

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

Aplicações de Mecânica Estatística à
Especiação Simpátrica e à Inferência Aproximativa

Fabiano Lemes Ribeiro

*Tese apresentada ao
Instituto de Física da Uni-
versidade de São Paulo
para a obtenção do título de
Doutor em Ciências*

Orientadores :

Prof. Dr. Nestor Caticha
Prof. Dr. Manfred Opper

Banca Examinadora:

Prof. Dr. Nestor Caticha (IF-USP)
Prof^a Dr^a Carmen Pimentel Cintra do Prado (IF-USP)
Prof^a Dr^a Carla Goldman (IF-USP)
Prof. Dr. Renato Vicente (EACH-USP)
Prof. Dr. Thadeu Josino Pereira Penna (UFF)

SÃO PAULO
2009

FICHA CATALOGRÁFICA

Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Ribeiro, Fabiano Lemes

Aplicações de Mecânica Estatística a Especiação Simpátrica e Inferência Aproximativa - São Paulo, 2009.

Tese (Doutorado) - Universidade de São Paulo.
Instituto de Física - Departamento de Física Geral

Orientador: Prof. Dr. Nestor Caticha
Área de Concentração: Física Geral

Unitermos:

1. Modelos de mecânica estatística;
2. Especiação simpátrica;
3. Acasalamento preferencial;
4. Inferência Bayesiana.

USP/IF/SBI-051/2009

Às minhas pequenas,
Luma e Isabela

AGRADECIMENTOS

Esta tese é mais que a realização de um trabalho. Ela é a concretização de um sonho que se inicia muito antes do meu próprio nascimento. Esse sonho se inicia da vontade de muitas gerações de meus antepassados que sequer tiveram oportunidade ou acesso à alfabetização. Muitos deles levando subvidas sobre o cabo da enxada, sol a sol para ganhar um salário miserável. Faço parte da primeira geração de minha família que teve não só acesso à Universidade, mas a primeira geração que teve a oportunidade de sentar em uma cadeira escolar.

Até os meus pais, o máximo conseguido foi segurar o lápis para “desenhar” os seus nomes. No entanto, mesmo sem ter o conhecimento formal, minha mãe foi doutora pela escola da vida. Ela soube como nenhuma outra pessoa me ensinar sobre dignidade e amor. A D. Benedita sabia da necessidade dos estudos para seus filhos e foi aí que o sonho de toda as gerações anteriores começa a ganhar forma. Mas claro, não foi fácil. Foi preciso muita faxina em casa alheia; foi preciso lavar muita roupa “pra fora” para conseguir pagar as despesas de meus estudos. Sou inteiramente grato à minha mãe por ter me dado esta oportunidade, mesmo a custo de tanto suor e lágrimas.

O meu doutorado começou em São Carlos, onde realizei as disciplinas. Fui para São Paulo, onde realizei o trabalho com o Professor Nestor Caticha. Enfim fui para Leipzig realizar o curso de Alemão e depois para Berlim trabalhar com o Professor Manfred Oppen. Encerrei em Alfenas, onde desfruterei da tranquilidade e do aconchego mineiro necessários para a redação da tese. Enfim, cinco cidades em 4 anos, daí dá para imaginar a quantidade de pessoas que conheci e me aproximei, mas de quem, por dura realidade da vida, tive que me separar para prosseguir viagem. Inúmeras são essas pessoas; muitas delas nunca mais verei, embora tenham sido fundamentais. Outras ficam para a vida toda, embora geograficamente distantes.

Algumas pessoas, porém, necessitam ser citadas explicitamente. São estas a minha família: meu pai Osvaldo, Kátia, Aparecido, Marcelo, Luciane e especialmente, Luma e Isabela. Essas pessoas maravilhosas não só me ajudaram moral e espiritualmente, mas sacrificaram muito de suas vidas para me proporcionar uma ajuda financeira e mesmo para suportar a minha longa ausência em tempos que tanto necessitaram da minha presença.

Gostaria de manifestar também a minha inteira admiração pelo Professor Nestor Caticha pela sua orientação verdadeira. Ao Professor Renato Vicente pelas conselhos

e principalmente pelas entusiasmadas discussões científicas. À minha grande amiga Karen Bargoa, não só por ter me ensinado sobre Especiação Simpátrica, mas por toda a amizade que construímos aqui no Brasil e que depois, por uma incrível surpresa do destino, iríamos continuar em Berlim. À Aparecida Donizetti Paes, pela simpática e competente ajuda com a revisão da linguagem da tese. Sou também inteiramente grato ao Professor Manfred Opper, pela generosidade e inspiração. Quero agradecer também ao CNPq e ao DAAD, que tornaram financeiramente viável a realização dos meus estudos.

Para finalizar, gostaria de manifestar a minha gratidão a Deus por ter sido meu guia nesta longa caminhada.

RESUMO

Apresentam-se nesta tese os resultados de aplicações do formalismo da Mecânica Estatística em dois problemas independentes. O primeiro diz respeito a um modelo para Evolução do Acasalamento Preferencial no processo de Especiação Simpátrica; enquanto que o segundo refere-se ao desenvolvimento de um algoritmo de aprendizado por meio de Inferência Aproximativa.

No problema biológico estudado, cada indivíduo em um modelo de agentes é composto por dois traços. Enquanto um é responsável pela ecologia do indivíduo, o outro dita uma aparência física descorrelacionada com a adaptabilidade. Esses traços são expressos por diferentes loci que estão ligados entre si por uma taxa de recombinação. O modelo inclui também a possibilidade de evolução da preferência sexual dos indivíduos. Foi construído para esse modelo um diagrama de fases no espaço dos parâmetros que descrevem o ambiente como, por exemplo, quantidades de recursos e deficiência do indivíduo híbrido. Foram encontradas três fases de equilíbrio: (i) emergência de Acasalamento Preferencial; (ii) extinção de um dos alelos do locus responsável pela ecologia e (iii) equilíbrio Hardy-Weinberg. Foi verificado que o acasalamento preferencial pode emergir ou mesmo ser perdido (e vice-versa) em resposta a mudanças no ambiente. Além disso, o sistema apresenta memória de seu estado anterior, caracterizando uma histerese. Histerese é uma característica típica de transições de primeira ordem, o que permitiu a descrição desse sistema biológico por meio do arcabouço da Mecânica Estatística.

Em relação à Inferência Aproximativa, está-se interessado na construção de um algoritmo de aprendizado supervisionado por meio da técnica de *Propagação de Expectativas*. Mais especificamente, pretende-se inferir os parâmetros que compõem um *Perceptron* Professor a partir do conjunto de pares - entradas e saídas - que formam o conjunto de dados disponíveis. A estimativa desses parâmetros será feita pela substituição de uma distribuição Posterior original, geralmente intratável, por uma distribuição aproximativa tratável. o algoritmo *Propagação de Expectativas* foi adotado para a atualização, passo a passo, dos termos que compõem essa distribuição aproximativa. Essa atualização deve ser repetida até que a convergência

seja atingida. Utilizando o Teorema do Limite Central e o método de Cavidade, foi possível obter um algoritmo genérico e que apresentou desempenho bastante eficiente em dois modelos estudados: o modelo do *Perceptron Binário* e o modelo do *Perceptron Gaussiano*, com desempenho ótimo em ambos os casos.

ABSTRACT

This thesis presents applications of the framework of Statistical Mechanics to two independent problems. The first corresponds to a computational model for the evolution of Assortative Mating in the Sympatric Speciation process; and the second a learning algorithm built by means of a Bayesian Inference approach.

In the biological problem each individual in an agent-based model is composed of two traits. One trait, called the *ecological trait*, is directly related with the fitness; the other, called the marker trait, has no bearing on the fitness. The traits are determined by different loci which are linked by a recombination rate. There is also the possibility of evolution of mating preferences, which are inherited from the mother and subject to random variations. The study of the phase diagram in the space of parameters describing the environment (like carrying capacity and disruptive selection) reveals the existence of three phases: (i) assortative mating; (ii) extinction of one allele from ecological loci; and (iii) Hardy-Weinberg equilibrium. It was verified that the assortative mating can emerge or even be lost (and vice-versa) according with the environmental changes. Moreover, the system shows “memory” of the initial condition, characterising a hysteresis. Hysteresis is the signature of first order phase transition, which allows the description of the system by means of the Statistical Mechanics framework.

In relation to the Bayesian Inference, a supervised learning algorithm was constructed by means of the Expectation Propagation approach. The idea is to estimate the parameters which compose a Teacher Perceptron by the substitution of the original posterior distribution, intractable, by a tractable approximative distribution. The step-by-step update of the terms composing the approximative distribution was performed by using the Expectation Propagation algorithm. The update must be repeated until the convergence occurs. Using the Central Limit Theorem and the Cavity Approach, it was possible to get a generic algorithm that has shown a very good performance in two application scenarios: The Binary Perceptron Model and the Gaussian Perceptron Model.

Sumário

1	Apresentação e Estrutura da Tese	1
I	Especiação Simpátrica	4
2	A Especiação Simpátrica: uma Introdução	5
3	O Modelo	12
3.1	A Estrutura genética dos Indivíduos	12
3.2	O Ambiente tipo Levene	13
3.3	Competição entre Indivíduos	14
3.4	Versão Pan-mítica do Modelo	16
3.5	Taxa de Recombinação entre os loci	19
3.6	Acasalamento Preferencial	20
3.7	A Dinâmica	20
3.8	Os Parâmetros de Ordem	23
4	Diagrama de Fases	26
4.1	O Algoritmo do Modelo de Agentes	26
4.2	Análise de caso	28
4.3	Transição de Fase	30
4.4	Diagrama de Fases	30
5	Mudanças no Ambiente e Reversão de Especiação	37
5.1	Mudanças Abruptas do Ambiente	38
5.2	Robustez do Isolamento Reprodutivo	40
6	Discussões sobre Transição de Fase	43
6.1	A Histerese	44

6.2	Minimização da Energia Livre	44
6.3	Balço Energia-Entropia	47
7	Conclusões	49
II	Algoritmos de Aprendizado e Inferência Aproximativa	51
8	Introdução	52
9	Inferência Aproximativa e Algoritmos de Aprendizado	56
9.1	O Aprendizado Bayesiano	56
9.2	Propagação de Expectativas	62
10	Algoritmo de Aprendizado via Propagação de Expectativas	66
10.1	Desenvolvimento da Teoria	66
10.2	Atualização dos elementos das matrizes A e B	70
10.3	Cálculo da Função de Partição	72
10.4	Resumo geral dos resultados	74
11	Aplicações	76
11.1	O Perceptron Binário	76
11.1.1	Versão <i>on-line</i>	78
11.1.2	A versão <i>off-line</i>	79
11.2	O Perceptron Gaussiano	82
11.2.1	A versão <i>On-line</i>	84
11.2.2	Versão <i>Off-line</i>	86
12	Conclusões	87
13	Considerações finais e Perspectivas Futuras	89
	Apêndice A: Glossário de Termos Biológicos.	93
	Apêndice B: Algumas Funções Matemáticas	97
B.1	Divergência Kullback-Leibler	97
B.2	Integrais Gaussianas	97
B.3	A função de Heaviside	98

B.4 A função característica	98
Apêndice C: Teorema do Limite Central	101
Apêndice D: Família Exponencial	105
D.1 Propriedade dos momentos	105
D.2 Propriedade da Minimização da divergência KL	106
Apêndice E: Erro de Generalização Ótimo para o Perceptron Binário	109
E.1 Espaço de Versões	109
E.2 Truque de Réplicas	111
E.3 Ansatz Réplica Simétrica	114
Apêndice F: Artigo Publicado	117
Bibliografia	119

Lista de Figuras

2.1	Figura esquemática com os dois tipos mais importantes de especiação.	6
2.2	Fotografias de duas <i>Chrysoperla carnea</i> e suas colorações assumidas ao longo do ano: verde claro (primavera, verão e inverno) e marrom (outono).	8
2.3	Fotografias de duas espécies de peixes <i>stickleback</i> . Fonte: [27].	9
3.1	Fronteira entre os regimes de Polimorfismo (abaixo de cada curva) e Monomorfismo (acima de cada curva) para diferentes escolhas de η .	18
4.1	Evolução da densidade de genótipos e do desequilíbrio de ligação dos haplotipos de uma população. Neste particular experimento, realizado com $N = 1000$, $\mu = 0.01$, $r = \frac{1}{2}$, $\kappa = 1$, $\eta = 1$ e $s = 0.7$, a população bifurca em dois genótipos: $AAbb$ e $aaBB$, implicando um excesso de haplotipos de repulsão ($D < 0$). Uma simulação desse tipo (com 1000 gerações) demanda cerca de 30 segundos de tempo de cpu para ser realizada.	29
4.2	Evolução da aceitabilidade sexual das fêmeas homozigotas AA (acima) e aa (abaixo). Fêmeas AA desenvolvem rejeição aos machos que exibem um traço aparente BB ($\langle \alpha(\mathcal{A} BB, AA\xi_{\mathbb{B}}) \rangle \rightarrow 0$), enquanto que fêmeas aa desenvolvem uma rejeição a machos bb ($\langle \alpha(\mathcal{A} bb, aa\xi_{\mathbb{B}}) \rangle \rightarrow 0$). Essa associação entre os dois traços (ecológico e aparente) e consequentemente entre os dois loci \mathbb{A} e \mathbb{B} acontece mesmo com ausência de ligação entre eles ($r = \frac{1}{2}$). Esse resultado provém do mesmo experimento numérico apresentado na figura (4.1).	29
4.3	Média de $\hat{\alpha}$ (aceitabilidade em relação ao traço oposto) e sua respectiva variância em função da deficiência do heterozigoto. Cada ponto representa a média sobre 30 simulações independentes com $r = 0.5$, $\kappa = 1$ e $\eta = 1$ fixos. O pico da variância em torno de $s = 0.6$ evidencia a transição de fase de $\hat{\alpha}$.	31

- 4.4 Diagrama de fase s - κ para três valores (fixos) de r . Cada ponto desses gráficos representa uma simulação realizada com os respectivos valores dos parâmetros do diagrama e com $N = 1000$ e $\eta = 1$. A linha $s = \frac{2}{4\kappa - (\eta + 1)} \Big|_{\eta=1}$ separa a fase de Extinção das duas fases polimórficas, enquanto que as fases polimórficas são separadas por uma linha horizontal (não apresentada) de transição de primeira ordem. O tamanho da região da fase de Acasalamento Preferencial é totalmente dependente de r , sendo maior quanto menor o valor da taxa de recombinação. 32
- 4.5 Diagrama de fases $s - r$ para diferentes valores (fixos) de κ . Assim como a figura (4.4), cada ponto nestes gráficos representa uma simulação realizada com os respectivos valores dos parâmetros do diagrama e com $N = 1000$ e $\eta = 1$. A linha preta representa a transição de primeira ordem entre as fases tipo HW e Acasalamento Preferencial. Essa linha foi obtida a partir da média de 30 experimentos numéricos independentes para cada r . Uma linha horizontal (não apresentada) separa a fase de extinção das fases polimórficas. Pode-se perceber que quanto maior κ (menor seleção dependente da frequência) menor é a região de emergência de preferência sexual. 34
- 4.6 Representação tridimensional do diagrama de fases no espaço (r, κ, s) . As três fases acessíveis são separadas por duas superfícies. A superfície superior, dada pela equação (4.2), é independente de r . Experimentos realizados com configurações dos parâmetros que se encontram acima dessa superfície conduzem a população à extinção de um dos alelos de \mathbb{A} . A curva mais abaixo é o resultado do ajuste ($s = ar^b + c$), descrito na figura (4.5). A emergência de preferência sexual acontece em experimentos realizados com configurações dos parâmetros que estão entre essas duas superfícies (acima da superfície inferior e abaixo da superior). Nos experimentos realizados com configurações abaixo desses dois planos, a população mantém o polimorfismo inicial, porém sem emergência de preferência sexual, i.e. fase tipo HW. 35

- 5.1 Evolução de uma população sujeita a um ambiente que apresenta mudanças abruptas na deficiência do heterozigoto. Os demais parâmetros são mantidos fixos com $r = 0.5$, $\kappa = 1$ e $\eta = 1$. A população reage às mudanças do ambiente se dividindo em duas subpopulações de homozigotos duplos ou reavendo o equilíbrio HW. 39
- 5.2 Emergência e perda da preferência sexual entre os indivíduos da população mediante as mudanças na deficiência do heterozigoto. Esses gráficos referem-se ao mesmo experimento descrito na figura (5.1). . . 39
- 5.3 Processo esquemático do fluxo entre fases acessíveis de uma população sujeita a perturbações do ambiente. 40
- 5.4 Diagrama de fases para diferentes taxas de recombinação de uma população inicialmente em isolamento reprodutivo. Cada ponto nestes diagramas representam uma simulação particular em que uma população é posta inicialmente a evoluir com a configuração ($s_0 = 0.8$, $\kappa_0 = 1$), o que permite a emergência do isolamento reprodutivo. Logo em seguida os parâmetros do ambiente são alterados abruptamente para o novo par (s , κ) representado no diagrama. A população então reage a esta mudança, adquirindo a fase apresentada nos diagramas. . . . 42
- 6.1 Observação da histerese em três valores fixos da taxa de recombinação. Os resultados foram obtidos a partir da média sobre 30 simulações independentes. Na parte superior da figura, apresentam-se as médias do parâmetro de ordem \hat{a} em função da deficiência s . A curva superior descreve os experimentos que a deficiência heterozigota inicia em $s = 0$ e vai sofrendo aumentos (com incremento $\Delta s = 0.033$) a cada intervalo de tempo (gerações); esse intervalo de tempo entre mudanças é suficientemente grande para que a população possa reaver o equilíbrio. A curva inferior é obtida pelo processo inverso, iniciando em $s = 1$. O valor $N = 1000$, $\kappa = 1$ e $\eta = 1$, são mantidos constantes. Na parte inferior da figura, estão as correspondentes flutuações das médias do parâmetro de ordem \hat{a} . A posição do pico, i.e. a transição de fase, depende de s estar aumentando ou diminuindo. 45

