmente prováveis, poderíamos ingenuamente concluir que a probabilidade de aprisionamento (trapping probability) é $p_t = P(x_k < x_{k-1}) = 1/2$, já que x_k e x_{k-1} são variáveis independentes e identicamente distribuídas (i.i.d.). Contrariando esta expectativa, simulações numéricas revelaram que esta probabilidade é de fato $p_t = 2/3$.

Para entender este resultado, inicialmente mostraremos que a probabilidade $P_v(r)$ do turista revisitar um sítio específico s_k é proporcional ao posto r ocupado pela distância associada x_k (entre os sítios s_k e s_{k+1}) quando se reordenam as distâncias x_1, x_2, ..., $x_{\tilde{n}-2}$ em ordem decrescente. Usaremos a notação $x_{(r)}$ para designar a r-ésima maior distância. Em seguida, mostraremos que a probabilidade $P_t(r)$ do turista ser aprisionado ao revisitar o sítio s_k é proporcional a $r-1$. Finalmente, usando $P_v(r)$ e $P_t(r)$ provaremos que $p_t = 2/3$.

4.2.1.1 Probabilidade de Revisita

Novamente, considere que o turista tenha percorrido a trajetória s_1, s_2, ..., $s_{\tilde{n}}$ (sem revisita a sítio algum). Calculemos a probabilidade $P_v(r)$ dele revisitar, no próximo passo, o sítio $s_{(r)} = s_k$, cuja distância associada $x_{(r)} = x_k$ é a r-ésima maior dentre as distâncias x_i. Como cada distância x_i tem pdf potência $g(x) = (N-2)(1-x)^{N-3}$ dada pela Eq. D.23 e distribuição acumulada $G(x) = 1 - (1-x)^{N-2}$ dada pela Eq. D.24, a pdf de $x_{(r)}$ é dada pela Eq. D.27

$$h_r(x) = \frac{\tilde{n}!}{(r-1)!(\tilde{n}-r)!} \left[1 - (1-x)^{N-2}\right]^{\tilde{n}-r} \left[(1-x)^{N-2}\right]^{r-1} (N-2)(1-x)^{N-3} \, .$$

Como $s_{(r)}$ é mais próximo dentre os sítios permitidos (devido à regra determinista da caminhada), a distância $z_{\tilde{n},(r)}$ é o mínimo de $N-2$ variáveis com distribuição uniforme e segue uma pdf potência dada pela Eq. D.23. Como o turista se moveu do sítio $s_{(r)} = s_k$ para o sítio s_{k+1} no $(k+1)$-ésimo passo, a distância $z_{\tilde{n},(r)}$ é certamente maior que $x_{(r)}$.

Assim,

$$
\begin{aligned}
P_v(r) &\propto P(z_{\tilde{n},(r)} > x_{(r)}) \\
&= \frac{\tilde{n}!}{(r-1)!(\tilde{n}-r)!} \int_0^1 \mathrm{d}x \left[1-(1-x)^{N-2}\right]^{\tilde{n}-r} \left[(1-x)^{N-2}\right]^{r-1} (N-2)(1-x)^{N-3} \times \\
&\quad \int_x^1 \mathrm{d}z (N-2)(1-z)^{N-3}.
\end{aligned}
$$

Calculando a integral em z e substituindo $y = (1-x)^{N-2}$ a integral resultante é identificada com a função beta (Eq. B.10):

$$
P_v(r) \propto \frac{\tilde{n}!}{(r-1)!(\tilde{n}-r)!} \mathrm{B}(\tilde{n}-r+1, r+1) = \frac{\tilde{n}!}{(r-1)!(\tilde{n}-r)!} \frac{(\tilde{n}-r)!\, r!}{(\tilde{n}+1)!} = \frac{r}{\tilde{n}+1}.
$$

Esta expressão *não* é a probabilidade P_v em si. Ao invés disto, ela apenas dá a dependência de P_v para com r.

Normalizando P_v sobre $1 \leq r \leq \tilde{n}-2$, tem-se

$$
P_v(r) = \frac{r}{\sum_{k=1}^{\tilde{n}-2} k} = \frac{2r}{(\tilde{n}-1)(\tilde{n}-2)}\ , \tag{4.10}
$$

onde $\tilde{n}-2$ é o número de sítios disponíveis à revisita (os sítios $s_{\tilde{n}}$ e $s_{\tilde{n}-1}$ são proibidos pela memória).

O resultado da Eq. 4.10 não contradiz a Eq. 4.6, pois a Eq. 4.6 fornece a probabilidade aproximada do turista revisitar um sítio específico s_k, independentemente de sua distância associada $x_k = x_{(r)}$, enquanto que a Eq. 4.10 fornece a probabilidade condicional do turista "escolher" o sítio $s_{(r)}$ de posto r durante uma revisita após de ter explorado $\tilde{n}$ sítios distintos.

4.2.1.2 Probabilidade de Aprisionamento

Uma vez que o turista revisitou o sítio s_k (ou equivalentemente $s_{(r)}$), a probabilidade $P_t(r)$ dele ser aprisionado também depende do posto r. A condição de aprisionamento é: x_{k-1} deve ser maior que x_k. Como $x_k = x_{(r)}$ é a r-ésima maior distância, restam somente $r-1$ distâncias (dentre $\tilde{n}-3$) maiores que x_k. Considerando que todas as ordenações das distâncias x_i são equiprováveis, temos

$$
P_t(r) = \frac{r-1}{\tilde{n}-3}\ . \tag{4.11}
$$

Combinando as Eqs. 4.10 e 4.11, a probabilidade do turista ser aprisionado ao revisitar

um sítio específico $s_{(r)}$ é:

$$P_v(r)P_t(r) = \frac{2r(r-1)}{(\tilde{n}-1)(\tilde{n}-2)(\tilde{n}-3)} \ .$$

Logo, a probabilidade do turista ser aprisionado ao revisitar um sítio qualquer é $p_t = \sum_{r=1}^{\tilde{n}-2} P_v(r)P_t(r)$. Substituindo $m = \tilde{n} - 2$ chegamos à probabilidade de aprisionamento

$$p_t = \sum_{r=1}^{m} P_v(r)P_t(r) = \frac{2}{m(m^2-1)} \sum_{r=1}^{m} r(r-1) = \frac{2}{m(m^2-1)} \frac{m(m^2-1)}{3} = \frac{2}{3} \ . \quad (4.12)$$

Vale mencionar que este resultado foi obtido sem aproximação alguma, e simulações numéricas concordam com ele mesmo para pequenos valores de N.

4.2.2 Probabilidades de Captura e de Subsistência

Combinando a probabilidade $\tilde{p}_{rl}$ do turista revisitar um sítio s_k específico (Eq. 4.6) e a probabilidade de aprisionamento p_t (Eq. 4.12), obtemos a probabilidade p_{rl} do turista revisitar s_k **e** ser aprisionado:

$$p_{rl} = \tilde{p}_{rl} p_t = \frac{1}{2N}\frac{2}{3} = \frac{1}{3N} \ . \quad (4.13)$$

Como no j-ésimo passo há $j-3$ sítios disponíveis à revisita, as probabilidade de captura (isto é, revisitar um sítio qualquer **e** ser aprisionado) e subsistência (isto é, explorar um novo sítio qualquer **ou** revisitar um sítio qualquer e não ser aprisionado) no j-ésimo passo são:

$$p_{2,rl}^{(N)}(j) = \frac{j-3}{3N} = 1 - q_{2,rl}^{(N)}(j)$$

e a distribuição acumulada complementar para o número n de sítios explorados pelo turista em toda a trajetória é

$$F_{2,rl}^{(N)}(n) = \prod_{j=4}^{n} q_{2,rl}^{(N)}(j) = \frac{\Gamma(3N)}{\Gamma(3N-n+3)(3N)^{n-3}} \ , \quad (4.14)$$

que é análoga à Eq. 4.7.

4.2.2.1 Comparação com o modelo Random Map

A expressão da Eq. 4.14 é similar àquela obtida para o modelo RM com memória $\mu = 0$ [30]: $F_{0,rm}^{(N)}(n) = \Gamma(N)/[\Gamma(N-n)N^n]$. Este resultado explica a equivalência não-trivial observada entre o modelo RL com N pontos e memória $\mu = 2$ (efeito da memória)

e o modelo RM com $3N$ pontos e memória $\mu = 0$ (efeito da quebra da simetria das distâncias), quando se comparam as distribuições do número total n de sítios explorados pelo turista.

Note que, tomando ambos os modelos com N pontos cada, no RL com $\mu = 2$, a cada passo, a probabilidade do turista revisitar um sítio específico e ser aprisionado é $p_{rl} \approx 1/(3N)$; enquanto que no RM com $\mu = 0$, esta probabilidade é $p_{rm} = 1/N$. Portanto, tomando o RL com N pontos e o RM com $3N$ pontos estas probabilidades se igualam, justificando a equivalência entre estes modelos.

4.2.2.2 Forma Exponencial

No limite $N \gg 1$, a probabilidade de subsistência é reescrita como $q_{2,rl}^{(N)}(j) = [1 - 1/(3N)]^{j-3}$ e obtém-se a forma exponencial da Eq. 4.14:

$$F_{2,rl}^{(N)}(n) = \prod_{j=4}^{n} q_{2,rl}^{(N)}(j) = \left(1 - \frac{1}{3N}\right)^{\omega} \approx e^{-\omega/(3N)} \ ,$$

com $\omega = (n-2)(n-3)/2$.

Em vez de derivar $F_{2,rl}^{(N)}(n)$, a distribuição $S_{2,rl}^{(N)}(n)$ do número n de sítios explorados em toda a trajetória é mais precisamente obtida impondo que o turista explore n sítios distintos e seja capturado no próximo passo (isto é, revisite **qualquer** sítio e seja aprisionado):

$$S_{2,rl}^{(N)}(n) = F_{2,rl}^{(N)}(n)\, p_{2,rl}^{(N)}(n+1) = \frac{n-2}{3N}\, e^{-\frac{(n-2)(n-3)/2}{3N}} \ . \tag{4.15}$$

Para $n \gg 1$, substituindo $y = \sqrt{3N} S_{2,rl}^{(N)}(n)$ e $x = (n - n_{min})/\sqrt{3N}$, com $n_{min} = \mu + 1 = 3$, obtém-se a curva universal para este sistema:

$$y = x\, e^{-x^2/2} \ , \tag{4.16}$$

com $x \geq 0$. Como o m-ésimo momento recai numa integral Gaussiana de ordem $m+1$ (dada pela Eq. A.3), resultando em $\langle x^m \rangle = 2^{m/2}\Gamma(m/2+1)$, a normalização é garantida por $\langle x^0 \rangle = 1$, a média é $\langle x \rangle = \sqrt{\pi/2}$ e a variância $\langle x^2 \rangle - \langle x \rangle^2 = 2 - \pi/2$. O gráfico da Fig. 9 ilustra a comparação da Eq. 4.16 com dados gerados via simulação numérica do modelo considerado. A partir desta figura, ou calculando analiticamente, obtém-se que a moda é $x = 1$.

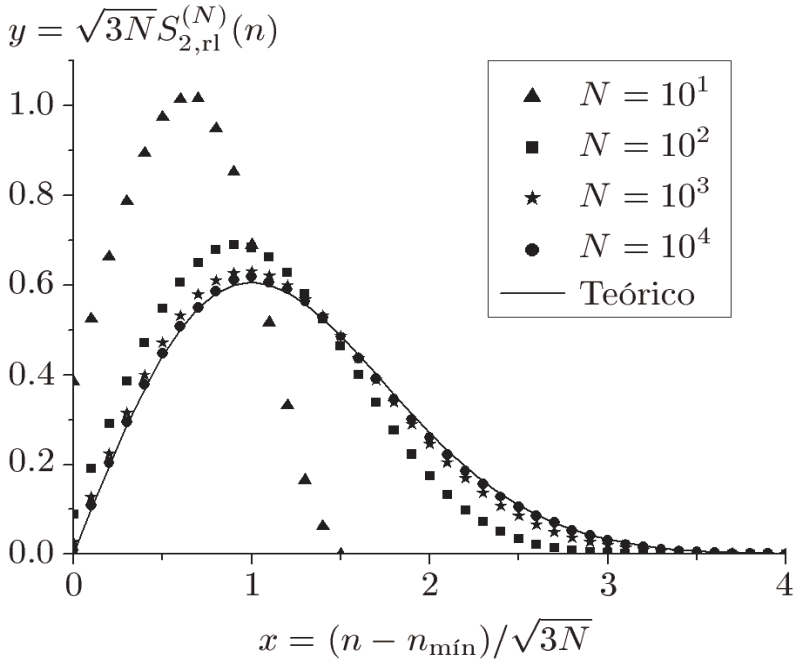


Figura 9: Efeito de tamanho finito para a distribuições e convergência para a curva universal: $y = x\, e^{-x^2/2}$.

4.3 Distribuição Conjunta de Transiente e Período

A distribução conjunta $S_{2,rl}^{(N)}(t,p)$ de transiente e período pode ser obtida de maneira similar à Eq. 4.15, impondo que turista explore n sítio distintos e revisite o sítio **específico** s_k (ao invés de **qualquer sítio**) e seja aprisionado, gerando uma trajetória com transiente $t = k - 1$ e período $p = n - k + 1$. Logo, $S_{2,rl}^{(N)}(t,p)$ é obtida multiplicando $F_{2,rl}^{(N)}(t+p)$ por p_{rl} (Eq. 4.13) [ou por $\tilde{p}_{rl}$ (Eq. 4.6) no caso $t = 0$, já que o turista é incondicionalmente aprisionado ao revisitar o sítio s_1]:

$$S_{2,rl}^{(N)}(t,p) = \frac{1}{(3-\delta_{t,0})N}\, e^{-\frac{(t+p-2)(t+p-3)/2}{3N}} \tag{4.17}$$

onde $\delta_{i,j}$ é o delta de Kronecker. Enfatizamos que a variável relevante é $t + p = n$. O gráfico da Fig. 10 exibe um traçado da Eq. 4.17 para $N = 1\,000$ pontos.

4.3.1 Distribuição Marginal de Tempos de Transiente

A distribuição de tempos de transiente é calculada somando Eq. 4.17 sobre todos os possíveis períodos, ou seja, $S_{2,rl}^{(N)}(t) = \sum_{p=3}^{N} S_{2,rl}^{(N)}(t,p)$. No limite $N \gg 1$, este somatório

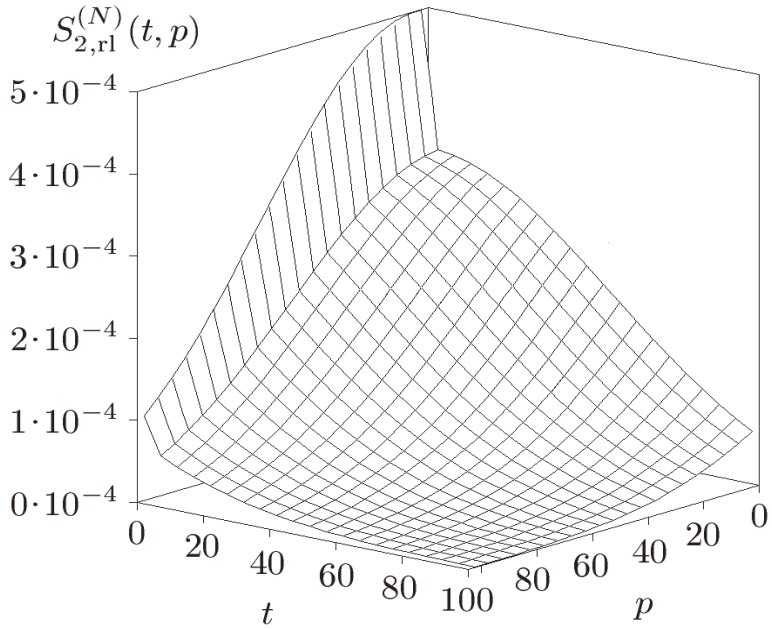


Figura 10: Distribuição conjunta de transiente e período (Eq. 4.17) para um mapa com $N = 1\,000$ pontos no modelo RL com $\mu = 2$.

pode ser aproximado pela integral

$$S^{(N)}_{2,rl}(t) = \int_{5/2}^{\infty} \mathrm{d}p\; S^{(N)}_{2,rl}(t,p) = \left(1 + \frac{\delta_{t,0}}{2}\right)\sqrt{\frac{\pi}{6N}}\,\mathrm{erfc}\left(\frac{t}{\sqrt{6N}}\right) , \tag{4.18}$$

onde o limite inferior 5/2 deve-se à correcão de continuidade de Yates[1] (que além de melhorar a aproximação da integral, torna mais simples a forma analítica) e o limite superior foi extendido ao infinito para facilitar o cálculo (pois a Eq. 4.17 produz valores próximos a 0 para $p > N$) e $\mathrm{erfc}(x)$ é a função erro complementar, dada pela Eq. B.2.

4.3.2 Distribuição Marginal de Períodos dos Ciclos

Similarmente, a distribução de períodos dos ciclos é

$$\begin{aligned} S^{(N)}_{2,rl}(p) &= \sum_{t=0}^{N-3} S^{(N)}_{2,rl}(t,p) = \int_{-1}^{\infty} \mathrm{d}t\; \frac{1}{3N}\, e^{-\frac{(t+p-2)(t+p-3)/2}{3N}} = \sqrt{\frac{\pi}{6N}}\,\mathrm{erfc}\left(\frac{p-7/2}{\sqrt{6N}}\right) \\ &\approx \frac{e^{-p^2/(6N)}}{p} , \end{aligned}$$

onde o limite inferior -1 deve-se tanto à correção de continuidade de Yates quanto a uma compensação para a parcela extra em $t = 0$ (representada pela função delta de Kronecker na Eq. 4.17). O período médio é $\langle p\rangle = \sqrt{3\pi N/8}$ e o desvio-padrão é $\sigma_p = \sqrt{(2 - 3\pi/8)N}$.

[1] A correção de continuidade de Yates é utilizada para amenizar o erro que surge ao se aproximar uma distribuição discreta por uma distribuição contínua (envelope).

4.4 Conclusão

Neste Capítulo, obtivemos analiticamente as distribuções estatísticas para a caminhada determinista do turista com memória $\mu = 2$ no modelo RL. A distribuição para o número de sítios explorados antes da primeira revisita foi comparada àquela previamente obtida para o caso $\mu = 1$, elucidando o mecanismo que aumenta fortemente o comportamento exploratório do turista. Este mecanismo é explicado como segue. De um lado, para $\mu = 1$ as distâncias percorridas a cada passo devem obedecer à ordenação $x_1 > x_2 > \ldots$, levando a uma exploração localizada. No limite termodinâmico, o número médio de sítios explorados é $\langle n \rangle = e = 2,71828\ldots$ e probabilidade de exploration no j-ésimo passo ($j \geq 4$) é $1/(j-1)$. A estatística de extremos está embutida nesta dinâmica. Por outro lado, para $\mu = 2$ as distâncias $x_1, x_2, \ldots$ não estão sujeitas a restrição alguma, levando a uma exploração extendida: $\langle n \rangle$ é proporcional a $N^{1/2}$ e a probabilidade de exploração no j-ésimo passo (com $j \ll N$) tende a 1, à medida que $N \to \infty$. A estatística combinatorial está embutida nesta dinâmica. O fator $(\tilde{n}-1)!$ da Eq. 4.2 representa a mudança da estatística de extremos para a estatística combinatorial, responsável por tornar a distribuição $\delta_{p,2}$ do caso $\mu = 1$ numa distribuição dispersa ($\propto 1/p$) do caso $\mu \geq 2$.

Através da probabilidade de aprisionamento $p_t = 2/3$ (cujo valor é contra-intuitivo), obtivemos as probabilidades de captura e de subsistância, assim como uma forma fechada para a distribuição acumulada complementar para o número de sítios explorados em toda a trajetória. Esta distribuição é análoga àquela obtida para o modelo RM com $\mu = 0$. Este resultado explica a equivalência entre estes modelos de campo médio (RL com N pontos e memória $\mu = 2$; e RM com $3N$ pontos e memória $\mu = 0$). Para um grande número de sítios ($N \gg 1$) no meio desordenado, a distribuição $S_{2,rl}^{(N)}(n)$ de se ter n sítios distintos visitados pelo turista com memória $\mu = 2$ no modelo random link é dada pela curva universal $y = x\, e^{-x^2/2}$ com $y = \sqrt{3N} S_{2,rl}^{(N)}(n)$ e $x = (n-3)/\sqrt{3N}$.

A distribuição conjunta $S_{2,rl}^{(N)}(t,p) = e^{[(t+p-3)^2/(3N)]/2}/[N(3-\delta_{t,0})]$ de tempo de transiente t e período de ciclo p foi obtida notando-se que a variável relevante é aproximadamente dada por $t + p = n$.

As distribuições marginais também são universais. Para o tempo de transiente tém-se: $y = [1+\delta(x)/2]\mathrm{erfc}(x)$ com $y = \sqrt{6N/\pi} S_{2,rl}^{(N)}(t)$ e $x = t/\sqrt{6N}$ e para o de período de ciclo: $y = \mathrm{erfc(x)}$, com $y = \sqrt{6N/\pi} S_{2,rl}^{(N)}(p)$ e $x = (p-7/2)/\sqrt{6N}$. Mostramos que a discrepância entre os transientes nulos ($t = 0$) e os trasientes subseqüentes ($t > 0$) deve-se à maior probabilidade de aprisionamento que o sítio de partida s_1 tem [a saber, $\tilde{p}_{rl} = 1/(2N)$] em relação aos outros sítios [$p_{rl} = 1/(3N)$].

5 *Sistemas Unidimensionais*

Neste Capítulo, obtemos resultados analíticos para a caminhada do turista com memória μ arbitrária no Sistema Unidimensional. Iniciamos com o meio semi-infinito, obtendo uma forma exata e outra aproximada para a distribuição do número de sítios explorados. Através da distribuição acumulada complementar, calculamos as probabilidades de exploração e de recuo para este meio. A seguir, abordamos o meio finito, determinando a probabilidade de percolação e mostrando a existência de uma memória crítica que separa o comportamento do caminhante em dois regimes. Justificamos analiticamente a equivalência entre estes dois meios, observada previamente em simulações numéricas.

5.1 Meio Semi-infinito

O meio desordenado semi-infinito pode ser representado por infinitos pontos distribuídos aleatoriamente e de maneira independente sobre uma semi-reta com uma densidade média de r pontos por unidade de comprimento. A Fig. 11 representa este meio, onde as distâncias x_k entre sítios consecutivos são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) segundo uma função de densidade de probabilidade exponencial, cuja pdf é dada pela Eq. D.15. A seguir obteremos distribuições estatísticas relativas à caminhada determinista do turista com memória μ arbitrária sobre o meio semi-infinito.

5.1.1 Distribuição do Número de Sítios Explorados

Nesta Seção obteremos analiticamente a probabilidade $S^{(\infty)}_{\mu,si}(n)$ do turista, com memória μ, visitar n pontos de um meio semi-infinito. Calcularemos inicialmente o resultado exato, que será usado para justificar uma aproximação simples de campo médio.

5.1.1.1 Resultado Exato

Considere que o turista parta do sítio s_1, localizado na origem da semi-reta da Fig. 11. As condições para o turista explorar $n \geq n_{min} = \mu + 1$ sítios distintos são as seguintes:

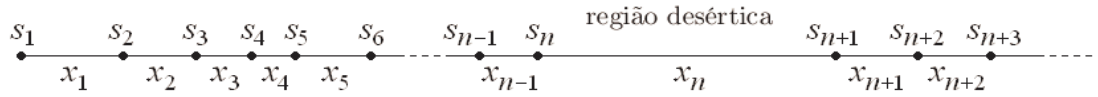


Figura 11: Meio semi-infinito. O sítio s_1 está localizado na origem da semi-reta.

1. as primeiras μ distâncias x_1, x_2, ..., x_μ podem assumir qualquer valor no intervalo $[0, \infty)$, pois a memória μ proíbe que o turista retroceda nos μ primeiros passos, garantindo a visitação aos $\mu + 1$ primeiros sítios s_1, s_2, ..., s_μ, $s_{\mu+1}$;

2. para que novos sítios sejam visitados nos passos seguintes, cada uma das distâncias $x_{\mu+1}$, $x_{\mu+2}$, ..., x_{n-1} deve ser menor que a soma das distâncias percorridas nos μ passos anteriores, até que o turista chegue, enfim, ao sítio s_n;

3. a distância x_n deve ser maior que a soma das distâncias percorridas nos μ passos anteriores, para que o turista, após alcançar o sítio s_n, recue ao sítio $s_{n-\mu}$ (ao invés de explorar s_{n+1}) e encerre a exploração de novos sítios.

Uma vez que o turista tenha recuado ao sítio $s_{n-\mu}$, nos próximos passos ele poderá migrar de volta ao sítio de partida s_1, ficar preso num atrator, ou até mesmo retornar ao sítio s_n; mas nenhuma manobra o fará transpor a região desértica entre s_n e s_{n+1}. A exploração de novos sítios estará, de fato, encerrada. Reunindo estas condições, a probabilidade do turista visitar n pontos distintos é

$$S_{\mu,si}^{(\infty)}(n) \;=\; \prod_{j=1}^{\mu} \int_0^\infty \mathrm{d}x_j r e^{-rx_j} \prod_{j=\mu+1}^{n-1} \int_0^{\sum_{k=j-\mu}^{j-1} x_k} \mathrm{d}x_j r e^{-rx_j} \int_{\sum_{k=n-\mu}^{n-1} x_k}^{\infty} \mathrm{d}x_n r e^{-rx_n} \,. \tag{5.1}$$

A dificuldade de se calcular $S_{\mu,si}^{(\infty)}(n)$ deve-se ao fato de que as n integrais são encadeadas, e o processo de integração deve começar pela integral em x_n. Podemos simplificar seus integrandos aplicando as substituições

$$y_j = e^{-rx_j} \;\Rightarrow\; \mathrm{d}y_j = -\mathrm{d}x_j r e^{-rx_j} \,, \qquad j = 1, 2, \ldots, n \,. \tag{5.2}$$

obtendo

$$S_{\mu,si}^{(\infty)}(n) \;=\; \prod_{j=1}^{n} \mathcal{I}_j \,, \tag{5.3}$$

onde a forma de cada funcional $\mathcal{I}_j$ depende de j:

$$\mathcal{I}_j = \begin{cases} \int_0^1 \mathrm{d}y_j \ , & \text{para } 1 \leq j \leq \mu \\ \int_{\tilde{y}_j}^1 \mathrm{d}y_j \ , & \text{para } \mu+1 \leq j \leq n-1 \\ \int_0^{\tilde{y}_j} \mathrm{d}y_j \ , & \text{para } j = n \end{cases} \tag{5.4}$$

e cada limite de integração $\tilde{y}_j = \prod_{k=j-\mu}^{j-1} y_k$ liga $\mathcal{I}_j$ às μ integrais precedentes. Isto significa que a Eq. 5.3 deve ser calculada de $\mathcal{I}_n$ para $\mathcal{I}_1$. Observe que as substituições de variáveis da Eq. 5.2 eliminaram a grandeza r, indicando que o número de sítios explorados pelo turista não depende da densidade do meio.

O esquema da Fig. 12 ilustra o cálculo da Eq. 5.3 para o caso particular $\mu = 3$ e $n = 8$. O valor de $S_{3,si}^{(\infty)}(8)$ é obtido somando-se todas as frações da última coluna à direita.