- 6.2 Energia livre de um sistema probabilístico qualquer em função do parâmetro de ordem α para diferentes valores do parâmetro de controle s . Como a dinâmica de minimização da energia livre depende de s , então mudanças nesse parâmetro podem levar a mudanças abruptas ou contínuas na posição do mínimo dessa função. Cada mínimo da energia livre está associado a uma fase acessível ao sistema. Aqui, tem-se um exemplo de uma transição abrupta, ou seja, uma transição de primeira ordem. De (a) para (f) s está crescendo. Em (a), existe apenas um mínimo. O sistema pode permanecer nesse mínimo (ou fase) até o valor de s que está na figura (e). Em (f), o primeiro mínimo desaparece e apenas o segundo mínimo existe. Caso o parâmetro s seja alterado de volta, a partir do estado de (f), o sistema pode continuar nesse segundo mínimo até o valor de s que está na figura (b). Apenas em (a), o sistema vai reaver o primeiro mínimo novamente. O sistema apresenta histerese, pois a população pode ser encontrada em duas fases distintas para um dado valor de s 46
- 9.1 Esquema para a atualização da distribuição posteriori via algoritmo bayesiano *on-line*. A posteriori inicia com $p(\mathbf{w}|\mathbf{A}_0)$, escolhido de forma aleatória mas com o vínculo de pertencer ao espaço \mathcal{A} ; a apresentação do dado \mathbf{x}^1 leva à posteriori $p(\mathbf{w}|\mathbf{A}_0, \mathbf{x}^1)$, que não necessariamente pertence ao espaço \mathcal{A} . Em seguida, projeta-se essa distribuição no espaço \mathcal{A} para se obter a posteriori aproximada $p(\mathbf{w}|\mathbf{A}_1)$. Esse processo é feito iterativamente com todos os P dados disponíveis. 59
- 11.1 Evolução *on-line* dos parâmetros de ordem Q e R (figura a) e do erro de generalização (figura b) em função de α . Nesses gráficos, é apresentada a predição teórica do algoritmo bayesiano estudado por Winther e Solla juntamente com os resultados de simulações com o algoritmo PE. As simulações foram realizadas com $N = 1000$, $\kappa = 0$ e com médias realizadas sobre 100 experimentos numéricos independentes. 80
- 11.2 Gráfico da Entropia e do erro de generalização para a previsão de réplica simétrica para o algoritmo ótimo e os resultados do algoritmo PE. As simulações com o algoritmo PE utilizaram $N = 1000$, $\kappa = 0$ e cada ponto gerado pela média sobre 30 experimentos numéricos independentes. 81

-
- 11.3 Gráfico da entropia em função do overlap R para o algoritmo ótimo. Para um dado α , o máximo dessa entropia corresponde ao valor mais provável de R 82
- 11.4 Resultados de simulações com o algoritmo PE nas versões *on-line* e *off-line* em comparação com o algoritmo ótimo. Essas simulações foram realizadas utilizando-se $N = 1000$. Na versão *on-line*, os resultados foram obtidos através da média sobre 100 experimentos numéricos independentes; as barras de erro desses experimentos são menores que a largura da curva. Na versão *off-line* as médias foram realizadas sobre 30 experimentos numéricos independentes. 85

Lista de Tabelas

3.1	Proporção de haplotipos produzidos por dois tipos de heterozigotos duplos em função da taxa de recombinação r	19
9.3	Resumo de notações.	58

1

APRESENTAÇÃO E ESTRUTURA DA TESE

O trabalho que será apresentado nesta tese segue a linha de aplicações da *Mecânica Estatística* (ME) a objetos externos aos campos tradicionais da física, fato que tem se tornado bastante comum nas últimas décadas. Isso se deve, em grande parte, ao fato de que os princípios teóricos e a estrutura matemática da ME são equivalentes à *Teoria da Informação* desenvolvida principalmente por Shannon e Jaynes. A equivalência matemática desses dois campos aparentemente ortogonais possibilitou o aumento da abrangência do conceito de entropia introduzido por Clausius, fazendo dessa entidade a ferramenta básica para lidar com inferência e processamento de informação. Nesse contexto, a ME chega ao século XXI como um instrumento importante no entendimento teórico de problemas típicos na área de Ciência da Computação, Inteligência Artificial, Neurociência e Biologia.

Serão apresentados aqui dois problemas independentes, porém ambos analisados sob o arcabouço da ME. O primeiro pertence ao campo da Biologia Evolutiva, enquanto que o segundo refere-se a um algoritmo de aprendizado em Redes Neurais. Na primeira parte da tese, serão tratados os resultados obtidos sob a orientação e colaboração do Professor Nestor Caticha da Universidade de São Paulo (USP), a respeito da evolução da preferência sexual no processo de Especiação Simpátrica ¹. Já na segunda parte, serão tratados os resultados da parceria com o Professor Manfred Opper da Technische Universität Berlin (TU-Berlin), que resultou no estudo de algoritmos de aprendizado Bayesiano por meio de métodos de aproximação.

A respeito da Especiação Simpátrica, estamos interessados na construção e análise

¹Todos os termos técnicos apresentados aqui serão devidamente definidos no decorrer da tese.

de um modelo matemático-computacional que possa promover em uma população a emergência e perda (reversão) de acasalamento preferencial como resposta a alterações ambientais. Ver-se-á que as configurações do ambiente em que acontecem a emergência e as configurações em que há perda de preferência sexual diferem entre si, evidenciando uma histerese. Histerese é a assinatura de um sistema metaestável sujeito a transições de fase descontínuas (ou de primeira ordem), o que permite a interpretação desse sistema biológico por meio do arcabouço da ME. Uma típica transição de fase descontínua acontece, por exemplo, quando se observa em casa o surgimento de bolhas na água fervente. O aumento da temperatura promove uma mudança descontínua na densidade de moléculas, o que faz com que a água vá, descontinuamente, de uma fase líquida (alta densidade) para uma fase gasosa (baixa densidade) em resposta à mudança na temperatura. Pela formulação da ME, a fase que apresenta uma energia livre menor (ou uma entropia maior) é a mais estável e conseqüentemente, dado um tempo suficientemente grande, o sistema vai convergir para essa fase.

A respeito de Redes Neurais, tem-se verificado ao longo das últimas décadas que a ME é uma técnica bastante poderosa para sua formulação e descrição. Em Redes Neurais, transições de primeira ordem também são observadas. Em modelos de aprendizado, por exemplo, uma rede neural pode se encontrar em uma fase em que não há aprendizado (generalização ruim) ou em uma fase de aprendizado total (generalização perfeita) a respeito de uma certa regra que se quer inferir. Em alguns modelos, há uma transição descontínua entre essas duas fases à medida que são apresentadas novas informações sobre essa regra. Dessa forma, mais uma vez somos remetidos ao formalismo da ME das transições de fase. Nesse contexto, estamos particularmente interessados na formulação de um algoritmo de aprendizado genérico por meio de uma técnica de inferência aproximativa chamada *Propagação de Expectativas*. Esse algoritmo deve ser aplicado no contexto de um aprendizado supervisionado, com o objetivo de inferir, por meio de informações disponíveis, uma regra gerada por uma rede neural professor. Será verificado que o algoritmo desenvolvido apresenta desempenho ótimo em dois cenários estudados e com custo computacional bastante razoável.

Organização da Tese

A primeira parte deste trabalho, que trata da Especiação Simpátrica, está contida entre os capítulos 2 e 7. Uma introdução a respeito desse tipo de Especiação, juntamente com uma apresentação de algumas motivações naturais e alguns modelos teóricos presentes na literatura serão apresentados no capítulo 2. O modelo que foi construído por nós será apresentado no capítulo 3 e os resultados obtidos estão descritos nos capítulos 4 e 5. Entre os resultados, destacam-se a identificação de uma região muito bem definida no espaço de parâmetros em que acontece emergência de preferência sexual entre os indivíduos; além disso, o processo reverso da especiação pôde ser observado e com presença de histerese. O capítulo 6 se dedica a uma discussão detalhada dessa histerese e de sua contextualização com o formalismo da ME. As conclusões desse estudo são apresentadas no capítulo 7. Um pequeno glossário a respeito dos termos biológicos utilizados pode ser encontrado no apêndice A; enquanto que o artigo publicado na revista *Journal of Theoretical Biology* sobre os resultados obtidos está no apêndice F.

A segunda parte deste trabalho, que se refere ao estudo de Inferência aproximativa e Algoritmos de Aprendizado em Redes Neurais, está contida nos capítulos 8 e 12. Uma pequena introdução a respeito de Inferência Bayesiana, Algoritmos de Aprendizado e Inferência Aproximativa, em especial o algoritmo *Propagação de Expectativas*, são apresentados nos capítulos 8 e 9. O capítulo 10 apresenta o algoritmo Bayesiano desenvolvido por nós mediante as técnicas do *Propagação de Expectativas*. Os resultados do algoritmo desenvolvido estão apresentados no capítulo 11. Entre os resultados, destacam-se o desempenho ótimo deste algoritmo em sua versão *off-line* nos dois casos estudados: os modelos do *Perceptron Binário* e *Perceptron Gaussiano*. As conclusões desse estudo estão apresentadas no capítulo 12.

Finalmente no capítulo 13 será apresentado um levantamento geral a respeito deste trabalho. Serão discutidas também as perspectivas futuras e alguns possíveis trabalhos e propostas que podem surgir a partir dos resultados obtidos e apresentados aqui.

Parte I

Especiação Simpátrica

2

A ESPECIAÇÃO SIMPÁTRICA: UMA INTRODUÇÃO

Neste capítulo será apresentada uma pequena revisão da literatura do processo de Especiação Simpátrica. Será apresentada primeiramente a definição deste processo e, em seguida, algumas motivações biológicas para o modelo teórico que será proposto no capítulo seguinte.

Entende-se por especiação os processos evolutivos que estão envolvidos no surgimento de novas espécies biológicas. Uma espécie mãe pode dar origem a duas ou mais espécies filhas quando existe quebra de intercâmbio gênico entre subgrupos de indivíduos. A forma com que se dá esta “quebra” caracteriza o tipo de especiação. Entre os tipos mais importantes, destacam-se dois:

- **Especiação Alopátrica ou Geográfica:** Ocorre quando subpopulações de uma mesma espécie divergem para áreas distintas como consequência de uma separação espacial (alopatria), ocasionada pelo surgimento de uma barreira geográfica (montanha, rio, deserto, etc.). Um desenho esquemático desse tipo de especiação é apresentado na figura (2.1-a). Se as áreas diferirem em suas características ecológicas e/ou as subpopulações ficarem separadas por um tempo suficiente, então deve ocorrer divergência genética entre os indivíduos, propiciando o surgimento de espécies distintas.
- **Especiação Simpátrica (ES):** Especiação sem separação geográfica. Ocorre quando duas ou mais subpopulações, inicialmente da mesma espécie, coexistem em um mesmo território, sem se inter cruzarem (veja figura (2.1-b)). Nesse tipo de especiação, acontece uma modificação genética que impede o cruzamento

entre alguns dos indivíduos da mesma população, criando assim subpopulações reprodutivamente isoladas dentro do mesmo território. Caso essa segregação sexual se mantenha por um tempo suficiente, os indivíduos de diferentes subpopulações começam a apresentar divergência genética e conseqüentemente há o surgimento de novas espécies.

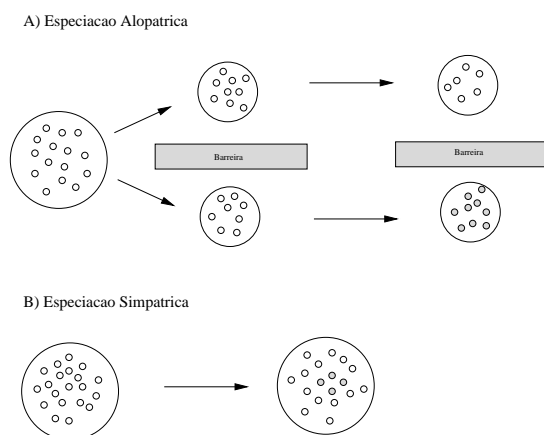


Figura 2.1: Figura esquemática com os dois tipos mais importantes de especiação.

Por um longo tempo, prevaleceu a ideia de que a ES, primeiramente sugerida por Walsh em 1864 [1], não poderia ocorrer ou então deveria ser um evento bastante raro [10]. Porém uma série de estudos teóricos [5, 8, 13, 15] e experimentais [2, 3, 4, 5] têm sugerido a existência desse tipo de especiação. Maynard Smith [8] apontou que a ES pode ocorrer quando acontece emergência de algum tipo de preferência sexual entre os indivíduos, levando-os a um isolamento sexual entre subpopulações. Essa preferência sexual poderia surgir como resposta evolutiva a um ambiente em que indivíduos híbridos apresentam algum tipo de deficiência adaptativa.

Um cenário que reproduz essas características é o chamado *Modelo de Levene* [6, 7]. Nesse modelo, uma população ocupa, simultaneamente, dois nichos ecológicos; uma subpopulação é bem adaptada a um nicho e uma outra subpopulação é bem adaptada ao outro, enquanto que indivíduos híbridos não são adaptados a nenhum deles. Um exemplo canônico dessa ecologia é o ambiente em que vivem uma especialidade de pássaros das ilhas de Galápagos, já descritos por Darwin. Alguns desses pássaros possuem bicos pequenos adaptados a sementes pequenas; outros possuem bicos grandes adaptados a sementes grandes. Um indivíduo híbrido dessas duas variedades, porém, é portador de um bico de tamanho intermediário, não adaptado

a nenhuma das sementes [23]. Nesse cenário, um acasalamento preferencial evitaria o desperdício “energético” do cruzamento entre indivíduos adaptados a diferentes nichos e que produziria híbridos mal adaptados.

Um outro cenário que também poderia promover uma deficiência adaptativa entre híbridos é aquele conhecido na literatura como *Modelo de Competição* [7, 12, 15, 16, 17, 38]. Nesse cenário, os recursos disponíveis estão distribuídos (normalmente uma gaussiana) ao longo do eixo de possíveis traços contínuos na população. Nesse contexto, há duas forças agindo. A primeira faz com que a competição seja menos intensa entre indivíduos com fenótipos opostos. Geralmente os indivíduos apresentam uma intensidade de competição que obedece a uma gaussiana da diferença entre fenótipos. A segunda faz com que o valor médio do *fitness* da população deva evoluir para aquele relacionado com maior abundância de recursos. Nesse cenário a deficiência adaptativa do híbrido deve acontecer quando a variância da competição entre os indivíduos for menor que a variância da distribuição dos recursos. A deficiência do fenótipo intermediário prevalece até que haja variação genética dentro da população que promova o dimorfismo sexual e o consequente isolamento reprodutivo.

Um exemplo natural de ES diz respeito a um inseto chamado de *chrysoperla* ou, como é chamado em inglês, *green lacewings*. Duas espécies de *lacewings*, a *Chrysoperla carnea* e *C. downesi*, que coexistem na América do Norte, são sexualmente separadas pela estação de acasalamento. Há evidências que corroboram o fato de que essas espécies nunca foram segregadas no passado [23]. A *C. carnea* acasala-se no inverno e no verão, enquanto que *C. downesi* acasala-se somente na primavera. As espécies possuem diferenças em sua coloração. A *C. carnea* é verde claro, mas muda para o marrom no outono (veja figura (2.2)), enquanto que a *C. downesi* é verde escuro o ano todo. Essa coloração é um resultado da adaptação das espécies a seu habitat. A *C. downesi* vive principalmente em coníferas e a *C. carnea* vive em árvores decíduas. As coníferas são sempre verdes, pois perdem suas folhas ao longo do ano, em vez de perdê-las todas de uma só vez como acontece com as árvores decíduas. Dessa forma, a coloração faz com que essas espécies se camuflem com a cor de seus habitats, confundindo seus predadores.

O trabalho de Tauber e Tauber [3, 23] determinou que essas espécies, que hibridam com sucesso, são diferentes em três loci. Um deles controla a coloração e, os dois outros, segundo Tauber, estão correlacionados com a estação de acasalamento e consequentemente com o isolamento reprodutivo entre as duas espécies. A coloração da *C. carnea* é expressada por um homozigoto *AA* e a coloração da *C. downesi* por



Figura 2.2: Fotografias de duas *Chrysoperla carnea* e suas colorações assumidas ao longo do ano: verde claro (primavera, verão e inverno) e marrom (outono).

outro homozigoto aa ; o heterozigoto Aa possui uma coloração intermediária, não adaptado nem a coníferas, nem a decíduas. Dois fatos, o primeiro de que esse heterozigoto nunca foi encontrado na natureza e, o segundo, de que ele pode ser gerado a partir de acasalamentos artificiais, sugerem que o isolamento reprodutivo entre as *lacewings* provavelmente foi uma resposta evolutiva à deficiência adaptativa do heterozigoto.

Outro estudo, realizado por Henry [23], mostrou que além da diferença entre as estações de acasalamento entre essas espécies, o isolamento reprodutivo também é influenciado por sons de cortejo emitido pelos machos. Porém não se sabe ao certo se esses dois mecanismos de isolamento sexual (estação de acasalamento e sons de cortejo) evoluíram simultaneamente ou se um deles surgiu primeiro que o outro. Caso um deles tenha evoluído primeiro, o outro apenas complementou o isolamento já estabelecido entre as duas espécies.