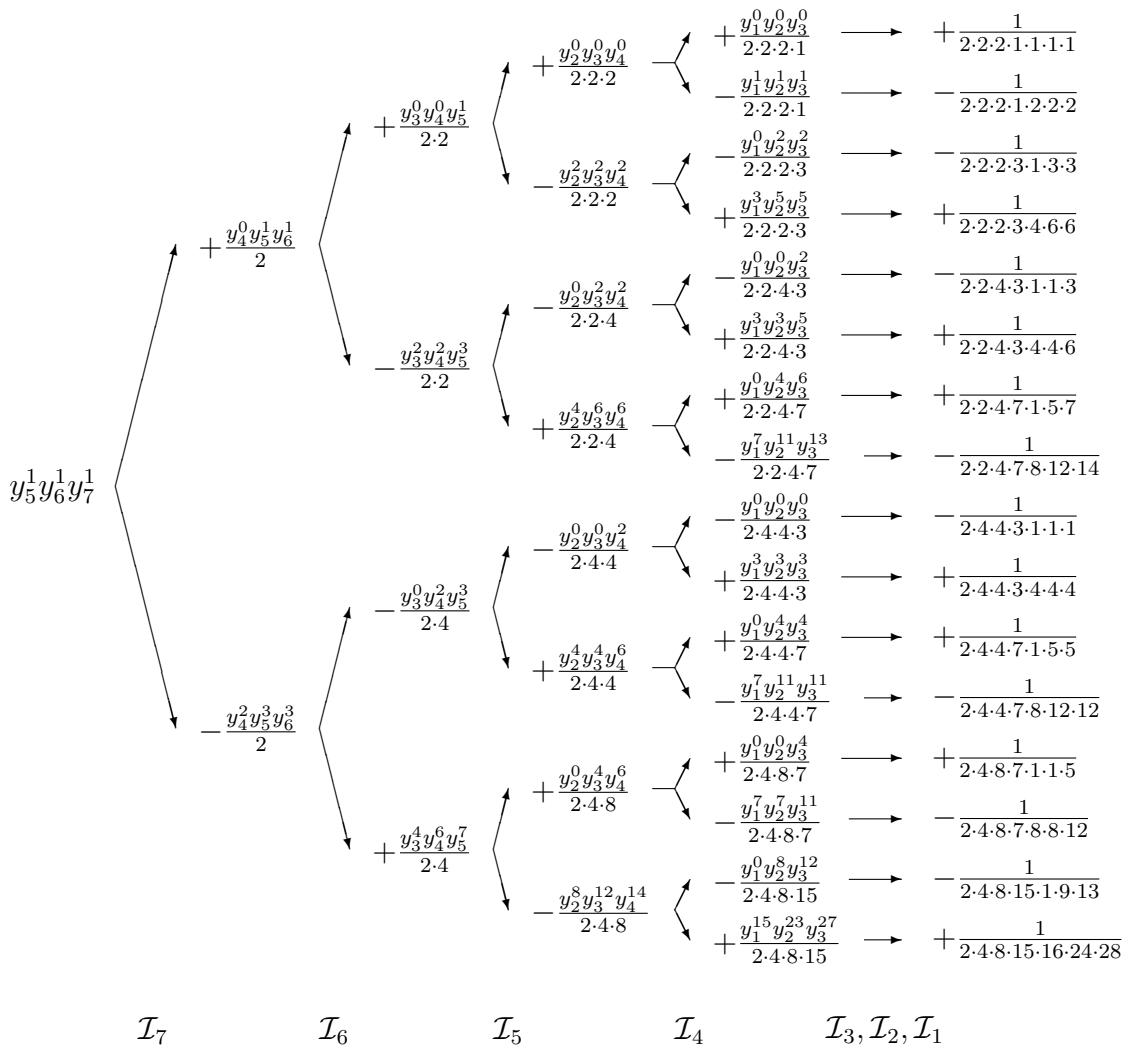


Figura 12: Exemplo de cálculo das integrais encadeadas da Eq. 5.3 para o caso particular $\mu = 3$ e $n = 8$.

O processo de integração consiste basicamente em três etapas, que correspondem às sentenças da Eq. 5.3.

1. O cálculo da primeira integral $\mathcal{I}_8$ (terceira sentença da Eq. 5.3) é trivial e resulta em seu limite superior $\tilde{y}_8$, produzindo o nó raiz $y_5^1 y_6^1 y_7^1$ da árvore, com todos os integrandos elevados à primeira potência. Denotemos estes expoentes por a_1, a_2, ..., a_μ, sendo que, em particular, a_μ é o expoente do integrando do passo de integração corrente.

2. Cada bifurcação da árvore representa uma integral de $\mathcal{I}_7$ a $\mathcal{I}_4$ (segunda sentença da Eq. 5.3). Esta etapa consiste basicamente em:

   (a) incrementar o expoente a_μ e tomá-lo como um novo fator $a_\mu+1$ no denominador do passo de integração corrente [o que equivale à integral elementar $\int y^a \mathrm{d}y = y^{a+1}/(a+1)$];

   (b) no ramo superior de cada bifurcação, a nova variável y é elevada a 0 e as demais mantém seus expoentes com os valores do passo de integração anterior;

   (c) no ramo inferior, invertemos o sinal da fração (em relação ao passo anterior) e o expoente de cada variável y é obtido somando-se $a_\mu + 1$ ao expoente da variável correspondente no ramo superior.

3. A última etapa representa o cálculo das integrais de $\mathcal{I}_3$ a $\mathcal{I}_1$ (primeira sentença da Eq. 5.3), onde todos os expoentes a_1, a_2, ..., a_μ são incrementados e se tornam novos fatores no denominador.

Observe que todas as variáveis y_1, y_2, ..., y_7 desaparecem após todas as etapas de integração. Apenas seus expoentes são relevantes, pois deles dependem os denominadores de cada parcela final. O processo pode ser mais facilmente sistematizado se focalizarmos a atenção nestes expoentes. Por exemplo, podemos representar a bifurcação

$$\int_{y_2 y_3 y_4}^{1} \mathrm{d}y_5 \, \frac{y_3^4 y_4^6 y_5^7}{2 \cdot 4} \quad = \quad \frac{y_2^0 y_3^4 y_4^6}{2 \cdot 4 \cdot 8} - \frac{y_2^8 y_3^{12} y_4^{14}}{2 \cdot 4 \cdot 8}$$

em notação vetorial como

$$\frac{(4,\ 6,\ 7)}{2 \cdot 4} \quad \mapsto \quad \frac{(0,\ 4,\ 6)}{2 \cdot 4 \cdot 8} - \frac{(8,\ 12,\ 14)}{2 \cdot 4 \cdot 8}$$

e as integrais terminais

$$\int_0^1 \mathrm{d}y_1 \int_0^1 \mathrm{d}y_2 \int_0^1 \mathrm{d}y_3 \, \frac{y_1^{15} y_2^{23} y_3^{27}}{2 \cdot 4 \cdot 8 \cdot 15} \quad = \quad \frac{1}{2 \cdot 4 \cdot 8 \cdot 15 \cdot 16 \cdot 24 \cdot 28}$$

como

$$\frac{(15,\ 23,\ 27)}{2\cdot 4\cdot 8\cdot 15} \quad \mapsto \quad \frac{1}{2\cdot 4\cdot 8\cdot 15\cdot 16\cdot 24\cdot 28}\ ,$$

de modo que o esquema da Fig. 12 pode ser sintetizado por:

$$S_{3,si}^{(\infty)}(n) \;=\; f_3(n,\ 1,\ 1,\ 1)\ , \qquad n = 4, 5, \ldots, \infty\ ,$$

$$f_3(k,\ a_1,\ a_2,\ a_3) \;=\; \begin{cases} \frac{f_3(k-1,\,0,\,a_1,\,a_2)-f_3(k-1,\,0+a_3+1,\,a_1+a_3+1,\,a_2+a_3+1)}{a_3+1} & \text{se } k > 4\ , \\ \frac{1}{(a_1+1)(a_2+1)(a_3+1)} & \text{se } k = 4\ . \end{cases}$$

Generalizando este procedimento para μ e n arbitrários, o desenvolvimento da Eq. 5.3 nos leva à seguinte fórmula recursiva:

$$S_{\mu,si}^{(\infty)}(n) \;=\; f_\mu[n,\ \vec{1}]\ , \qquad n = \mu+1, \mu+2, \ldots, \infty\ , \tag{5.5}$$

$$f_\mu[j, \vec{a}] \;=\; \begin{cases} \frac{f_\mu[j-1,\,\mathrm{shift}(\vec{a})]-f_\mu[j-1,\,\mathrm{shift}(\vec{a})+(a_\mu+1)\cdot\vec{1}]}{a_\mu+1} & \text{se } j > \mu+1\ , \\ 1/\prod_{k=1}^{\mu}(a_k+1) & \text{se } j = \mu+1\ , \end{cases} \tag{5.6}$$

onde $\vec{1} = (1, 1, 1, \ldots, 1)$ e $\vec{a} = (a_1, a_2, a_3, \ldots, a_\mu)$ são vetores μ-dimensionais, $\mathrm{shift}(\vec{a}) = (0, a_1, a_2, \ldots, a_{\mu-1})$ é a operação de deslocamento acíclico de coordenadas e j é o passo de integração, utilizado também como condição de parada. Observe que a condição inicial da Eq. 5.5 e a primeira e segunda sentenças da Eq. 5.6 representam respectivamente a terceira, segunda e primeira sentenças da Eq. 5.4.

O menor ciclo permitido na caminhada determinista do turista tem período $p_{min} = \mu + 1$. Como a memória μ garante que o turista visite no mínimo $n_{min} = \mu + 1$ sítios, é conveniente definir o número *extra* de sítios explorados como sendo $n_e = n - \mu - 1$ a fim de que as distribuições $S_{\mu,si}^{(\infty)}$ iniciem todas no valor $n_e = 0$, independentemente do valor de μ. Observe que, em particular, para $\mu = 1$ o número extra de sítios explorados n_e é *numericamente* igual ao tempo de transiente t (isto não significa que sejam o mesmo trecho da trajetória), e neste caso a Eq. 5.5 se reduz à forma fechada [30] $S_{1,si}^{(\infty)}(n_e) = (n_e + 1)/(n_e + 2)!$.

As linhas contínuas do gráfico da Fig. 13 foram geradas pela Eq. 5.5 para alguns valores de μ. Pode-se observar que o decaimento é praticamente exponencial, com pequenas deformações (efeito de borda) nos primeiros $\mu + 1$ valores de n_e que se tornam mais

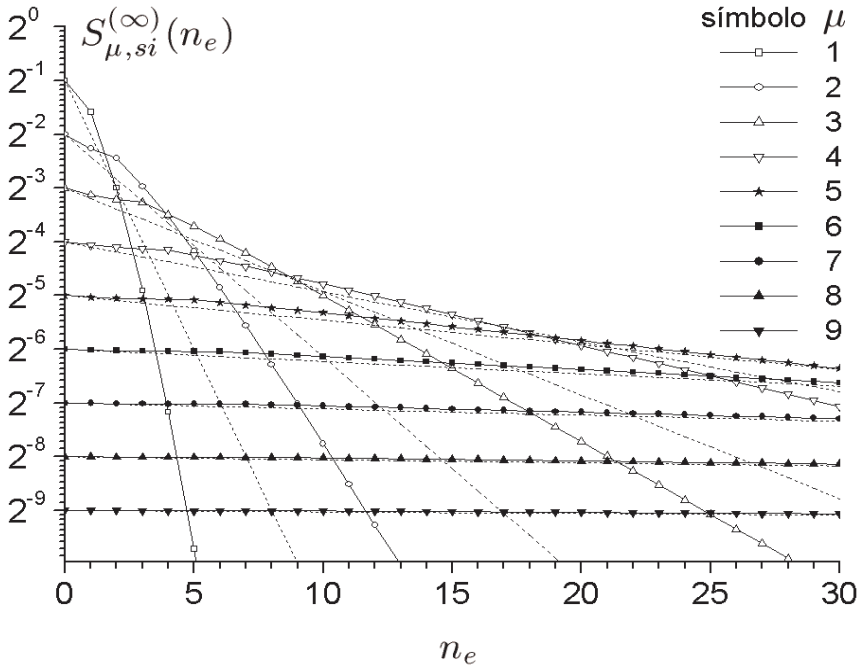


Figura 13: Distribuição de n_e para alguns valores de μ. As linhas contínuas referem-se à forma exata da Eq. 5.5 e as linhas pontilhadas referem-se à forma aproximada da Eq. 5.9.

amenas para $\mu \gg 1$. Como se pode observar nesta figura, uma propriedade notável é

$$S^{(\infty)}_{\mu,si}(n_e = 0) = \frac{1}{2^\mu} \tag{5.7}$$

para todo μ. Esta é exatamente a probabilidade de se ter um transiente nulo e um ciclo de período mínimo p_{min} na caminhada do turista unidimensional.

Não é possível obter uma expressão analítica fechada, simples e exata para $S^{(\infty)}_{\mu,si}(n)$. Apesar de ser exata, a forma recursiva da Eq. 5.6 não é prática para ser manipulada algebricamente. Até mesmo para cálculo computacional é desvantajosa, pois o tempo de processamento cresce exponencialmente em função de n_e, devido às bifurcações (vide esquema da Fig. 12). Tal dependência exponencial limitou o traçado das Figs. 13 e 14 aos valores de $n_e \leq 30$ e de $j \leq 30$, respectivamente. A propriedade da Eq. 5.7 possibilitará aproximar a Eq. 5.5 por uma distribuição geométrica.

5.1.1.2 Aproximação de Campo Médio

A recursividade da Eq. 5.5 foi herdada das integrais encadeadas da Eq. 5.3. Entretanto, para $\mu \gg 1$ pode-se usar uma aproximação de campo médio para desconectar tais integrais. Esta aproximação consiste em substituir os produtos $\tilde{y}_j$ por seus valores médios.

A Eq. D.28 do Apêndice D.4 fornece a pdf do produto $\tilde{y}$ de n variáveis y_j com dis-

tribuição uniforme. Neste Apêndice, a Eq. D.29 fornece o m-ésimo momento desta pdf: $\mathrm{E}(\tilde{y}^m) = (m+1)^{-n}$.

As variáveis y_j defininas pela Eq. 5.2 são i.i.d. com distribuição uniforme no intervalo $(0;1]$. A primeira sentença ($1 \leq j \leq \mu$) da Eq. 5.4 declara que as variáveis y_1, y_2, ..., y_μ podem variar livremente de 0 a 1. Considerando que para $\mu \gg 1$ o produto $\tilde{y}_{\mu+1} = \prod_{k=1}^{\mu} y_k$ tem pequena variância, este pode ser aproximado pelo seu valor médio $\mathrm{E}(\tilde{y}_{\mu+1}) = 2^{-\mu}$.

No tocante ao próximo produto $\tilde{y}_{\mu+2} = \prod_{k=2}^{\mu+1} y_k$, as variáveis y_2, y_3, ..., $y_{\mu+1}$ não são todas i.i.d., pois $y_{\mu+1}$ está restrita ao intervalo $[2^{-\mu}, 1]$. No entanto, para $\mu \gg 1$ o intervalo $[2^{-\mu}, 1]$ se torna próximo a $[0, 1]$, permitindo que $\tilde{y}_{\mu+2}$ também seja aproximado por seu valor médio $2^{-\mu}$. Este raciocínio pode, por indução, ser aplicado aos demais limites de integração $\tilde{y}_j$. A Eq. 5.4 é aproximada para

$$\mathcal{I}_j = \begin{cases} \int_0^1 \mathrm{d}y_j = 1 \ , & \text{para } 1 \leq j \leq \mu \\ \int_{2^{-\mu}}^1 \mathrm{d}y_j = 1 - 2^{-\mu} \ , & \text{para } \mu + 1 \leq j \leq n-1 \\ \int_0^{2^{-\mu}} \mathrm{d}y_j = 2^{-\mu} \ , & \text{para } j = n \end{cases} \tag{5.8}$$

e a Eq. 5.3 resulta em

$$S_{\mu,si}^{(\infty)}(n) \quad = \quad 2^{-\mu}(1-2^{-\mu})^{n-\mu-1} \ , \tag{5.9}$$

com $n = \mu+1, \mu+2, \ldots, \infty$. A expressão da Eq. 5.9 é justamente uma distribuição geométrica (Eq. D.7) com média $\mathrm{E}(n) = 2^\mu + \mu$, que pode ser interpretada como o *alcance característico* da caminhada), e variância $\mathrm{Var}(n) = 2^{2\mu} - 2^\mu$. As linhas pontilhadas do gráfico da Fig. 13 representam esta aproximação para $1 \leq \mu \leq 9$.

5.1.2 Probabilidades de Exploração e de Recuo

A motivação de calcular as probabilidades de exploração e de retorno é obter a Eq. 5.9 através de uma dedução alternativa e desenvolver argumentos simples para obter a probabilidade de percolação para um meio desordenado finito. Através de um argumento similar àquele usado para obter a Eq. 5.6, podemos calcular a distribuição acumulada complementar

$$F_{\mu,si}^{(\infty)}(n) \quad = \quad \sum_{k=n}^{\infty} S_{\mu,si}^{(\infty)}(k) \ ,$$

que fornece a probabilidade do turista visitar *no mínimo* n sítios distintos. Há uma única modificação a ser feita: uma vez que o turista tenha explorado n sítios ao chegar ao sítio

s_n, é indiferente se recuar ou prosseguir a caminhada. Portanto, a última integral (em x_n) da Eq. 5.1 não é mais necessária, resultando em

$$F_{\mu,si}^{(\infty)}(n) = \prod_{j=1}^{n-1} \mathcal{I}_j \ , \tag{5.10}$$

onde cada funcional $\mathcal{I}_j$ é dado pela Eq. 5.4. O nó raiz da Fig. 12 é alterado para 1 (ou, equivalentemente, $y_5^0 y_6^0 y_7^0$), que leva a

$$F_{\mu,si}^{(\infty)}(n) \ = \ f_\mu[n, \vec{0}] \ , \qquad n = \mu+1, \mu+2, \ldots, \infty \ , \tag{5.11}$$

onde $\vec{0} = (0, 0, 0, \ldots, 0)$ é o vetor nulo μ-dimensional e f_μ é dada pela Eq. 5.6. Observe que $F_{\mu,si}^{(\infty)}(n)$ usa a mesma estrutura recursiva da Eq. 5.6, mas com uma condição inicial diferente ($\vec{0}$ no lugar de $\vec{1}$). Se usarmos a Eq. 5.8 como uma aproximação para a Eq. 5.10, prontamente obtemos a distribuição acumulada complementar (dada pela Eq. D.9) do número de sítios explorados

$$F_{\mu,si}^{(\infty)}(n) = (1 - 2^{-\mu})^{n-\mu-1} \ . \tag{5.12}$$

Esta forma aproximada para a distribuição acumulada desempenhará um papel central na obtenção da probabilidade de percolação de um meio finito.

Em particular, para $\mu = 1$ a Eq. 5.11 assume a forma analítica fechada

$$F_{1,si}^{(\infty)}(n) \ = \ \frac{1}{(n-1)!} \tag{5.13}$$

que está de acordo com a Ref. [30].

A memória μ garante que o turista, partindo do sítio s_1, avance nos primeiros μ passos e visite os sítios s_2, s_3, ..., $s_{\mu+1}$. A partir daí, os demais passos são incertos, pois o turista pode avançar e explorar um novo sítio ou recuar e encerrar a exploração do meio. Com o intuito de fazer uma analogia à distribuição geométrica, é conveniente definir as probabilidades de avanço $q_{\mu,si}^{(\infty)}(j)$ (tomado como fracasso) como a probabilidade do turista explorar um novo sítio no j-ésimo passo incerto.

Dessa forma, a probabilidade de recuo $p_{\mu,si}^{(\infty)}(j)$ (tomado como sucesso) no j-ésimo passo incerto é igual à probabilidade de se explorar exatamente $n = \mu + j$ sítios condicionada ao fato do turista já ter explorado $n = \mu + j$ sítios, isto é

$$p_{\mu,si}^{(\infty)}(j) = \frac{S_{\mu,si}^{(\infty)}(n = \mu + j)}{F_{\mu,si}^{(\infty)}(n = \mu + j)} = \frac{f(\mu + j, \vec{1})}{f(\mu + j, \vec{0})} \tag{5.14}$$

onde f_μ é dada pela Eq. 5.6. Para $\mu = 1$, as probabilidades de recuo e de avanço se confundem, respectivamente, com as probabilidades de captura e de subsistência, que assume a expressão fechada

$$q_{1,si}^{(\infty)}(j) = \frac{1}{j+1} \ .$$

A análise destas probabilidades tem se mostrado frutífera. Para $\mu = 1$ permitiu a generalização da distribuição de transientes para dimensionalidade arbitrária [30].

O gráfico da Fig. 14 mostra os valores de $p_{\mu,si}^{(\infty)}(j)$ para os 30 primeiros passos incertos, com μ variando de 1 a 9.

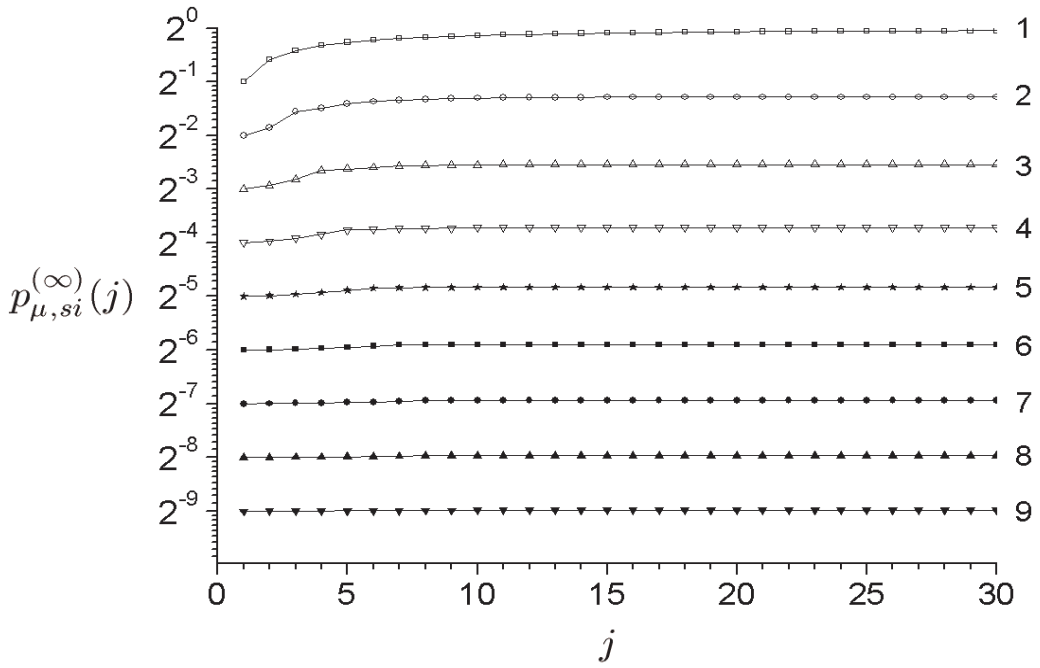


Figura 14: Probabilidade de recuo a cada passo da caminhada para μ variando entre 1 e 9, obtida a partir da Eq. 5.14. As linhas contínuas servem de guia para os olhos.

Analisando-se este gráfico, pode-se observar que, ao longo da caminhada:

1. para $\mu = 1$, a probabilidade de recuo se aproxima assintoticamente de 1;

2. para os demais valores de μ, a probabilidade de recuo também aumenta, mas tende a um valor menor que 1;

3. para $\mu \gg 1$ (na base 2), a probabilidade de recuo é praticamente constante e igual ao seu valor inicial $p_{\mu,si}^{(\infty)}(1) = 1/2^\mu$, dado pela Eq. 5.7.

Dessa forma, pode-se verificar empiricamente que para $\mu \gg 1$ as probabilidades de

retorno e de exploração podem ser tomada como $p_{\mu,si}^{(\infty)} = 2^{-\mu}$ e $q_{\mu,si}^{(\infty)} = 1 - 2^{-\mu}$ para todos os passos. Esta aproximação também pode ser obtida dividindo-se a Eq. 5.9 pela Eq. 5.13.

5.1.3 Dedução Alternativa

As expressões aproximadas para as probabilidades de avanço e de recuo também podem ser obtidas analiticamente através de uma dedução mais direta. Considere novamente que o turista parta do sítio s_1 localizado na origem do meio semi-infinito ilustrado na Fig. 11.

Como a memória μ garante a exploração dos primeiros $\mu + 1$ sítios, as distâncias x_1, x_2, ..., x_μ podem assumir qualquer valor no intervalo real $[0, \infty)$. A probabilidade de exploração $q_{\mu,si}^{(\infty)}(1)$ para o primeiro passo incerto pode ser obtida impondo que a distância $x_{\mu+1}$ seja menor que a soma $\tilde{x}_1 = \sum_{k=1}^{\mu} x_k$. Como as variáveis x_1, x_2, ..., x_μ são i.i.d. com distribuição exponencial (Eq. D.15), $\tilde{x}_1$ tem distribuição gama (Eq. D.18). Logo

$$q_{\mu,si}^{(\infty)}(1) = \int_0^\infty \frac{\mathrm{d}\tilde{x}_1 r^\mu \tilde{x}_1^{\mu-1} e^{-r\tilde{x}_1}}{(\mu-1)!} \cdot \int_0^{\tilde{x}_1} \mathrm{d}x_{\mu+1} r e^{-rx_{\mu+1}} = 1 - 2^{-\mu} \ .$$

A probabilidade de exploração $q_{\mu,si}^{(\infty)}(2)$ para o segundo passo incerto não é exatamente igual a $q_{\mu,si}^{(\infty)}(1)$. Como a distância $x_{\mu+1}$ está confinada ao intervalo $[0, \tilde{x}_1]$, as variáveis x_2, x_3, ..., $x_{\mu+1}$ não são todas independentes e conseqüentemente $\tilde{x}_2 = \sum_{k=2}^{\mu+1} x_k$ não segue exatamente uma distribuição gama. Entretanto, para $\mu \gg 1$, $x_{\mu+1}$ raramente excede $\tilde{x}_1$. De fato, a probabilidade disto acontecer é $P(x_{\mu+1} > \tilde{x}_1) = 1 - q_{\mu,si}^{(\infty)}(1) = 2^{-\mu}$. Isto significa que, para $\mu \gg 1$, há uma fraca correlação entre as distâncias x_i. Portanto, podemos aproximar a pdf de $\tilde{x}_2$ por uma distribuição gama e considerar que $q_{\mu,si}^{(\infty)}(2) \approx q_{\mu,si}^{(\infty)}(1)$. O mesmo argumento pode ser usado para as distâncias subseqüentes.

Quando o sítio s_n for alcançado, o turista deverá recuar, encerrando a exploração do meio. Utilizando $q_{\mu,si}^{(\infty)}(1)$ como probabilidade de exploração para todos os passos, a probabilidade de retorno é $p_\mu = 1 - q_\mu = 2^{-\mu}$ e então $S_{\mu,si}^{(\infty)}(n) = 2^{-\mu}(1 - 2^{-\mu})^{n_e}$, que é o resultado da Eq. 5.9.

5.2 Meio Finito

O meio finito é constituído por N pontos cujas coordenadas z_k, foram geradas aleatória e independentemente segundo uma distribuição uniforme (Eq. D.2) no intervalo $[0, 1]$. Simulações numéricas revelaram que os resultados analíticos obtidos para o meio semi-

infinito também se aplicam ao meio finito. Esta equivalência não é trivial, pois *todos* os resultados obtidos para o meio semi-infinito assumiram que as distâncias entre sítios consecutivos são variáveis i.i.d. e seguem uma distribuição exponencial (Eq. D.15), podendo, em princípio, assumir qualquer valor no intervalo $[0, \infty)$, o que obviamente não é verdade para o meio finito.

Porém, tal equivalência entre estes dois meios pode ser demonstrada comparando as distribuições das abscissas dos pontos desses meios. Por um lado, a abscissa do ponto que ocupa o k-ésimo posto no meio finito segue uma distribuição beta (Eq. D.25). Por outro lado, se o meio semi-infinito for restrito ao comprimento das primeiras $N + 1$ distâncias e normalizado para se ajustar ao intervalo $[0, 1]$, então a abscissa do ponto que ocupa o k-ésimo posto também terá distribuição beta, como num autêntico meio finito (veja Sec. D.3.2.3 do Apêndice D). Esta normalização não afeta a caminhada do turista, pois nesta caminhada somente os postos de vizinhança são relevantes, e não as distâncias em si.