Outro exemplo de especiação simpátrica é o caso dos peixes *stickleback* (*Gasterosteus aculeatus*) que apresentam duas espécies reprodutivamente isoladas, a *Benthic* e *Limnetic*, que coexistem em 5 lagos Canadenses [31]. Essas duas espécies são diferentes entre si essencialmente pelo tamanho e pela forma, como pode ser visto na figura (2.3). O *Benthic* é maior em estatura, mas apresenta mandíbulas menores que o *Limnetic*. As diferenças morfológicas entre eles são o resultado evolutivo da adaptação a diferentes nichos ecológicos que eles ocupam [28]. Durante o período de acasalamento, eles habitam a mesma região, com as fêmeas apresentando preferência sexual relacionada ao tamanho do macho [29]. Os híbridos, intermediários no tamanho, representam apenas uma parcela ínfima (cerca de 2%) da população presente na natureza [31]. O fato de que acasalamentos realizados em laboratório



Figura 2.3: Fotografias de duas espécies de peixes *stickleback*. Fonte: [27].

mostram que os híbridos são férteis e viáveis [30] sugerem que eles possuem algum tipo de desvantagem no ambiente natural.

Recentemente, no entanto, alguns estudos apresentados em [31] [32] [33] identificaram um processo de hibridização desses peixes em dois dos lagos em que eles coexistem. A razão desse fenômeno ainda não está clara. Uma hipótese seria o resultado da introdução artificial de um bagre (*Ameiurus nebulosus*) nas águas desses lagos. O fato pode ter criado dois tipos de cenários possíveis. No primeiro, o bagre teria mudado o nicho ecológico, aumentando o *fitness* do *stickleback* híbrido. No segundo, o bagre teria sido responsável pelo aumento da turbidez das águas, diminuindo a acurácia das fêmeas *stickleback* no reconhecimento dos machos e conseqüentemente diminuindo o acasalamento preferencial. Dessa forma, os peixes canadenses *stickleback* apresentam perda (ou diminuição) de preferência sexual que os está levando a uma processo de reversão da especiação simpátrica. Fenômeno similar ao dos peixes *stickleback* também foi observado nos peixes ciclídeos no Lago Vitória e reportado em [34].

Neste trabalho, está-se interessado na elaboração de um modelo estatístico-computacional para analisar a evolução do isolamento sexual entre subpopulações que vivem num mesmo espaço físico. É claro que apenas o isolamento sexual não é suficiente para se dizer que as subpopulações são espécies distintas, uma vez que nesse estágio os indivíduos de diferentes subpopulações ainda conseguem gerar descendentes férteis ¹, embora com pouca frequência. Nesse caso, diz-se que há um

¹Considera-se aqui que duas espécies são distintas quando não conseguem gerar descendentes férteis. Mas claro, esta definição só se aplica à espécies sexuadas.

isolamento sexual pré-zigótico, pois as subpopulações estão sexualmente segregadas, mas há a possibilidade de geração de híbridos férteis. No entanto, se essas subpopulações conseguem se manter isoladas sexualmente por um tempo suficientemente grande, os indivíduos começam a apresentar divergência genética. Nesse momento, eles não conseguem mais gerar híbridos férteis, ou seja, diz-se que as subpopulações apresentam *isolamento sexual pós-zigótico*. Agora, sim, o processo de especiação simpátrica está completo.

Por simplificação, a escala de tempo e as configurações do modelo que serão abordadas neste trabalho permitem que uma população chegue no máximo ao estágio de isolamento sexual, não apresentando divergência genética. Porém, se trabalhar com a hipótese de que nenhum evento extraordinário acontece ao longo do curso evolutivo das subpopulações, a emergência de isolamento sexual será suficiente para concluir que as subpopulações darão origem a espécies distintas.

Modelos Teóricos para a Especiação Simpátrica

Existe na literatura uma vasta gama de modelos matemáticos e computacionais que descrevem emergência de acasalamento preferencial entre os indivíduos de uma população. Um guia bastante prático desses modelos, classificados por tipos e estruturas, pode ser encontrado em [36]

O ingrediente básico do modelo que se abordará neste trabalho é o fato de que os indivíduos apresentam dois fenótipos (ou traços), cada um deles associado a um respectivo genótipo. Enquanto um dos traços é responsável pela ecologia do indivíduo, isto é, por sua capacidade adaptativa, o outro estabelece uma aparência física ao indivíduo sem influenciar sua ecologia. O primeiro será chamado de *traço ecológico* e, o segundo, de *traço aparente*. Os dois genótipos que expressam esses fenótipos estão ligados com uma *taxa de recombinação* r ; essa quantidade pode ser pensada como a distância entre os loci desses genótipos no cromossomo ².

O primeiro trabalho que incluiu esses elementos foi apresentado em [9] e recebeu posteriormente o nome de *Modelo de Udovic*. Nesse modelo, cada indivíduo possui uma probabilidade $1 - \alpha$ de acasalar com indivíduos com o mesmo traço aparente; com probabilidade α , eles são indiferentes a esse traço e se acasalam de forma aleatória. Esse tipo de acasalamento preferencial é baseado na suposição de que há uma relação genética entre o traço aparente e suas preferências sexuais.

²No apêndice é apresentado um glossário com alguns termos técnicos utilizados neste trabalho.

Modelos que incluem essa suposição são conhecidos na literatura por *Assortment Models* [36]. Por outro lado, modelos mais realísticos, conhecidos por *trait-preference models* [36, 37], desconsideram essa relação genética entre preferência sexual e traço aparente.

No modelo de Udovic, que é do tipo Levene, os heterozigotos apresentam uma deficiência adaptativa s em relação aos homozigotos. Foi mostrado pelo autor que, dado que toda a população apresenta uma preferência α constante ao longo das gerações, então existe condição para especiação simpátrica quando $\alpha < 1 - \frac{2r(1-s)(2-s)}{s+4r(1-s)}$. O processo de especiação é, portanto, mais fácil quando há forte ligação entre os loci (r pequeno) e seleção forte (s grande).

A consideração de que a preferência sexual é fixa não é realística; então, algum tipo de sofisticação deveria ser introduzida ao modelo de Udovic. Nos trabalhos de Dieckmann e Doebeli [17] foi introduzido em um modelo *multi-loci* a possibilidade de evolução dessa preferência sexual. Os autores verificaram que uma população inicialmente panmítica evolui para uma população com acasalamento preferencial, formando assim uma barreira sexual que, por sua vez, possibilita a emergência de correlações entre os traços ecológico e aparente. Outros modelos que apresentam emergência de subpopulações sexualmente isoladas a partir de uma população mãe incluem [11, 13, 14, 15].

3

O MODELO

Neste capítulo será proposto um modelo teórico para a análise da evolução de um isolamento sexual de uma população num processo simpátrico. A segregação sexual é o passo fundamental no processo de especiação simpátrica. Esse modelo será apresentado em duas formas, ambas equivalentes, mas com uma representação diferente da mesma hipótese. Essa diferença consiste na abordagem matemática em que cada uma delas é baseada. A primeira forma é um modelo probabilístico que, em geral, é intratável sob o ponto de vista analítico. Já a segunda, diz respeito a um modelo computacional que pode ser realizado por meio de simulações usando técnica de Monte Carlo e descrito em linguagem de *Modelos de Agentes*.

Este capítulo está organizado da seguinte forma. Na seção (3.1) será apresentado a estrutura genética dos indivíduos que compõem a população, sendo eles formados por dois loci, responsáveis, cada um deles, por uma expressão fenotípica específica. Nas seções (3.2) e (3.3) serão descritos o tipo de ambiente que esses indivíduos estão inseridos e de que forma o heterozigoto difere, ecologicamente, dos homozigotos. Na seção (3.4) será apresentada uma versão panmítica do modelo, que possibilitará uma análise de estabilidade. Nas seções (3.5) e (3.6) serão introduzidas as regras para a recombinação entre os loci e da hereditariedade da preferência sexual, respectivamente. Finalmente, a dinâmica imposta pelas regras do modelo será descrita em (3.7). Na última seção, a (3.8), será apresentada a forma com que este modelo de biologia evolutiva se conecta com o arcabouço da Mecânica Estatística.

3.1 A Estrutura genética dos Indivíduos

No modelo para Especiação Simpátrica que será proposto aqui, a população é composta por indivíduos formados por dois loci Mendelianos. O primeiro locus, que será denotado por A , contém os alelos A ou a e expressa o traço responsável pela ecologia

(fitness) do indivíduo (traço ecológico). Já o segundo locus, que será denotado por \mathbb{B} e que contém os alelos B ou b , expressa o traço responsável pela aparência física do indivíduo (traço aparente). Por simplicidade, não serão considerados efeitos de dominância entre alelos.

Como se considera que cada indivíduo possui um genoma constituído por pelo menos dois loci (\mathbb{A} e \mathbb{B}), então pode-se dizer que esses indivíduos possuem quatro haplotipos (ou gametas) possíveis: AB , Ab , aB e ab . AB e ab são chamados de *haplotipos de atração*, enquanto Ab e aB , *haplotipos de repulsão*. Como se está considerando indivíduos diplóides, o genótipo desses indivíduos pode ser representado por uma simbologia baseada ou nos alelos ou nos haplotipos. Por exemplo, um indivíduo formado pelos haplotipos de atração AB e ab é um indivíduo duplamente heterozigoto, que pode ser representado na linguagem dos alelos por $AaBb$ ou na linguagem dos haplotipos por AB/ab [23]. A linguagem dos alelos é menos robusta, uma vez que pode levar a ambiguidades. Por exemplo a representação por alelos do heterozigoto duplo $AaBb$ não informa se esse indivíduo é composto pelos gametas $Ab + aB$ ou por $AB + ab$. Na linguagem dos haplotipos, porém, essa ambiguidade é eliminada.

3.2 O Ambiente tipo Levene

Considera-se nesse modelo um ambiente do tipo Levene, com os homozigotos AA e aa igualmente adaptados aos seus respectivos nichos e o heterozigoto Aa com uma deficiência adaptativa s em relação a eles. É conveniente definir essa quantidade, que também é conhecida por parâmetro de *disruptive selection*, como um número entre 0 e 1 e que expressa a diferença adaptativa entre homozigotos e heterozigotos. Nesse sentido, pode-se definir os fitness ecológicos dos indivíduos como

$$f_{AA}^{(s)} = f_{aa}^{(s)} = 1, \quad f_{Aa}^{(s)} = 1 - s, \quad (3.1)$$

para os homozigotos e heterozigotos, respectivamente.

Considerar-se-á $\xi_{\mathbb{A}}$ uma variável que assume um dos três possíveis genótipos do locus \mathbb{A} : AA , Aa ou aa . De forma análoga, $\xi_{\mathbb{B}}$ assume um dos três possíveis genótipos de \mathbb{B} : BB , Bb ou bb . Como cada genótipo de \mathbb{A} ou \mathbb{B} está diretamente correlacionado ao seu respectivo fenótipo (devido à ausência tanto de dominância entre os genes quanto de interação entre os locus), vai-se muitas vezes abusar da linguagem e se referir $\xi_{\mathbb{A}}$ ou $\xi_{\mathbb{B}}$ como fenótipos.

Está-se interessado na probabilidade de sobrevivência de um certo indivíduo dado que ele possua o genótipo $\xi_{\mathbb{A}}$. Esta probabilidade será denotada por $P(\mathcal{S}|\xi_{\mathbb{A}})$, que lê-se “Probabilidade de que o indivíduo sobreviva até a idade de acasalamento, dado que este possui o genótipo $\xi_{\mathbb{A}}$ ”; aqui, a asserção \mathcal{S} significa: “*O indivíduo sobrevive até a idade de acasalamento*”. Note-se que essa probabilidade é independente do traço aparente $\xi_{\mathbb{B}}$. Na verdade, essa probabilidade é dependente de dois fatores, ambos relacionados ao traço ecológico. O primeiro é o *fitness* ecológico definido em (3.1) e, o segundo, é a competição por recursos disponíveis. Voltar-se-á a falar sobre essa competição na seção (3.3).

Como ilustração desse modelo, pode-se considerar o seguinte cenário. Considere-se uma população de pássaros em que os indivíduos apresentam em seu genoma o locus \mathbb{A} responsável pelo tamanho do seu bico. Esses pássaros vivem em um ambiente em que apenas dois tipos de sementes são viáveis para a sua alimentação; em outras palavras, há apenas dois nichos ecológicos disponíveis. Os pássaros com genótipo AA se alimentam principalmente de um dos tipos de semente, enquanto que os de genótipo aa se alimentam principalmente do outro tipo de semente. Ambos os pássaros homozigotos se alimentam com mesma eficiência, porém os heterozigoto não são adaptados a nenhum dos dois tipos de semente. Ainda dentro desse exemplo, pode-se imaginar que esses pássaros carregam também um outro locus responsável pelos genótipos BB , Bb e bb que dão origem à cor de suas penas (traço aparente). Por exemplo, os homozigotos podem ter penas amarelas ou azuis (cores primárias) e o heterozigoto penas verdes (cor secundária). Obviamente a cor da penagem não possui qualquer relação com a capacidade de coletar ou comer sementes.

3.3 Competição entre Indivíduos

É sensato considerar que a quantidade de recursos é escassa de forma que os indivíduos de uma população devem competir entre si por alimentos disponíveis. O efeito dessa competição deve depender da densidade de indivíduos que usufrui de um certo nicho, com alta competição intranicho mas com baixa competição internicho, em que, sendo essa última suficientemente baixa, pode ser desconsiderada. Em outras palavras, um certo indivíduo compete apenas com outro indivíduo de mesmo traço ecológico, pois ambos utilizam do mesmo nicho. Nesse raciocínio, o *fitness* ligado à competição sentida por um indivíduo com genoma $\xi_{\mathbb{A}}$ será definido como

$$f_{\xi_A}^{(c)} = (1 - \sigma_{\xi_A}), \quad (3.2)$$

em que o *parâmetro de competição* σ_{ξ_A} , um número entre 0 e 1, representa a competição intranicho.

A probabilidade de que o indivíduo com genótipo ξ_A chegue à maturidade sexual e seja hábil para reprodução é o resultado tanto do fitness ecológico (3.1) quanto do fitness de competição (3.2) de modo que

$$P(\mathcal{S}|\xi_A) \propto F_{\xi_A} \equiv f_{\xi_A}^{(s)} f_{\xi_A}^{(c)} = f_{\xi_A}^{(s)}(1 - \sigma_{\xi_A}). \quad (3.3)$$

Aqui foi introduzido o *fitness efetivo* (não normalizado) F_{ξ_A} do indivíduo portador do genótipo ξ_A .

Considere-se a densidade n_{AA} , n_{aa} e n_{Aa} da população dos homozigotos AA , aa e dos heterozigotos Aa , respectivamente. Essas densidades obedecem à normalização $n_{AA} + n_{Aa} + n_{aa} = 1$. Considere-se também que a população efetiva presente em um dos nichos é $n_{AA} + \eta \frac{n_{Aa}}{2}$ e $n_{aa} + \eta \frac{n_{Aa}}{2}$ no outro. Aqui o fator $\frac{1}{2}$ aparece porque se está considerando que os heterozigotos estão igualmente distribuídos entre os dois nichos (uma vez que eles não possuem preferências por nenhum nicho). No entanto um heterozigoto pode (ou não) competir de forma menos intensa que o homozigoto, devido à sua deficiência adaptativa. Dessa forma pode-se assumir que, enquanto um dado homozigoto compete com um fator 1, um heterozigoto compete com um fator η . Esse parâmetro, definido no intervalo $[0, 1]$, representa a eficiência dos heterozigotos na competição por recursos. Enfim, pode-se definir que

$$\sigma_{AA} = \frac{n_{AA} + \eta \frac{n_{Aa}}{2}}{\kappa} \quad (3.4)$$

e

$$\sigma_{aa} = \frac{n_{aa} + \eta \frac{n_{Aa}}{2}}{\kappa}, \quad (3.5)$$

em que $\kappa \equiv \frac{K}{N}$, uma fração do *carrying capacity* K por indivíduo, satisfaz $1 \leq \kappa < \infty$. Aqui, vai-se considerar que K é o *carrying capacity* em termos dos recursos e N regulado por algo mais, como espaço, sítios disponíveis para o ninho, etc.. Por simplicidade vai-se considerar que os dois nichos apresentam o mesmo *carrying capacity*.

Uma possível escolha para o impacto de competição η dos heterozigotos é escrevê-lo em termos da própria deficiência do heterozigoto, sendo este menor quanto maior

a deficiência desse indivíduo; por exemplo $\eta = 1 - s$. Por outro lado, a escolha $\eta = 1$, i.e. mesmo impacto de competição que os homozigotos, faz com que os parâmetros de competição possam ser escritos apenas em termos da frequência de alelos A , dada por $p = n_{AA} + \frac{n_{Aa}}{2}$ (e conseqüentemente $1 - p$ é a frequência do alelo a). Nesse caso tem-se

$$\sigma_{AA} = \frac{p}{\kappa} \quad (3.6)$$

e

$$\sigma_{aa} = \frac{1 - p}{\kappa}. \quad (3.7)$$

Com relação à competição sentida pelo heterozigoto, vai-se considerá-la como a média simples da competição sentida pelos homozigotos, de modo que

$$\sigma_{Aa} = \frac{\sigma_{AA} + \sigma_{aa}}{2}. \quad (3.8)$$

Essa é uma simplificação que acontece se se considerar que os dois homozigotos possuem a mesma população. Essa formulação introduz uma seleção dependente da frequência que favorece indivíduos que se encontram em um nicho menos denso (*negative frequency-dependent selection* [23]). Para o caso particular $\kappa = 1$ e $\eta = 1$, este modelo recupera o modelo de seleção dependente da frequência, estudado e apresentado em [39] e [40].

3.4 Versão Pan-mítica do Modelo

Até o presente momento, ainda não foram definidas as regras da evolução da preferência sexual entre os indivíduos da população. Antes dessa introdução, porém, serão analisadas brevemente as conseqüências dinâmicas desse modelo no caso de acasalamento aleatório e sem possibilidade de evolução da preferência sexual. Essa versão será denominada de *Modelo Pan-mítico*. Claro que em pan-mixia, como já foi discutido nos primeiros trabalhos teóricos sobre especiação simpátrica [8, 9], nenhuma bifurcação da população deve acontecer. Dessa forma, não se deve esperar a especiação simpátrica nesse modelo simples. No entanto, a observação desse contexto será útil para análises de estabilidade.