5.2.1 Probabilidade de Percolação

A probabilidade $P_N(\mu)$ do turista explorar todo o meio finito com N pontos pode ser deduzida impondo que o turista avance nos primeiros $N - \mu - 1$ passos incertos e, quando o último ponto s_N for alcançado, não há necessidade de impor um retorno no próximo passo. Portanto a probabilidade de percolação é

$$P_N(\mu) = q_\mu^{N-\mu-1} = (1 - 2^{-\mu})^{N-\mu-1} \ . \tag{5.15}$$

É interessante notar que a probabilidade de percolação está diretamente relacionada com a distribuição acumulada complementar (Eq. 5.12). Particularmente para $\mu = 1$ a probabilidade de percolação assume uma forma exata fechada muito simples

$$P_N(1) = F_{1,si}^{(\infty)}(N) = \frac{1}{(N-1)!}$$

obtida a partir da Eq. 5.13.

O gráfico da Fig. 15 mostra a superfície de transição para a percolação, obtida a partir da Eq. 5.15.

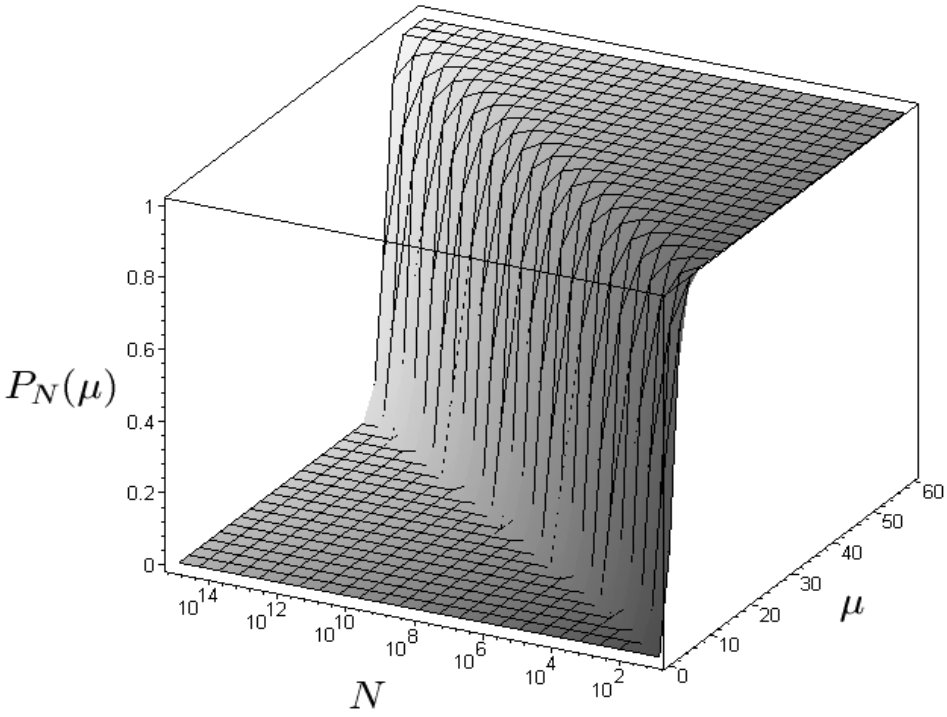


Figura 15: Probabilidade de percolação em função de μ e de N.

5.2.2 Memória Crítica

O gráfico da Fig. 16 é uma vista lateral da superfície da Fig. 15 e mostra uma comparação entre valores obtidos através da Eq. 5.15 e resultados de simulações de Monte Carlo. Pode-se observar que, para cada valor de N, há um valor crítico de μ para o qual a probabilidade de percolação aumenta vigorosamente de quase 0 a quase 1, dentro de uma região de transição bem definida.

Em analogia a uma transição de fase de primeira ordem, definimos o ponto de transição como o valor máximo da derivada de $P_N(\mu)$ em relação a μ. Isto significa que a segunda derivada se anula neste máximo $\mathrm{d}^2_\mu P_N(\mu)\big|_{\mu_1^{(c)}} = 0$, levando a uma equação transcendental, que não pode ser resolvida analiticamente para obter $\mu_1^{(c)}$. Uma estimativa μ_1 deste valor crítico pode ser calculada considerando $N \gg \mu \gg 1$ e aproximando a Eq. 5.15 para

$$P = (1 - 2^{-\mu})^N \tag{5.16}$$

O valor crítico de μ pode ser determinado através do ponto de inflexão da curva. Pela regra da cadeia, a primeira derivada é

$$\frac{\mathrm{d}P}{\mathrm{d}\mu} = N \cdot (1 - 2^{-\mu})^{N-1} \cdot 2^{-\mu} \cdot \ln 2 \tag{5.17}$$

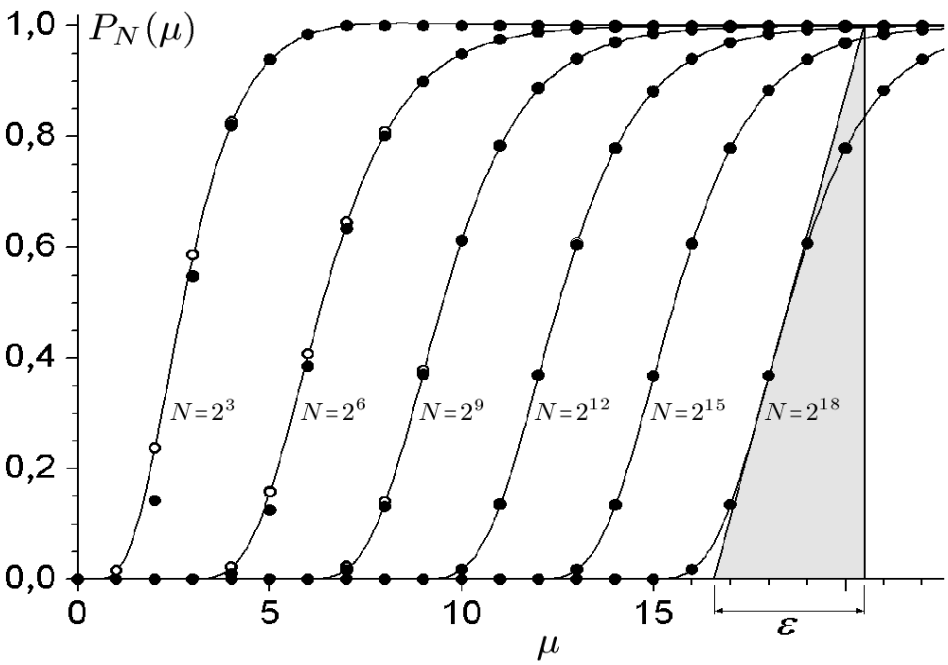


Figura 16: Probabilidade de percolação para alguns valores de N. Os círculos vazios referem-se à Eq. 5.15 e os círculos cheios representam simulações numéricas ($M = 100\,000$ mapas para cada valor de N e de μ). As barras de erro são menores que o tamanho dos círculos e por isso foram omitidas. As linhas contínuas servem apenas como guias, pois μ é uma variável discreta.

a e segunda derivada

$$\frac{\mathrm{d}^2 P}{\mathrm{d}\mu^2} = N \cdot \ln^2 2 \cdot \left[(N-1)(1-2^{-\mu})^{N-2} \cdot 4^{-\mu} - (1-2^{-\mu})^{N-1} \cdot 2^{-\mu}\right] .$$

Rearranjando e igualando a zero, temos

$$N \cdot \ln^2 2 \cdot (1-2^{-\mu_1})^{N-1} \cdot 2^{-\mu_1} \cdot \underbrace{\left[(N-1)(1-2^{-\mu_1})^{-1} \cdot 2^{-\mu_1} - 1\right]}_{=0} = 0 .$$

Como cada fator fora dos colchetes é certamente diferente de zero, é necessário que

$$(1-2^{-\mu_1})^{-1} \cdot 2^{-\mu_1} = \frac{1}{N-1} ,$$

de onde obtemos o valor crítico

$$\mu_1 = \log_2 N . \tag{5.18}$$

Para calcular a largura da região de transição, substituímos μ_1 na Eq. 5.17 para

obtermos a inclinação da curva no ponto de inflexão

$$\left.\frac{\mathrm{d}P}{\mathrm{d}\mu}\right|_{\mu_1} = N \cdot (1 - 2^{-\log_2 N})^{N-1} \cdot 2^{-\log_2 N} \cdot \ln 2 \, .$$

Simplificando, temos

$$\left.\frac{\mathrm{d}P}{\mathrm{d}\mu}\right|_{\mu_1} = \underbrace{\left(1 - \frac{1}{N}\right)^{N-1}}_{1/e} \cdot \ln 2 \, .$$

Considerando $N \gg 1$, surge o limite fundamental do cálculo

$$\left.\frac{\mathrm{d}P}{\mathrm{d}\mu}\right|_{\mu_1} = \frac{\ln 2}{e} \, .$$

A largura ε da região de transição destacada na Fig. 16 pode ser determinada por trigonometria elementar

$$\tan\alpha = \frac{1}{\varepsilon} = \frac{\ln 2}{e} = \left.\frac{\mathrm{d}P}{\mathrm{d}\mu}\right|_{\mu_1} \, .$$

Logo

$$\varepsilon = \frac{e}{\ln 2} \approx 3,92 \, .$$

Este resultado mostra que a transição para a percolação ocorre numa estreira faixa de memória, independentemente do tamanho do sistema. Como a memória crítica μ_1 cresce lentamente ($\log_2 N$) em função do tamanho N do sistema e região de transição é independente de N, tem-se uma transição abrupta no limite termodinâmico ($N \gg 1$). Enfatizamos que as aproximações empregadas geraram resultados satisfatórios até mesmo para pequenos valores de N e μ, como mostra a Fig. 16.

5.3 Conclusão

Nosso principal resultado é que o turista não necessita ter uma memória da ordem de N para explorar todo o meio; uma pequena memória (da ordem de $\ln N$) já é suficiente. Todos os resultados aqui obtidos concordam com o caso particular $\mu = 1$ tratado na Ref. [30]. Um resultado exato interessante é que a probabilidade de se obter uma trajetória com transiente nulo $t = 0$ e período mínimo $p = p_{min} = \mu + 1$ é igual a $2^{-\mu}$.

6 Probabilidade de Vizinhança em Ambientes Poissônicos: Fórmula de Cox

A probabilidade $P_{m,n}^{(d)}$ de um evento arbitrário, num processo poissônico de dimensionalidade d, ser o m-ésimo vizinho mais próximo de seu n-ésimo vizinho mais próximo tem atraído a atenção de pesquisadores desde os estudos pioneiros de Clark e Evans [34] e Clark [35] sobre alguns aspectos de padrões espaciais em populações biológicas. Eles criaram o termo vizinhos reflexivos para o caso $m = n$ e a probabilidade de posto de vizinhança reflexiva por eles calculada foi corrigida por Dacey [36] ($m > 1$) no contexto de análise geográfica e então generalizada (para $m \neq n$) por Cox [21], que nós denominamos de *fórmula de Cox*. Em particular, a probabilidade $P_{1,1}^{(d)}$ representa a densidade de casais num ambiente poissônico d-dimensional e assumiu um papel central na parametrização da dimensionalidade d do meio na distribuição de transientes para a caminhada do turista sem memória [30].

Neste Capítulo, obtemos a fórmula de Cox através de uma dedução alternativa [37] usando somente a distribuição de Poisson em vez de usar as distribuições de Poisson, binomial e gama como no artigo original [21]. Assim como nos cálculos de Cox, nós escrevemos as probabilidades no limite termodinâmico $N \to \infty$. Porém, ao contrário de Cox, nós as escrevemos em termos de distribuições (multinomial, binomial e hipergeométrica) e funções conhecidas (em vez de escrevê-las em termos de uma integral). Com o uso de funções especiais consideramos o limite de alta dimensionalidade, que nos leva à probabilidade de vizinhança no modelo RL. Usando argumentos similares, deduzimos a probabilidade de vizinhança para sistemas RL de tamanho finito. Finalmente, escrevemos as probabilidades de Cox para os dois casos considerados do modelo RM. Todos os resultados analíticos foram comparados e validados por simulações numéricas de Monte Carlo.

6.1 Dedução Alternativa da Fórmula de Cox

Num ambiente poissônico d-dimensional ilimitado com densidade média de r pontos por unidade de volume, a probabilidade de que um volume V_d (de forma arbitrária, até mesmo desconexo) contenha k pontos é dada pela distribuição de Poisson (Eq. D.14):

$$P_{Po}(k) \;=\; \frac{\lambda^k e^{-\lambda}}{k!}\;,$$

com $k = 0, 1, 2, \ldots, \infty$, onde $\lambda = \langle k \rangle = rV_d$ é o número esperado de pontos dentro do volume V_d. Observe que o único parâmetro desta distribuição é λ (a dimensionalidade d do meio não é relevante, mas está representada implicitamente pelo volume V_d).

Considere dois pontos I e J de um ambiente poissônico d-dimensional distantes $D_{I,J} = R$ entre si. O volume $V_d(R)$ da hiperesfera de raio R centrada em I (e que, portanto, passa por J) é dado pela Eq. C.8. Evidentemente este também é o volume da hiperesfera centrada em J que passa por I (também de raio R). A Fig. 17 mostra um exemplo para $d = 2$.

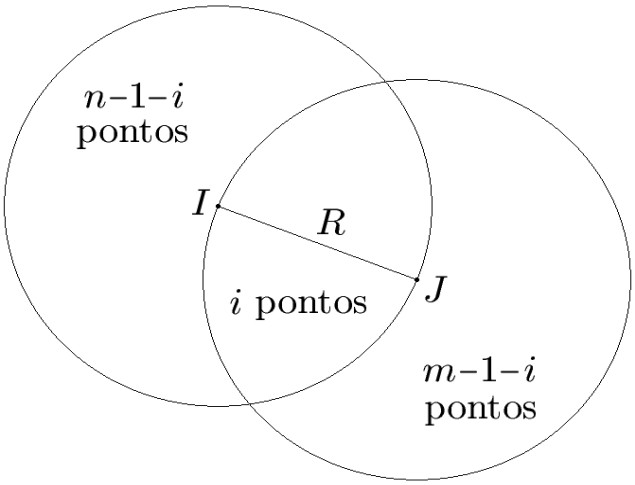


Figura 17: Processo poissônico bidimensional. Os círculos centrados nos pontos I e J têm área $V_2 = \pi R^2$ e a intersecção tem área $V_{\cap,2} = V_2(1 - p_2) = (2\pi/3 - \sqrt{3}/2)R^2$. Há i pontos na intersecção dos círculos e os crescentes de I e J contêm $n - 1 - i$ e $m - 1 - i$ pontos, respectivamente.

O volume relativo p_d de um crescente (em relação a uma das hiperesferas) é dado pela Eq. C.10 para o caso particular $x = 1$

$$p_d = p_d(1) = \mathrm{I}_{1/4}\left(\frac{1}{2}, \frac{d+1}{2}\right)\;, \tag{6.1}$$

onde $\mathrm{I}_z(a, b)$ é a função beta incompleta normalizada, definida pela Eq. B.14.

Para que I seja o m-ésimo vizinho mais próximo de J e J seja o n-ésimo vizinho mais próximo de I, as seguintes condições devem ser satisfeitas:

1. deve haver i pontos intersecção das duas hiperesferas, com i variando de 0 a $\text{mín}(m-1, n-1)$, o número esperado de pontos é $\lambda(1-p_d)$,

2. deve haver $m-1-i$ pontos no crescente de J, o número esperado de pontos é λp_d,

3. deve haver $n-1-i$ pontos no crescente de I, o número esperado de pontos também é λp_d,

4. a distância R entre I e J pode assumir qualquer valor no intervalo $[0, \infty)$ e, consequentemente, o volume $V_d(R)$ e o número esperado de pontos $\lambda = rV_d(R)$ dentro dele também podem variar de 0 a ∞.

Reunindo estas condições, obtemos a seguinte expressão para a probabilidade $P^{(d)}_{m,n} = P^{(d)}_{n,m}$:

$$P^{(d)}_{m,n} = \int_0^\infty \mathrm{d}\lambda \sum_{i=0}^{\text{mín}(m-1,n-1)} \frac{[\lambda(1-p_d)]^i e^{-\lambda(1-p_d)}}{i!} \cdot \frac{(\lambda p_d)^{m-1-i} e^{-\lambda p_d}}{(m-1-i)!} \cdot \frac{(\lambda p_d)^{n-1-i} e^{-\lambda p_d}}{(n-1-i)!} \, .$$

Observe que as variáveis i, m e n são discretas (pois representam "número de pontos"), mas λ é uma variável contínua (pois representa "número *esperado* de pontos"). Colocando os fatores que não dependem de λ fora do integrando, temos

$$P^{(d)}_{m,n} = \sum_{i=0}^{\text{mín}(m-1,n-1)} \frac{(1-p_d)^i p_d^{m-1-i} p_d^{n-1-i}}{i!(m-1-i)!(n-1-i)!} \int_0^\infty \mathrm{d}\lambda \, \lambda^{m+n-2-i} e^{-\lambda(1+p_d)}$$

e a integral remanescente pode ser expressa em termos da função gama (Eq. B.3):

$$\int_0^\infty \mathrm{d}\lambda \, \lambda^{m+n-2-i} e^{-\lambda(1+p_d)} = \frac{\Gamma(m+n-1-i)}{(1+p_d)^{m+n-1-i}}$$

obtendo-se assim a forma original da fórmula de Cox:

$$P^{(d)}_{m,n} = \sum_{i=0}^{\text{mín}(m-1,n-1)} \frac{(m+n-2-i)!}{i!(m-1-i)!(n-1-i)!} \cdot \frac{(1-p_d)^i p_d^{m-1-i} p_d^{n-1-i}}{(1+p_d)^{m+n-1-i}} \, , \tag{6.2}$$

onde $m = 1, 2, \ldots, \infty$ e $n = 1, 2, \ldots, \infty$. Fazendo i variar de 1 a $\text{mín}(m, n)$ e rearranjando os termos, temos:

$$P^{(d)}_{m,n} = \frac{1}{1+p_d} \sum_{i=1}^{\text{mín}(m,n)} \frac{(m+n-1-i)!}{(i-1)!(m-i)!(n-i)!} \left(\frac{1-p_d}{1+p_d}\right)^{i-1} \left(\frac{p_d}{1+p_d}\right)^{m-i} \left(\frac{p_d}{1+p_d}\right)^{n-i}$$

A expressão do somatório pode ser identificada com a distribuição multinomial (Eq. D.6):

$$\frac{P_{m,n}^{(d)}}{P_{1,1}^{(d)}} = \sum_{i=1}^{\text{mín}(m,n)} \text{Mult}\left(i-1, m-i, n-i; \frac{1-p_d}{1+p_d}, \frac{p_d}{1+p_d}, \frac{p_d}{1+p_d}\right) \tag{6.3}$$

$$P_{1,1}^{(d)} = \frac{1}{1+p_d}\,, \tag{6.4}$$

onde $P_{1,1}^{(d)}$ é a densidade de vizinhos mutuamente mais próximos (casais).

Observe que a dimensionalidade do meio considerado aparece indiretamente na fórmula de Cox através da grandeza p_d (Eq. 6.1), denotada por p na Referência [21]. Esta é a grandeza relevante que parametriza as distribuições de tempos de transiente na caminhada do turista sem memória [30].

É importante ressaltar que a distribuição de probabilidade de Cox *não* é uma distribuição conjunta nas variáveis m e n, pois o somatório $\sum_{m,n=1}^{\infty} P_{m,n}(d)$ diverge. Para cada m a distribuição está normalizada, pois $\sum_{n=1}^{\infty} P_{m,n}(d) = 1$. Considerando m um parâmetro e n a variável da distribuição, obtemos a média $\mathrm{E}(n) = m + p_d$ e a variância $\mathrm{Var}(n) = (2m + p_d - 1)p_d$.

6.2 Probabilidades de Vizinhança no Modelo Random Link e em Alta Dimensionalidade

O limite de alta dimensionalidade pode ser obtido diretamente da fórmula de Cox. Neste procedimento, pode-se facilmente obter a correção de primeira ordem das probabilidades de vizinhança no modelo Random Link.

6.2.1 Limite Termodinâmico

Consideremos a situação de alta dimensionalidade ($d \gg 1$), que corresponde a tomar $b = (d+1)/2 \gg a = 1/2$ na Eq. 6.1 e usar a aproximação da Eq. B.12 na Eq. B.14 para obter

$$\mathrm{I}_z(a,b) \approx \frac{b^a}{\Gamma(a)} \int_0^z \mathrm{d}t\, t^{a-1}(1-t)^b\,.$$

Como $t \leq z = 1/4$ implica $t \ll 1$, a aproximação $(1-t)^b = e^{b\ln(1-t)} \approx e^{-bt}$ produz

$$\mathrm{I}_z(a,b) \approx \frac{\gamma(a,bz)}{\Gamma(a)}\,,$$

onde $\gamma(a,b)$ é a função gama incompleta dada pela Eq. B.5. Usando a relação dada pela Eq. B.6 e o valor particular $\Gamma(1/2) = \sqrt{\pi}$,

$$I_z(a,b) \approx \frac{\gamma(1/2, bz)}{\Gamma(1/2)} = \mathrm{erf}(\sqrt{bz})$$

e Eq. 6.1 pode ser re-escrita como:

$$p_d \approx \mathrm{erf}\left(\sqrt{\frac{d}{8}}\right) ,$$

onde a dimensionalidade característica $d_0 = 8$ emerge naturalmente da análise.

Mais uma aproximação pode ser feita notando que $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$ e usando a forma assintótica da Eq. B.2, de forma que para $|z| \gg 1$, pode-se escrever:

$$p_d \approx 1 - \mathrm{erfc}\left(\sqrt{\frac{d}{8}}\right) = 1 - \alpha_d , \tag{6.5}$$

onde

$$\alpha_d = \frac{1}{\sqrt{\pi}} \frac{e^{-d/8}}{\sqrt{d/8}} \left(1 - \frac{4}{d} + \cdots\right) \tag{6.6}$$

é uma aproximação para o volume relativo da intersecção de duas hiperesferas.

Usando $1 - p_d \approx \alpha_d$ e $p_d \approx 1$, a soma da Eq. 6.2 dá as probabilidades de Cox como uma série de potências em α_d para sistemas de alta dimensionalidade:

$$P_{m,n}^{(d\gg 1)} = P_{m,n}^{(rl)} + \frac{2^{2-(m+n)}}{B(m-1, n-1)} \alpha_d + \cdots , \tag{6.7}$$

sendo que na aproximação para o random link ($d \to \infty$) esta probabilidade é escrita em termos da distribuição binomial (Eq. D.4)

$$\frac{P_{m,n}^{(rl)}}{P_{1,1}^{(rl)}} = \frac{1}{2^{m+n-2}} \frac{\Gamma(m+n-1)}{\Gamma(m)\Gamma(n)} = \mathrm{Binom}\left(m-1, n-1; \frac{1}{2}, \frac{1}{2}\right) \tag{6.8}$$

$$P_{1,1}^{(rl)} = \frac{1}{2} , \tag{6.9}$$

onde $P_{1,1}^{(rl)}$ é a densidade de casais. Expressões simples podem ser obtidas, tais como: $P_{1,n}^{(rl)} = 1/2^n$, $P_{2,n}^{(rl)} = n/2^{n+1}$.

No limite de alta dimensionalidade $d \to \infty$, o volume relativo do crescente (Eq. 6.1) tende a 1 ($p_d \to 1$) e o número esperado de pontos $\lambda(1-p_d)$ dentro da intersecão se anula. Como $\lim_{p_d \to 1}[(1-p_d)/(1+p_d)]^i = \delta_{i,0}$, onde $\delta_{i,j}$ é o delta de Kronecker, a distribuição multinomial da Eq. 6.3 se torna a distribuição binomial da Eq. 6.8. Isto pode ser facilmente

visto se considerarmos uma hiperesfera de raio r dentro de um hipercubo de aresta $2r$. À medida que a dimensionalidade aumenta o volume da hiperesfera diminui relativamente ao volume do hipercubo e a diferença de volumes aumenta, indicando que todos os pontos caem no volume externo à hiperesfera [38].

Valores numéricos relativos ao caso de alta dimensionalidade são mostrados na Tab. 3.

d	p_d	$P_{1,1}^{(d)}$	$P_{1,2}^{(d)}$	$P_{2,2}^{(d)}$
0	1/3	3/4	3/16	15/32
1	1/2	2/3	2/9	10/27
2	$\frac{2\pi+3\sqrt{3}}{6\pi}$	$\frac{6\pi}{8\pi+3\sqrt{3}}$	$\frac{6\pi(2\pi+3\sqrt{3})}{(8\pi+3\sqrt{3})^2}$	$\frac{6\pi(40\pi^2+12\sqrt{3}\pi+27)}{(8\pi+3\sqrt{3})^3}$
3	11/16	16/27	176/729	6032/19683
$\vdots$		$\vdots$	$\vdots$	$\vdots$
$\gg 1$	$1-\alpha_d$	$(1+p_d)^{-1}$	$p_d(1+p_d)^{-2}$	$(1+p_d^2)(1+p_d)^{-3}$
∞ (RL)	1	1/2	1/4	1/4

Tabela 3: Alguns valores das probabilidades de vizinhança. Para baixa dimensionalidade, utiliza-se a Eq. 6.2. Um caso-limite interessante é $d = 0$, que produz: $p_0 = \int_0^{1/4} \mathrm{d}t/[\pi\sqrt{t(1-t)}] = 1/3$. Para $d \gg 1$, utiliza-se a Eq. 6.5 e para o modelo "Random Link", onde $d \to \infty$, utiliza-se a Eq. 6.8.

6.2.2 Sistema com Tamanho Finito

O limite de alta dimensionalidade $d \to \infty$ corresponde ao modelo RL, no qual todas as distâncias se tornam variáveis aleatórias i.i.d. Como as distâncias euclidianas são apenas um meio de obter as probabilidades para os postos de vizinhança, estas probabilidades são independentes da escolha particular para pdf destas distâncias [30]. Por simplicidade, consideraremos uma distribuição uniforme no intervalo $[0, 1]$ (Eq. D.2) para as distâncias entre os N pontos.

Como antes, para que I seja o m-ésimo vizinho mais próximo de J e J seja o n-ésimo vizinho mais próximo de I, devem ser satisfeitas as seguintes condições:

1. a distância x entre I e J pode assumir qualquer valor no intervalo $[0, 1]$,

2. as distâncias de J a cada um de seus $m - 1$ vizinhos mais próximos devem ser menores que x e

3. as distâncias de J a cada um de seus $N-m-1$ vizinhos mais distantes devem ser maiores que x, assim como

4. as distâncias de I a cada um de seus $n-1$ vizinhos mais próximos devem ser menores que x e

5. as distâncias de I a cada um de seus $N-n-1$ vizinhos mais distantes devem ser maiores que x.

A Fig. 18 ilustra a situação. Observe que os $N-2$ vizinhos de I são os mesmos $N-2$ vizinhos de J, ou seja, os quatro conjuntos de pontos não são disjuntos.

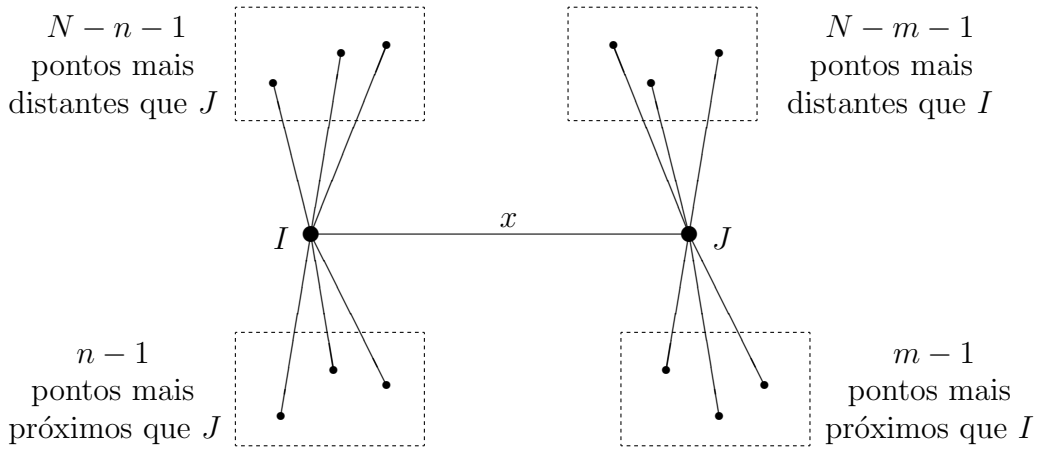


Figura 18: Ilustração esquemática dos pontos I e J e seus vizinhos no modelo "Random Link" com N pontos.

Também devemos observar que:

1. ao escolher um ponto I arbitrário, seu m-ésimo vizinho mais próximo está automaticamente determinado, e para isto existem $N-1$ possibilidades,

2. devem ser contadas todas as combinações de distribuição dos $N-2$ vizinhos de J entre os $m-1$ "mais próximos" e os $N-m-1$ "mais distantes" que J,

3. a mesma contagem deve ser feita para os $N-2$ vizinhos de I.

Combinando estas três contagens com as cinco restrições às distâncias, temos:

$$\begin{aligned} P_{m,n}^{(rl,N)} &= (N-1)\cdot\frac{(N-2)!}{(m-1)!(N-m-1)!}\cdot\frac{(N-2)!}{(n-1)!(N-n-1)!}\cdot \\ &\cdot\int_0^1 \mathrm{d}x\left[\int_0^x \mathrm{d}y\right]^{m-1}\left[\int_x^1 \mathrm{d}y\right]^{N-m-1}\left[\int_0^x \mathrm{d}y\right]^{n-1}\left[\int_x^1 \mathrm{d}y\right]^{N-n-1} \end{aligned}$$

ou

$$P_{m,n}^{(rl,N)} = (N-1)\begin{pmatrix} N-2 \\ m-1 \end{pmatrix}\begin{pmatrix} N-2 \\ n-1 \end{pmatrix}\int_0^1 \mathrm{d}x\, x^{m+n-2}(1-x)^{2N-m-n-2} \quad (6.10)$$

A integral é identificada com a função beta (Eq. B.10):

$$\begin{aligned}\int_0^1 \mathrm{d}x\, x^{m+n-2}(1-x)^{2N-m-n-2} &= B(m+n-1, 2N-m-n-1) \\ &= \frac{\Gamma(m+n-1)\Gamma(2N-m-n-1)}{\Gamma(2N-2)} \\ &= \frac{(m+n-2)!(2N-m-n-2)!}{(2N-3)(2N-4)!} \\ &= \frac{1}{(2N-3)\begin{pmatrix} 2N-4 \\ m+n-2 \end{pmatrix}}\end{aligned}$$

Substituindo na Eq. 6.10

$$P_{m,n}^{(rl,N)} = \frac{N-1}{2N-3}\cdot\frac{\begin{pmatrix} N-2 \\ m-1 \end{pmatrix}\begin{pmatrix} N-2 \\ n-1 \end{pmatrix}}{\begin{pmatrix} 2N-4 \\ m+n-2 \end{pmatrix}}, \qquad \begin{matrix} m = 1,2,3,\ldots,N-1 \\ n = 1,2,3,\ldots,N-1 \end{matrix}.$$

O primeiro fator pode ser identificado com a densidade de casais, e o segundo, com a distribuição hipergeométrica (Eq. D.19):

$$\frac{P_{m,n}^{(rl,N)}}{P_{1,1}^{(rl)}} = \mathrm{HipGeom}(N-2, N-2; m-1, n-1) \quad (6.11)$$

$$P_{1,1}^{(rl)} = \frac{N-1}{2N-3} \quad (6.12)$$

No limite $N \gg 1$, as Eqs. 6.11 e 6.12 se reduzem às Eqs. 6.8 e 6.9. Este limite corresponde justamente ao processo de tomar uma população infinita, tratado na Sec. D.3.1.1, de forma que a distribuição hipergeométrica se torna a distribuição binomial.

6.3 Probabilidades de Vizinhança no Modelo Random Map

Nesta Seção, calculamos as probabilidades de Cox para os dois casos do modelo RM. No caso em que a restrição $D_{i,i} = 0, \forall i$ é preservada, se um ponto arbitrário I é escolhido, seu m-ésimo vizinho J é automaticalmente determinado, porém cada um dos outros $N-1$ pontos tem igual probabilidade de ser o n-ésimo vizinho de J, pois as distâncias são totalmente independentes. Então, a probabilidade $P_{m,n}^{(rm)}$ do ponto I ser o n-ésimo vizinho de seu m-ésimo vizinho é simplesmente:

$$P_{m,n}^{(rm)} = \frac{1}{N-1} \ ,$$

onde $m = 1, 2, \ldots, N-1$ e $n = 1, 2, \ldots, N-1$.

Por outro lado, no caso em que $D_{i,i} \neq 0$ é permitido, a probabilidade $P_{m,n}^{(rm)}$ é duas vezes maior para vizinhos reflexivos do que para não-reflexivos, pois devemos considerar que cada ponto é sempre seu próprio m-ésimo vizinho mais próximo, para algum m. Portanto:

$$P_{m,n}^{(rm)} = \frac{1+\delta_{m,n}}{N+1} \ ,$$

onde $\delta_{m,n}$ é o delta de Kronecker, $m = 1, 2, \ldots, N$ e $n = 1, 2, \ldots, N$.

Note que no limite termodinâmico $N \gg 1$, estes casos ainda são distinguíveis devido à presença do fator 2 para os vizinhos reflexivos.

6.4 Conclusão

Usando apenas a distribuição de Poisson, obtivemos a fórmula de Cox através de uma dedução simples e a identificamos com a distribuição multinomial. Ao escrever o parâmetro de dimensionalidade p_d em termos da função beta incompleta normalizada, pudemos obter a aproximação de alta dimensionalidade para as probabilidades de vizinhança de um processo poissônico (PPA, por exemplo) pela distribuição binomial. Estas probabilidades de vizinhança podem ser ferramentas úteis na dedução de distribuições estatísticas para a caminhada do turista com memória $\mu \geq 2$ em meios desordenados com dimensionalidade d arbitrária.

Usando a mesma linha de raciocínio, foram obtidas as probabilidades de vizinhança para sistemas RL de tamanho finito. Neste caso, as probabilidades foram identificadas

com a distribuição hipergeométrica. Por fim, expressões simples foram obtidas para o modelo RM.

7 Conclusão

O objetivo principal deste trabalho foi resolver analiticamente o problema do turista com memória ($\mu \geq 2$). Para tanto, utilizamos métodos e conceitos similares àqueles utilizados no problema do turista sem memória ($\mu = 1$), tais como as probabilidades de avanço e recuo, de exploração e revisita, e de subsistência e captura. Devido a essa similaridade entre métodos, obtivemos para o caso $\mu = 2$ expressões analíticas análogas ao caso $\mu = 1$. A comparação destas formas analíticas possibilitaram compreender o principal mecanismo que torna as situação $\mu = 1$ e $\mu \geq 2$ tão distintas.

No Cap. 2 descrevemos o meio desordenado e apresentamos o modelo de distâncias aleatórias como limite de alta dimensionalidade para espaços euclidianos e o modelo de mapeamento aleatório como caso limite para as redes de Kauffman. Também descrevemos a caminhada determinista do turista e relacionamos os principais resultados já obtidos para o caso $\mu = 1$.

No Cap. 3 discutimos alguns aspectos computacionais envolvidos na geração do meio desordenado. Para o problema de pontos aleatórios, apresentamos o recurso da máscara, que reduz o consumo de memória da ordem de $O(N^2)$ (obtido na implementação direta) para $O(N)$. A seguir abordamos o modelo Random Link e enfatizamos que a simetria da métrica euclidiana impossibilita o uso da máscara para este modelo. Apresentamos então dois algoritmos que permitem gerar mapas do modelo Random Link com o mesmo consumo de memória $O(N)$ obtido com o uso da máscara. Concluímos que o algoritmo multiple-seed é o mais eficiente, pois mantém a dependência temporal em $O(N^2)$ (obtida na implementação direta).

No Cap. 4 tratamos analiticamente a caminhada determinista do turista com memória $\mu = 2$ sobre o modelo Random Link. Iniciamos calculando a distribuição do número de sítios explorados antes da primeira revisita, e obtivemos expressões análogas ao caso $\mu = 1$. Este resultado nos permitiu compreender o efeito da memória no comportamento exploratório do turista. A seguir calculamos a probabilidade de aprisionamento e enfatizamos

que o resultado obtido é contra-intuitivo. Combinando a probabilidade de aprisionamento com os resultados anteriores obtivemos a distribuição do número de sítios explorados em toda a caminhada. Através de uma analogia à distribuição geométrica, pudemos calcular as probabilidades de captura e de subsistência e estabelecer uma equivalência não-trivial entre o modelo Random Link com memória $\mu = 2$ e o modelo Random Map com memória $\mu = 0$. Também calculamos a distribuição conjunta dos tempos de transiente e período de atratores, e mostramos que a variável relevante é $n = t + p$. Obtivemos expressões simples para as distribuições marginais.

No Cap. 5 tratamos analiticamente a caminhada do turista com memória arbitrária em sistemas unidimensionais. Abordamos inicialmente o meio semi-infinito, calculando a distribuição do número de sítios explorados. Como resultado exato, obtivemos uma fórmula recursiva, que concorda com o caso $\mu = 1$ obtido previamente. Através de uma aproximação de campo médio, pudemos calcular as probabilidades de exploração e de recuo, e a distribuição do número de sítios explorados pode então ser aproximada por uma distribuição geométrica. A seguir justificamos a equivalência não-trivial entre o meio semi-infinito e o meio finito. Explorando esta equivalência, calculamos a probabilidade de percolação para o meio finito com N pontos e mostramos a existência de uma memória crítica $\mu_1 = \log_2 N$ para este meio.

No Cap. 6 obtivemos uma dedução alternativa para a fórmula de Cox, apresentando os resultados finais em termos de distribuições estatísticas conhecidas. Também obtivemos estas probabilidades de vizinhança para o modelo Random Link (no limite termodinâmico e para sistemas finitos) e para o modelo Random Map (nos casos $D_{i,i} = 0$ e $D_{i,i} \neq 0$).

O problema do turista se revelou como um terreno fértil para o estudo didático de várias distribuições estatísticas conhecidas. No contexto de meios desordenados, estas distribuições adquirem significados intuitivos, possibilitando inferir várias relações entre elas. Tais distribuições foram reunidas no Apêndice D.

Como perpectivas para trabalhos futuros, podemos citar alguns pontos ainda em aberto:

- não foi possível obter as probabilidades de vizinhança em ambientes poissônicos com dimensionalidade e tamanho finitos;
- os resultados analíticos obtidos para o modelo Random Link se aplicam apenas ao limite de baixa memória ($\mu \ll N$) e não explicam o comportamento caótico observado em simulações com memórias mais altas;

- outras estatísticas podem ser exploradas para os sistemas unidimensionais, tais como tempo de transiente e período dos atratores;
- ao contrário do caso $\mu = 1$, os resultados obtidos para o caso $\mu = 2$ não puderam ser imediatamente estendidos para sistemas com dimensionalidade arbitrária.

APÊNDICE A – Integrais Gaussianas

Neste apêndice explicitamos o cálculo das integrais Gaussianas, que serão utilizadas no Apêndice C como ferramenta no cálculo do volume de uma hiperesfera.

A.0.1 Ordem Zero

A integral gaussiana de ordem zero é definida por [39]

$$\mathcal{I}_0 = \int_{-\infty}^{\infty} \mathrm{d}x\, e^{-x^2} = \sqrt{\pi}\ , \tag{A.1}$$

e representa a área sob a curva $y = e^{-x^2}$ ao longo de todo o eixo real. O truque para calcular esta integral consiste em elevá-la ao quadrado

$$\mathcal{I}_0^2 = \int_{-\infty}^{\infty} \mathrm{d}x\ e^{-x^2} \int_{-\infty}^{\infty} \mathrm{d}y\ e^{-y^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}x\ \mathrm{d}y\ e^{-(x^2+y^2)}\ . \tag{A.2}$$

Geometricamente, esta expressão representa o volume sob a superfície $z = e^{-(x^2+y^2)}$ varrendo-se todo o plano cartesiano, como mostra a Fig. 19.

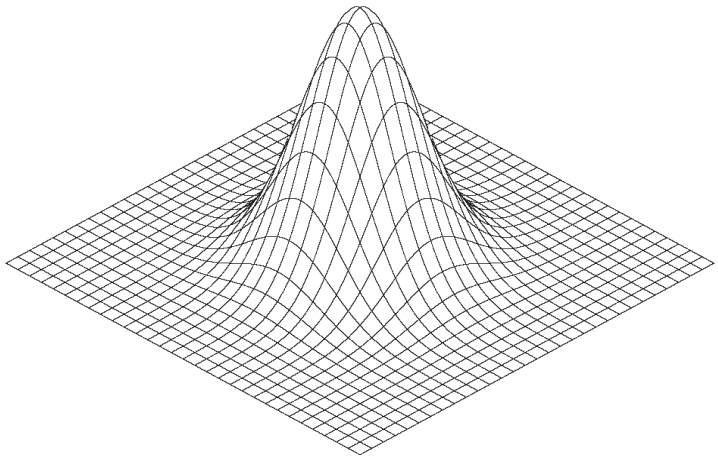


Figura 19: Superfície gerada pela função $z = e^{-(x^2+y^2)}$ em coordenadas retangulares.

A identidade $R^2 = x^2 + y^2$ revela que o integrando da Eq. A.2 tem apenas dependência radial. Por isso, é conveniente aplicar uma transformação de coordenadas retangulares para coordenadas polares

$$\mathcal{I}_0^2 = \int_0^\infty \int_0^{2\pi} \mathrm{d}R \, \mathrm{d}\phi \, R e^{-R^2} ,$$

como mostra a Fig. 20.

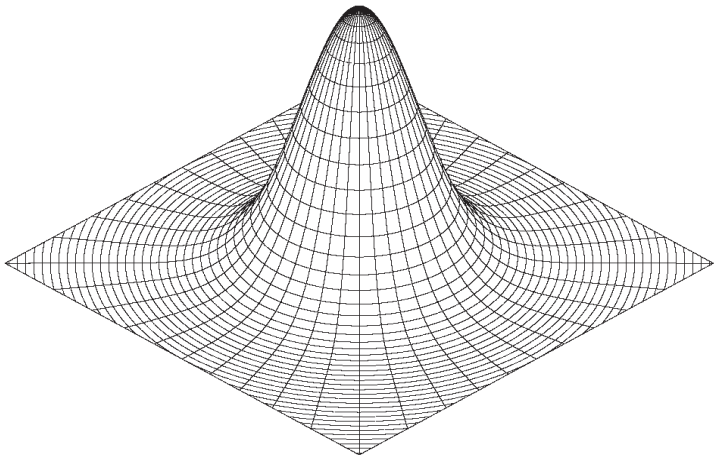


Figura 20: Superfície gerada pela função $z = e^{-R^2}$ em coordenadas polares (cilíndricas).

A integral angular é prontamente calculada, $\int_0^{2\pi} \mathrm{d}\phi = 2\pi$, e a integral radial remanescente é calculada por substituição de variáveis, obtendo-se

$$\mathcal{I}_0^2 = 2\pi \int_0^\infty \mathrm{d}R \, R e^{-R^2} = \pi \int_0^\infty \mathrm{d}R^2 \, e^{-R^2} = \pi \left[-\frac{e^{-R^2}}{2} \right]_0^\infty = \pi .$$

Extraindo a raiz quadrada chega-se ao resultado $\mathcal{I}_0 = \sqrt{\pi}$.

A.0.2 Ordem Arbitrária

A integral gaussiana de ordem n é definida por [39]

$$\mathcal{I}_n = \int_0^\infty \mathrm{d}R \, R^n e^{-R^2} = \frac{1}{2} \, \Gamma\left(\frac{n+1}{2} \right) , \tag{A.3}$$

onde $\Gamma(z)$ é a função gama, definida pela Eq. B.3. Observe que esta definição nos remete à Eq. B.4 e que esta integral está relacionada com o n-ésimo momento da distribuição Meia-Gaussiana, dada pela Eq. D.30.

APÊNDICE B – Algumas Funções Especiais

Reunimos neste Apêndice algumas funções especiais que foram amplamente utilizadas ao longo desta Tese. Apresentamos também algumas propriedades, curiosidades e aproximações destas funções.

Iniciamos com as funções do tipo erro, cujos integrandos são uma função Gaussiana. Passaremos em seguida às funções do tipo gama, cujos integrandos são o produto entre uma lei de potência e uma função exponencial. Continuamos com as funções do tipo beta, cujos integrandos são o produto de duas leis de potência. Finalizamos com as funções do tipo hipergeométrica.

B.1 Funções do Tipo Erro

A função erro [40] é definida como

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{d}t \; e^{-t^2} = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k \cdot z^{2k+1}}{k!(2k+1)} \tag{B.1}$$

e a função erro complementar [40] definida por

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} \mathrm{d}t \; e^{-t^2} \; .$$

Como conseqüência destas definições, tem-se

$$\mathrm{erf}(z) + \mathrm{erfc}(z) = 1.$$

Para $|z| \gg 1$, a função erro complementar tem a seguinte forma assintótica [40]

$$\mathrm{erfc}(z) = \frac{e^{-z^2}}{z\sqrt{\pi}} \left(1 - \frac{z^2}{2} + \cdots \right) \; . \tag{B.2}$$

Observe que estas funções estão relacionadas com a distribuição Gaussiana acumulada.

B.2 Funções do Tipo Gama

A função gama [40] (complete gamma function) é definida por

$$\Gamma(a) \;=\; \int_0^\infty \mathrm{d}t\; t^{a-1}e^{-t}\;, \tag{B.3}$$

onde o argumento a é um complexo qualquer, excetuando-se os inteiros não-positivos. Aplicando-se a substituição $t' = e^{-t}$, podemos escrever a função gama em termos do logaritmo natural

$$\Gamma(a) \;=\; \int_0^1 \mathrm{d}t\; \ln^{a-1}\left(\frac{1}{t}\right)$$

ou, substituindo $t' = \sqrt{t}$, em termos do produto de uma potência por uma Gaussiana

$$\Gamma(a) \;=\; 2\int_0^\infty \mathrm{d}t\; t^{2a-1}e^{-t^2}\;, \tag{B.4}$$

que para $a = 1/2$ assume o valor particularmente importante $\Gamma(1/2) = \sqrt{\pi}$ e se torna a integral Gaussiana de ordem zero $\mathcal{I}_0 = \Gamma(1/2)$ dada pela Eq. A.1. Sua principal propriedade

$$\Gamma(a+1) \;=\; a\,\Gamma(a)$$

(obtida integrando-se por partes a Eq. B.3) a torna uma generalização do fatorial

$$\Gamma(n+1) \;=\; n!\;,$$

para n natural não-nulo.

A função gama incompleta [40] (lower incomplete gamma function) é definida por

$$\gamma(a,b) \;=\; \int_0^b \mathrm{d}t\; t^{a-1}e^{-t} \tag{B.5}$$

e apresenta a seguinte relação para com a função erro (dada pela Eq. B.1)

$$\gamma(1/2,x) \;=\; \sqrt{\pi}\,\mathrm{erf}(\sqrt{x})\;. \tag{B.6}$$

A função gama incompleta complementar [40] (upper incomplete gamma function) é escrita como

$$\Gamma(a,b) \;=\; \int_b^\infty \mathrm{d}t\; t^{a-1}e^{-t}\;. \tag{B.7}$$

Observe que, por definição,

$$\gamma(a,b) + \Gamma(a,b) \quad = \quad \Gamma(a)$$

e que os símbolos γ (minúsculo) e Γ (maiúsculo) são mnemônicos para os termos "lower tail" e "upper tail", respectivamente.

A função gama incompleta normalizada (regularized lower gamma function) é definida por

$$\mathrm{P}(a,b) \quad = \quad \frac{\gamma(a,b)}{\Gamma(a)} \tag{B.8}$$

e a função gama incompleta complementar normalidada (regularized upper gamma function) é definida por

$$\mathrm{Q}(a,b) \quad = \quad \frac{\Gamma(a,b)}{\Gamma(a)} \ . \tag{B.9}$$

Observe que, por definição,

$$\mathrm{P}(a,b) + \mathrm{Q}(a,b) \quad = \quad 1$$

e que a função gama $\Gamma(a)$ assume o papel de fator de normalização de $\mathrm{P}(a,b)$ e de $\mathrm{Q}(a,b)$. Chama-se a atenção para o fato do integrando destas funções serem o produto de uma lei de potência por uma exponencial.

B.3 Funções do Tipo Beta

A função beta [40] é definida como

$$\mathrm{B}(a,b) \quad = \quad \int_0^1 \mathrm{d}t \ t^{a-1}(1-t)^{b-1} \tag{B.10}$$

e guarda a seguinte relação com a função gama

$$\mathrm{B}(a,b) \quad = \quad \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \ . \tag{B.11}$$

Para a e b naturais não-nulos, a função beta apresenta a seguinte relação com o binômio de Newton

$$\frac{1}{\mathrm{B}(a,b)} \quad = \quad \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{(a+b-1)!}{(a-1)!(b-1)!} = (a+b-1)\begin{pmatrix} a+b-2 \\ a-1 \end{pmatrix}$$

e para $b \gg a$ a Eq. B.11 pode ser aproximada por

$$\mathrm{B}(a,b) \approx \frac{\Gamma(a)}{b^a} \ . \tag{B.12}$$

A função beta incompleta é definida por

$$\mathrm{B}_z(a,b) = \int_0^z \mathrm{d}t \ t^{a-1}(1-t)^{b-1}$$

e definimos a função beta incompleta complementar por

$$\mathrm{B}'_z(a,b) = \int_z^\infty \mathrm{d}t \ t^{a-1}(1-t)^{b-1} \ , \tag{B.13}$$

apesar de não ser comumente abordada na literatura. Observe que, por definição,

$$\mathrm{B}_z(a,b) + \mathrm{B}'_z(a,b) \ = \ 1 \ .$$

A função beta incompleta normalizada [40] é definida como

$$\mathrm{I}_z(a,b) \ = \ \frac{\mathrm{B}_z(a,b)}{\mathrm{B}(a,b)} \ , \tag{B.14}$$

onde a função beta $\mathrm{B}(a,b)$ é o fator de normalização de $\mathrm{I}_z(a,b)$.

Observe que os integrandos das funções do tipo beta são um produto de duas leis de potência.

B.4 Funções do Tipo Hipergeométrica

A função hipergeométrica generalizada é definida por

$${}_p\mathrm{F}_q \left[\begin{array}{cccc} a_1 & a_2 & \dots & a_p \\ b_1 & b_2 & \dots & b_q \end{array} ; x \right] = \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \cdots (a_p)_n}{(b_1)_n (b_2)_n \cdots (b_q)_n} \frac{x^n}{n!}$$

onde

$$(a)_n \ = \ a(a+1)(a+2)\cdots(a+n-1) = \frac{(a+n-1)!}{(a-1)!} = \frac{\Gamma(a+n)}{\Gamma(a)}$$

é a notação de Pochhammer para o fatorial ascendente, cuja aproximação $(a)_n \approx a^n$ é válida para $a \gg n$. O caso particularmente importante $p = 2$ e $q = 1$

$${}_2\mathrm{F}_1(a,b;c;x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{t_n x^n}{n!} \ , \tag{B.15}$$

onde

$$\frac{t_{k+1}}{t_k} = \frac{(k+a)(k+b)}{k+c} \qquad \text{e} \qquad t_0 = 1 \ ,$$

é conhecido como função hipergeométrica de Gauss.

APÊNDICE C – Áreas de Hipersuperfícies, Volume de Hiperesferas, de Hipercalotas e de Outras Hiperformas

Apesar do resultado do hipervolume de uma hiperesfera ser muito utilizado, raramente encontramos a sua dedução nos livros textos. Neste Apêndice, apresentamos este cálculo e de formas afins que são utilizadas nesta Tese.

O volume de uma hiperesfera d-dimensional de raio R é definido como

$$V_d(R) = \int\int\cdots\int_{x_1^2+x_2^2+\cdots+x_d^2\leq R^2} \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\,R^d\,. \tag{C.1}$$

Aplicando as substituições $x_j = y_jR \;\Rightarrow\; \mathrm{d}x_j = \mathrm{d}y_jR$, com $j = 1, 2, \ldots, d$, torna-se evidente que o volume $V_d(R)$ é proporcional a R^d, ou seja,

$$V_d(R) = B_dR^d\,, \tag{C.2}$$

onde o fator de proporcionalidade

$$B_d = \int\int\cdots\int_{y_1^2+y_2^2+\cdots+y_d^2\leq 1} \mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_d \tag{C.3}$$

depende exclusivamente da dimensionalidade d e representa o volume de uma hiperesfera d-dimensional de raio unitário, isto é, $B_d = V_d(1)$. Portanto, o cálculo de $V_d(R)$ se resume em obter B_d. A seguir, calcularemos $V_d(R)$ de três maneiras distintas: através de coordenadas retangulares, de coordenadas polares e de integrais gaussianas.

C.1 Coordenadas Retangulares

As integrais da Eq. C.1 somam todos os elementos de volume $\mathrm{d}V = \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d$ que estejam a uma distância máxima de R do centro da hiperesfera. A Fig. 21 ilustra os elementos de volume para o caso $d = 3$. Apesar de cada elemento de volume ser simples,

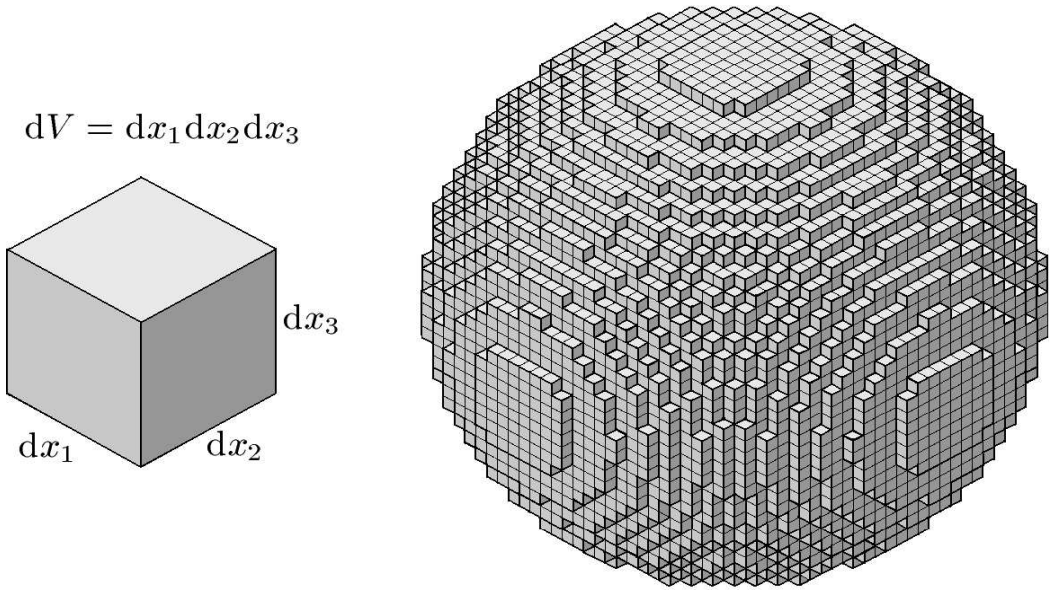


Figura 21: Um elemento de volume e a construção de uma hiperesfera tridimensional.

a restrição imposta pela Eq. C.1 aos limites de integração é complicada. Porém, além de hipercubos, outros elementos de volume podem ser considerados. Nesta seção utilizaremos hipercilindros como elementos de volume.