Em uma população pan-mítica a frequência do acasalamento entre dois genótipos está relacionada com a frequência desses genótipos na população. Considerando que a população está na geração g em equilíbrio Hardy-Weinberg (HW), pode-se escrever

a frequência genotípica em termos da frequência dos alelos pela forma $n_{AA} = p^2$, $n_{aa} = (1-p)^2$ e $n_{Aa} = 2p(1-p)$. Valendo-se do *fitness* efetivo introduzido em (3.3), a densidade de indivíduos com genótipo AA e Aa na próxima geração ($g+1$) será

$$n'_{AA} = \frac{1}{\langle \mathcal{F} \rangle^2} \left(n_{AA}^2 F_{AA}^2 + n_{AA} n_{Aa} F_{AA} F_{Aa} + \frac{1}{4} n_{Aa}^2 F_{Aa}^2 \right) \quad (3.9)$$

$$n'_{Aa} = \frac{1}{\langle \mathcal{F} \rangle^2} \left(n_{AA} n_{Aa} F_{AA} F_{Aa} + 2n_{AA} n_{aa} F_{AA} F_{aa} + \frac{1}{2} n_{Aa}^2 F_{Aa}^2 + n_{Aa} n_{aa} F_{Aa} F_{aa} \right), \quad (3.10)$$

respectivamente. Aqui foi introduzido $\langle \mathcal{F} \rangle$ como o *fitness* médio da população.

Considerando $p' = n'_{AA} + \frac{n'_{Aa}}{2}$ como a densidade do alelo A da geração seguinte, pode-se determinar o incremento $\Delta p = p' - p$ de uma geração para outra, que será

$$\Delta p = -\frac{1}{2} \frac{p(3p-s-1-2p^2-10p^3s\eta-2s\eta p+8s\eta p^2-6psk+4p^2sk+4p^4s\eta+5ps+2sk-10sp^2+10p^3s-4p^4s)}{-1+3p+k-3p^2-\eta p+\eta p^2-4p^3s\eta+2s\eta p^2-2psk+2p^2sk+2p^4s\eta+ps-3sp^2+4p^3s-2p^4s} \quad (3.11)$$

Pode-se verificar facilmente por testes de estabilidade de (3.11) que esse sistema apresenta três pontos fixos não triviais. Esses pontos, que satisfazem $\Delta p = 0$, são

$$p_1^* = \frac{1}{2} \quad (3.12)$$

$$p_{2,3}^* = \frac{1}{2} \frac{s\eta - s \pm \sqrt{-s^2 + s^2\eta^2 + 2s\eta - 4s^2\eta k - 2s + 4s^2k}}{s(\eta - 1)}. \quad (3.13)$$

Todos eles são estáveis quando obedecem a inequação

$$s < \frac{2}{4\kappa - (\eta + 1)}. \quad (3.14)$$

Porém, apenas $p_1^* = \frac{1}{2}$ apresenta solução real nessa região de estabilidade e portanto apenas ele interessa.

Estabilidade em $p = \frac{1}{2}$ significa que o sistema se mantém em polimorfismo, pois todos os alelos estão presentes. A não obediência da inequação (3.14) leva a população para as soluções triviais $p = 0$ ou $p = 1$. Nesse caso, um dos alelos se extingue, dando origem a uma população monomórfica. Além disso, $\Delta p = 0$ com $p = \frac{1}{2}$ significa que não há alterações gênicas na população e, portanto, se o sistema estava em equilíbrio HW na geração g , ele também estará na geração $g+1$.

O gráfico da figura (3.1) apresenta a curva que delimita os regimes de polimorfismo e monomorfismo do locus \mathbf{A} , dada pela inequação (3.14). Nota-se que há apenas uma diferença quantitativa entre as curvas. O resultado qualitativo é o mesmo, independentemente da escolha de η .

Sintetizando estas ideias, pode-se concluir que o polimorfismo é mantido desde que κ seja suficientemente pequeno ou, em outras palavras, desde que haja pouco alimento e, conseqüentemente, forte competição entre os indivíduos. No entanto, para κ suficientemente grande (pouca competição entre indivíduos), basta uma pequena desvantagem do heterozigoto para levar um dos alelos de \mathbf{A} à extinção. Essa característica mostra a importância da competição para a manutenção do polimorfismo em uma população sujeita à desvantagem heterozigota.

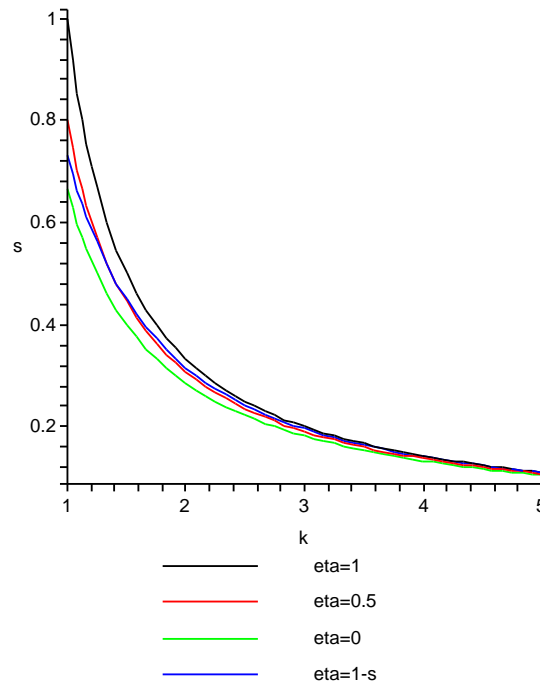


Figura 3.1: Fronteira entre os regimes de Polimorfismo (abaixo de cada curva) e Monomorfismo (acima de cada curva) para diferentes escolhas de η .

Para finalizar, pode-se pensar que a manutenção do polimorfismo é o resultado da competição entre duas forças: a deficiência adaptativa do heterozigoto, que tende a destruir o polimorfismo, e a seleção dependente da frequência, que tende a preservar o polimorfismo. Vale enfatizar que os alelos do locus \mathbf{B} , por não influenciarem no

fitness dos indivíduos, permanecem (nesta versão pan-mítica) em equilíbrio HW, indiferentes à dinâmica dos alelos do locus \mathbb{A} . Dessa forma, em uma população finita, um dos alelos de \mathbb{B} pode muito bem ser perdido (se extinguir) por deriva genética. Outro fato interessante é que a fronteira de estabilidade dessa versão pan-mítica se mantém importante, mesmo quando se enriquece o modelo com uma taxa de recombinação entre os loci ou com possibilidade de emergência de preferências sexuais entre os indivíduos. Detalhes serão apresentados nos capítulos seguintes.

3.5 Taxa de Recombinação entre os loci

A probabilidade de que os genes dos gametas sejam segregados pelo *crossover* é dado pela *taxa de recombinação* r . Esse parâmetro, compreendido entre 0 e $\frac{1}{2}$, é diretamente relacionado com a distância entre os loci \mathbb{A} e \mathbb{B} no cromossomo. No caso extremo $r = 0$ os dois loci estão pareados um com o outro no cromossomo, o que faz com que os alelos extraloci (alelos que pertencem a loci diferentes) nunca se segreguem por *crossover*; nesse caso, diz-se que os alelos extraloci estão fortemente ligados. No outro caso extremo $r = \frac{1}{2}$, que é o limite de ligação fraca, os dois loci estão muito distantes entre si no cromossomo ou mesmo pertencem a cromossomos diferentes; nesse caso, os genes extraloci segregam-se independentemente.

A recombinação entre os alelos extraloci só acontece quando envolve indivíduos duplamente heterozigotos. Esses indivíduos podem gerar todos os quatro tipos possíveis de gametas com probabilidades que dependem de r (veja tabela 3.1). Pode-se verificar facilmente que os outros tipos de indivíduos produzem apenas seus próprios gametas, independentemente da taxa de recombinação.

	AB	Ab	aB	ab
AB/ab	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$
Ab/aB	$\frac{1}{2}r$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$

Tabela 3.1: Proporção de haplotipos produzidos por dois tipos de heterozigotos duplos em função da taxa de recombinação r .

3.6 Acasalamento Preferencial

Os indivíduos da população apresentam, além dos traços ecológico e aparente, uma expressão fenotípica que governará sua preferência sexual. Um certo indivíduo, que se chamará de “ f ” e que se encontra na idade de acasalamento, deve escolher um parceiro sexual. Porém, a única informação que ele possui a respeito dos candidatos são os seus respectivos traços aparentes. f , que possui genótipos $\xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f$, deve aceitar ou não um candidato com traço aparente $\xi_{\mathbb{B}}$ de acordo com sua *aceitabilidade sexual*

$$\alpha^f(\mathcal{A}|\xi_{\mathbb{B}}, \xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f). \quad (3.15)$$

Essa expressão fenotípica, compreendida no intervalo $[0, 1]$, representa a probabilidade condicional de que f aceite o acasalamento com um candidato sexual com traço aparente $\xi_{\mathbb{B}}$. A asserção \mathcal{A} significa: “*indivíduo aceita o acasalamento*”.

Como existem três traços aparentes possíveis na população, cada indivíduo apresenta três aceitabilidades sexuais (uma para cada traço aparente), guardadas no *vetor aceitabilidade sexual*

$$\boldsymbol{\alpha}^f(\mathcal{A}|\xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f) = \left(\alpha^f(\mathcal{A}|BB, \xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f), \alpha^f(\mathcal{A}|Bb, \xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f), \alpha^f(\mathcal{A}|bb, \xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f) \right). \quad (3.16)$$

Vale explicitar que nesse modelo cada indivíduo terá o seu próprio vetor aceitabilidade sexual.

Esse tipo de modelagem para a preferência sexual pertence à classe de modelos *trait preference model*, uma vez que os indivíduos apresentam uma preferência sexual específica para cada um dos traços aparentes presentes na população.

3.7 A Dinâmica

Após a devida introdução das regras da preferência sexual especificar-se-ão detalhes sobre a dinâmica da população. Considere-se

$$P(\xi_{\mathbb{A}} \xi_{\mathbb{B}} \boldsymbol{\alpha} | t) d\boldsymbol{\alpha} \quad (3.17)$$

como a probabilidade de nascimento (ou densidade) de indivíduos com genótipo $\xi_{\mathbb{A}} \xi_{\mathbb{B}}$ e vetor preferência sexual compreendido entre $\boldsymbol{\alpha}$ e $\boldsymbol{\alpha} + \Delta \boldsymbol{\alpha}$ na geração t . Note-se que a representação $\xi_{\mathbb{A}} \xi_{\mathbb{B}} \boldsymbol{\alpha}$, à qual se denominará *genoma*, especifica completamente um

indivíduo. Aqui, as variáveis $\xi_{\mathbb{A}}$ e $\xi_{\mathbb{B}}$ são discretas, enquanto que as componentes do vetor α são contínuas. O diferencial presente em (3.17), que é o elemento de volume do espaço tri-dimensional que contém o vetor α , pode ser escrito também como:

$$d\alpha = d^3\alpha = d\alpha(\mathcal{A}|BB, \xi_{\mathbb{A}}\xi_{\mathbb{B}})d\alpha(\mathcal{A}|Bb, \xi_{\mathbb{A}}\xi_{\mathbb{B}})d\alpha(\mathcal{A}|bb, \xi_{\mathbb{A}}\xi_{\mathbb{B}}). \quad (3.18)$$

A probabilidade de que dois indivíduos, a que se chamará de “ f ” e “ m ” de genomas $\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\alpha^f$ e $\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\alpha^m$, respectivamente, sobrevivam até a maturidade sexual e que se encontrem será

$$P(\mathcal{S}|\xi_{\mathbb{A}}^f)P(\mathcal{S}|\xi_{\mathbb{A}}^m) P(\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\alpha^f|t) P(\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\alpha^m|t)d^3\alpha^m d^3\alpha^f, \quad (3.19)$$

onde $P(\mathcal{S}|\xi_{\mathbb{A}})$ é dado pela equação (3.3). Mas apenas o encontro não é suficiente para o acasalamento, uma vez que os indivíduos possuem preferências sexuais e podem ou não aceitar um ao outro para o ato sexual. Dessa forma, a probabilidade de que estes dois indivíduos se acasalem será dada em termos da preferência sexual, introduzida em (3.16), equacionada como

$$P(\mathcal{M}|\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\alpha^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\alpha^m) = \alpha^f(\mathcal{A}|\xi_{\mathbb{B}}^m, \xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f)\alpha^m(\mathcal{A}|\xi_{\mathbb{B}}^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m), \quad (3.20)$$

onde a asserção \mathcal{M} significa “acontece o acasalamento”.

Até o momento, não se fez qualquer distinção entre machos e fêmeas. No entanto, tal distinção deve ser introduzida, pois aumenta a plausibilidade do modelo. Na natureza, na maioria das vezes, os machos são sexualmente receptíveis, enquanto que as fêmeas apresentam forte seleção sexual [3]. Nas simulações do modelo baseado em agentes, que se descreverá na próxima seção, cada indivíduo recebe um *label* (escolhido de forma aleatória) por meio do qual se diz que ele é uma fêmea (“ f ”) ou um macho (“ m ”). Caso o indivíduo seja um macho, ele será totalmente receptível a qualquer fêmea. Com essa consideração, a probabilidade de acasalamento (3.20) toma a forma mais simples

$$P(\mathcal{M}|\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\alpha^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\alpha^m) = \alpha^f(\mathcal{A}|\xi_{\mathbb{B}}^m, \xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f), \quad (3.21)$$

pois depende apenas da aceitabilidade da fêmea.

Uma vez que o acasalamento entre f e m se consumou, a probabilidade de que haja a geração dos genótipos $\xi_{\mathbb{A}}\xi_{\mathbb{B}}$ será denotada por

$$P_{cross}(\xi_{\mathbb{A}}\xi_{\mathbb{B}}|\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m). \quad (3.22)$$

Essa probabilidade depende do *crossover* entre os haplotipos e, portanto, está diretamente relacionada à taxa de recombinação r , quando pelo menos um dos seus genitores é heterozigoto duplo; para qualquer outro tipo de acasalamento que não envolva heterozigotos duplos, essa probabilidade é puramente mendeliana.

Mutação

Na escala de tempo de que se tratará, apenas as mutações da preferência sexual serão consideradas. Isso significa que as mutação dos genes A , a , B , b e da taxa de recombinação r são tão raras (se comparadas com as mutações das preferências sexuais) que podem ser desconsideradas.

Considerar-se-á que a probabilidade de que f e m gerem um filho com vetor preferência sexual α seja denotada por

$$P_{pref}(\alpha|\alpha^f, \alpha^m). \quad (3.23)$$

Além disso, restringir-se-á à análise de que a preferência sexual é herdada apenas da mãe para filho ou filha; ou seja, o macho não influencia a aceitabilidade dos filhos. Essa consideração leva à simplificação

$$P_{pref}(\alpha|\alpha^f, \alpha^m) = P_{pref}(\alpha|\alpha^f) \quad (3.24)$$

da probabilidade (3.23).

Para cada indivíduo, há uma probabilidade $1 - \mu$ de que a sua aceitabilidade sexual seja idêntica à de sua mãe e uma probabilidade μ de que essa aceitabilidade tenha sofrido uma mutação. O parâmetro μ (tipicamente $\mathcal{O}(10^{-2})$) é a *taxa de mutação*. No modelo de agentes, quando um indivíduo é um mutante, este recebe uma aceitabilidade sexual gerada aleatoriamente.

Equação de Recorrência

Com todos os ingredientes introduzidos, pode-se escrever a densidade de indivíduos com o genoma $\xi_A \xi_B \alpha$ na geração $t + 1$ em termos da população prévia. Dado que não há superposição de gerações, o modelo probabilístico para especiação simpátrica proporciona a seguinte equação de atualização

$$\begin{aligned}
P(\xi_{\mathbb{A}}\xi_{\mathbb{B}}\boldsymbol{\alpha}|t+1) &\propto \sum_{\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m} P_{cross}(\xi_{\mathbb{A}}\xi_{\mathbb{B}}|\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m) \times \\
&\times \left(\int P_{pref}(\boldsymbol{\alpha}|\boldsymbol{\alpha}^f, \boldsymbol{\alpha}^m)P(\mathcal{M}|\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\boldsymbol{\alpha}^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\boldsymbol{\alpha}^m) \times \right. \\
&\times \left. P(\mathcal{S}|\xi_{\mathbb{A}}^f)P(\mathcal{S}|\xi_{\mathbb{A}}^m)P(\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\boldsymbol{\alpha}^f|t)P(\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\boldsymbol{\alpha}^m|t)d^3\alpha^f d^3\alpha^m \right) \quad (3.25)
\end{aligned}$$

Considerando as simplificações discutidas, a equação de recorrência acima adquire a forma

$$\begin{aligned}
P(\xi_{\mathbb{A}}\xi_{\mathbb{B}}\boldsymbol{\alpha}|t+1) &= \sum_{\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m} P_{cross}(\xi_{\mathbb{A}}\xi_{\mathbb{B}}|\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f, \xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m) \\
&\times \int d^3\alpha_f P_{pref}(\boldsymbol{\alpha}|\boldsymbol{\alpha}_f)\alpha_f(\mathcal{A}|\xi_{\mathbb{B}}^f\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^m) \\
&\times P(\mathcal{S}|\xi_{\mathbb{A}}^f)P(\mathcal{S}|\xi_{\mathbb{A}}^m)P(\xi_{\mathbb{A}}^f\xi_{\mathbb{B}}^f\boldsymbol{\alpha}_f|t)P(\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m|t), \quad (3.26)
\end{aligned}$$

onde se introduz a distribuição marginal $P(\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m|t) = \int d\boldsymbol{\alpha}^m P(\xi_{\mathbb{A}}^m\xi_{\mathbb{B}}^m\boldsymbol{\alpha}^m|t)$.

A equação de recorrência acima será simulada por um modelo de agentes. Tendo fixado os parâmetros r , s e κ , a população deve evoluir para um estado de equilíbrio. No próximo capítulo, será abordado com detalhes esse modelo de agentes e seus resultados, bem como os diferentes estados de equilíbrio para essa dinâmica.

3.8 Os Parâmetros de Ordem

Será introduzida agora uma grandeza que proporcionará a interpretação desse problema de biologia evolutiva por meio do arcabouço da Mecânica Estatística (ME). Em um sistema sujeito à transições de fase de primeira ordem, objeto da teoria da ME, cada fase acessível pode ser caracterizada ou identificada a partir da medida de uma certa quantidade chamada de *Parâmetro de Ordem* [20, 21]. Nesse tipo de transições (de primeira ordem), o parâmetro de ordem pode ser facilmente identificado, pois ele sofre mudanças descontínuas quando as configurações do sistema passeiam por regiões entre fases.

No caso da água, por exemplo, o parâmetro de ordem é a sua densidade, que assume um certo valor na fase líquida e um valor bem mais baixo na fase gasosa.

No momento da transição líquido-gás (de primeira ordem) essa densidade apresenta uma descontinuidade. Um outro exemplo é o caso de materiais magnéticos. Nesses materiais, o magnetismo é produzido por spins e outros momentos magnéticos que podem se alinhar espontaneamente para produzir magnetismo, mesmo na ausência de um campo magnético externo. Esses materiais apresentam duas fases acessíveis: um estado em que o valor médio da magnetização aponta para cima (apresenta sinal positivo) e, um outro, em que esse valor médio aponta para baixo (apresenta sinal negativo). Dessa forma, como a magnetização média apresenta valores bem específicos para cada fase acessível do sistema e, portanto, possibilita a identificação da fase do sistema pela sua medida, então ela é o parâmetro de ordem de um sistema magnético.

Nesse presente problema biológico, deve-se identificar uma quantidade que possa agir como o parâmetro de ordem. Por meio desse parâmetro pode-se identificar se a população apresenta acasalamento preferencial (uma fase) ou se reproduz em panmixia (outra fase). Essas duas fases mencionadas estão diretamente correlacionadas com a preferência sexual; portanto a preferência média entre os indivíduos deve agir como parâmetro de ordem, como sugerido em [11].

Antes de se definir com clareza esse parâmetro de ordem, definir-se-á a seguinte notação para o que se chamará de *homozigoto oposto*. O homozigoto oposto de BB é bb e vice versa. O heterozigoto não tem homozigoto oposto. Representar-se-á, então, por $\bar{\xi}_{\mathbb{B}}$ o homozigoto oposto de $\xi_{\mathbb{B}}$; isso significa, por exemplo, que se $\xi_{\mathbb{B}} = bb$, então $\bar{\xi}_{\mathbb{B}} = BB$.

Pode-se, assim, introduzir o parâmetro de ordem

$$\hat{\alpha} \equiv \sum_{\xi_{\mathbb{A}}} \sum_{\xi_{\mathbb{B}}=BB,bb} \int d^3\alpha P(\xi_{\mathbb{A}}\xi_{\mathbb{B}}|\alpha|t) \alpha(\mathcal{A}|\bar{\xi}_{\mathbb{B}}, \xi_{\mathbb{A}}\xi_{\mathbb{B}}) \quad (3.27)$$

para o modelo probabilístico, onde se usa a notação integral

$$\int d^3\alpha(\dots) = \int \int \int d\alpha(\mathcal{A}|BB, \xi_{\mathbb{A}}\xi_{\mathbb{B}}) d\alpha(\mathcal{A}|Bb, \xi_{\mathbb{A}}\xi_{\mathbb{B}}) d\alpha(\mathcal{A}|bb, \xi_{\mathbb{A}}\xi_{\mathbb{B}})(\dots). \quad (3.28)$$

Esse parâmetro de ordem nada mais é que o valor médio da aceitabilidade sexual de uma fêmea homozigota em \mathbb{B} em relação a um macho homozigoto oposto em \mathbb{B} . Em outras palavras, $\hat{\alpha}$ é a aceitabilidade sexual média de uma fêmea por um macho que possui um traço aparente oposto ao dela. A segregação sexual entre indivíduos de diferentes traços aparentes deve ser sinalizada quando $\hat{\alpha} \rightarrow 0$. Nesse caso, a condição para que a população sofra especiação simpátrica será atingida.

No caso do modelo baseado em agentes, $\hat{\alpha}$ pode ser medido usando-se

$$\hat{\alpha} \equiv \frac{\sum_f \alpha_f (\mathcal{A} | \xi_{\mathbb{B}}^f, \xi_{\mathbb{A}}^f \xi_{\mathbb{B}}^f)}{\sum_f 1_f(\xi_{\mathbb{B}}^f)}, \quad (3.29)$$

em que a somatória \sum_f age sobre todos as fêmeas de uma população composta por N indivíduos (machos e fêmeas); no denominador tem-se o número total de fêmeas homozigotas. A representação $1_f(\xi_{\mathbb{B}}^f)$ é uma função do evento “*indivíduo f é homozigoto em \mathbb{B}* ”, assumindo o valor 1 se o evento é verdadeiro ou 0 se não.

Será visto no próximo capítulo, onde serão apresentados os resultados de simulações, que existe uma outra quantidade que também funciona como parâmetro de ordem nesse sistema. Essa quantidade é o *desequilíbrio de ligação* entre os loci \mathbb{A} e \mathbb{B} , definido como

$$D \equiv g_{AB}g_{ab} - g_{aB}g_{Ab}, \quad (3.30)$$

onde g_{AB} , g_{ab} , g_{Ab} e g_{aB} , são as frequências dos quatro haplotipos possíveis na população. Quando $D > 0$, o gene A é mais frequentemente encontrado com o gene B ; da mesma forma, a com b . Nesse caso, o desequilíbrio de ligação informa que há excesso de haplotipos de atração. De forma análoga, $D < 0$ diz que há excesso de haplotipos de repulsão. Para o caso particular $D \approx 0$, não há nenhum excesso, o que significa que um certo gene do locus \mathbb{A} tem a mesma probabilidade de ser encontrado com qualquer outro do locus \mathbb{B} . Essa quantidade pode ser utilizada como parâmetro de ordem, porque o valor por ela assumido está diretamente correlacionado com a preferência sexual entre os indivíduos. É fácil perceber, que a situação $D \neq 0$ só pode acontecer quando o acasalamento preferencial está presente na população. A emergência de $\hat{\alpha} \rightarrow 0$ promove uma diminuição de heterozigotos duplos, juntamente com o aumento da população de homozigotos duplos; nessa configuração, haverá excesso de um dos tipos de haplotipos e, conseqüentemente, $D \neq 0$.

4

DIAGRAMA DE FASES

Neste capítulo, descrevem-se os resultados de simulações computacionais a respeito do modelo de agentes introduzido no capítulo anterior. Verifica-se, pela simulação desse modelo, que a emergência de acasalamento preferencial é possível; porém, apenas em uma específica região das configurações dos parâmetros do sistema. Dessa forma, serão apresentados os resultados de uma varredura pelo espaço de configurações dos parâmetros κ , r e s . Essa varredura foi realizada não só com o objetivo de identificar o regime de acasalamento preferencial, mas também de identificar outros regimes acessíveis à população.

Este capítulo está organizado da seguinte forma. Na seção (4.1), será apresentado o algoritmo para a simulação do modelo de agentes. Os resultados de um experimento realizado com este algoritmo e que promove emergência de acasalamento preferencial entre os indivíduos da população é apresentado na seção (4.2). Verifica-se na seção (4.3) que, ao percorrer o parâmetro de seleção s , mantendo os demais parâmetros fixos, acontece uma transição de fase da aceitabilidade sexual. Esse parâmetro de ordem vai de um regime sem acasalamento preferencial, quando s é pequeno, para um regime de acasalamento preferencial, quando s é suficientemente grande. Uma varredura a procura dos regimes acessíveis ao sistema, juntamente com a identificação da região do espaço de parâmetros onde o isolamento reprodutivo acontece, é apresentada na seção (4.4).

4.1 O Algoritmo do Modelo de Agentes

Modelos baseado em agentes tem como ideia principal simular as interações microscópicas entre indivíduos que compõem uma população (os “agentes”), de forma a se observar algum tipo de fenômeno coletivo. Nesses modelos, cada agente é uma entidade computacional que funciona de forma autônoma, munida de regras de in-

teração com o ambiente e com os demais agentes. Em suma, esses modelos servem como laboratório virtual para o estudo da emergência de fenômenos macroscópicos a partir de interações microscópicas.

O modelo de agentes que será tratado aqui apresenta uma população com tamanho fixo, formada por N indivíduos. Nos experimentos numéricos realizados, N é suficientemente grande para evitar perda de genes por deriva genética. As gerações são discretas e não há contato sexual ou geográfico entre indivíduos de diferentes gerações. O algoritmo para as simulações segue os seguintes passos:

- Passo 1 - Cria-se uma população de N indivíduos, todos eles formados por dois loci (\mathbb{A} e \mathbb{B}). Escolhem-se aleatoriamente os genes que irão formar esses indivíduos, de modo que a população inicial esteja em equilíbrio HW. Além disso, todos esses N indivíduos são sexualmente receptíveis, com $\alpha^f = 1$ para qualquer indivíduo f .
- Passo 2 - Durante a geração g , escolhem-se aleatoriamente dois indivíduos da população baseado na probabilidade de sobrevivência

$$P(\mathcal{S}|\xi_{\mathbb{A}}) = (1 - \sigma_{\xi_{\mathbb{A}}})f_{\xi_{\mathbb{A}}}, \quad (4.1)$$

que está relacionada apenas com o traço ecológico desses indivíduos. Arbitrariamente denomina-se o primeiro indivíduo como fêmea e o segundo como macho.

- Passo 3 - Decide-se se a fêmea aceita ou não o acasalamento com o macho. Essa aceitação será baseada na aceitabilidade sexual $\alpha_{\xi_{\mathbb{B}}} \in [0, 1]$ dessa fêmea em relação ao traço aparente $\xi_{\mathbb{B}}$ desse macho. Mais especificamente, a aceitação ocorre se um número R_{rand} ($0 \leq R_{rand} \leq 1$), gerado aleatoriamente, satisfaz $R_{rand} < \alpha_{\xi_{\mathbb{B}}}$. Existem as seguintes possibilidades:

- Caso a fêmea *aceite* o macho, dois novos indivíduos (filhos) são gerados; eles vão compor a população da geração $g + 1$. Os genótipos desses filhos serão construídos a partir dos genótipos paternos por meio do *crossover* e de regras mendelianas. A aceitabilidade sexual dos filhos será herdada da mãe, salvo sob mutações. Cada filho tem probabilidade $1 - \mu$ de que sua aceitabilidade sexual seja idêntica à de sua mãe; porém com probabilidade

μ esse indivíduo terá uma aceitabilidade gerada aleatoriamente entre 0 e 1.

- Caso a fêmea *não aceite* o macho, um novo indivíduo é sorteado, com base na probabilidade de sobrevivência. Esse indivíduo será um novo macho para cortejo. A fêmea continua a mesma. Repete-se o passo 3 até que essa fêmea aceite, finalmente, um macho para o acasalamento.

Obs. 1: Neste presente modelo, não se consideram custos para a escolha sexual da fêmea. Dessa forma, ela pode rejeitar quantos machos forem necessários até a sua escolha final.

Obs. 2: Um dado indivíduo pode ser escolhido mais de uma vez para o acasalamento (extração com reposição).

- Passo 4 - Vai-se para o passo 2, até que a população da nova geração ($g + 1$) complete N membros. Todos os indivíduos da geração g são deletados.

4.2 Análise de caso

Será visto logo adiante que existem configurações bem específicas dos parâmetros s , κ (relacionados ao ambiente) e r (relacionado aos indivíduos) que promovem emergência de preferência sexual. Antes dessa análise, porém, será apresentada uma simulação particular em que acontece essa emergência. Nesse experimento numérico, considerar-se-á que (i) os dois loci \mathbb{A} e \mathbb{B} segregam-se independentemente ($r = \frac{1}{2}$), (ii) existe alta competição por recursos ($\kappa = 1$) e (iii) os heterozigotos Aa são pouco adaptados, com $s = 0.7$, embora tenham um impacto competitivo idêntico ao dos homozigotos, i.e. $\eta = 1$.

As figuras (4.1) e (4.2) apresentam a evolução da população nessa simulação. Pode-se perceber que, em torno da geração $g = 250$, a população se divide em dois tipos de homozigotos duplos ($AAbb$ e $aaBB$), declinando a população dos heterozigotos. A figura (4.2) mostra que esta bifurcação é consequência da emergência da preferência sexual entre os indivíduos: fêmeas $AAbb$ adquirem uma rejeição a machos com traço aparente BB (i.e. $\langle \alpha(\mathcal{A}|BB, AAbb) \rangle \rightarrow 0$), enquanto que as fêmeas $aaBB$ rejeitam machos com traço aparente bb ($\langle \alpha(\mathcal{A}|bb, aaBB) \rangle \rightarrow 0$). Por outro lado, a aceitabilidade das fêmeas homozigotas em relação a machos heterozigotos não vai a

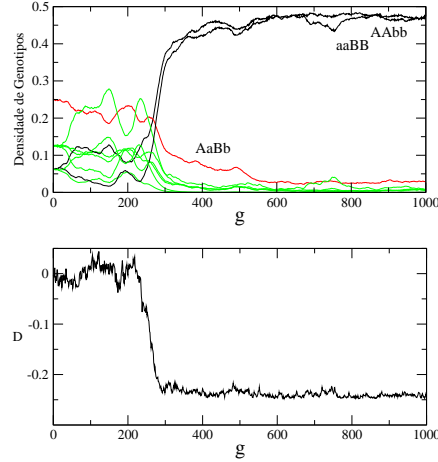


Figura 4.1: Evolução da densidade de genótipos e do desequilíbrio de ligação dos haplotipos de uma população. Neste particular experimento, realizado com $N = 1000$, $\mu = 0.01$, $r = \frac{1}{2}$, $\kappa = 1$, $\eta = 1$ e $s = 0.7$, a população bifurca em dois genótipos: $AAbb$ e $aaBB$, implicando um excesso de haplotipos de repulsão ($D < 0$). Uma simulação desse tipo (com 1000 gerações) demanda cerca de 30 segundos de tempo de cpu para ser realizada.

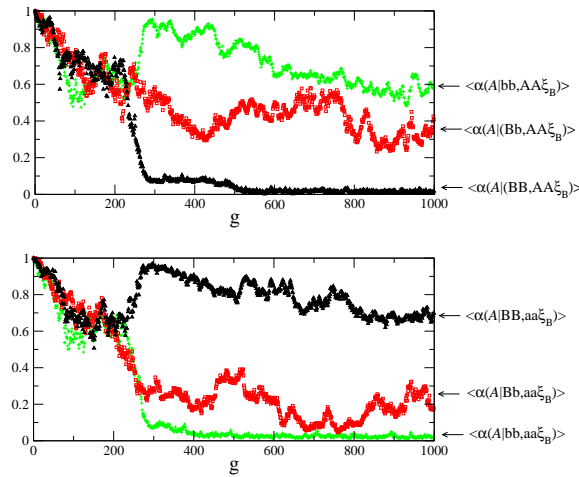


Figura 4.2: Evolução da aceitabilidade sexual das fêmeas homozigotas AA (acima) e aa (abaixo). Fêmeas AA desenvolvem rejeição aos machos que exibem um traço aparente BB ($\langle \alpha(\mathcal{A}|BB, AA\xi_B) \rangle \rightarrow 0$), enquanto que fêmeas aa desenvolvem uma rejeição a machos bb ($\langle \alpha(\mathcal{A}|bb, aa\xi_B) \rangle \rightarrow 0$). Essa associação entre os dois traços (ecológico e aparente) e conseqüentemente entre os dois loci \mathbb{A} e \mathbb{B} acontece mesmo com ausência de ligação entre eles ($r = \frac{1}{2}$). Esse resultado provém do mesmo experimento numérico apresentado na figura (4.1).

zero. Isso acontece, pois a população desses machos praticamente se anula, cessando assim a pressão evolutiva para que as fêmeas os rejeitem.

O fato interessante, já observado nos trabalhos [16] e [17], é que a emergência dessa preferência sexual leva ao surgimento de uma correlação entre dois loci, mesmo sendo eles geneticamente independentes. Dessa forma, a população vai do equilíbrio ($D \approx 0$) para o desequilíbrio de ligação com excesso de haplotipos de repulsão ($D < 0$).

Nesse particular experimento, a evolução de preferência sexual conduziu à predominância dos homozigotos duplos $AAbb$ e $aaBB$. Mas, claro, dependendo das condições iniciais, a emergência da população $AABB$ e $aabb$ é igualmente provável.

4.3 Transição de Fase

O experimento numérico descrito na seção anterior mostra a emergência de um regime de preferência sexual em uma população, dada uma particular configuração do ambiente. Uma vez observado que a emergência de preferência sexual é possível no modelo proposto, necessita-se verificar os limites desse regime, dado o espaço de configurações dos parâmetros. O que se pode fazer inicialmente é realizar experimentos com vários valores possíveis de um certo parâmetro, enquanto que os demais permanecem fixos. Por exemplo, nas simulações apresentadas na figura (4.3), foram fixados $\kappa = 1$, $r = 0.5$ e $\eta = 1$, percorrendo os valores possíveis de s . Cada ponto nesse gráfico representa a média (e a respectiva variância) da preferência sexual ao traço oposto, obtida sobre 30 simulações independentes. Nota-se a ausência de preferência sexual quando $s \lesssim 0.6$, que é caracterizada por $\hat{a} \neq 0$. A emergência de isolamento reprodutivo acontece em $s \gtrsim 0.6$, assinalada por $\hat{a} \approx 0$. O pico na variância evidencia uma transição de fase da aceitabilidade \hat{a} . Será mostrado nos dois capítulos seguintes que esta transição apresenta histerese, i.e. memória do seu estado anterior, sugerindo uma transição de primeira ordem na aceitabilidade sexual. Detalhes sobre esta histerese serão discutidos nos próximos capítulos.