Para tornar o processo mais intuitivo, calculemos inicialmente o "volume" de uma "hiperesfera unidimensional" ($d = 1$) de raio R (ou seja, o comprimento de um segmento de reta). Neste caso simples, a Eq. C.1 torna-se

$$V_1(R) \;=\; \int_{x_1^2\leq R^2} \mathrm{d}x_1 = \underbrace{\int_{-R}^{R} \mathrm{d}x_1}_{\mathcal{I}_1} = \underbrace{2}_{C_1} R\,,$$

como mostra a Fig. 22.

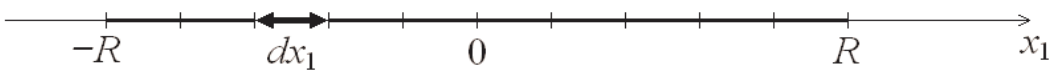
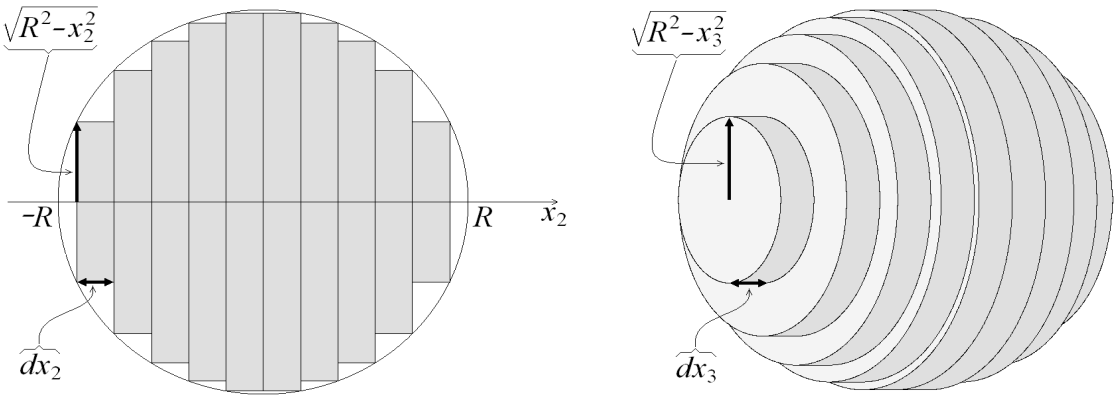


Figura 22: Construção de uma hiperesfera uni-dimensional.

Em vez de calcular o "volume" de uma "hiperesfera bidimensional" ($d = 2$) de raio

unitário (isto é, a área de um círculo) a partir da Eq. C.1, podemos imaginá-la como sendo composta por infinitos "cilindros bidimensionais" de mesma altura infinitesimal $\mathrm{d}x_2$ cujas bases são "hiperesferas unidimensionais" de raios $\sqrt{R - x_2^2}$, como ilustra a Fig. 23a.



(a) (b)

Figura 23: Construção das hiperesferas bi e tridimensional a partir da união de vários cilindros bi e tridimensionais, respectivamente.

Temos, então

$$V_2(R) = \int_{-R}^{R} \mathrm{d}x_2 V_1\left(\sqrt{R^2 - x_2^2}\right) = C_1 \underbrace{\int_{-R}^{R} \mathrm{d}x_2 (R^2 - x_2^2)^{\frac{1}{2}}}_{\mathcal{I}_2} = C_1 \underbrace{\frac{\pi}{2}}_{C_2} R^2 = \pi R^2 \ .$$

Nesta expressão, a integral $\mathcal{I}_2$ é prontamente determinada pelas técnicas usuais do cálculo (substituição de variável $x = R\sin\theta$, integração por partes em $\int \mathrm{d}\theta \cos^2\theta$ e identidade trigonométrica $\sin^2\theta = 1 - \cos^2\theta$).

Para calcular o volume de uma esfera ($d = 3$) de raio R, podemos somar os elementos de volume de infinitos cilindros de mesma altura infinitesimal $\mathrm{d}x_3$ cujas bases são círculos de raios $\sqrt{R^2 - x_3^2}$, como mostra a Fig. 23b (neste caso, os termos volume, esfera e cilindro são propriamente aplicados). Assim,

$$V_3(R) = \int_{-R}^{R} \mathrm{d}x_3 V_2\left(\sqrt{R^2 - x_3^2}\right) = C_1 C_2 \underbrace{\int_{-R}^{R} \mathrm{d}x_3 (R^2 - x_3^2)^{\frac{2}{2}}}_{\mathcal{I}_3} = C_1 C_2 \underbrace{\frac{4}{3}}_{C_3} R^3 = \frac{4\pi}{3} R^3 \ .$$

Nesta expressão, a integral $\mathcal{I}_3$ é trivialmente calculada.

Não é possível construir figuras que ilustrem fielmente a construção de uma hiperesfera com dimensionalidade $d \geq 4$, porém o processo algébrico continua válido: o volume uma hiperesfera quadri-dimensional pode ser calculado somando-se os elementos de volume de infinitos hipercilindros de mesma altura infinitesimal $\mathrm{d}x_4$ cujas bases são esferas de raios

$\sqrt{R^2 - x_4^2}$, ou seja,

$$V_4(R) = \int_{-R}^{R} \mathrm{d}x_4 V_3\left(\sqrt{R^2 - x_4^2}\right) = C_1C_2C_3 \underbrace{\int_{-R}^{R} \mathrm{d}x_4 (R^2 - x_4^2)^{\frac{3}{2}}}_{\mathcal{I}_4} = C_1C_2C_3 \underbrace{\frac{3\pi}{8}}_{C_4} R^4 = \frac{\pi^2}{2} R^4 \ .$$

Nesta expressão, a integral $\mathcal{I}_4$ também pode ser calculada pelas técnicas tradicionais do cálculo, porém é consideravelmente mais trabalhosa que $\mathcal{I}_2$.

Generalizando este raciocínio, o volume $V_d(R)$ de uma hiperesfera d-dimensional é dado pela fórmula recursiva $V_d(R) = \int_{-R}^{R} \mathrm{d}x_d V_{d-1}\left(\sqrt{R^2 - x_d^2}\right)$, que após d recursões resulta em $V_d(R) = C_1C_2C_3 \cdots C_d R^d$. O problema, então, se resume a determinar os coeficientes C_1, C_2, C_3, ..., C_d através do cálculo uma única integral genérica

$$\mathcal{I}_j = \int_{-R}^{R} \mathrm{d}x_j (R^2 - x_j^2)^{\frac{j-1}{2}} \ . \tag{C.4}$$

Para simplificar a notação, podemos suprimir o índice j da variável de integração x_j (pois trata-se de uma variável muda). Como o integrando é uma função par de x, podemos explorar a simetria para simplificar os limites de integração: $\mathcal{I}_j = 2\int_0^R \mathrm{d}x (R^2 - x^2)^{(j-1)/2}$. Colocando R^{j-1} em evidência, temos $\mathcal{I}_j = 2R^{j-1}\int_0^R \mathrm{d}x \left[1 - (x/R)^2\right]^{(j-1)/2}$. Aplicando a substituição de variáveis

$$\left(\frac{x}{R}\right)^2 = t \quad \Rightarrow \quad x = Rt^{\frac{1}{2}} \quad \Rightarrow \quad \mathrm{d}x = \frac{1}{2} R \, \mathrm{d}t \, t^{-\frac{1}{2}} \ , \tag{C.5}$$

a integral $\mathcal{I}_j$ pode ser expressa em termos da função beta (Eq. B.10)

$$\mathcal{I}_j = 2R^{j-1}\frac{1}{2}R\int_0^1 \mathrm{d}t \, t^{-\frac{1}{2}}(1-t)^{\frac{j-1}{2}} = \mathrm{B}\left(\frac{1}{2}, \frac{j+1}{2}\right) R^j$$

e o coeficiente C_j fica automaticamente determinado

$$\mathcal{I}_j = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{j+1}{2}\right)}{\Gamma\left(\frac{j}{2}+1\right)} R^j = \underbrace{\frac{\pi^{\frac{1}{2}}\Gamma\left(\frac{j+1}{2}\right)}{\Gamma\left(\frac{j}{2}+1\right)}}_{C_j} R^j \ . \tag{C.6}$$

Portanto, o volume $V_d(R)$ vale

$$\begin{aligned} V_d(R) &= C_1C_2C_3 \cdots C_d R^d \\ &= \frac{\pi^{\frac{1}{2}}\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)} \cdot \frac{\pi^{\frac{1}{2}}\Gamma\left(\frac{3}{2}\right)}{\Gamma(2)} \cdot \frac{\pi^{\frac{1}{2}}\Gamma(2)}{\Gamma\left(\frac{5}{2}\right)} \cdots \frac{\pi^{\frac{1}{2}}\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)} \cdot R^d = \frac{\pi^{\frac{d}{2}} R^d}{\Gamma\left(\frac{d}{2}+1\right)} \ . \end{aligned} \tag{C.7}$$

Alternativamente, $V_d(R)$ pode ser expresso em termos de $\mathcal{I}_d$

$$\begin{aligned} V_d(R) &= C_1C_2C_3\cdots C_{d-1}\mathcal{I}_d = \frac{\pi^{\frac{1}{2}}\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)}\cdot\frac{\pi^{\frac{1}{2}}\Gamma\left(\frac{3}{2}\right)}{\Gamma(2)}\cdot\frac{\pi^{\frac{1}{2}}\Gamma(2)}{\Gamma\left(\frac{5}{2}\right)}\cdots\frac{\pi^{\frac{1}{2}}\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)}\cdot\mathcal{I}_d \\ &= \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)}\cdot \mathrm{B}\left(\frac{1}{2},\frac{d+1}{2}\right)\cdot R^d\,. \end{aligned} \tag{C.8}$$

Esta expressão será útil na Sec. C.1.2 para determinar do volume relativo de um crescente.

C.1.1 Volume de uma Hipercalota Esférica

Considere a hipercalota esférica d-dimensional $\mathcal{C}$ determinada por um hiperplano que intercepta uma hiperesfera de raio R a uma distância $y \in [0, R]$ de seu centro. A Fig. 24 ilustra o caso particular $d = 2$.

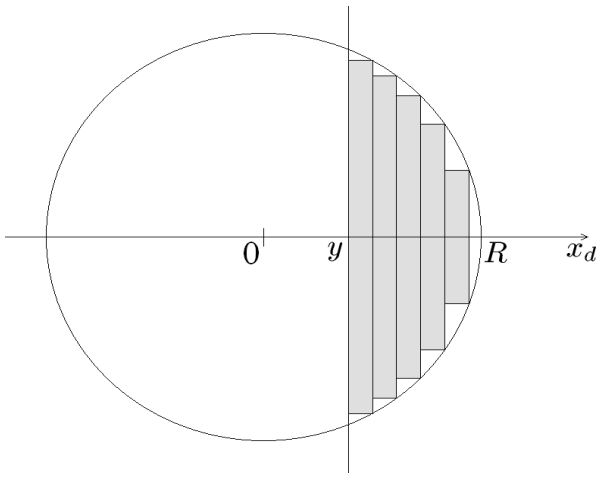


Figura 24: Hipercalota esférica bidimensional. Observe que o intervalo de integração de x_d é de y a R.

Para obter esta hipercalota, podemos reunir infinitos hipercilindros de mesma altura infinitesimal $\mathrm{d}x_d$ cujas bases são hiperesferas $(d-1)$-dimensionais de raios $\sqrt{R^2 - x_d^2}$, com x_d variando de y a R. Portanto o volume desta hipercalota é

$$V_{\mathcal{C},d}(R) = \int_y^R \mathrm{d}x_d V_{d-1}\left(\sqrt{R^2 - x_d^2}\right) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)}\int_y^R \mathrm{d}x(R^2 - x^2)^{\frac{d-1}{2}}\,,$$

onde aplicamos o resultado da Eq. C.7. Observe que o integrando é o mesmo de $\mathcal{I}_j$ (Eq. C.4), apenas os limites de integração são diferentes. Colocando R^{d-1} em evidência e

aplicando a mesma substituição de variáveis da Eq. C.5, temos

$$V_{\mathcal{C},d}(R) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} R^{d-1} \frac{1}{2} R \int_{(y/R)^2}^{1} \mathrm{d}t\, t^{-\frac{1}{2}} (1-t)^{\frac{d-1}{2}} \ .$$

Simplificando, obtemos a seguinte expressão para o volume da hipercalota

$$V_{\mathcal{C},d}(R) = \frac{1}{2} \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} \mathrm{B}'_{(y/R)^2}\left(\frac{1}{2}, \frac{d+1}{2}\right) R^d \ , \tag{C.9}$$

onde $\mathrm{B}'_z(a,b)$ é a função beta incompleta complementar, definida pela Eq. B.13.

C.1.2 Volume Relativo de um Crescente

Considere duas hiperesferas idênticas de raios iguais a R, cujos centros distam xR entre si, com x variando de 0 (hiperesferas concêntricas) a 2 (hiperesferas tangentes externamente). O volume da intersecção destas duas hiperesferas pode ser calculado simplesmente multiplicando-se a Eq. C.9 por 2.

Denominamos por *crescente* o conjunto de pontos de uma hiperesfera que não pertencem à intersecção. Obviamente, o volume de um desses crescentes é dado por $V_d(R) - 2V_{\mathcal{C},d}(R)$. Então, a fração $p_d(x)$ do volume que este crescente ocupa na hiperesfera é dada por

$$p_d(x) = \frac{V_d(R) - 2V_{\mathcal{C},d}(R)}{V_d(R)} = \frac{\frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} R^d \left[\mathrm{B}\left(\frac{1}{2}, \frac{d+1}{2}\right) - \mathrm{B}'_{(x/2)^2}\left(\frac{1}{2}, \frac{d+1}{2}\right)\right]}{\frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} R^d \mathrm{B}\left(\frac{1}{2}, \frac{d+1}{2}\right)} = \frac{\mathrm{B}_{(x/2)^2}\left(\frac{1}{2}, \frac{d+1}{2}\right)}{\mathrm{B}\left(\frac{1}{2}, \frac{d+1}{2}\right)} \ ,$$

onde usamos os resultados das Eqs. C.8 e C.9.

Logo, o volume relativo $p_d(x)$ do crescente pode ser expresso em termos da função beta incompleta normalizada (Eq. B.14)

$$p_d(x) = \mathrm{I}_{(x/2)^2}\left(\frac{1}{2}, \frac{d+1}{2}\right) \ . \tag{C.10}$$

Observe que $p_d(x)$ não depende do raio R da hiperesfera, já que representa uma proporção.

C.2 Coordenadas Esféricas

Uma forma alternativa de calcular volume de uma hiperesfera, dado pela Eq. C.1, é através do uso de coordenadas esféricas generalizadas [41]

$$
\begin{aligned}
x_1 &= \rho \sin\phi_1 \sin\phi_2 \sin\phi_3 \sin\phi_4 \cdots \sin\phi_{d-1} \\
x_2 &= \rho \cos\phi_1 \sin\phi_2 \sin\phi_3 \sin\phi_4 \cdots \sin\phi_{d-1} \\
x_3 &= \rho \cos\phi_2 \sin\phi_3 \sin\phi_4 \cdots \sin\phi_{d-1} \\
&\vdots \\
x_{d-2} &= \rho \cos\phi_{d-3} \sin\phi_{d-2} \sin\phi_{d-1} \\
x_{d-1} &= \rho \cos\phi_{d-2} \sin\phi_{d-1} \\
x_d &= \rho \cos\phi_{d-1}
\end{aligned}
$$

com $\rho \in [0, R]$, $\phi_1 \in [0, 2\pi]$ e $\phi_j \in [0, \pi]$, $j = 2, 3, \ldots, d-1$. Esta transformação estabelece uma correspondência biunívoca entre as d-uplas $(x_1, x_2, \ldots, x_d)$ e $(\rho, \phi_1, \phi_2, \ldots, \phi_{d-1})$. No entanto, devemos determinar o elemento de volume neste novo sistema de coordenadas. A Fig. 25 ilustra, para o caso $d = 3$, os elementos de volume em coordenadas esféricas.

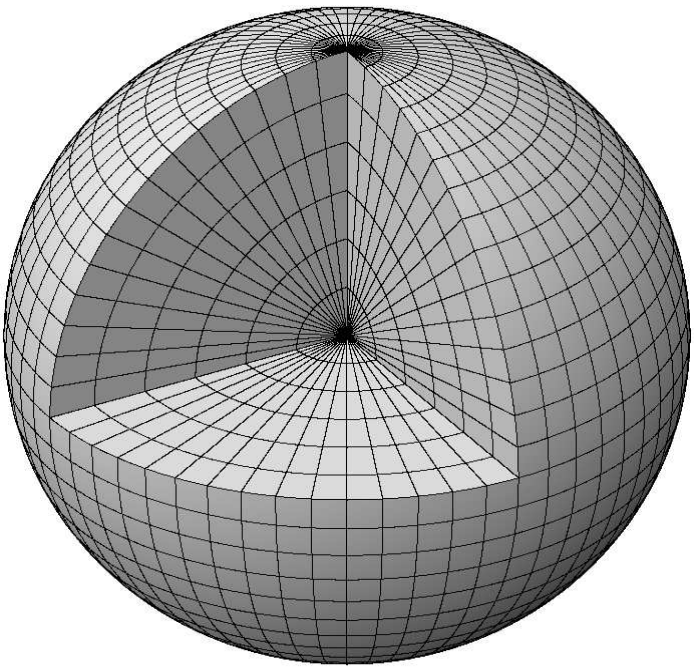


Figura 25: Elementos de volume de uma hiperesfera tridimensional em coordenadas esféricas. O corte no primeiro octante foi feito para que se possa visualizar os elementos de volume no interior da esfera.

Para uma dimensionalidade arbitrária d, o pré-fator do elemento de volume em coordenadas esféricas é dado pelo determinante jacobiano

$$\det\left(\frac{\partial(x_1,x_2,x_3,\ldots,x_d)}{\partial(\rho,\phi_1,\phi_2,\ldots,\phi_{d-1})}\right) = \begin{vmatrix} \frac{\partial x_1}{\partial\rho} & \frac{\partial x_1}{\partial\phi_1} & \frac{\partial x_1}{\partial\phi_2} & \cdots & \frac{\partial x_1}{\partial\phi_{d-1}} \\ \frac{\partial x_2}{\partial\rho} & \frac{\partial x_2}{\partial\phi_1} & \frac{\partial x_2}{\partial\phi_2} & \cdots & \frac{\partial x_2}{\partial\phi_{d-1}} \\ \frac{\partial x_3}{\partial\rho} & \frac{\partial x_3}{\partial\phi_1} & \frac{\partial x_3}{\partial\phi_2} & \cdots & \frac{\partial x_3}{\partial\phi_{d-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial\rho} & \frac{\partial x_d}{\partial\phi_1} & \frac{\partial x_d}{\partial\phi_2} & \cdots & \frac{\partial x_d}{\partial\phi_{d-1}} \end{vmatrix}.$$

Calculando as derivadas parciais e abreviando $\sin\phi_j = s_j$ e $\cos\phi_j = c_j$, temos:

$$\begin{vmatrix} s_1s_2s_3s_4\ldots s_{d-1} & \rho c_1s_2s_3s_4\ldots s_{d-1} & \rho s_1c_2s_3s_4\ldots s_{d-1} & \rho s_1s_2c_3s_4\ldots s_{d-1} & \ldots & \rho s_1s_2s_3s_4\ldots c_{d-1} \\ c_1s_2s_3s_4\ldots s_{d-1} & -\rho s_1s_2s_3s_4\ldots s_{d-1} & \rho c_1c_2s_3s_4\ldots s_{d-1} & \rho c_1s_2c_3s_4\ldots s_{d-1} & \ldots & \rho c_1s_2s_3s_4\ldots c_{d-1} \\ c_2s_3s_4\ldots s_{d-1} & 0 & -\rho s_2s_3s_4\ldots s_{d-1} & \rho c_2c_3s_4\ldots s_{d-1} & \ldots & \rho c_2s_3s_4\ldots c_{d-1} \\ c_3s_4\ldots s_{d-1} & 0 & 0 & -\rho s_3s_4\ldots s_{d-1} & \ldots & \rho c_3s_4\ldots c_{d-1} \\ c_4\ldots s_{d-1} & 0 & 0 & 0 & \ldots & \rho c_4\ldots c_{d-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{d-1} & 0 & 0 & 0 & \ldots & -\rho s_{d-1} \end{vmatrix}.$$

Para resolver este determinante, realizamos as seguintes manobras:

- dividimos a 2ª coluna por $\rho s_2s_3s_4\cdots s_{d-1}$, a 3ª por $\rho s_3s_4\cdots s_{d-1}$, e assim sucessivamente até a $(d-1)$ª coluna (que será dividida por ρs_{d-1}) e transportamos a 1ª coluna para a última posição à direita:

$$\rho^{d-1}s_2s_3^2s_4^3\ldots s_{d-1}^{d-2}\cdot\begin{vmatrix} c_1 & s_1c_2 & s_1s_2c_3 & s_1s_2s_3c_4 & \ldots & s_1s_2s_3s_4\ldots c_{d-1} & s_1s_2s_3s_4\ldots s_{d-1} \\ -s_1 & c_1c_2 & c_1s_2c_3 & c_1s_2s_3c_4 & \ldots & c_1s_2s_3s_4\ldots c_{d-1} & c_1s_2s_3s_4\ldots s_{d-1} \\ 0 & -s_2 & c_2c_3 & c_2s_3c_4 & \ldots & c_2s_3s_4\ldots c_{d-1} & c_2s_3s_4\ldots s_{d-1} \\ 0 & 0 & -s_3 & c_3c_4 & \ldots & c_3s_4\ldots c_{d-1} & c_3s_4\ldots s_{d-1} \\ 0 & 0 & 0 & -s_4 & \ldots & c_4\ldots c_{d-1} & c_4\ldots s_{d-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -s_{d-1} & c_{d-1} \end{vmatrix},$$

- com i variando de 1 a $d-1$, dividimos a $(i+1)$ª linha e multiplicamos a iª coluna por c_k, assim como dividimos as 1ª, 2ª, ..., iª linhas e multiplicamos as 1ª, 2ª, ...,

i^{a} colunas por s_k:

$$\rho^{d-1} s_2 s_3^2 s_4^3 \cdots s_{d-1}^{d-2} \cdot \begin{vmatrix} c_1^2 & c_2^2 & c_3^2 & c_4^2 & \cdots & c_{d-1}^2 & 1 \\ -s_1^2 & c_2^2 & c_3^2 & c_4^2 & \cdots & c_{d-1}^2 & 1 \\ 0 & -s_2^2 & c_3^2 & c_4^2 & \cdots & c_{d-1}^2 & 1 \\ 0 & 0 & -s_3^2 & c_4^2 & \cdots & c_{d-1}^2 & 1 \\ 0 & 0 & 0 & -s_4^2 & \cdots & c_{d-1}^2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -s_{d-1}^2 & 1 \end{vmatrix} ,$$

- de cada linha subtraímos os elementos da linha imediatamente abaixo.

$$\rho^{d-1} s_2 s_3^2 s_4^3 \cdots s_{d-1}^{d-2} \cdot \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -s_1^2 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -s_2^2 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -s_3^2 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -s_4^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -s_{d-1}^2 & 1 \end{vmatrix} .$$

Como a matriz resultante é triangular (todos os elementos acima da diagonal principal são nulos), o determinante é simplesmente o produto dos elementos da diagonal principal, ou seja, igual a 1, e a Eq. C.1 é escrita em coordenadas esféricas generalizadas

$$V_d(R) \;=\; \int_0^R \mathrm{d}\rho\, \rho^{d-1} \int_0^{2\pi} \mathrm{d}\phi_1 \prod_{j=2}^{d-1} \int_0^{\pi} \mathrm{d}\phi_j \sin^{j-1}\phi_j \;. \tag{C.11}$$

Cada integral $\mathcal{I}_j = \int_0^{\pi} \mathrm{d}\phi_j \;\sin^j \phi_j = 2\int_0^{\pi/2} \mathrm{d}\phi_j \;\sin^j \phi_j$ pode ser calculada aplicando a substituição $t = \cos^2\phi \;\Rightarrow\; \mathrm{d}t = -2\,\mathrm{d}\phi\,\cos\phi\,\sin\phi$, obtendo

$$V_d(R) \;=\; \frac{R^d}{d} 2\pi \prod_{j=2}^{d-1} C_j \;, \tag{C.12}$$

onde

$$C_j = 2\int_0^1 \frac{\mathrm{d}t(1-t)^{j/2}}{2t^{1/2}(1-t)^{1/2}} = \int_0^1 \mathrm{d}t\; t^{-1/2}(1-t)^{\frac{j-1}{2}} = \mathrm{B}\left(\frac{1}{2}\,,\frac{j+1}{2}\right)$$

e o cálculo recai na Eq. C.6.

C.3 Integrais Gaussianas

Na dedução da Eq. C.2, concluímos que o volume $V_d(R)$ de uma hiperesfera d-dimensional de raio R é proporcional a R^d, ou seja, $V_d(R) = B_d R^d$, onde o fator B_d depende apenas da dimensionalidade e representa o volume de uma hiperesfera d-dimensional de raio unitário, isto é, $B_d = V_d(1)$. Logo, a área $A_d(R)$ de uma superfície hiperesférica d-dimensional de raio R é dada por

$$A_d(R) = \frac{dV_d(R)}{dR} = dB_d R^{d-1} \tag{C.13}$$

e o elemento de volume de uma hipercasca esférica de raio R e espessura $\mathrm{d}R$ é dado por $\mathrm{d}V = A_d(R)\,\mathrm{d}R = dB_d R^{d-1}\,\mathrm{d}R$, como mostra a Fig. 26.

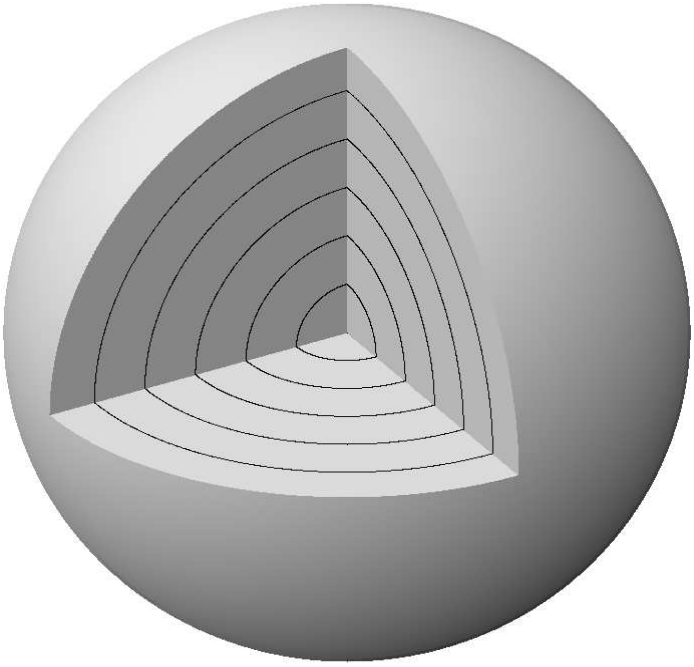


Figura 26: Cascas esféricas como elementos de volume de uma hiperesfera tridimensional. O corte no primeiro octante foi feito para se poder visualizar os elementos de volume no interior da esfera.