4.4 Diagrama de Fases

O que se pode fazer agora é estender essa varredura aos demais parâmetros do sistema. O objetivo dessa varredura é não só verificar os limites da emergência de isolamento reprodutivo mas também identificar outros regimes acessíveis. Referindo-

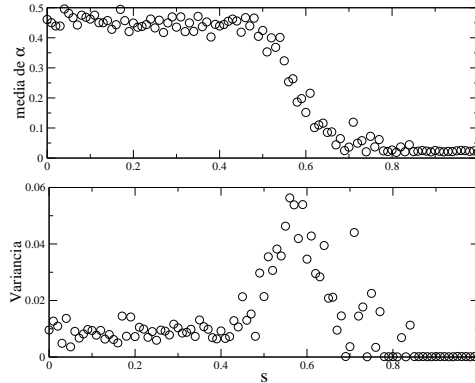


Figura 4.3: Média de $\hat{\alpha}$ (aceitabilidade em relação ao traço oposto) e sua respectiva variância em função da deficiência do heterozigoto. Cada ponto representa a média sobre 30 simulações independentes com $r = 0.5$, $\kappa = 1$ e $\eta = 1$ fixos. O pico da variância em torno de $s = 0.6$ evidencia a transição de fase de $\hat{\alpha}$.

se a esses regimes como *fases*, pode-se fazer analogia à termodinâmica e construir o chamado *Diagrama de Fases* do sistema.

Primeiramente serão analisados, separadamente, dois diagramas de fases bi-dimensionais: os diagramas $s-\kappa$ e $s-r$. Posteriormente, será apresentado um diagrama tridimensional $\kappa-s-r$. As simulações mostram que uma população, inicialmente sem preferência sexual, deve convergir para uma de três fases (ou regimes) possíveis, dependendo das configurações dos parâmetros s , κ e r . Essas fases são:

1. *Fase de Acasalamento Preferencial (AP)*: acontece emergência de isolamento reprodutivo, permitindo que a população se bifurque em duas sub-populações de homozigotos duplos sexualmente segregados. É nessa fase que a especiação simpátrica deve acontecer.
2. *Fase de Extinção*: um dos alelos do locus \mathbb{A} se extingue. Esse regime acontece quando a seleção dependente da frequência não é suficientemente forte para manter o polimorfismo. Não há emergência de preferência sexual nesse regime, pois o monomorfismo resultante da extinção faz com que todos os indivíduos apresentem um único traço ecológico. Sendo assim, como todos os indivíduos apresentam o mesmo *fitness*, não há pressão evolutiva para que se estabeleça uma preferência sexual.
3. *Fase tipo HW*: a população permanece em equilíbrio HW e sem emergência de isolamento reprodutivo. Esse regime acontece quando: (i) a seleção depen-

dente da frequência é forte o bastante para manter o polimorfismo, mas (ii) a pressão evolutiva imposta pelo ambiente não é suficiente para promover a emergência de acasalamento preferencial.

Diagrama $s - \kappa$

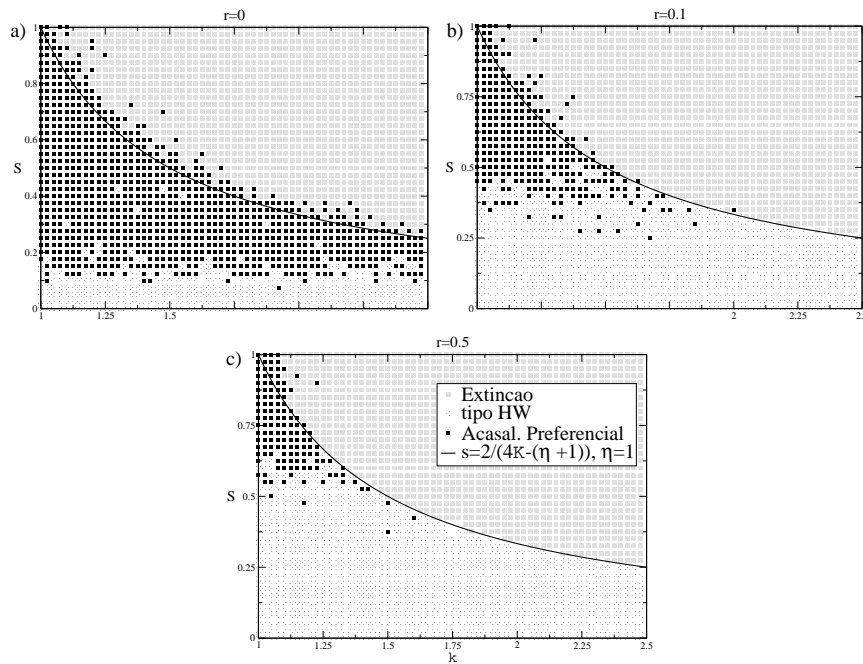


Figura 4.4: Diagrama de fase $s-\kappa$ para três valores (fixos) de r . Cada ponto desses gráficos representa uma simulação realizada com os respectivos valores dos parâmetros do diagrama e com $N = 1000$ e $\eta = 1$. A linha $s = \frac{2}{4\kappa - (\eta + 1)} \Big|_{\eta=1}$ separa a fase de Extinção das duas fases polimórficas, enquanto que as fases polimórficas são separadas por uma linha horizontal (não apresentada) de transição de primeira ordem. O tamanho da região da fase de Acasalamento Preferencial é totalmente dependente de r , sendo maior quanto menor o valor da taxa de recombinação.

A figura (4.4) apresenta o diagrama $s - \kappa$ com r fixo em diferentes valores. Pode-se perceber que na região acima da curva

$$s = \frac{2}{4\kappa - (\eta + 1)} \quad (4.2)$$

a população vai para a fase de Extinção. Essa curva, que separa a fase monomórfica das outras duas fases polimórficas, é a mesma apresentada na seção (3.4), onde foi

abordado o Modelo Pan-mítico. Essa similaridade dos resultados do modelo pan-mítico e do modelo com evolução de preferência sexual não pode ser uma coincidência. O que acontece é que, na região acima dessa curva, o processo de eliminação de um alelo é mais rápido que o estabelecimento de uma preferência sexual. Dessa forma a população age sempre em panmixia, o que explica a coerência entre as simulações desse modelo e a curva (4.2).

Simulações com parâmetros abaixo dessa curva apresentam duas fases polimórficas distintas que são separadas por uma linha de transição de primeira ordem da aceitabilidade $\hat{\alpha}$. A região mais abaixo é caracterizada por uma população que se mantém em reprodução pan-mítica (fase tipo HW), enquanto que na região mais acima há a emergência e a fixação do acasalamento preferencial. É nessa última região de configuração dos parâmetros que a especiação simpátrica pode ocorrer.

Os diagramas da figura (4.4) confirmam que uma maior ligação entre os loci (r pequeno) facilita a emergência de isolamento reprodutivo; “fácil” aqui significa um ambiente não tão hostil aos heterozigoto (s pequeno). Isso é bastante coerente, uma vez que quanto mais r se aproxima de zero (forte ligação entre loci), mais os indivíduos se comportam como se os dois traços (ecológico e aparente) fossem apenas um. Assim sendo, a fêmea têm acesso direto ao traço ecológico, o que não é possível no limite de ligação fraca entre os loci.

Um fato interessante é que a linha de transição descontínua é independente de κ . Isso pode ser entendido por meio do seguinte raciocínio. Análises de estabilidade dinâmica dentro da região polimórfica sugerem que a frequência p dos alelos A só é estável e com solução real e não trivial quando assume $p = \frac{1}{2}$. Esse ponto fixo é o mesmo nas duas fases polimórficas. Considerando então que na região polimórfica vale as igualdades $p = n_{AA} + \frac{n_{Aa}}{2} = \frac{1}{2}$ e $(1 - p) = n_{aa} + \frac{n_{Aa}}{2} = \frac{1}{2}$, pode-se usar as definições (3.4) e (3.5) para obter

$$\sigma_{\xi_A} = \frac{1}{2\kappa} [1 - n_{Aa}(1 + \eta)], \quad (4.3)$$

qualquer que seja ξ_A . Dessa forma, a intensidade da seleção dependente da frequência será a mesma para todos os genótipos, o que permite escrever o *fitness* efetivo pela forma $F_{AA} = F_{aa} = 1 \cdot (1 - \sigma_{\xi_A})$ e $F_{Aa} = (1 - s) \cdot (1 - \sigma_{\xi_A})$. Finalmente, a probabilidade de se sortear um indivíduo com genótipo ξ_A para o acasalamento deve ser proporcional ao *fitness* efetivo e frequência desse genótipo, i.e.

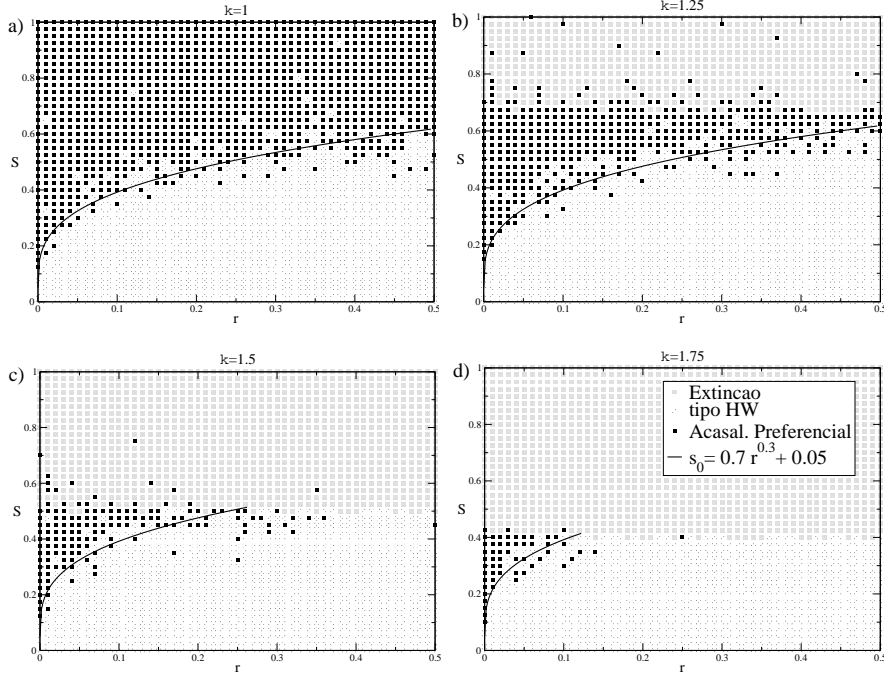


Figura 4.5: Diagrama de fases $s - r$ para diferentes valores (fixos) de κ . Assim como a figura (4.4), cada ponto nestes gráficos representa uma simulação realizada com os respectivos valores dos parâmetros do diagrama e com $N = 1000$ e $\eta = 1$. A linha preta representa a transição de primeira ordem entre as fases tipo HW e Acasalamento Preferencial. Essa linha foi obtida a partir da média de 30 experimentos numéricos independentes para cada r . Uma linha horizontal (não apresentada) separa a fase de extinção das fases polimórficas. Pode-se perceber que quanto maior κ (menor seleção dependente da frequência) menor é a região de emergência de preferência sexual.

$$P_{\xi_A} = \frac{n_{\xi_A} F_{\xi_A}}{C}, \quad (4.4)$$

onde C é uma constante obtida pela normalização $C = \sum_{\xi_A} n_{\xi_A} F_{\xi_A}$. Essa normalização faz com que o parâmetro κ se cancele em (4.4). Dessa forma, a transição de fase é devida à mudança de s e não de κ .

Em todos os experimentos realizados, foi assumido que os heterozigotos competem com a mesma intensidade que os homozigotos, ou seja $\eta = 1$. Na verdade, verificou-se que o valor desse parâmetro não afeta qualitativamente as fronteiras entre as três fases acessíveis ao sistema. Dessa forma, a escolha $\eta = 1$ é tão signifi-

cante quanto qualquer outra escolha.

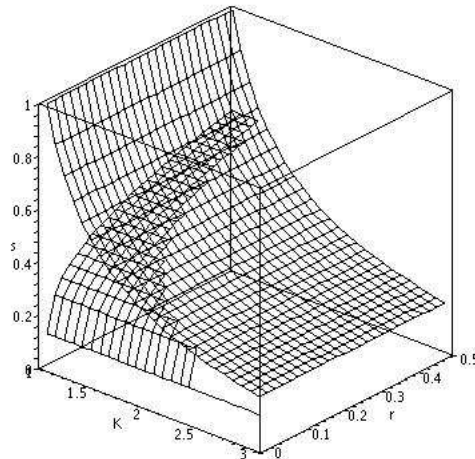


Figura 4.6: Representação tridimensional do diagrama de fases no espaço (r, κ, s) . As três fases acessíveis são separadas por duas superfícies. A superfície superior, dada pela equação (4.2), é independente de r . Experimentos realizados com configurações dos parâmetros que se encontram acima dessa superfície conduzem a população à extinção de um dos alelos de \mathbb{A} . A curva mais abaixo é o resultado do ajuste $(s = ar^b + c)$, descrito na figura (4.5). A emergência de preferência sexual acontece em experimentos realizados com configurações dos parâmetros que estão entre essas duas superfícies (acima da superfície inferior e abaixo da superior). Nos experimentos realizados com configurações abaixo desses dois planos, a população mantém o polimorfismo inicial, porém sem emergência de preferência sexual, i.e. fase tipo HW.

Diagrama $s - r$

O diagrama $s - r$ com diferentes valores de κ fixos e as respectivas fases acessíveis ao sistema está apresentado na figura (4.5). Note-se que a emergência de preferência sexual, assim como no diagrama $s - \kappa$, está sempre restrita a uma região bem determinada no espaço dos parâmetros. O tamanho dessa região diminui com o aumento de κ . Isso acontece, pois com o aumento de recursos disponíveis diminui a competição. Com pouca competição, a seleção dependente da frequência é fraca, o que torna mais viável a eliminação de um alelo, possibilitando, assim, o aumento da

região do espaço de fase que promove a fase de Extinção. Em uma outra situação extrema, no entanto, como é o caso de $\kappa = 1$, a seleção dependente da frequência é forte o bastante para manter o polimorfismo, independentemente de s ou de r . Um fato curioso apresentado por essas simulações é que, mesmo no limite de ligação forte entre os loci ($r = 0$), é necessário um limiar de deficiência do heterozigoto para que aconteça emergência de isolamento reprodutivo.

Diagrama $r - \kappa - s$

Finalmente, podem-se condensar as informações contidas nos dois diagramas apresentados em um único “volume” de fases (r, κ, s) , apresentado na figura (4.6). Nesse diagrama tridimensional, duas superfícies separam as fases acessíveis a uma população inicialmente sem preferência sexual. A emergência de preferência sexual, e conseqüentemente, a possibilidade de especiação simpátrica, acontece na região de parâmetros localizada entre as superfícies.

Nesse volume de fases, cada uma das superfícies é independente de um parâmetro. A superfície que separa a fase de Extinção das demais fases é independente da taxa de recombinação r ; ao mesmo tempo que a superfície que separa a fase com preferência sexual da fase tipo HW é independente de κ .

5

MUDANÇAS NO AMBIENTE E REVERSÃO DE ESPECIAÇÃO

Os resultados discutidos no capítulo anterior foram obtidos no contexto de uma população sujeita a um ambiente estável ao longo das gerações. Além disso, a população evoluía (nesse ambiente estável) a partir de uma população inicialmente sem preferência sexual. Neste capítulo, serão repetidos esses mesmos procedimentos, porém com ressalvas. Será escolhida inicialmente uma configuração de parâmetros do ambiente que possa conduzir a população ao isolamento reprodutivo; uma vez estabelecido esse regime, serão realizadas mudanças arbitrárias nas configurações do ambiente, deixando a população evoluir novamente. A população pode, ou não, permanecer no estado de isolamento reprodutivo de acordo com as novas configurações. Esse procedimento permitirá analisar a robustez da preferência sexual no modelo proposto.

Este capítulo está organizado em duas seções. Na primeira, a seção (5.1), será analisada a evolução de uma população sujeita a uma deficiência heterozigota que muda de tempos em tempos. Será verificado que a população pode adquirir, perder, ou mesmo readquirir a preferência sexual de acordo com as configurações do parâmetro de seleção. Na segunda, a seção (5.2), será analisada a persistência do acasalamento preferencial por meio de uma varredura pelo espaço de configurações do ambiente.

5.1 Mudanças Abruptas do Ambiente

A ideia por trás das simulações que se seguirão, consiste na mudança arbitrária dos parâmetros s e/ou κ , de tal modo que esses parâmetros cruzem as superfícies apresentadas na figura (4.6). Nesse sentido, as simulações devem partir de um ponto do espaço $s - k - r$ que promova emergência do Acasalamento Preferencial e, em seguida, ser alterado para outro ponto além das fronteiras dessa fase. Analisar-se-á, então, a possibilidade da persistência da fase Acasalamento Preferencial em regiões do volume de fases em que o isolamento reprodutivo não poderia acontecer, quando a população inicial se reproduz em pan-mixia.

Será analisado, inicialmente, o caso em que o parâmetro de seleção s se mantém estável por algumas gerações, mas sofre mudanças bruscas de tempos em tempos. Verificar-se-á de que forma a população reage às pressões evolutivas impostas por esse ambiente inconstante. As figuras (5.1) e (5.2) apresentam a evolução de uma população sujeita a $r = 0.5$ e $\kappa = 1$ fixo. Essa simulação inicia com $s = 0.7$, o que proporciona, rapidamente, a emergência do isolamento reprodutivo entre os indivíduos da população. No entanto, na geração $g = 1000$ o coeficiente de seleção é alterado para $s = 0.35$; como resposta, a população oscila, mas não perde o acasalamento preferencial. Foi visto no capítulo anterior que o isolamento reprodutivo não emerge quando uma população pan-mítica está sujeita a este valor da deficiência heterozigota. Porém, uma vez estabelecido o acasalamento preferencial, o isolamento sexual apresenta robustez em se manter, mesmo em regiões do espaço de fase que, *a priori*, não causaria sua emergência.