Para determinar o fator B_d, considere a integral Gaussiana de ordem zero $\mathcal{I}_0 = \int_{-\infty}^{\infty} \mathrm{d}x\; e^{-x^2}$. Observe que a expressão

$$\begin{aligned} \mathcal{I}_0^d &= \int_{-\infty}^{\infty} \mathrm{d}x_1\; e^{-x_1^2} \int_{-\infty}^{\infty} \mathrm{d}x_2\; e^{-x_2^2} \cdots \int_{-\infty}^{\infty} \mathrm{d}x_d\; e^{-x_d^2} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_d\; e^{-(x_1^2+x_2^2+\cdots+x_d^2)} \end{aligned} \tag{C.14}$$

varre todo o espaço d-dimensional. Tendo em vista e identidade $R^2 = x_1^2 + x_2^2 + \cdots +$

x_d^2, conclui-se que o integrando tem apenas dependência radial. Portanto, em vez de utilizar hipercubos como elementos de volume $\mathrm{d}V = \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d$, é mais conveniente utilizar hipercascas esféricas $\mathrm{d}V = dB_d R^{d-1}\ \mathrm{d}R$ e escrever a Eq. C.14 na forma $\mathcal{I}_0^d = \int_0^\infty \mathrm{d}R\ dB_d\ R^{d-1}e^{-R^2}$. Substituindo $\mathcal{I}_0 = \sqrt{\pi}$ (de acordo com a Eq. A.1), temos

$$(\sqrt{\pi})^d \quad = \quad dB_d \int_0^\infty \mathrm{d}R\ R^{d-1}e^{-R^2}\ . \tag{C.15}$$

Esta última é uma integral gaussiana de ordem $d-1$ dada pela Eq. A.3 e vale $\mathcal{I}_{d-1} = \Gamma(d/2)/2$. Isolando B_d na Eq. C.15, resulta

$$B_d \quad = \quad \frac{\pi^{d/2}}{d}\cdot\frac{2}{\Gamma(d/2)} = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}\ ,$$

que substituído na Eq. C.2 nos leva à Eq. C.1.

APÊNDICE D – Distribuições Estatísticas em Meios Desordenados

Neste Apêndice reunimos algumas distribuições estatísticas que emergiram naturalmente no estudo da caminhada determinista do turista. Diferentemente das apresentações tradicionais dos livros-texto que, em geral, consideram estes processos como sendo temporais (provavelmente devido ao estudo de teoria de filas), apresentamos aqui a emergência destas distribuições em sistemas espaciais, considerando a desordem que quebra a homogeneidade do meio. Nosso objetivo não é fornecer um manual completo sobre tais distribuições, mas dar a elas significados intuitivos à luz de meios desordenados. A Tab. 4 mostra um panorama deste Apêndice, relacionando cada distribuição com sua(s) variável(eis) e o(s) meio(s) de onde emerge.

Variável	Meio			
	Ilimitado		Finito	
	Discreto	Contínuo	Discreto	Contínuo
Nº de pontos num intervalo	Binomial	Poisson	Hipergeométrica	Binomial
Distância entre pontos consecutivos (ou abscissa do 1º sítio)	Geométrica	Exponencial	Hipergeométrica Multivariada	Potência
Distância entre pontos não-consecutivos (ou abscissa do k-ésimo sítio)	Pascal	Gama	Hipergeométrica Multivariada	Beta

Tabela 4: Distribuições estatísticas com suas respectivas variáveis e meios desordenados.

Iniciamos a apresentação com a distribuição de Bernoulli e a pdf uniforme. Estas são as distribuições elementares a partir das quais se obtêm distribuições mais complexas. Passamos então, aos meios desordenados. Começamos considerando os meios ilimitados nas versões discreta e contínua. Para meios discretos, obteremos as distribuições binomial, multinomial, geométrica e Pascal. Em seguida faremos o limite de meios contínuos onde

obteremos as distribuições de Poisson e as funções densidade de probabilidade exponencial e gama. Continuamos com a consideração de meios finitos, também nas versões discreta e contínua. Esta restrição implica na modificação das distribuições e funções densidade de probabilidade obtidas previamente. Para os meios discretos obteremos as distribuições hipergeométrica e hipergeométrica multivariada. No limite dos meios contínuos consideramos novamente a distribuição binomial e introduziremos as pdf's potência e beta. Apresentamos uma generalização da distribuição beta que nos remete à estatística de ordem. Enfatizamos a estreita relação das distribuições acumuladas com as funções especiais obtidas no Apêndice B e mostramos as relações entre tais distribuições, tomando-se casos-limites ou casos-particulares.

Encerramos com a pdf do produto de variáveis uniformes. Apesar desta pdf não emergir naturalmente de um meio desordenado, é utilizada no Cap. 5 e por isso a incluímos aqui.

D.1 Distribuições-base

Nesta seção descrevemos a distribuição de bernoulli e a pdf uniforme, que serão usadas para gerar outras distribuições mais complexas.

D.1.1 Distribuição de Bernoulli

Denomina-se prova de Bernoulli um experimento aleatório com apenas dois resultados possíveis, denominados genericamente sucesso e fracasso. Um exemplo de prova de Bernoulli é o lançamento de uma moeda (honesta ou viciada). Sejam p a probabilidade de ocorrência de sucesso e $q = 1 - p$ a probabilidade de ocorrência de fracasso. Nestas condições, a variável aleatória k definida por

$$k = \begin{cases} 1, & \text{se o resultado for sucesso} \\ 0, & \text{se o resultado for fracasso} \end{cases}$$

segue a distribuição de Bernoulli, dada por

$$P_{Be}(k) = \begin{cases} p, & \text{se } k = 1 \\ q, & \text{se } k = 0 \\ 0, & \text{se } k \neq 0 \text{ e } k \neq 1 \end{cases}$$

ou, de forma abreviada, $P_{Be}(k) = p\delta_{k,1} + q\delta_{k,0}$, onde $\delta_{a,b}$ é o delta de Kronecker, ou ainda $P_{Be}(k) = p^k q^{1-k}$, com $k = 0, 1$.

A média e a variância da distribuição de Bernoulli são

$$\begin{aligned} \mathrm{E}(k) &= p \\ \mathrm{Var}(k) &= pq \; . \end{aligned} \tag{D.1}$$

Apesar de sua simplicidade, a distribuição de Bernoulli assume um papel importante como geradora de outras distribuições.

D.1.2 Distribuição Uniforme (ou Retangular)

Considere que a abscissa y de um ponto seja aleatoriamente gerada de maneira homogênea num intervalo real $[a, b]$. Nestas condições, a variável y tem distribuição uniforme, dada por

$$f_{un}(y)\,\mathrm{d}y = \frac{1}{b-a}\,\mathrm{d}y \; , \qquad y \in [a, b] \; .$$

Em particular, para o intervalo real $[0, 1]$, esta distribuição assume a seguinte forma

$$f_{un}(y)\,\mathrm{d}y = \mathrm{d}y \; , \qquad y \in [0, 1] \; . \tag{D.2}$$

A média e a variância da distribuição uniforme são

$$\begin{aligned} \mathrm{E}(y) &= (a+b)/2 \\ \mathrm{Var}(y) &= (b-a)^2/12 \; . \end{aligned}$$

Observe que a média $\mathrm{E}(y)$ equivale ao ponto médio do intervalo $[a, b]$.

D.2 Meios Ilimitados

Ilustramos como podemos obter as distribuições de probabilidades conhecidas a partir de meios desordenados. A seguir, consideraremos um meio desordenado ilimitado nas versões discreta e contínua. Obteremos as distribuições de algumas variáveis relacionadas a este meio. Por simplicidade, utilizaremos um meio unidimensional. Porém, os resultados também se aplicam a meios com dimensionalidade superior.

D.2.1 Meios Discretos

Na versão discreta, um meio desordenado ilimitado unidimensional pode ser representado por uma reta composta por infinitos segmentos com mesmo comprimento ℓ_0. Cada segmento pode estar ativado (com probabilidade p) ou desativado (com probabilidade $q = 1 - p$).

D.2.1.1 Número de Pontos num Intervalo – Distribuição Binomial

Considere que sejam realizadas N provas de Bernoulli independentes com as mesmas probabilidades p de sucesso e $q = 1 - p$ de fracasso. Como as provas são independentes, a probabilidade de se obter k sucessos seguidos de $N - k$ fracassos (nesta ordem) é igual a

$$\underbrace{p \cdot p \cdots p}_{k \text{ vezes}} \cdot \underbrace{q \cdot q \cdots q}_{N-k \text{ vezes}} = p^k q^{N-k} \ . \tag{D.3}$$

Para determinar a probabilidade ocorrerem k sucessos e $N - k$ fracassos em qualquer ordem, devemos considerar todas as ordenações possíveis. Esta é justamente a distribuição binomial, dada por

$$P_{bi}(k) = \begin{pmatrix} N \\ k \end{pmatrix} p^k q^{N-k} \ , \qquad k = 0, 1, 2, \ldots, N \ , \tag{D.4}$$

onde o binomial

$$\begin{pmatrix} N \\ k \end{pmatrix} = \frac{N!}{k!(N-k)!} \tag{D.5}$$

realiza a contagem e dá nome à distribuição. Dessa forma, a soma de N variáveis com distribuição de Bernoulli tem distribuição binomial.

A média e a variância da distribuição binomial são

$$\begin{aligned} \mathrm{E}(k) &= Np \\ \mathrm{Var}(k) &= Npq \ , \end{aligned}$$

ou seja, são respectivamente iguais à média e à variância da distribuição de Bernoulli (dadas pela Eq. D.1) multiplicadas pelo número N de provas.

A distribuição acumulada é convenientemente escrita em termos da função beta in-

completa normalizada (Eq. B.14)

$$F_{bi}(k) = \sum_{j=0}^{k} P_{bi}(j) = 1 - \mathrm{I}_p(k+1, N-k) \ .$$

Como aplicação física da distribuição binomial, considere um meio desordenado ilimitado discreto. A probabilidade de haver k segmentos ativados (sítios) num trecho de comprimento fixo $\ell = N\ell_0$ contendo N segmentos é dada pela distribuição binomial (ver Fig. 27).

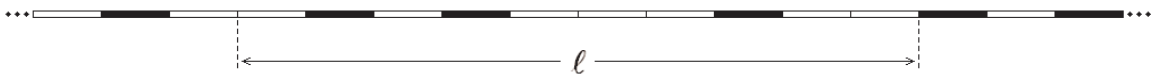


Figura 27: Meio desordenado ilimitado discreto formado por infinitos segmentos de comprimento ℓ_0. Se a probabilidade de cada segmento estar ativado é p, então a probabilidade de obter $k = 3$ segmentos ativados dentro de um trecho de comprimento $\ell = N\ell_0 = 10\ell_0$ com $N = 10$ segmentos é dada pela distribuição binomial (Eq. D.4).

Fisicamente, $\mathrm{E}(k)$ representa o número esperado (médio) λ de sítios dentro de um trecho de comprimento ℓ, de forma que a grandeza $r = \lambda/\ell$ representa a densidade média do meio.

D.2.1.2 Distribuição Multinomial (ou Polinomial)

Considere um experimento aleatório com n possibilidades enumeradas de 1 a n cujas probabilidades de ocorrência p_1, p_2, ..., p_n são constantes (com $p_1 + p_2 + \cdots + p_n = 1$). Seja tal experimento realizado N vezes de maneira independente. A probabilidade de se obter N_1 ocorrências do resultado 1, N_2 do resultado 2, ..., N_n ocorrências do resultado n (com $N_1 + N_2 + \cdots + N_n = N$) é dada pela distribuição multinomial

$$P_{mu}(N_1, N_2, \ldots, N_n) = \frac{N!}{N_1! N_2! \cdots N_n!} \cdot p_1^{N_1} \cdot p_2^{N_2} \cdots p_n^{N_n} \ , \tag{D.6}$$

que é uma generalização da distribuição binomial.

Como interpretação física da distribuição multinomial, considere um meio desordenado ilimitado discreto, no qual cada segmento pode assumir qualquer um dentre n estados possíveis (e não apenas dois estados – ativado e desativado – como na distribuição binomial). A distribuição multinomial fornece a probabilidade de haver, dentro de um trecho de comprimento fixo $\ell = N\ell_0$ com N segmentos, N_1 segmentos no estado 1, N_2 segmentos no estado 2, ..., N_n segmentos no estado n (ver Fig. 28).

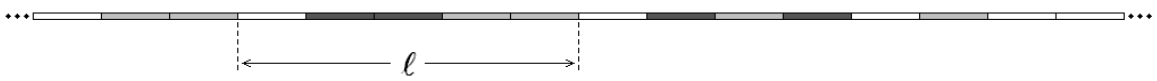


Figura 28: Meio desordenado ilimitado discreto. Se as probabilidades de cada segmento estar nos estados 1 (branco), 2 (cinza) e 3 (preto) são, respectivamente, p_1, p_2 e p_3, então a probabilidade de se ter $N_1 = 1$, $N_2 = 2$ e $N_3 = 2$ segmentos nos estados 1, 2, e 3 dentro de um trecho de comprimento ℓ com $N = 5$ segmentos é dada pela distribuição multinomial (Eq. D.6).

D.2.1.3 Distância entre Pontos Consecutivos – Distribuição Geométrica

Considere que sejam realizadas sucessivas e independentes provas de Bernoulli com as mesmas probabilidades p de sucesso e $q = 1 - p$ fracasso. A distribuição geométrica fornece a probabilidade do primeiro sucesso aparecer justamente na N-ésima realização. Como as $N - 1$ realizações anteriores são fracassos e as provas são independentes, temos simplesmente

$$P_{ge}(N) = q^{N-1}p \ , \qquad N = 1, 2, 3, \ldots, \infty \ . \tag{D.7}$$

Observe que esta expressão é o caso particular da Eq. D.3 em que $k = 1$, pois desejamos que em N tentativas ocorram $N - 1$ fracassos e, em seguida, 1 sucesso (nesta ordem).

A média e a variância da distribuição geométrica são

$$\begin{aligned} \mathrm{E}(N) &= 1/p \\ \mathrm{Var}(N) &= q/p^2 \end{aligned} \tag{D.8}$$

e a distribuição acumulada é dada por

$$F_{ge}(N) = \sum_{j=1}^{N} P_{ge}(j) = 1 - q^N \ . \tag{D.9}$$

A interpretação física da distribuição geométrica é imediata. Ela representa a probabilidade de um caminhante ter de percorrer N passos de um meio discreto para encontrar o primeiro sítio (que poderia ser uma armadilha, onde a caminhada terminaria). Se a caminhada reiniciar-se logo após uma armadilha, a distribuição geométrica fornece a distribuição de probabilidade da distância $\ell = N\ell_0$ (discreta, neste caso) entre armadilhas consecutivas (Fig. 29).

Em termos de grandezas físicas, a Eq. D.7 é escrita como

$$P_{ge}(\ell) = (1 - r\ell_0)^{\frac{\ell}{\ell_0} - 1} r\ell_0 \ , \qquad \ell = \ell_0, 2\ell_0, 3\ell_0, \ldots, \infty \ , \tag{D.10}$$

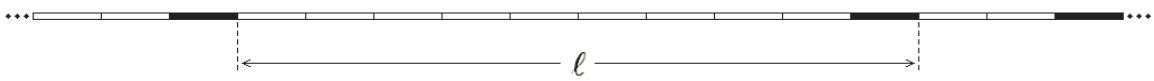


Figura 29: Meio desordenado ilimitado discreto. Se a probabilidade de cada segmento estar ativado é p, então a distância discreta $\ell = N\ell_0$ entre dois segmentos ativados consecutivos segue a distribuição geométrica (Eq. D.7).

onde ℓ_0 é o comprimento de cada segmento e r é a densidade média do meio. Assim como a binomial, a distribuição geométrica é uma distribuição discreta, e também o meio físico é discreto.

D.2.1.4 Distância entre Pontos Não-consecutivos – Distribuição de Pascal (ou Binomial Negativa)

Considere que sejam realizadas sucessivas e independentes provas de Bernoulli com as mesmas probabilidades p de sucesso e $q = 1 - p$ de fracasso. A distribuição de Pascal fornece a probabilidade do k-ésimo sucesso acontecer justamente na N-ésima realização. Observe que:

1. as provas são independentes,
2. a N-ésima realização deve ser sucesso (a probabilidade é p) e
3. deve haver $k - 1$ sucessos dentre as $N - 1$ realizações anteriores (como todas as ordenações possíveis devem ser consideradas, a probabilidade é dada pela distribuição binomial).

Portanto

$$P_{Pa}(N) = p \cdot \begin{pmatrix} N-1 \\ k-1 \end{pmatrix} p^{k-1} q^{(N-1)-(k-1)} = \begin{pmatrix} N-1 \\ k-1 \end{pmatrix} p^k q^{N-k} \ . \tag{D.11}$$

com $N = k, k+1, k+2, \ldots, \infty$. Dessa forma, a soma de k variáveis com distribuição geométrica tem distribuição de Pascal.

A média e a variância da distribuição de Pascal são

$$\begin{aligned} \mathrm{E}(N) &= k/p \\ \mathrm{Var}(N) &= kq/p^2 \ , \end{aligned}$$

ou seja, são respectivamente iguais à média e à variância da distribuilção geométrica (Eq. D.8) multiplicadas por k.

A distribuição acumulada é dada por

$$F_{Pa}(N) = \sum_{j=0}^{N} P_{Pa}(j) = \mathrm{I}_p(k, N-k+1) \ ,$$

onde $\mathrm{I}_z(a,b)$ é a função beta incompleta normalizada (Eq. B.14).

Fisicamente, a distribuição de Pascal fornece a distribuição da distância discreta ℓ entre sítios não-consecutivos de um meio desordenado discreto.

Figura 30: Meio desordenado ilimitado discreto. Se a probabilidade de cada segmento estar ativado é p, então a distância discreta $\ell = N\ell_0$ entre segmentos ativados não-consecutivos de ordem k segue a distribuição de Pascal (Eq. D.11).

Em termos das grandezas físicas r, ℓ e ℓ_0, temos

$$P_{Pa}(\ell) = \begin{pmatrix} \frac{\ell}{\ell_0} - 1 \\ k-1 \end{pmatrix} (r\ell_0)^k (1 - r\ell_0)^{\frac{\ell}{\ell_0} - k} \ , \qquad \ell = k\ell_0, (k+1)\ell_0, (k+2)\ell_0, \ldots, \infty \ . \quad \text{(D.12)}$$

A distribuição de Pascal (Eq. D.11) é semelhante à distribuição binomial (Eq.D.4). Na distribuição binomial, o número de tentativas é um parâmetro e o numero de sucessos é a variável. Na distribuição de Pascal, estes papéis são permutados.

D.2.2 Meios Contínuos

Na versão contínua, um meio desordenado ilimitado unidimensional pode ser representado por infinitos pontos distribuídos de maneira aleatória e independente sobre uma reta segundo uma densidade média de r pontos por unidade de comprimento.

D.2.2.1 Número de Pontos num Intervalo – Distribuição de Poisson

A distribuição de Poisson corresponde ao caso limite da distribuição binomial em que $N \to \infty$ e $p \to 0$, de forma que $\mathrm{E}(k) = Np = \lambda$ permaneça constante. A interpretação física desta situação é que o meio desordenado se torna cada vez menos discretizado: o tamanho ℓ_0 de cada segmento é cada vez menor e o número N de segmentos dentro de um trecho de comprimento fixo ℓ é cada vez maior, de forma que a densidade média r se mantém fixa (o número médio λ de sítios dentro de um trecho de comprimento fixo ℓ

se mantém constante). Dessa forma, a distribuição de Poisson fornece a probabilidade de um trecho de comprimento ℓ conter k sítios. Compare as Figs. 27 e 31 e observe que em ambas há $k = 3$ segmentos ativados dentro de um trecho de comprimento $\ell = N\ell_0$.

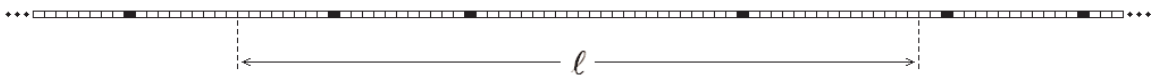


Figura 31: Um meio ilimitado contínuo como caso-limite de um meio ilimitado discreto. O comprimento ℓ_0 e a probabilidade p de um segmento estar ativado tendem a 0, mantendo constante a densidade média r de segmentos ativados. A probabilidade de haver k segmentos ativados dentro de um trecho de comprimento fixo ℓ é dada pela distribuição de Poisson (Eq. D.14).

Observe que substituindo $\ell = N\ell_0$ em $\mathrm{E}(k) = Np = \lambda = r\ell$ obtém-se $p = r\ell_0$, ou seja, a probabilidade p de um segmento estar ativado é proporcional ao seu comprimento ℓ_0. No limite $\ell_0 \to 0$, o meio se torna contínuo e os segmentos ativados se tornam pontos.

$$P_{Po}(k) = \lim_{N\to\infty} \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \tag{D.13}$$

Como N e p não são grandezas estáveis no processo de discretização do meio, é necessário expressar a Eq. D.13 em termos da grandeza relevante $\lambda = r\ell$. Expandindo $\frac{N!}{(N-k)!}$ e substituindo p por $\frac{\lambda}{N}$

$$P_{Po}(k) = \lim_{N\to\infty} \frac{1}{k!} \cdot N \cdot (N-1) \cdot (N-2) \cdots (N-k+1) \cdot \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^{N-k} ,$$

colocando N^k em evidência

$$P_{Po}(k) = \lim_{N\to\infty} \frac{1}{k!} \cdot N^k \cdot 1 \cdot \left(1 - \frac{1}{N}\right) \cdot \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{k-1}{N}\right) \cdot \frac{1}{N^k} \cdot \lambda^k \left(1 - \frac{\lambda}{N}\right)^{N-k}$$

obtemos

$$P_{Po}(k) = \lim_{N\to\infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k}$$

Nesta expressão, $\lim_{N\to\infty} \left(1 - \frac{\lambda}{N}\right)^N = e^{-\lambda}$ é um limite fundamental do cálculo e $\lim_{N\to\infty} \left(1 - \frac{\lambda}{N}\right)^{-k} = 1$. Obtemos assim a distribuição de Poisson

$$P_{Po}(k) = \frac{\lambda^k e^{-\lambda}}{k!} , \tag{D.14}$$

com $k = 0, 1, 2, \ldots, \infty$, cujas média e variância são

$$\begin{aligned} \mathrm{E}(k) &= \lambda \\ \mathrm{Var}(k) &= \lambda \, . \end{aligned}$$

É intuitivo que a soma de variáveis com distribuição de Poisson (com médias arbitrárias) tem distribuição de Poisson (com as médias somadas). A distribuição acumulada é dada por

$$F_{Po}(k) = \sum_{j=0}^{k} P_{Po}(j) = \frac{\Gamma(k+1, \lambda)}{\Gamma(k+1)} = \mathrm{Q}(k+1, \lambda) \, ,$$

onde $\Gamma(a, b)$ é a função gama incompleta complementar (Eq. B.7), $\Gamma(a)$ é a função gama completa (Eq. B.3) e $\mathrm{Q}(a, b)$ é a função gama incompleta complementar normalizada (Eq. B.9).

Observe que as distribuições binomial e de Poisson são ambas discretas, pois a variável k comum a ambas representa uma contagem: o número de sítios dentro de um trecho de comprimento ℓ. A diferença reside no substrato: o meio físico considerado é discreto no caso binomial e contínuo no caso de Poisson. No entanto, em ambos os casos, o meio é ilimitado.

D.2.2.2 Distância entre Pontos Consecutivos – Distribuição Exponencial

Na seção anterior, mostramos que a distribuição de Poisson é o equivalente contínuo da binomial. Analogamente, a distribuição exponencial corresponde ao limite contínuo da distribuição geométrica: o número N de segmentos num trecho de comprimento fixo ℓ tende ao infinito e a probabilidade p de um segmento estar ativado tende a zero, de forma que $\lambda = Np$ se mantém constante.

Novamente, esta situação representa fisicamente um meio contínuo como caso-limite de um meio discreto. Assim, a distribuição exponencial fornece a distribuição de probabilidade da distância ℓ entre pontos consecutivos, como mostra a Fig. 32.

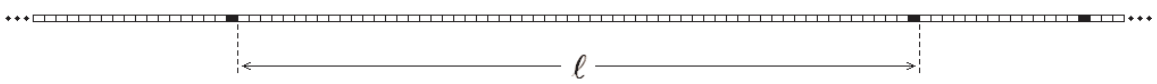


Figura 32: Um meio ilimitado contínuo com densidade média r de segmentos ativados. A distância contínua ℓ entre segmentos ativados consecutivos segue a distribuição exponencial (Eq. D.15).

Ao contrário das distribuições binomial, Poisson e geométrica (que são discretas), a distribuição exponencial é contínua (pois sua variável ℓ é contínua) e portanto descrita por uma função densidade de probabilidade. Partindo da Eq. D.10, temos

$$f_{ex}(\ell)\,\mathrm{d}\ell = \lim_{\ell_0\to 0}(1-r\ell_0)^{\frac{\ell}{\ell_0}-1} r\ell_0 \ .$$

Rearranjando os termos obtemos $f_{ex}(\ell)\,\mathrm{d}\ell = r\lim_{\ell_0\to 0}(1-r\ell_0)^{\ell/\ell_0}\ell_0/(1-r\ell_0)$ e surge o limite fundamental do cálculo. Como $\ell_0 \to 0$, podemos identificá-lo como o elemento de comprimento $\mathrm{d}\ell$, obtendo assim a distribuição exponencial

$$f_{ex}(\ell)\,\mathrm{d}\ell = re^{-r\ell}\,\mathrm{d}\ell \ , \qquad \ell \in [0,\infty) \ . \tag{D.15}$$

A média e a variância da distribuição exponencial são

$$\begin{aligned} \mathrm{E}(\ell) &= 1/r \\ \mathrm{Var}(\ell) &= 1/r^2 \end{aligned} \tag{D.16}$$

e a distribuição acumulada é dada por

$$F_{ex}(\ell) = \int_0^{\ell} \mathrm{d}\ell' f_{ex}(\ell') = 1 - e^{-r\ell} \ .$$

Observe que também poderíamos ter obtido a distribuição exponencial a partir da distribuição de Poisson (Eq. D.14). A distribuição acumulada da distância entre dois pontos consecutivos A e B pode ser obtida impondo-se a condição de que deve haver no mínimo 1 ponto dentro do trecho de comprimento ℓ (ou seja, pelo menos o próprio ponto B):

$$F_{ex}(\ell) = \sum_{j=1}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} = 1 - \frac{(r\ell)^0 e^{-r\ell}}{0!} = 1 - e^{-r\ell} \tag{D.17}$$

como mostra a Fig. 33.

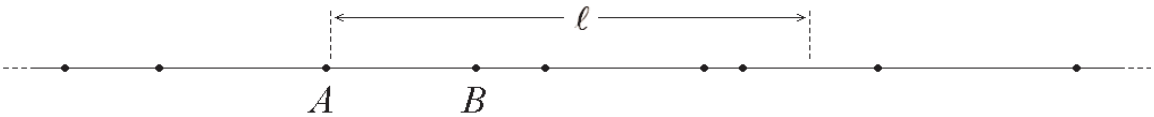


Figura 33: Distribuição exponencial obtida a partir da distribuição de Poisson.

Derivando a Eq. D.17 em relação a ℓ obtemos a pdf exponencial

$$f_{ex}(\ell)\,\mathrm{d}\ell = \frac{\partial F_{ex}(\ell)}{\partial \ell}\,\mathrm{d}\ell = \frac{\partial(1-e^{-r\ell})}{\partial \ell}\,\mathrm{d}\ell = re^{-r\ell}\,\mathrm{d}\ell \ .$$

Observe que este resultado é o mesmo da Eq. D.15. A distribuição exponencial é a única distribuição sem memória e tem um único parâmetro: a densidade média r do meio.

D.2.2.3 Distância entre Pontos Não-consecutivos – Distribuição Gama

A distribuição gama é o equivalente contínuo da distribuição de Pascal. Novamente, o comprimento ℓ_0 de cada segmento e a probabilidade p dele estar ativado tendem a zero, de forma que, num trecho de comprimento fixo ℓ, o número total N de segmentos cresce indefinidamente, mas o número esperado $\lambda = Np$ de segmentos ativados permanece constante.