De qualquer forma, existe limite a essa robustez. A população reverte à pan-mixia quando se reduz o parâmetro de seleção para $s = 0.2$. Nessa nova configuração do ambiente, o isolamento reprodutivo é perdido, pois ele não é mais viável sob o ponto de vista evolutivo. Aqui, da mesma forma que vem acontecendo com os peixes *stickleback*, descritos no capítulo 2, a população realizou o processo reverso da especiação. Essa hibridização é uma resposta evolutiva a um ambiente menos hostil aos heterozigotos.

Ainda referindo-se aos experimentos descritos nas figuras (5.1) e (5.2), o parâmetro de seleção continua sofrendo alterações. Agora, acontece uma mudança para $s = 0.5$; a preferência sexual não emerge. Note-se que, entre as gerações 1000 e 2000, o coeficiente de seleção era menor, no entanto o isolamento reprodutivo estava presente, como herança do estado anterior. Isso mostra que a população apresenta

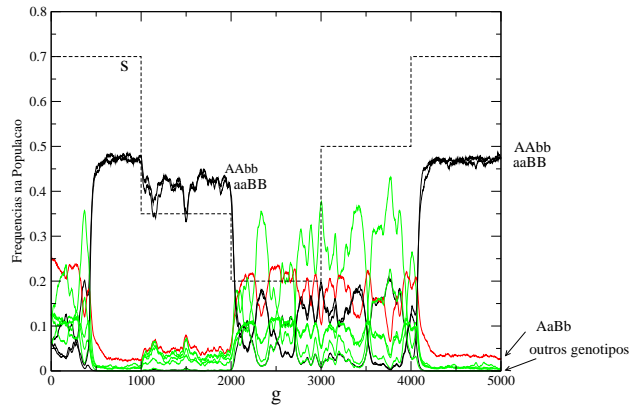


Figura 5.1: Evolução de uma população sujeita a um ambiente que apresenta mudanças abruptas na deficiência do heterozigoto. Os demais parâmetros são mantidos fixos com $r = 0.5$, $\kappa = 1$ e $\eta = 1$. A população reage às mudanças do ambiente se dividindo em duas subpopulações de homozigotos duplos ou reavendo o equilíbrio HW.

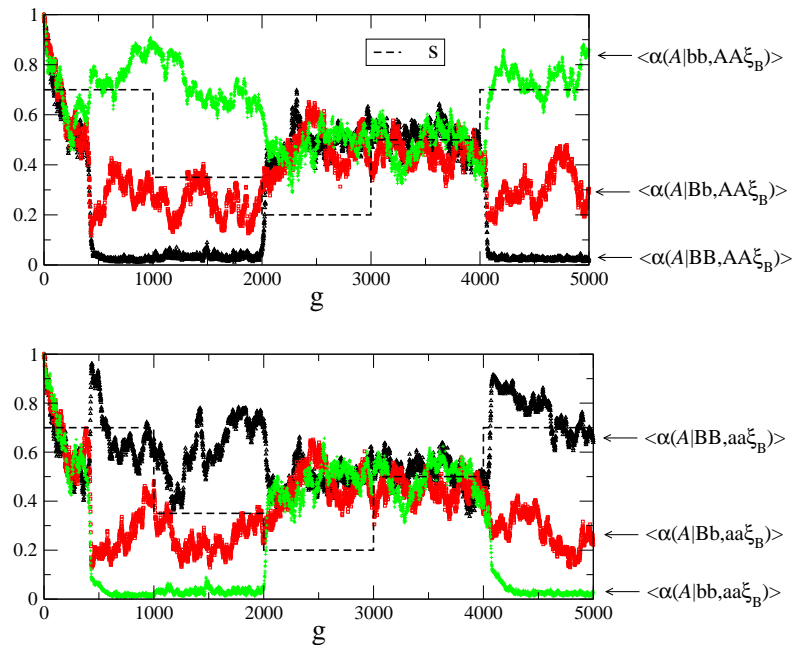


Figura 5.2: Emergência e perda da preferência sexual entre os indivíduos da população mediante as mudanças na deficiência do heterozigoto. Esses gráficos referem-se ao mesmo experimento descrito na figura (5.1).

uma “memória” que influencia o estado de equilíbrio, em resposta a perturbações no ambiente. Em outras palavras, diz-se que o sistema apresenta *histerese*, que é a assinatura de uma transição de primeira ordem. Essa característica será descrita com mais detalhes no próximo capítulo.

Esses resultados permitem concluir que o sistema apresenta acesso entre as duas fases polimórficas. O isolamento reprodutivo pode emergir e desaparecer (ou vice-versa) dependendo da deficiência heterozigota. É claro que esse acesso não é permitido quando se trata da fase de Extinção, pois, uma vez monomórfica, a população não readquire o alelo perdido, pelo menos não na escala de tempo aqui considerada. O processo de extinção de um alelo é um processo irreversível. Um desenho esquemático dos fluxos do sistema, sujeito a perturbações na fronteira entre fases acessíveis, pode ser visualizado pela figura (5.3).

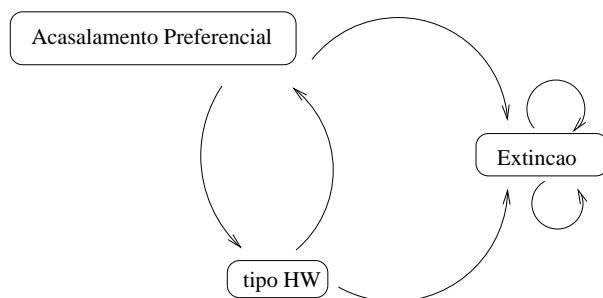


Figura 5.3: Processo esquemático do fluxo entre fases acessíveis de uma população sujeita a perturbações do ambiente.

5.2 Robustez do Isolamento Reprodutivo

Nesta seção, será analisada a robustez do isolamento reprodutivo em função das diversas configurações possíveis do ambiente. Com esse objetivo, foi construído o diagrama $s - \kappa$, para a descrição de simulações que obedecem aos seguintes procedimentos. Uma particular simulação inicia com o par de parâmetros (s_0, κ_0) , escolhidos de tal forma que seja possível a emergência de isolamento reprodutivo; por exemplo $(s_0, \kappa_0) = (0.8, 1)$. Após a emergência da segregação, um novo par de parâmetros (s, κ) é escolhido; a população, em resposta, pode ou não perder o acasalamento preferencial.

Os resultados dos regimes assumidos pelas populações no diagrama $s - \kappa$ estão apresentados na figura (5.4). Note-se que a região com isolamento reprodutivo avança consideravelmente para regiões em que não seria possível sua emergência em uma população previamente pan-mítica. Ela avança tanto para s menores, onde se esperaria uma fase tipo HW, quanto para κ grande, onde se esperaria uma fase de extinção. Dessa forma, a fase de extinção não acontece nas imediatas vizinhanças superiores da curva (4.2), como se espera em uma população previamente pan-mítica; isso se deve ao fato de que as duas sub-populações sexualmente isoladas interagem fracamente. Essa interação é muito mais forte quando a população se reproduz em pan-mixia, o que leva à extinção de um dos alelos. Para s (> 0.3) fixo, a probabilidade de que a preferência sexual persista decresce com o aumento de κ e vai para zero quando κ é suficientemente grande ($> 12.$). Para $s < 0.3$, os heterozigotos não são penalizados o suficiente e o comportamento da população independe se esta se encontra previamente com ou sem isolamento reprodutivo: a população sempre convergirá para a fase tipo HW.

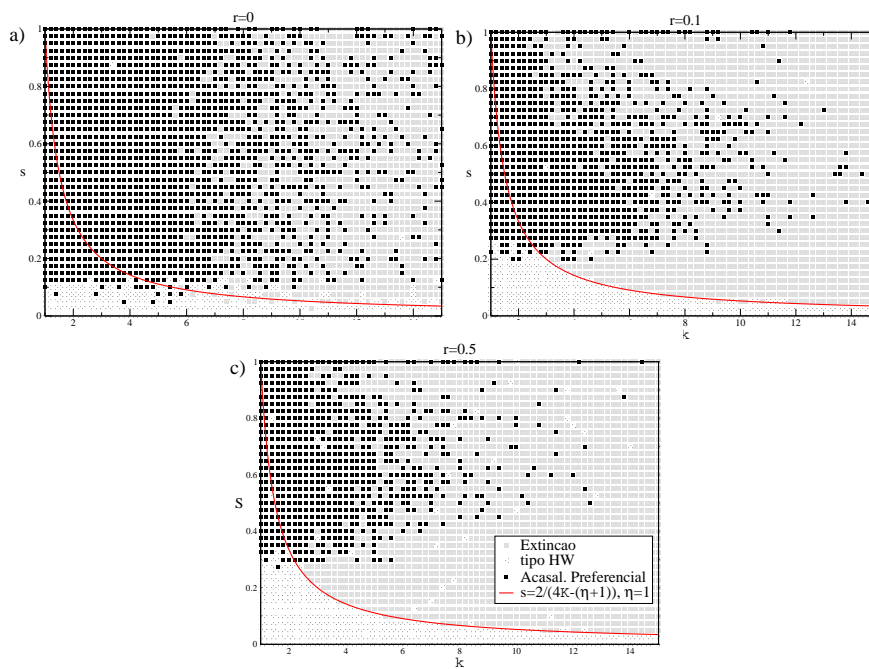


Figura 5.4: Diagrama de fases para diferentes taxas de recombinação de uma população inicialmente em isolamento reprodutivo. Cada ponto nestes diagramas representam uma simulação particular em que uma população é posta inicialmente a evoluir com a configuração ($s_0 = 0.8$, $\kappa_0 = 1$), o que permite a emergência do isolamento reprodutivo. Logo em seguida os parâmetros do ambiente são alterados abruptamente para o novo par (s, κ) representado no diagrama. A população então reage a esta mudança, adquirindo a fase apresentada nos diagramas.

6

DISCUSSÕES SOBRE TRANSIÇÃO DE FASE

Foi visto nos capítulos prévios que as simulações do modelo proposto apresenta uma população com memória do seu estado anterior. Essa característica faz com que o sistema permaneça no regime de isolamento reprodutivo, mesmo em regiões do espaço de fase que não poderia promover a emergência dessa segregação. Dessa forma, pode-se dizer que o isolamento reprodutivo é suficientemente robusto para manter a população segregada. Neste capítulo, será quantificada com mais detalhes essa robustez do acasalamento preferencial. Neste sentido, será apresentado os resultado de simulações de uma população sujeita a um ambiente em que a deficiência do heterozigoto apresenta mudanças infinitésimas e constantes ao longo das gerações. O objetivo é quantificar os valores da deficiência heterozigota que proporcionam emergência e os valores que proporcionam perda da preferência sexual. A diferença entre esses dois valores informa a histerese do sistema.

A histerese é a assinatura de transições de primeira ordem e, por isso, uma explicação da evolução da preferência sexual por meio dos conceitos da Mecânica Estatística é uma escolha imediata. Serão abordados aqui alguns conceitos relevantes da Teoria de Transições de Fase da Mecânica Estatística e suas aplicações ao modelo para Especiação Simpátrica. Mais detalhes dessa teoria podem ser encontrados em [20] ou [21].

Este capítulo está organizado da seguinte forma. Na seção (6.1) estão apresentados os resultados de simulações que mostram a histerese no modelo para Especiação Simpátrica. A interpretação dessa histerese por meio de uma minimização da função energia livre é apresentada na seção (6.2). Nessa seção, fazer-se-á também uma analogia entre este modelo biológico e a transição de fases da água. O capítulo finaliza na seção (6.3), com uma interpretação para a transição de fase.

6.1 A Histerese

A histerese é uma característica típica de sistemas que apresentam transições de primeira ordem, como o que está sendo estudado aqui. O gráfico apresentado pela figura (6.1) mostra com clareza a presença de histerese. Nessa figura, estão apresentados os resultados médios da preferência sexual ao traço oposto frente a mudanças graduais no parâmetro de seleção s ; os demais parâmetros permanecem fixos. O intervalo de tempo (gerações) entre as mudanças do ambiente é grande o bastante para que a população possa reaver o equilíbrio. O valor de s que acontece a transição depende de a deficiência do heterozigoto estar diminuindo ou aumentando. A posição exata da transição pode ser determinada pela flutuação (variância) do valor médio do parâmetro de ordem. Além disso, o tamanho da histerese é nitidamente dependente da taxa de recombinação r .

6.2 Minimização da Energia Livre

Modelos probabilísticos como este podem sempre ser pensados por meio de uma dinâmica de minimização de uma função chamada *energia livre* no contexto da Mecânica Estatística. A figura (6.2.a-f) apresenta, de forma ilustrativa, uma função energia livre, cuja forma depende de um parâmetro (de controle) que se assumirá s . Cada um dos seis gráficos refere-se a um diferente valor desse parâmetro. A dinâmica de um sistema probabilístico age de forma a minimizar a função energia livre em relação a um parâmetro (de ordem) que se assumirá α . Note-se que a curva da energia livre tem inicialmente apenas um mínimo e, logo em seguida, desenvolve dois mínimos, à medida que s muda; cada mínimo está associado a uma fase acessível ao sistema e, conseqüentemente, cada fase, associada a um valor específico de α . A profundidade desses mínimos também se altera com a mudança de s . Variando-se s , o primeiro mínimo se torna menor e, simultaneamente, o segundo mínimo se torna maior. Isso acontece até o desaparecimento completo do primeiro mínimo; a energia livre volta a ter um único mínimo, mas agora em um valor de α diferente do que era inicialmente. A mudança de um mínimo para outro caracteriza a transição de fase. À medida que s varia, o valor do mínimo promove um salto descontínuo do parâmetro de ordem, de um mínimo para outro. O ponto exato da transição acontece quando os dois mínimos apresentam a mesma profundidade.

O processo de minimização dessa energia livre apresenta memória de sua condição

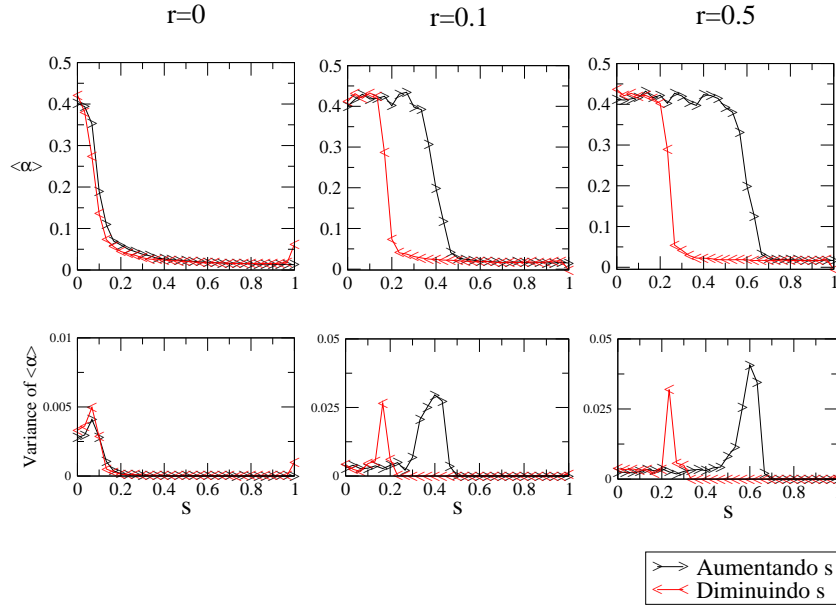


Figura 6.1: Observação da histerese em três valores fixos da taxa de recombinação. Os resultados foram obtidos a partir da média sobre 30 simulações independentes. Na parte superior da figura, apresentam-se as médias do parâmetro de ordem $\hat{\alpha}$ em função da deficiência s . A curva superior descreve os experimentos que a deficiência heterozigota inicia em $s = 0$ e vai sofrendo aumentos (com incremento $\Delta s = 0.033$) a cada intervalo de tempo (gerações); esse intervalo de tempo entre mudanças é suficientemente grande para que a população possa reaver o equilíbrio. A curva inferior é obtida pelo processo inverso, iniciando em $s = 1$. O valor $N = 1000$, $\kappa = 1$ e $\eta = 1$, são mantidos constantes. Na parte inferior da figura, estão as correspondentes flutuações das médias do parâmetro de ordem $\hat{\alpha}$. A posição do pico, i.e. a transição de fase, depende de s estar aumentando ou diminuindo.

inicial. Com a mudança de s , a partir de (a), o sistema continua no primeiro mínimo; o sistema continua preso aí, mesmo quando esse mínimo não é mais o mínimo global. Apenas com o desaparecimento completo do mínimo original (em (f)), a dinâmica converge completamente para o mínimo global. Quando o valor de s retrocede sua trajetória (de (f) para (a)), o sistema permanece no segundo mínimo até que este mínimo desapareça por completo. Dessa forma, o sistema se mantém em um dado mínimo, i.e. numa determinada fase, de acordo com o seu estado anterior. Isso explica a “memória” que os sistemas probabilísticos guardam; tem-se aí a histerese. A histerese implica que o sistema, uma vez que a transição ocorreu, não retorne ao estado anterior, quando se realizam pequenas alterações de volta no parâmetro

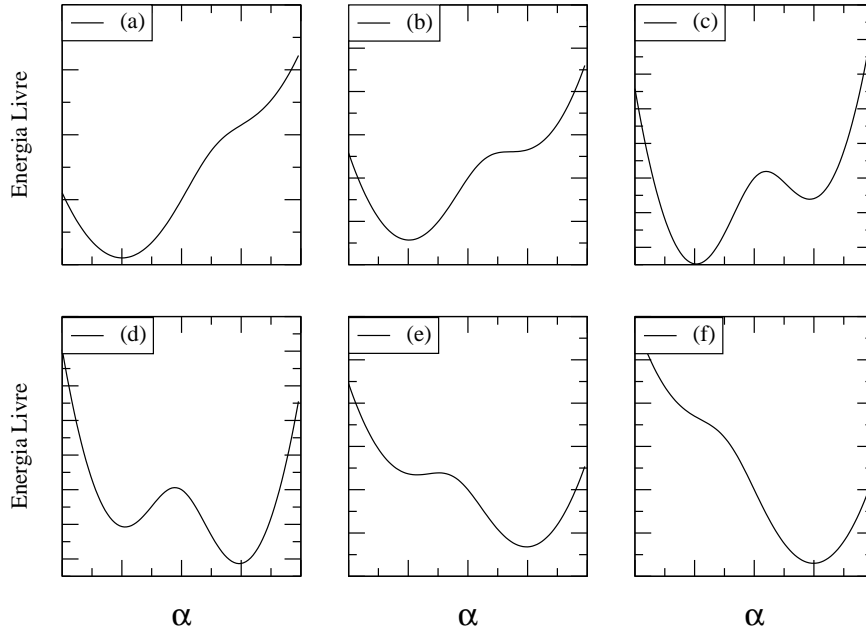


Figura 6.2: Energia livre de um sistema probabilístico qualquer em função do parâmetro de ordem α para diferentes valores do parâmetro de controle s . Como a dinâmica de minimização da energia livre depende de s , então mudanças nesse parâmetro podem levar a mudanças abruptas ou contínuas na posição do mínimo dessa função. Cada mínimo da energia livre está associado a uma fase acessível ao sistema. Aqui, tem-se um exemplo de uma transição abrupta, ou seja, uma transição de primeira ordem. De (a) para (f) s está crescendo. Em (a), existe apenas um mínimo. O sistema pode permanecer nesse mínimo (ou fase) até o valor de s que está na figura (e). Em (f), o primeiro mínimo desaparece e apenas o segundo mínimo existe. Caso o parâmetro s seja alterado de volta, a partir do estado de (f), o sistema pode continuar nesse segundo mínimo até o valor de s que está na figura (b). Apenas em (a), o sistema vai reaver o primeiro mínimo novamente. O sistema apresenta histerese, pois a população pode ser encontrada em duas fases distintas para um dado valor de s .

de controle. Transições de fase que apresentam histerese proporcionam mudanças descontínuas (transição de primeira ordem) no parâmetro de ordem. As transições contínuas (de segunda ordem) falham em exibir histerese.