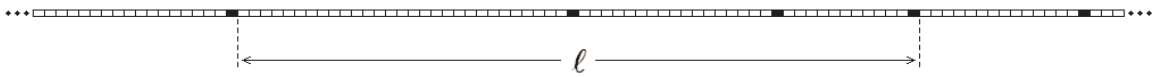


Figura 34: Um meio ilimitado contínuo com densidade média r de segmentos ativados. A distância contínua ℓ entre segmentos ativados não-consecutivos de ordem k segue a distribuição gama (Eq. D.18).

Partindo da Eq. D.12

$$f_{ga}(\ell)\mathrm{d}\ell = \lim_{\ell_0\to 0} \frac{\left(\frac{\ell}{\ell_0}-1\right)!}{(k-1)!\left(\frac{\ell}{\ell_0}-k\right)!}(r\ell_0)^k(1-r\ell_0)^{\frac{\ell}{\ell_0}-k} ,$$

expandindo o fatorial

$$f_{ga}(\ell)\mathrm{d}\ell = \lim_{\ell_0\to 0} \frac{1}{(k-1)!}\cdot\left[\frac{\ell}{\ell_0}-1\right]\cdot\left[\frac{\ell}{\ell_0}-2\right]\cdots\left[\frac{\ell}{\ell_0}-k+1\right]\cdot r^k\ell_0^k\cdot(1-r\ell_0)^{\frac{\ell}{\ell_0}-k} ,$$

multiplicando e dividindo por ℓ_0

$$f_{ga}(\ell)\mathrm{d}\ell = \lim_{\ell_0\to 0} \frac{1}{(k-1)!}\cdot\underbrace{[\ell-\ell_0]\cdot[\ell-2\ell_0]\cdots[\ell-(k-1)\ell_0]}_{\ell^{k-1}}\cdot\underbrace{\left(\frac{1}{\ell_0}\right)^{k-1}\ell_0^k}_{\ell_0=\mathrm{d}\ell}\cdot r^k\cdot\underbrace{\frac{\overbrace{(1-r\ell_0)^{\frac{\ell}{\ell_0}}}^{e^{-r\ell}}}{(1-r\ell_0)^k}}_{1}$$

e aplicando o limite fundamental do cálculo, obtemos a distribuição gama

$$f_{ga}(\ell)\mathrm{d}\ell = \frac{\ell^{k-1}e^{-r\ell}r^k}{(k-1)!}\mathrm{d}\ell , \qquad \ell\in[0,\infty) . \tag{D.18}$$

É intuitivo que a soma de k variáveis com distribuição exponencial (com mesmo parâmetro r) tem distribuição gama (com parâmetro r). A média e a variância da dis-

tribuição gama são

$$\begin{aligned} \mathrm{E}(\ell) &= k/r \\ \mathrm{Var}(\ell) &= k/r^2 \ , \end{aligned}$$

ou seja, são respectivamente iguais à média e à variância da distribuição exponencial (Eq. D.16) multiplicadas por k.

A distribuição acumulada é dada por

$$F_{ga}(\ell) = \int_0^\ell \mathrm{d}\ell' f_{ga}(\ell') = \frac{\gamma(k, r\ell)}{\Gamma(k)} = \mathrm{P}(k, r\ell) \ ,$$

onde $\gamma(a, b)$ é a função gama incompleta (Eq. B.5), $\Gamma(a)$ é a função gama completa (Eq. B.3) e $\mathrm{P}(a, b)$ é a função gama incompleta normalizada (Eq B.8).

Assim como a distribuição exponencial, a distribuição gama também pode ser obtida a partir da distribução de Poisson (Eq. D.14). A distribuição acumulada da distância ℓ entre dois pontos não consecutivos A e B de ordem k pode ser obtida impondo-se a condição de que deve haver no mínimo k pontos dentro do trecho de comprimento ℓ (ou seja, pelo menos os pontos entre A e B e o próprio ponto B):

$$F_{ga}(\ell) = \sum_{j=k}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} = 1 - \sum_{j=0}^{k-1} \frac{\lambda^j e^{-\lambda}}{j!} = 1 - e^{-r\ell} \sum_{j=0}^{k-1} \frac{(r\ell)^j}{j!} = \frac{\gamma(k, r\ell)}{\Gamma(k)} = \mathrm{P}(k, r\ell)$$

como mostra a Fig. 35.

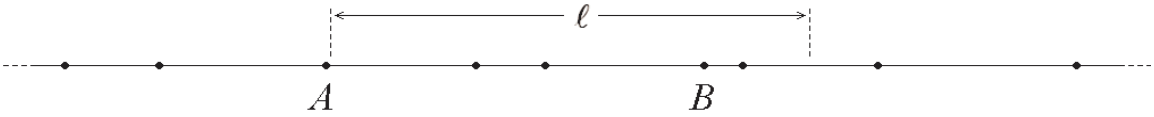


Figura 35: Distribuição gama obtida a partir da distribuição de Poisson.

Se for atribuída a abscissa $y = 0$ a um ponto qualquer do meio desordenado, escolhido como origem do meio, a distribuição gama tem como variável a abscissa y_k do k-ésimo ponto a partir da origem.

D.3 Meios Finitos

A seguir, consideraremos um meio desordenado finito nas versões discreta e contínua. Obteremos as distribuições de algumas variáveis relacionadas a este meio.

D.3.1 Meios Discretos

Na versão discreta, um meio desordenado finito unidimensional pode ser representado por um segmento de reta composto por M segmentos com mesmo comprimento ℓ_0.

D.3.1.1 Número de Pontos num Intervalo – Distribuição Hipergeométrica

Considere uma população finita com M elementos, dos quais M_1 têm uma certa característica (e $M_2 = M - M_1$ elementos não a têm). Dessa população extrai-se uma amostra aleatória com N elementos distintos. A probabilidade de haver, na amostra, N_1 elementos com a característica considerada (e conseqüentemente $N_2 = N - N_1$ sem tal característica) é dada pela distribuição hipergeométrica e pode ser obtida a partir da definição clássica de probabilidade:

- O número total de amostras é igual ao binomial $\begin{pmatrix} M \\ N \end{pmatrix}$, pois devemos escolher N elementos quaisquer dentre os M elementos disponíveis.

- O número de amostras que satisfazem a condição imposta é dada pelo produto $\begin{pmatrix} M_1 \\ N_1 \end{pmatrix}\begin{pmatrix} M_2 \\ N_2 \end{pmatrix}$, pois deve haver N_1 dos M_1 elementos com a característica considerada e N_2 dos M_2 elementos sem tal característica.

Portanto, a distribuição hipergeométrica é dada por

$$P_{hi}(N_1) = \frac{\begin{pmatrix} M_1 \\ N_1 \end{pmatrix}\begin{pmatrix} M_2 \\ N_2 \end{pmatrix}}{\begin{pmatrix} M \\ N \end{pmatrix}}, \qquad N_1 = 0, 1, 2, \ldots, \text{mín}(N, M_1)\ . \tag{D.19}$$

Como interpretação física da distribuição hipergeométrica, considere um meio discreto finito com M segmentos. Cada segmento tem comprimento $\ell_0 = 1/M$, de forma que o meio tenha comprimento unitário. Deste meio, escolhem-se aleatoriamente M_1 segmentos para serem ativados. Conseqüentemente, $M_2 = M - M_1$ segmentos permanecerão desativados.

A probabilidade de haver N_1 segmentos ativados (sítios) num trecho de comprimento fixo $\ell = N\ell_0$ contendo N segmentos é dada pela distribuição hipergeométrica (ver Fig. 36).

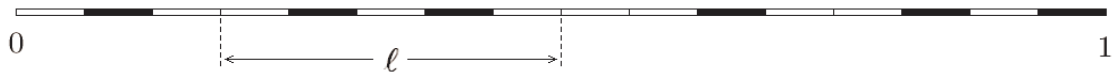


Figura 36: Meio desordenado finito discreto de comprimento unitário formado por $M = 16$ segmentos, dos quais escolhem-se aleatoriamente $M_1 = 6$ segmentos para serem ativados. A probabilidade de haver $N_1 = 2$ segmentos ativados dentro de um trecho de comprimento ℓ com $N = 5$ segmentos é dada pela distribuição hipergeométrica (Eq. D.19).

Denominando

$$p = \frac{M_1}{M} \qquad \text{e} \qquad q = \frac{M_2}{M} \ ,$$

a média e a variância da distribuição hipergeométrica são

$$\begin{aligned} \mathrm{E}(N_1) &= Np \\ \mathrm{Var}(N_1) &= Npq(M-N)/(M-1) \ . \end{aligned}$$

Em inferência estatística, o termo $(M-N)/(M-1)$ é conhecido como "fator de correção para população finita".

No caso limite $M, M_1, M_2 \to \infty$, mantendo constantes N e as proporções p e q, a distribuição hipergeométrica se torna a distribuição binomial. De fato, partindo da Eq. D.19

$$P_{bi}(N_1) = \lim_{M\to\infty} \frac{\frac{M_1!}{N_1!(M_1-N_1)!} \cdot \frac{M_2!}{N_2!(M_2-N_2)!}}{\frac{M!}{N!(M-N)!}} \ ,$$

isolando os valores amostrais

$$P_{bi}(N_1) = \lim_{M\to\infty} \frac{N!}{N_1!N_2!} \cdot \frac{\frac{M_1!}{(M_1-M_1)!} \cdot \frac{M_2!}{(M_2-N_2)!}}{\frac{M!}{(M-N)!}} \ ,$$

expandindo os fatoriais

$$P_{bi}(N_1) = \lim_{M\to\infty} \frac{N!}{N_1!N_2!} \cdot \frac{\overbrace{M_1(M_1-1)\cdots(M_1-N_1+1)}^{N_1 \text{ fatores}} \cdot \overbrace{M_2(M_2-1)\cdots(M_2-N_2+1)}^{N_2 \text{ fatores}}}{\underbrace{M(M-1)\cdots(M-N+1)}_{N \text{ fatores}}}$$

e rearranjando os fatores

$$P_{bi}(N_1) = \lim_{M\to\infty} \frac{N!}{N_1!N_2!} \cdot \underbrace{\frac{M_1}{M}}_{p} \cdot \underbrace{\frac{M_2}{M-1}}_{q} \cdot \underbrace{\frac{M_1-1}{M-2}}_{p} \cdot \underbrace{\frac{M_2-1}{M-3}}_{q} \cdot \underbrace{\frac{M_1-2}{M-4}}_{p} \cdot \underbrace{\frac{M_2-2}{M-5}}_{q} \cdots$$

$$\cdots \underbrace{\frac{M_1 - N_1 + 1}{M - N + 2}}_{p} \cdot \underbrace{\frac{M_2 - N_2 + 1}{M - N + 1}}_{q} ,$$

obtemos

$$P_{bi}(N_1) = \frac{N!}{N_1! N_2!} \cdot p^{N_1} \cdot q^{N_2} , \qquad N_1 = 0, 1, 2, \ldots, N.$$

A interpretação geométrica deste limite é que o meio se expande indefinidamente, tornando-se ilimitado, como no caso da Fig. 27. Porém tal meio permanece discreto.

D.3.1.2 Abscissa do k-ésimo Posto – Distribuição Hipergeométrica Multivariada

A distribuição hipergeométrica pode ser generalizada para o caso em que os elementos podem ser classificados em mais de dois tipos. Considere uma população finita que tem M_1 elementos do tipo 1, M_2 do tipo 2, ..., M_n do tipo n, totalizando $M = M_1 + M_2 + \cdots + M_n$ elementos. A probabilidade de uma amostra (extraída desta população finita) com N elementos conter N_1 elementos do tipo 1, N_2 do tipo 2, ..., N_n do tipo n é dada pela distribuição hipergeométrica multivariada.

$$P_{hm}(N_1, N_2, \ldots, N_n) = \frac{\begin{pmatrix} M_1 \\ N_1 \end{pmatrix} \begin{pmatrix} M_2 \\ N_2 \end{pmatrix} \cdots \begin{pmatrix} M_n \\ N_n \end{pmatrix}}{\begin{pmatrix} M \\ N \end{pmatrix}} . \tag{D.20}$$

Como interpretação física, considere um meio discreto finito com M segmentos. Cada segmento tem comprimento $\ell_0 = 1/M$, de forma que o meio tenha comprimento unitário. Deste meio, escolhem-se aleatoriamente M_1 segmentos para o estado 1, M_2 para o estado 2, ..., M_n para o estado n. A probabilidade de haver N_1 segmentos no estado 1, N_2 no estado 2, ..., N_n no estado n, dentro de um trecho de comprimento fixo $\ell = N\ell_0$ contendo $N = N_1 + N_2 + \cdots + N_n$ segmentos é dada pela distribuição hipergeométrica multivariada (ver Fig. 37).

A distribuição hipergeométrica multivariada também admite uma outra interpretação física em meios desordenados. Considere um meio discreto finito com M segmentos. Este meio finito discreto é particionado em n trechos de comprimentos $\ell_1 = M_1\ell_0$, $\ell_2 = M_2\ell_0$, ..., $\ell_n = M_n\ell_0$, contendo respectivamente M_1, M_2, ..., M_n segmentos, de forma que $M_1 + M_2 + \cdots + M_n = M$ e $\ell_1 + \ell_2 + \cdots + \ell_n = 1$.

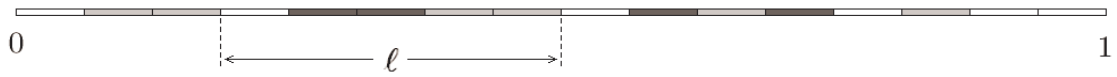


Figura 37: Meio desordenado finito discreto formado por $M = 16$ segmentos, dos quais escolhem-se aleatoriamente $M_1 = 6$ (branco), $M_2 = 6$ (cinza) e $M_3 = 4$ (preto) segmentos para os estados 1, 2 e 3, respectivamente. A probabilidade de haver $N_1 = 1$, $N_2 = 2$ e $N_3 = 2$ segmentos nos estados 1, 2 e 3 dentro de um trecho de comprimento ℓ com $N = 5$ segmentos é dada pela distribuição hipergeométrica multivariada (Eq. D.20).

Deste meio são escolhidos de maneira aleatória e independente N segmentos para serem ativados. A probabilidade de haver N_1 segmentos ativados (sítios) no trecho 1, N_2 no trecho 2, ..., N_n no trecho n (com $N_1 + N_2 + \cdots + N_n = N$) é dada pela distribuição hipergeométrica multivariada (ver Fig 38).

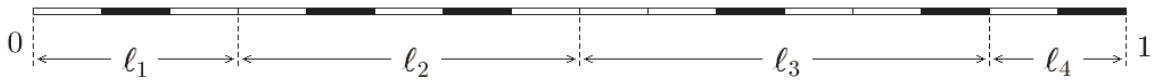


Figura 38: Meio desordenado finito discreto formado por $M = 16$ segmentos, dos quais $N = 6$ foram escolhidos para serem ativados. O meio foi particionado em $n = 4$ trechos de comprimentos ℓ_1, ℓ_2, ℓ_3 e ℓ_4 contendo respectivamente $M_1 = 3$, $M_2 = 5$, $M_3 = 6$ e $M_4 = 2$ segmentos. A distribuição hipergeométrica multivariada (Eq. D.20) fornece a probabilidade de haver $N_1 = 1$, $N_2 = 2$, $N_3 = 2$ e $N_4 = 1$ segmentos ativados dentro dos trechos 1, 2, 3 e 4, respectivamente.

Para o caso particular $n = 3$, com $M_2 = 1$, a distribuição hipergeométrica multivariada fornece a distribuição da posição j (ou da abscissa $y = j/M$) do k-ésimo segmento ativado, como mostra a Fig. 39.

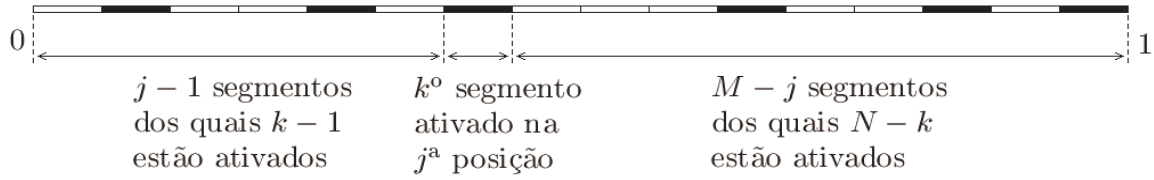


Figura 39: Meio desordenado da Fig. 38 particionado em 3 trechos com $M_1 = 6$, $M_2 = 1$ e $M_3 = 9$ segmentos. A distribuição hipergeométrica multivariada fornece a distribuição da posição j do k-ésimo segmento ativado.

Neste caso, a Eq. D.20 se torna:

$$P_{hm}(j) = \frac{\begin{pmatrix} j-1 \\ k-1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} M-j \\ N-k \end{pmatrix}}{\begin{pmatrix} M \\ N \end{pmatrix}} = \frac{\begin{pmatrix} j-1 \\ k-1 \end{pmatrix}\begin{pmatrix} M-j \\ N-k \end{pmatrix}}{\begin{pmatrix} M \\ N \end{pmatrix}} \tag{D.21}$$

Observe que o meio desordenado é finito e discreto, qualquer que seja a interpretação dada à distribuição hipergeométrica multivariada.

D.3.2 Meios Contínuos

Na versão contínua, um meio desordenado finito unidimensional pode ser representado por um segmento de reta de comprimento unitário no qual M_1 pontos são distribuídos segundo uma distribuição uniforme (Eq. D.2).

D.3.2.1 Número de Pontos num Intervalo – Distribuição Binomial

A distribuição binomial também pode ser obtida a partir da distribuição hipergeométrica através do limite $M, N \to \infty$, mantendo-se constantes M_1 e N_1, com $0 \leq N_1 \leq M_1$, e as probabilidades de sucesso $p = N/M = N\ell_0$ e de fracasso $q = (M-N)/M$. Este limite corresponde a tomar população e amostra ambas infinitas, de modo que a amostra represente uma proporção p fixa da população. No exemplo da Fig. 40, a população é todo o meio desordenado de comprimento unitário e a amostra é trecho de comprimento ℓ em destaque.

Partindo da Eq. D.19

$$P_{bi}(N_1) = \frac{\frac{M_1!}{N_1!(M_1-N_1)!} \cdot \frac{M_2!}{N_2!(M_2-N_2)!}}{\frac{M!}{N!(M-N)!}} ,$$

isolando M_1 e N_1

$$P_{bi}(N_1) = \frac{M_1!}{N_1!(M_1-N_1)!} \cdot \frac{\frac{N!(M-N)!}{N_2!(M_2-N_2)!}}{\frac{M!}{M_2!}} ,$$

expandindo os fatoriais

$$\begin{aligned} P_{bi}(N_1) &= \frac{M_1!}{N_1!(M_1-N_1)!} \cdot \\ &\quad \frac{N(N-1)\cdots(N_2+1)\cdot(M-N)(M-N-1)\cdots(M_2-N_2+1)}{M(M-1)\cdots(M_2+1)} , \end{aligned}$$

rearranjando os fatores

$$\begin{aligned} P_{bi}(N_1) &= \frac{M_1!}{N_1!(M_1-N_1)!}\cdot\frac{N}{M}\cdot\frac{N-1}{M-1}\cdots\frac{N-N_1+1}{M-N_1+1}\cdot \\ &\quad\cdot\frac{M-N}{M-N_1}\cdot\frac{M-N-1}{M-N_1-1}\cdots\frac{(M-M_1)-(N-N_1)+1}{M-M_1+1} \end{aligned}$$

e substituindo devidamente cada fração por p ou q, obtemos a distribuição binomial

$$P_{bi}(N_1) = \frac{M_1!}{N_1!(M_1-N_1)!}\cdot p^{N_1}\cdot q^{M_1-N_1}\,, \qquad N_1 = 0,1,2,\ldots,M_1. \tag{D.22}$$

A interpretação física deste limite é que o meio continua finito (limitado ao intervalo $[0,1]$), porém se torna contínuo, como mostra a Fig. 40. Neste caso, a distribuição binomial fornece a probabilidade de haver N_1 pontos dentro de um trecho de comprimento ℓ.

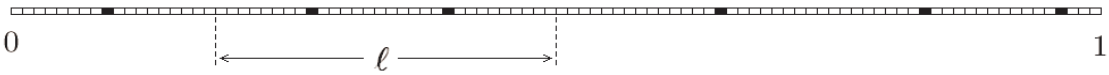


Figura 40: Um meio finito contínuo como caso-limite de um meio finito discreto. O número M de segmentos do meio cresce indefinidamente e o comprimento $\ell_0 = 1/M$ de cada segmento tende a 0, de forma que o meio tenha comprimento fixo (unitário). Deste meio escolhem-se $M_1 = 6$ segmentos para serem ativados. A probabilidade de haver $N_1 = 2$ segmentos ativados dentro de um trecho de comprimento fixo $\ell = N\ell_0$ é dada pela distribuição binomial (Eq. D.22).

Este resultado já era esperado, pois durante a distribuição dos N pontos no intervalo $[0,1]$, cada ponto tem probabilidade $p = \ell$ de pertencer ao trecho de comprimento $\ell = N\ell_0$ e probabilidade $q = 1-\ell$ de não pertencer. Como as abscissas dos N pontos são uniformemente distribuídas no intervalo $[0,1]$ e independentes, surge naturalmente a distribuição binomial.

D.3.2.2 Abscissa do Primeiro Posto – Distribuição Potência

Considere N variáveis aleatórias contínuas independentes X_1, X_2, ..., X_N; com suas respectivas pdf's $g_1(x_1)$, $g_2(x_2)$, ..., $g_N(x_N)$. A pdf $f(y)$ da variável aleatória Y, definida como o mínimo entre as variáveis X_1, X_2, ..., X_N, pode ser determinada como segue.

Para que Y assuma um dado valor y, é necessário que pelo menos uma das variáveis X_1, X_2, ..., X_N seja igual a y e que todas as demais sejam maiores ou iguais a y. Assim:

$$f(y) = g_1(y)\int_y^\infty \mathrm{d}x_2 g_2(x_2)\int_y^\infty \mathrm{d}x_3 g_3(x_3)\cdots\int_y^\infty \mathrm{d}x_N g_N(x_N) +$$

$$+g_2(y)\int_y^\infty \mathrm{d}x_1 g_1(x_1)\int_y^\infty \mathrm{d}x_3 g_3(x_3)\cdots\int_y^\infty \mathrm{d}x_N g_N(x_N)+\cdots$$
$$\cdots+g_N(y)\int_y^\infty \mathrm{d}x_1 g_1(x_1)\int_y^\infty \mathrm{d}x_2 g_2(x_2)\cdots\int_y^\infty \mathrm{d}x_{N-1} g_{N-1}(x_{N-1})\ .$$

De maneira mais compacta:

$$f(y) \;=\; \sum_{i=1}^{N}\left[g_i(y)\cdot\prod_{j\,(\neq i)=1}^{N}\int_y^\infty \mathrm{d}x_j g_j(x_j)\right]\ .$$

Se X_1, X_2, ... , X_N têm todas a mesma pdf $g(x)=g_1(x_1)=g_2(x_2)=\ldots=g_N(x_N)$, temos para Y:

$$f(y) \;=\; Ng(y)\left[\int_y^\infty \mathrm{d}x g(x)\right]^{N-1} = Ng(y)\left[1-G(y)\right]^{N-1}\ ,$$

onde

$$G(y) \;=\; \int_{-\infty}^{y}\mathrm{d}x g(x)$$

é a distribuição acumulada de $g(x)$. Em particular, se a pdf $g(x)$ for uma distribuição uniforme entre 0 e 1 (dada pela Eq. D.2), a função de densidade de probabilidade de Y se torna a pdf potência, dada por:

$$f_{pt}(y) \;=\; N\cdot\left[\int_y^1 \mathrm{d}x\right]^{N-1} = N\left(1-y\right)^{N-1}\ , \tag{D.23}$$

com y variando de 0 a 1.

A média e a variância da distribuição potência são

$$\begin{aligned}\mathrm{E}(y) &= 1/(N+1)\\ \mathrm{Var}(y) &= N/[(N+1)^2(N+2)]\end{aligned}$$

e a distribuição acumulada é dada por

$$F_{pt}(y)=\int_0^y \mathrm{d}y'\, f_{pt}(y') = 1-(1-y)^N\ . \tag{D.24}$$

Como interpretação física, considere um meio contínuo finito com N pontos. A distribuição potência fornece a abscissa do ponto que ocupa o primeiro posto, isto é, o ponto de menor abscissa, como mostra a Fig. 41.

Observe que a pdf potência (Eq. D.23) pode ser obtida a partir da Eq. D.21 com $k=1$ e considerando o limite $M, j\to\infty$, de forma que a abscissa $y=j/M$ pernaceça constante.

Figura 41: Meio finito contínuo de comprimento unitário com $N = 16$ pontos distribuídos uniforme e independentemente. A abscissa y do ponto que ocupa a primeira posição segue a distribuição potencia (Eq. D.23).

D.3.2.3 Abscissa do k-ésimo Posto – Distribuição Beta

Considere N pontos cujas coordenadas foram geradas aleatoriamente de maneira independente segundo uma distribuição uniforme no intervalo real $[0, 1]$. A distribuição beta fornece a distribuição da abscissa y do ponto que ocupa a k-ésima posição (com $k = 1, 2, \ldots, N$). A Fig. 42 ilustra a situação.

Figura 42: Meio finito contínuo de comprimento unitário com $N = 16$ pontos distribuídos uniforme e independentemente. A abscissa y do ponto que ocupa a k-ésima posição segue a distribuição beta (Eq. D.25).

Para determinar sua pdf, observe que

1. há N candidatos ao k-ésimo posto
2. deve haver $k - 1$ pontos à esquerda de y
3. os $N - k$ pontos restantes devem estar à direita de y.

A probabilidade de haver $k - 1$ pontos à esquerda de y e $N - k$ pontos à direita de y é dada pela distribuição binomial.

$$f_{bt}(y)\,\mathrm{d}y = N\frac{(N-1)!}{(k-1)!(N-k)!}y^{k-1}(1-y)^{N-k}\,\mathrm{d}y$$

Logo, a pdf da distribuição beta é dada por

$$f_{bt}(y)\,\mathrm{d}y = \frac{y^{k-1}(1-y)^{N-k}}{\mathrm{B}(k, N-k+1)}\,\mathrm{d}y\ , \qquad y \in [0, 1]\ , \tag{D.25}$$

onde $\mathrm{B}(a, b)$ é a função beta (Eq. B.10). Para o caso particular $k = 1$, a pdf beta se reduz à pdf potência (Eq. D.23). Observe que a pdf da distribuição beta (Eq. D.25) é semelhante à distribuição binomial (Eq. D.4). Porém, na distribuição binomial, N é um parâmetro e y é a variável; na distribuição beta, estes papéis são permutados.

A média e a variância da distribuição beta são

$$\begin{aligned} \mathrm{E}(y) &= k/(N+1) \\ \mathrm{Var}(y) &= k(N-k+1)/[(N+1)^2(N+2)] \end{aligned}$$

e a distribuição acumulada é convenientemente expressa em termos da função beta incompleta normalizada (Eq. B.14):

$$F_{bt}(y) = \int_0^y \mathrm{d}y' f_{bt}(y') = \mathrm{I}_y(k, N-k+1) \ .$$

A distribuição beta (Eq. D.25) também pode ser obtida a partir da distribuição hipergeométrica. Partindo da Eq. D.21, tomando $M \to \infty$ e $j \to \infty$ de forma que $y = j/M$ e $1 - y = (M-j)/M$ permaneçam fixos, temos

$$\begin{aligned} f_{bt}(y)\,\mathrm{d}y &= \frac{(j-1)!}{(k-1)!(j-k)!} \cdot \frac{(M-j)!}{(N-k)!(M-j-N+k)!} \cdot \frac{N!(M-N)!}{M!} \\ &= \frac{N!}{(k-1)!(N-k)!} \cdot \frac{\overbrace{(j-1)\cdots(j-k+1)}^{k-1 \text{ fatores}} \cdot \overbrace{(M-j)\cdots(M-j-N+k+1)}^{N-k \text{ fatores}}}{\underbrace{M(M-1)(M-2)\cdots(M-N+1)}_{N \text{ fatores}}} \\ &= \frac{1}{B(k, N-k+1)} \cdot \underbrace{\frac{j-1}{M-1} \cdots \frac{j-k+1}{M-k+1}}_{y^{k-1}} \cdot \underbrace{\frac{M-j}{M-k} \cdots \frac{M-j-N+k+1}{M-k-N+k+1}}_{(1-y)^{N-k}} \cdot \underbrace{\frac{1}{M}}_{\mathrm{d}y} \\ &= \frac{y^{k-1}(1-y)^{N-k}}{\mathrm{B}(k, N-k+1)}\,\mathrm{d}y \end{aligned}$$

Este limite representa uma transição de um meio finito discreto (Fig. 39) para um meio finito contínuo (Fig. 42). O número de segmentos cresce ilimitadamente ($M \to \infty$) e o comprimento de cada segmento diminui proporcionalmente ($\ell = 1/M \to 0$), de forma que o meio terá comprimento unitário.

A distribuição beta também surge naturalmente no meio semi-infinito (introduzido na Sec. 5.1), no qual infinitos pontos são distribuídos de maneira uniforme e independente sobre uma semi-reta, com uma densidade média de r pontos por comprimento unitário. Neste meio, a distância x entre pontos consecutivos segue uma distribuição exponencial (Eq. D.15). Podemos normalizar as abscissas dos $N+2$ primeiros pontos deste meio, como mostra a Fig 43.

Com esta normalização, as abscissas dos pontos extremos são 0 e 1. Excluindo estes

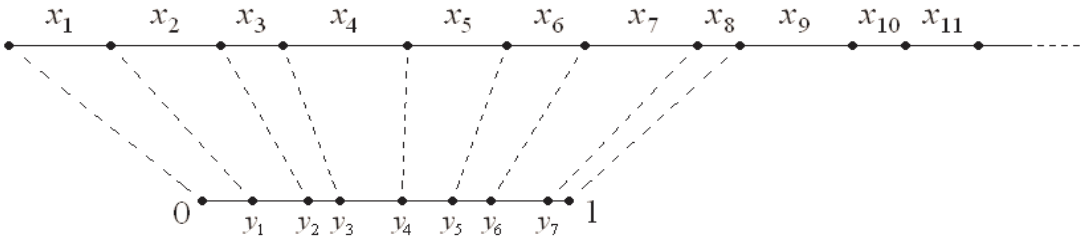


Figura 43: Equivalência entre o meio semi-infinito e o meio finito. As distâncias x_k entre pontos consecutivos têm distribuição exponencial e a abscissa y_k do ponto que ocupa o k-ésimo posto tem distribuição gama.

pontos extremos, restam N pontos no interior do intervalo $[0,1]$. A abscissa y do ponto que ocupa a k-ésima posição é $y = \ell_1/(\ell_1+\ell_2)$, onde $\ell_1 = x_1 + x_2 + \cdots + x_k$ e $\ell_2 = x_{k+1} + x_{k+2} + \cdots + x_N + x_{N+1}$ têm distribuição gama, dada pela Eq. D.18. A pdf conjunta de ℓ_1 e ℓ_2 é

$$g(\ell_1,\ell_2)\,\mathrm{d}\ell_1\mathrm{d}\ell_2 = \frac{\ell_1^{k-1}e^{-r\ell_1}r^k}{(k-1)!}\cdot\frac{\ell_2^{N-k}e^{-r\ell_2}r^{N-k+1}}{(N-k)!}\,\mathrm{d}\ell_1\mathrm{d}\ell_2 = \frac{\ell_1^{k-1}\ell_2^{N-k}e^{-r(\ell_1+\ell_2)}r^{N+1}}{(k-1)!(N-k)!}\,\mathrm{d}\ell_1\mathrm{d}\ell_2\ .$$

Seja $u = \ell_1 + \ell_2$. Nestes termos $\ell_1 = uy$ e $\ell_2 = u(1-y)$. O Jacobiano é

$$J\left(\frac{\ell_1,\ell_2}{u,y}\right) = \begin{vmatrix} y & u \\ 1-y & -u \end{vmatrix} = -u$$

então

$$\begin{aligned} h(u,y)\,\mathrm{d}u\mathrm{d}y &= g(\ell_1,\ell_2)\,\mathrm{d}\ell_1\mathrm{d}\ell_2 = g(\ell_1,\ell_2)u\,\mathrm{d}u\mathrm{d}y = \frac{(uy)^{k-1}[u(1-y)]^{N-k}e^{-ru}r^{N+1}u}{(k-1)!(N-k)!}\,\mathrm{d}u\mathrm{d}y \\ &= \frac{e^{-ru}u^N y^{k-1}(1-y)^{N-k}r^{N+1}}{(k-1)!(N-k)!}\,\mathrm{d}u\mathrm{d}y \end{aligned}$$

A pdf da variável y é dada pela distribuição marginal de $h(u,y)$

$$\begin{aligned} f_{bt}(y) &= \int_0^\infty \mathrm{d}u\ h(u,y) = \int_0^\infty \frac{\mathrm{d}u\ e^{-ru}u^N y^{k-1}(1-y)^{N-k}r^{N+1}}{(k-1)!(N-k)!} \\ &= \frac{y^{k-1}(1-y)^{N-k}}{(k-1)!(N-k)!}\int_0^\infty \mathrm{d}u\ e^{-ru}u^N r^{N+1} \end{aligned}$$

a integral é identificada com a função gama

$$\int_0^\infty e^{-ru}u^N r^{N+1}\mathrm{d}u = \int_0^\infty e^{-ru}(ru)^N\mathrm{d}(ru) = \Gamma(N+1)$$

portanto

$$f_{bt}(y) \;=\; \frac{y^{k-1}(1-y)^{N-k}}{\Gamma(k)\Gamma(N-k+1)}\Gamma(N+1) = \frac{y^{k-1}(1-y)^{N-k}}{B(k,N-k+1)}$$

que é a distribuição beta.

Generalização da Distribuição Beta – Estatística de Ordem A distribuição beta pode ser generalizada para o caso em que as coordenadas y não seguem, necessariamente, uma pdf uniforme. Considere N variáveis aleatórias y_1, y_2, ..., y_N independentes e identicamente distribuídas (i.i.d.) segundo uma pdf $g(y)$ arbitrária. Se as reordenarmos em ordem crescente, de forma que $y_{(1)} < y_{(2)} < \ldots < y_{(N)}$, onde o índice k entre parênteses é o posto de $y_{(k)}$, então a pdf de $y_{(k)}$ é dada por

$$f_k(y) = \frac{[G(y)]^{k-1}[1-G(y)]^{N-k}g(y)}{\mathrm{B}(k,N-k+1)}\ , \tag{D.26}$$

com $k = 1, 2, \ldots, N$. Para o caso particular em que $g(y)$ tem pdf uniforme, a Eq. D.26 se reduz à Eq. D.25, obtendo-se a pdf beta usual.

Se as N variáveis y_1, y_2, ..., y_N forem reordenadas em ordem decrescente $y_{(1)} > y_{(2)} > \ldots > y_{(N)}$, então a pdf de $y_{(k)}$ é dada por

$$f_k(y) = \frac{[G(y)]^{N-k}[1-G(y)]^{k-1}g(y)}{\mathrm{B}(k,N-k+1)}\ , \tag{D.27}$$

também com $k = 1, 2, \ldots, N$.

D.4 Distribuição do Produto de Variáveis Uniformes

Considere n variáveis aleatórias y_1, y_2, ..., y_n independentes e identicamente distribuídas (i.i.d.) segundo uma distribuição uniforme (Eq. D.2) no intervalo $[0, 1]$. Determinemos a pdf $f_{pr}(\tilde{y})$ da variável $\tilde{y}$ definida como o produto das variáveis y_1, y_2, ..., y_n, isto é,

$$\tilde{y} = \prod_{j=1}^{n} y_j\ .$$

Para tal, apliquemos sobre $\tilde{y}$ a transformação de variáveis:

$$\tilde{x} = -\ln\tilde{y} = -\ln\left(\prod_{j=1}^{n} y_j\right) = \sum_{j=1}^{n}(-\ln y_j) = \sum_{j=1}^{n} x_j\ ,$$

onde as variáveis $x_j = -\ln y_j$, com $j = 1, 2, \ldots, n$, são i.i.d. e seguem uma distribuição exponencial (Eq. D.15) com parâmetro $r = 1$.

Como a soma de variáveis com distribuição exponencial tem distribuição gama, a pdf de $\tilde{x}$ é

$$f_{ga}(\tilde{x}) = \frac{\tilde{x}^{n-1} e^{-\tilde{x}}}{(n-1)!} \, .$$

Aplicando-se a *lei fundamental de transformação de probabilidades* [42] dada por $|f_y(y)\,\mathrm{d}y| = |f_x(x)\,\mathrm{d}x|$, obtemos

$$f_{pr}(\tilde{y}) = \frac{\tilde{x}^{n-1} e^{-\tilde{x}}}{(n-1)!} \cdot \frac{1}{\tilde{y}} \, .$$

Substituindo $\tilde{x} = -\ln \tilde{y}$, a distribuição de $\tilde{y}$ é dada por

$$f_{pr}(\tilde{y}) = \frac{(-\ln \tilde{y})^{n-1}}{(n-1)!} \, . \tag{D.28}$$

O gráfico da Fig.44 mostra a distribuição de $\tilde{y}$ para alguns valores de n.

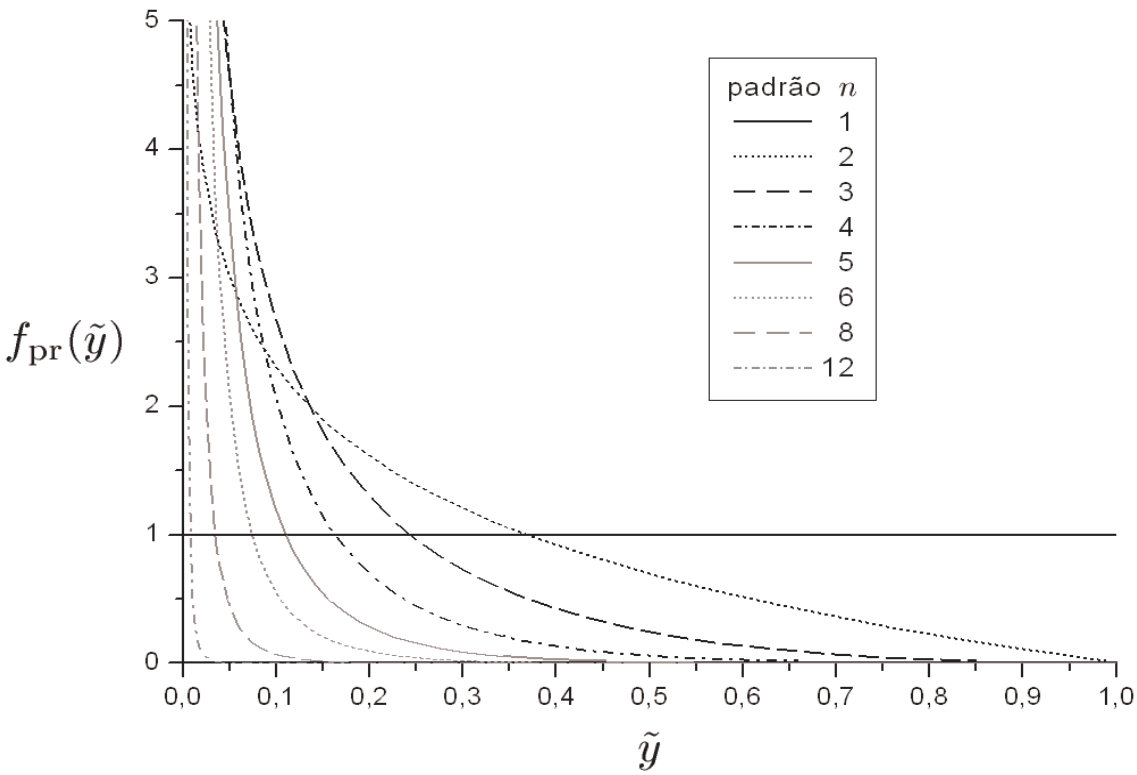


Figura 44: Distribuição do produto de n variáveis com distribuição uniforme. À medida que n aumenta, a distribuição fica mais concentrada em torno de $\tilde{y} = 0$.

O m-ésimo momento $\mathrm{E}(\tilde{y}^m) = \int_0^1 \mathrm{d}\tilde{y}(-\ln \tilde{y})^{n-1}\tilde{y}^m/(n-1)!$ da pdf da Eq. D.28 é

calculado aplicando a substituição de variáveis $z = -(m+1)\ln\tilde{y}$:

$$\mathrm{E}(\tilde{y}^m) = \frac{1}{(n-1)!} \cdot \int_0^\infty \frac{\mathrm{d}z\; e^{-z/(m+1)}}{m+1} \left(\frac{z}{m+1}\right)^{n-1} \cdot e^{-mz/(m+1)} \; .$$

Simplificando, temos

$$\mathrm{E}(\tilde{y}^m) = \frac{1}{\Gamma(n)} \cdot \frac{1}{(m+1)^n} \cdot \int_0^\infty \mathrm{d}z\; z^{n-1} e^{-z}$$

e a integral é identificada com a função gama, resultando em

$$\mathrm{E}(\tilde{y}^m) = (m+1)^{-n} \; . \tag{D.29}$$

Logo, a média e a variância desta distribuição são

$$\begin{aligned} \mathrm{E}(\tilde{y}) &= 2^{-n} \\ \mathrm{Var}(\tilde{y}) &= 3^{-n} - 4^{-n} \end{aligned}$$

e a distribuição acumulada é dada por

$$F_{pr}(\tilde{y}) = \int_0^y \mathrm{d}y'\; f_{pr}(y') = \mathrm{Q}(n, -\ln y) \; ,$$

onde $\mathrm{Q}(a,b)$ é a função gama incompleta complementar normalidada (Eq. B.9).

Devemos chamar a atenção para um detalhe: a pdf da Eq. D.28 *não é* a pdf log-gama. De fato, se uma variável x tem pdf gama, então a variável $y = \ln x$ tem pdf log-gama, dada por

$$f_{lg}(y) = \frac{e^{ky - e^y}}{(k-1)!} \; .$$

Em contrapartida, a Eq. D.28 fornece a pdf da variável $\tilde{y} = e^{-\tilde{x}}$, onde $\tilde{x}$ tem pdf gama. Contudo, para $n \gg 1$ a pdf da Eq. D.28 pode ser aproximada por uma pdf log-normal, dada pela Eq. D.30.

D.5 Distribuição Gaussiana (ou Normal)

Apesar da distribuição Gaussiana não emergir diretamente de um meio desordenado, a incluímos aqui devido à sua importância como distribuição-limite. Uma variável aleatória X tem distribuição Gaussiana se sua pdf é descrita por

$$f_{Ga}(x) = A \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \; ,$$

onde $x \in (-\infty, \infty)$, μ é um parâmetro de localização, σ^2 é um parâmetro de escala e $A = 1/\sqrt{2\pi\sigma^2}$ é o fator de normalização. A média e a variância da distribuição Gaussiana são

$$\begin{aligned} \mathrm{E}(x) &= \mu \\ \mathrm{Var}(x) &= \sigma^2 \end{aligned}$$

e a distribuição acumulada é dada por

$$F_{Ga}(x) = \int_{-\infty}^{x} \mathrm{d}x' \, f_{Ga}(x') = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right] ,$$

onde $\mathrm{erf}(z)$ é a função erro dada pela Eq. B.1.

A distribuição Gaussiana está presente, pelo menos de forma aproximada, em muitos fenômenos naturais, devido a um resultado surpreendente conhecido como *Teorema do Limite Central de Gauss* (TLCG). Em sua essência, o TLCG pode ser anunciado como segue. Dadas n variáveis aleatórias independentes Y_i, cada qual seguindo uma distribuição arbitrária $f_i(y_i)$ com média $\mathrm{E}(y_i)$ e variância finita $\mathrm{Var}(y_i)$, então a distribuição da variável $X = \sum_{i=1}^{n} Y_i$ tende a uma pdf Gaussiana, cujos parâmetros são $\mu = \sum_{i=1}^{n} \mathrm{E}(y_i)$ e variância $\sigma^2 = \sum_{i=1}^{n} \mathrm{Var}(y_i)$.

Observe que as variáveis aleatórias da maioria das distribuições estatísticas abordadas neste Apêndice podem ser escritas como a soma de outras variáveis aleatórias. Por exemplo,

- a soma de variáveis com distribuição de Bernoulli tem distribuição binomial,
- a soma de variáveis com distribuição geométrica tem distribuição de Pascal,
- a soma de variáveis com pdf exponencial tem pdf gama,
- a soma de variáveis com distribuição de Poisson tem distribuição de Poisson,
- a pdf beta está relacionada com somas de variáveis com pdf gama.

É justamente devido a este fato que a pdf Gaussiana é a distribuição-limite de várias distribuições contínuas e é também o envelope de várias distribuições discretas, tomando-se os limites apropriados. A distribuição multinomial, por sua vez, tem como envelope a pdf Gaussiana Multivariada.

Em contrapartida, existem distribuições estatísticas cujas variáveis aleatórias não podem ser escritas como uma soma de outras variáveis aleatórias e, portanto, não tendem a

uma pdf Gaussiana. Como exemplo, podemos citar a distribuição geométrica (problemas de primeira passagem), a pdf exponencial e a pdf potência (pdf do mínimo de outras variáveis aleatórias).

D.5.1 Distribuição Meia-Gaussiana

Uma variável aleatória X tem distribuição Meia-Gaussiana se sua pdf é descrita por

$$f_{mG}(x) = A \exp\left[-\frac{x^2}{\pi\theta^2}\right] ,$$

onde $x \in [0, \infty)$, θ é um parâmetro de escala e $A = 2/(\pi\theta)$ é o fator de normalização. A média e a variância da distribuição Meia-Gaussiana são

$$\begin{aligned} \mathrm{E}(x) &= \theta \\ \mathrm{Var}(x) &= (\pi - 2)\theta^2/2 \end{aligned}$$

e a distribuição acumulada é dada por

$$F_{mG}(x) = \int_0^x \mathrm{d}x' \, f_{mG}(x') = \mathrm{erf}\left(\frac{x}{\theta\sqrt{\pi}}\right) ,$$

onde $\mathrm{erf}(z)$ é a função erro dada pela Eq. B.1.

D.5.2 Distribuição Log-Normal

Dada uma variável X com distribuição normal, a variável $Y = e^X$ segue uma pdf log-normal dada por

$$f_{ln}(y) = A\frac{\exp[-(\ln y - M)^2/(2S^2)]}{y} , \tag{D.30}$$

onde $y \in [0, \infty)$, M é um parâmetro de localização, S^2 é um parâmetro de escala e $A = 1/\sqrt{2\pi S^2}$ é o fator de normalização. A média e a variância da distribuição log-normal são

$$\begin{aligned} \mathrm{E}(y) &= e^{M+S^2/2} \\ \mathrm{Var}(y) &= e^{S^2+2M}(e^{S^2} - 1) \end{aligned}$$

e a distribuição acumulada é dada por

$$F_{ln}(x) = \int_0^y \mathrm{d}y' \, f_{ln}(y') = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\ln y - M}{\sqrt{2S^2}}\right)\right] ,$$

onde $\mathrm{erf}(z)$ é a função erro dada pela Eq. B.1.

Observe que, como X tem distribuição normal, pelo TLCG podemos escrever X como uma soma de variáveis aleatórias. Considerando que $Y = e^X$ e a propriedade $e^{a+b} = e^a e^b$, concluímos que Y pode ser escrita como um produto de variáveis aleatórias. Portanto, a pdf log-normal é a distribuição-limite do produto de variáveis aleatórias, assim como a pdf Gaussiana é a distribuição-limite da soma de variáveis aleatórias.

Referências

[1] R. Metzler and J. Klafter, "The random walk's guide to anomalous diffusion: a fractional dynamics approach," *Phys. Rep.*, vol. 339, pp. 1–77, 2000.

[2] D. S. Fisher, "Random walks in random environments," *Phys. Rev. A*, vol. 30, pp. 960–964, August 1984.

[3] G. T. Barkema, P. Biswas, and H. van Beijeren, "Diffusion with random distribution of static traps," *Phys. Rev. Lett.*, vol. 87, p. 170601, 22 October 2001.

[4] H. Freund and P. Grassberger, "The Red Queens walk," *Physica A*, vol. 190, pp. 218–237, December 1992.

[5] D. Gale, J. Propp, S. Sutherland, and S. Troubetzkoy, "Further travels with my ant," *Math. Inteligencer*, vol. 17, no. 3, pp. 48–56, 1995.

[6] L. A. Bunimovich, "Deterministic walks in random environments," *Physica D*, vol. 187, pp. 20–29, 2004.

[7] L. Lam, "Active walks: The first twelve years (part ii)," *Int. J. Chaos Bifurc.*, vol. 16, pp. 239–268, FEB 2006.

[8] M. C. Santos, D. Boyer, O. Miramontes, G. M. Viswanathan, E. P. Raposo, J. L. Mateos, and M. G. E. da Luz, "Origin of power-law distributions in deterministic walks: The influence of landscape geometry," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 75, no. 6, p. 061114, 2007.

[9] D. Boyer, O. Miramontes, G. Ramos-Fernandez, J. L. Mateos, and G. Cocho, "Modeling the searching behavior of social monkeys," *Physica A*, vol. 342, pp. 329–335, October 2004.

[10] D. Boyer and H. Larralde, "Looking for the right thing at the right place: Phase transition in an agent model with heterogeneous spatial resources," *Complexity*, vol. 10, pp. 52–55, January-February 2005.

[11] D. Boyer, G. Ramos-Fernandez, O. M. J. L. Mateos, G. Cocho, H. Larralde, H. Ramos, and F. Rojas, "Scale-free foraging by primates emerges from their interaction with a complex environment," *Proc. R. Soc. Lond. B*, vol. 273, pp. 1743–1750, JUL 22 2006.

[12] G. F. Lima, A. S. Martinez, and O. Kinouchi, "Deterministic walks in random media," *Phys. Rev. Lett.*, vol. 87, p. 010603, Jun 2001.

[13] H. E. Stanley and S. V. Buldyrev, "Statistical physics - the salesman and the tourist," *Nature (London)*, vol. 413, pp. 373–374, 27 September 2001.

[14] O. Kinouchi, A. S. Martinez, G. F. Lima, G. M. Lourenço, and S. Risau-Gusman, "Deterministic walks in random networks: an application to thesaurus graphs," *Physica A*, vol. 315, pp. 665–676, December 2002.

[15] M. G. Campiteli, P. D., O. Kinouchi, and A. S. Martinez, "Deterministic walks as an algorithm of pattern recognition," *Phys. Rev. E*, vol. 74, no. 2, p. 026703, 2006.

[16] M. G. Campiteli, A. S. Martinez, and O. M. Bruno, "An image analysis methodology based on deterministic tourist walks," *Lect. Note Comput. Sci.*, vol. 4140, pp. 159–167, 2006.

[17] A. R. Backes, O. M. Bruno, M. G. Campiteli, and A. S. Martinez, "Deterministic tourist walks as an image analysis methodology based," *Lect. Note Comput. Sci.*, vol. 4225, pp. 784–793, 2006.

[18] S. Risau-Gusman, A. S. Martinez, and O. Kinouchi, "Escaping from cycles through a glass transition," *Phys. Rev. E*, vol. 68, p. 016104, Jul 2003.

[19] A. S. Martinez, O. Kinouchi, and S. Risau-Gusman, "Exploratory behavior, trap models, and glass transitions," *Phys. Rev. E*, vol. 69, no. 1, p. 017101, 2004.

[20] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *J. Theor. Biol.*, vol. 22, p. 437, 1969.

[21] T. F. Cox, "Reflexive nearest neighbours," *Biometrics*, vol. 37, pp. 367–369, June 1981.

[22] M. Mézard and G. Parisi, "Mean-field equations for the matching and travelling salesman problem," *Europhys. Lett.*, vol. 2, pp. 913–918, December 1986.

[23] A. G. Percus and O. C. Martin, "Finite size and dimensional dependence in the Euclidean travelling salesman problem," *Phys. Rev. Lett.*, vol. 76, pp. 1188–1191, February 1996.

[24] N. J. Cerf, J. H. B. de Monvel, O. Bohigas, O. C. Martin, and A. G. Percus, "The random link approximation for the Euclidean travelling salesman problem," *J. Phys. I (France)*, vol. 7, pp. 117–136, January 1997.

[25] A. G. Percus and O. C. Martin, "Scaling universalities of kth nearest neighbor distances in closed manifolds," *Advances Appl. Maths.*, vol. 21, pp. 424–436, October 1998.

[26] A. G. Percus and O. C. Martin, "The stochastic traveling salesman problem: finite size scaling and the cavity prediction," *J. Stat. Phys.*, vol. 94, pp. 739–758, March 1999.

[27] C. A. S. Terçariol and A. S. Martinez, "An efficient algorithm to generate random uncorrelated euclidean distances: The random link model," *Braz. J. Phys.*, vol. 36, pp. 232–236, March 2006.

[28] B. Derrida and H. Flyvbjerg, "The random map model - a disordered model with deterministic dynamics," *J. Phys. (Paris)*, vol. 48, pp. 971–978, June 1987.

[29] B. Harris, "Probability-distributions related to random mappings," *Ann. Math. Stat.*, vol. 31, no. 4, pp. 1045–1062, 1960.

[30] C. A. S. Terçariol and A. S. Martinez, "Analytical results for the statistical distribution related to a memoryless deterministic walk: Dimensionality effect and mean-field models," *Phys. Rev. E*, vol. 72, no. 2, p. 021103, 2005.

[31] C. A. S. Terçariol, R. S. Gonzalez, W. T. R. Oliveira, and A. S. Martinez, "Deterministic and random partially self-avoiding walks in random media," *Physica A*, 2007.

[32] C. A. S. Terçariol and A. S. Martinez, "Influence of memory in deterministic walks in random media: Analytical calculation within a mean-field approximation," *Phys. Rev. E*, vol. 78, no. 3, p. 031111, 2008.

[33] C. A. S. Terçariol, R. S. Gonzalez, and A. S. Martinez, "Analytical calculation for the percolation crossover in deterministic partially self-avoiding walks in one-dimensional random media," *Phys. Rev. E*, vol. 75, no. 6, p. 061117, 2007.

[34] P. J. Clark and F. C. Evans, "On some aspects of spatial pattern in biological populations," *Science*, vol. 121, pp. 397–398, March 1955.

[35] P. J. Clark, "Grouping in spatial distributions," *Science*, vol. 123, pp. 373–374, September 1956.

[36] M. F. Dacey, "Proportion of reflexive nth-order neighbours in a spatial distribution," *Geographical Analysis*, vol. 1, pp. 385–388, 1969.

[37] C. A. S. Terçariol, F. de Moura Kiipper, and A. S. Martinez, "An analytical calculation of neighbourhood order probabilities for high dimensional poissonian processes and mean field models," *J. Phys. A*, vol. 40, no. 9, pp. 1981–1989, 2007.

[38] R. S. González, "Difusão anômala: transição entre os regimes localizado e estendido na caminhada do turista unidimensional," Master's thesis, Universidade de São Paulo, Ribeirão Preto, Setembro 2006. http://www.teses.usp.br/teses/disponiveis/59/59135/tde-02022007-115428/.

[39] S. R. A. Salinas, *Introdução à Física Estatística*. Edusp, 1999.

[40] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*. New York: Dover Publications, 1972.

[41] D. J. Smith and M. K. Vanamamurthy, "How small is a unit ball?," *Mathematics Magazine*, vol. 62, no. 2, pp. 101–107, 1989.

[42] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Science Computing*. Cambridge: Cambridge University Press, 1989.