Com essa descrição, pode-se estabelecer uma analogia entre o modelo para emergência de preferência sexual e a Mecânica Estatística da água. No contexto da transição líquido-gás da água, o parâmetro de ordem α está associado a sua densidade, en-

quanto que o parâmetro de controle s ao o inverso da temperatura. À medida que se varia a temperatura, a água pode ir de uma fase líquida, caracterizada por uma densidade alta, para uma fase gasosa, com baixa densidade. A fase em que o sistema se encontra depende do processo de minimização da função energia livre, cuja forma está diretamente relacionada com a temperatura. No modelo de especiação simpátrica, a dinâmica da população leva à minimização dessa energia livre com respeito à preferência sexual $\hat{\alpha}$. A população converge para um valor de $\hat{\alpha}$ que minimize essa energia livre. Além disso, a forma dessa energia livre é diretamente dependente da deficiência s do heterozigoto. O processo de minimização pode levar o sistema tanto para uma fase de acasalamento pan-mítico, onde $\hat{\alpha} \neq 0$ minimiza a função energia livre, quanto para a fase Acasalamento Preferencial, cuja minimização acontece em $\hat{\alpha} = 0$.

6.3 Balanço Energia-Entropia

É interessante descrever uma transição de fase, ou mais especificamente, a forma da função energia livre, como uma competição entre dois efeitos: *Energia* e *Entropia*. Nesse modelo para evolução de preferência sexual, o *fitness* age como o termo da energia; a entropia é dita pelo logaritmo do número de genótipos possíveis, como consequência da definição original de entropia dada por Boltzmann - logaritmo do número de estados acessíveis. A parte entrópica é constante, mas a parte da energia (*fitness*) aumenta com s . Para s suficientemente grande, a parte da energia domina e sua minimização acontece, quando os híbridos são eliminados; isso proporciona emergência de preferência sexual para evitar acasalamento entre indivíduos de diferentes genótipos. Para s suficientemente pequeno, o termo entrópico vence e, como consequência, os híbridos permanecem. Pode-se dizer que a população repousa na fase ($\hat{\alpha}$) associada ao efeito (energia ou entropia) dominante. Para s pequeno, o número de formas possíveis de se gerar híbridos (entropia) domina sobre a deficiência do híbrido - essa é a fase tipo HW. Mas isso não acontece para todos os valores de s e, eventualmente, quando essa deficiência passa a dominar sobre o número de formas de se gerar híbridos a transição de fases ocorre.

Quando se mantém a água líquida sobre o fogo, as bolhas e a temperatura se mantêm fixa até que todo o líquido tenha se transformado em vapor. A energia livre que deve ser minimizada nesse sistema é $F = E - TS$, onde E é a energia, S a entropia e T a temperatura. Para baixas temperaturas (análogas a um s grande),

a minimização de F acontece, essencialmente, pela minimização de E (termo da energia vence sobre a entropia). Para temperaturas suficientemente altas, há a transição para a fase gasosa, obtida pela minimização da entropia. Isso significa que se vai de uma fase mais ordenada (a líquida), com baixa energia, para uma fase mais desordenada (a gasosa), com alta entropia. No sistema biológico, baixos valores de s significa que o sistema está em uma fase desordenada (acasalamento panmítico). Com s grande, o sistema se encontra em uma fase ordenada (acasalamento preferencial).

7

CONCLUSÕES

Foram reportados neste trabalho os resultados do estudo de um modelo para evolução de preferência sexual e da conseqüente Especiação Simpátrica. Nesse modelo, foi considerado que os indivíduos são compostos por dois loci. Um deles (\mathbb{A}) é de importância ecológica e com deficiência adaptativa do heterozigoto Aa . O outro (\mathbb{B}) está associado ao traço aparente e sem relevância ecológica.

Foram estudados neste modelo os diferentes estados possíveis à população, mediante mudanças na adaptabilidade do híbrido, na quantidade de recursos do ambiente e na taxa de recombinação. Verificou-se que, dependendo dos valores escolhidos para os parâmetros que modelam o ambiente, a emergência de preferência sexual pode ou não ocorrer, levando ao aparecimento de uma forte correlação entre os alelos dos dois loci. O acasalamento preferencial estabelece uma barreira sexual entre subpopulações, o que conduz à Especiação Simpátrica. Aqui, trabalha-se com a hipótese de que duas espécies serão distintas, quando o cruzamento entre elas não puder produzir híbridos viáveis.

Foram encontradas regiões muito bem definidas do espaço de parâmetros onde acontece essa emergência. Além disso, o modelo apresenta reversão de especiação quando o *fitness* do heterozigoto, controlado pelo parâmetro de seleção s , é restabelecido em valores maiores. Este processo apresenta histerese, isto é, o valor da deficiência heterozigota em que acontece a transição do acasalamento pan-mítico para o acasalamento preferencial não é o mesmo da transição reversa. Como foi discutido no capítulo anterior, a histerese está associada a transições de primeira ordem.

No curso das simulações, foram assumidas condições simétricas, como por exemplo, nichos de mesmo tamanho e mesmo *fitness* para diferentes homozigotos. A quebra dessa simetria pode mudar os resultados. Por exemplo, a fase de extinção

pode não ser simétrica nos dois alelos do locus A . Para pequenas mudanças nessa simetria, espera-se que as fronteiras entre as fases apresentadas mudem, mas a estrutura global deve permanecer qualitativamente similar. Uma questão relevante é se as condições para essa transição são robustas sob o aumento de realismo nas hipóteses do modelo. Por exemplo, não foram considerados explicitamente custos à escolha preferencial por parte das fêmeas; isso poderia reduzir a região de acasalamento preferencial no diagrama de fases. Porém, a aceitabilidade sexual das fêmeas inicialmente não apresenta diversidade. Inicialmente, as fêmeas aceitam qualquer macho e, dessa forma, ao menos indiretamente, um custo para uma escolha preferencial foi considerado. Outra complexidade que poderia ser introduzida seria o aumento do número de loci, como o que foi feito nos trabalhos [16, 17]. A contribuição da entropia aumenta com o aumento do número de loci presentes no modelo, pois aumenta também o número de configurações possíveis. Isso afeta as características do sistema, da mesma forma que o aumento da taxa de recombinação r também afeta - o aumento dessa taxa reduz a região de acasalamento preferencial no diagrama $s - \kappa$. Pela introdução da taxa de recombinação, foi encontrado o giro de histerese e o tamanho dessa histerese diminui à medida que essa taxa decresce a zero.

Foram considerados também alguns exemplos naturais, nos quais a especiação reversa, ou hibridização, podem estar ocorrendo. Esses são exemplos que podem ser candidatos a serem descritos pelo modelo aqui apresentado. No entanto, é bem provável que somente em laboratório se consiga autonomia suficiente dos parâmetros de controle de modo que o giro de histerese possa ser observado em uma população natural. Porém é um avanço importante saber que o reestabelecimento das condições iniciais pode implicar que o sistema não necessariamente retorne ao estado inicial.

Parte II

Algoritmos de Aprendizado e Inferência Aproximativa

8

INTRODUÇÃO

Os métodos de Inferência Bayesiana têm se tornado cada vez mais populares em muitas áreas de ciência e de tecnologia, em especial no campo de aprendizado em máquinas [49]. O propósito principal da Inferência Bayesiana, assim como qualquer teoria Estatística, é extrair informações contidas em conjuntos de observações. Realizar inferência significa incorporar incertezas e conhecimento *a priori* de forma sistemática por meio de algum tipo de modelo. Esse modelo, por sua vez, deve ser constituído de parâmetros adaptáveis aos dados observados, de modo a representar o conhecimento a respeito do “mundo” mediante as informações disponíveis. A adaptação desses parâmetros, por meio da observação, é conduzido por regras simples de probabilidade.

Em Estatística Bayesiana, todas as incertezas são expressas usando-se probabilidades¹. Atribuem-se probabilidades aos diferentes estados possíveis do mundo e essas probabilidades representam o grau de crença de que cada um desses estados ocorra. A forma com que essas probabilidades são construídas é feita a partir do modelo que se escolhe para representar o mundo. No contexto bayesiano, o modelo é constituído de duas partes [61]. A primeira parte consiste na probabilidade *a priori* dos parâmetros do modelo. Essa probabilidade informa o grau de conhecimento prévio, anterior às observações e oriunda de um alguma outra fonte. A segunda parte, a *Verossimilhança*, é a probabilidade de que uma dada configuração dos parâmetros seja coerente com os dados observados. Pela regra de Bayes, pode-se determinar a probabilidade *a posteriori* dos parâmetros; esta probabilidade resume a crença (ou conhecimento) a respeito do mundo, mediante as informações disponíveis.

¹Probabilidades codificam incerteza tanto na estatística bayesiana, quanto na estatística clássica. A diferença é o significado de incerteza em uma (crença na ocorrência de um evento) e na outra (frequência assintótica de ocorrência de um evento).

Entretanto, para muitos modelos realísticos, inferências exatas são computacionalmente intratáveis. Isso se deve ao fato de que a probabilidade a posteriori é uma distribuição multivariada complicada, que precisa ser marginalizada para calcular valores esperados com respeito a essa distribuição. O processo de marginalização envolve a soma (no caso discreto) sobre todos os estados possíveis, o que, teoricamente, é possível realizar; porém, na prática, existe um número exponencial de estados que tornam o cálculo exato da marginalização proibitivamente caro [41].

Métodos de Inferência Aproximativa

O custo computacional da Inferência Bayesiana tem feito dos métodos de aproximação, chamados também de *Métodos de Inferência Aproximativa*, uma ferramenta bastante útil para lidar com o problema. Uma aproximação pode ser entendida como a substituição de uma média sobre a distribuição original, intratável, por outra tratável. Um exemplo bem simples e conhecido de aproximação é o *Método de Laplace*, que consiste numa aproximação Gaussiana para o ponto máximo de uma distribuição.

Uma classe de métodos de aproximação inclui os chamados *métodos determinísticos*. Esses métodos consistem, basicamente, da transferência do problema da marginalização Bayesiana para um problema de otimização [41, 42, 43]. A controvérsia desses métodos é que, embora sejam precisos e tratáveis, podem não promover convergência ou mesmo produzir resultados pouco confiáveis, necessitando de uma checagem final. Como exemplo desses métodos, pode-se citar o método sequencial *Assumed-density filtering* [44] e os métodos de *Sampling* e *Kalman filtering*, que utilizam refinamentos iterativos para a distribuição aproximativa [49]. Existe também uma família de métodos de aproximação determinística que inclui os métodos de *Inferência Variacional* [49, 50]. Entre esses métodos, destacam-se o *Expectation Consistent*, as aproximações *Bethe-Kikuchi* [51] e o método que será abordado no presente trabalho: o *Propagação de Expectativas*².

Outra classe de modelos de aproximação inclui os chamados *métodos estocásticos*. Esses métodos promovem estimativas exatas no limite de número grande de amostras. São métodos, entre os quais se inclui o algoritmo Markov Chain Monte Carlo (MCMC) [49, 46]; com a característica de serem, normalmente, bem mais lentos que os métodos determinísticos[41].

²Do Inglês *Expectation Propagation*.

O algoritmo Propagação de Expectativas

O algoritmo *Propagação de Expectativas* (PE) [52] tem sido considerado como um método aproximado de larga gama de aplicabilidade, amiúde mostrando um melhor desempenho que outros métodos. O PE é baseado numa dinâmica em que seus fatores, cujo produto forma uma distribuição aproximativa global e tratável, são iterativamente otimizados. Essa otimização acontece localmente pela minimização da divergência Kulback-Leibler entre a distribuição apropriadamente definida.

O PE é sempre escolhido para ser utilizado em Inferência Bayesiana aproximada, quando se necessita de eficiência e precisão. O PE apresenta maior acurácia que os algoritmos de aproximação existentes e com custo computacional similar [52]. Esse julgamento tem sido baseado, até agora, em estudos empíricos em conjuntos de dados específicos. Um exemplo é a comparação entre PE e MCMC para processos gaussianos apresentada em [46]. Nesse trabalho, mostrou-se que o PE funciona melhor em uma grande variedade de conjuntos de dados, tanto em relação à distribuição preditiva, quanto em se tratando da computação das marginais.

No entanto, embora esses estudos empíricos sejam de grande valor, não garantem que o PE funcione bem para outros conjuntos de dados ou outras classes de problemas. Na verdade, ainda não está claro como distinguir problemas em que esse tipo de aproximação é confiável dos problemas nos quais esse método pode levar a predições não precisas.

O Propagação de Expectativas como Algoritmo de Aprendizado

A proposta deste trabalho é a construção de um algoritmo de aprendizado supervisionado em Redes Neurais que seja genérico, robusto e preciso, mas sem um caro custo computacional. A eficiência descrita pela literatura em relação ao PE motivou a escolha dessa técnica para a construção deste algoritmo.

No aprendizado supervisionado, as regras são fixadas por um *Perceptron Professor* que emite uma resposta discreta ou linear mediante estímulos externos. Por meio do algoritmo PE proposto, deve ser possível inferir os parâmetros que compõem esse professor a partir dos dados disponíveis. Essa estimativa será feita utilizando-se uma distribuição a posteriori original, geralmente intratável, por uma outra, pertencente a uma classe de distribuições tratáveis. A distribuição aproximativa deve ser atualizada passo a passo, por intermédio das técnicas do PE. Esse procedimento

resultará, como se verá, em um algoritmo bastante eficiente e que apresentou desempenho ótimo em dois modelos estudados: o modelo do *Perceptron Gaussiano* e do *Perceptron Binário*. Os detalhes da formulação desse algoritmo, bem como a apresentação dos resultados obtidos, serão discutidos a partir da próxima seção.

9

INFERÊNCIA APROXIMATIVA E ALGORITMOS DE APRENDIZADO

Será abordado neste capítulo os elementos teóricos necessários para a elaboração de um algoritmo de aprendizado gerado pelo *Propagação de Expectativas* (PE). Ele está organizado da seguinte forma. Na seção (9.1), serão apresentadas com detalhes os objetivos da Inferência Bayesiana. Entre estas, a distribuição *a priori*, a *verossimilhança*, a distribuição *a posteriori*, a *previsibilidade*, etc. . Nessa seção, será descrita também a formulação dessa estatística para o contexto do *Aprendizado Supervisionado*. O capítulo será finalizado na seção seguinte (9.2), com a descrição do algoritmo PE.

9.1 O Aprendizado Bayesiano

Considere-se disponível um conjunto composto por P dados, representado por $\mathcal{D} = \{\mathbf{x}^\mu\}_{\mu=1\dots P}$. Esses dados são gerados por uma função estocástica chamada de *Professor* no contexto de Redes Neurais Artificiais [54] e que possui parâmetros contidos no *vetor professor* $\mathbf{w} \in R^N$. A geração dos dados a partir do professor pode ser quantificada pela probabilidade

$$p(\mathbf{x}^\mu | \mathbf{w}), \tag{9.1}$$

lida como “*probabilidade de que seja gerado o dado \mathbf{x}^μ , dado o vetor professor \mathbf{w}* ”. Essa probabilidade caracteriza a verossimilhança no contexto bayesiano. A verossimilhança, que também pode ser referida como o modelo escolhido para a representação dos dados, quantifica a relação entre os dados e a saída gerada pelo professor.

Considerando que os dados são gerados independentemente (i.i.d.), pode-se escrever a verossimilhança do conjunto dos dados através do produto

$$p(\mathcal{D}|\mathbf{w}) = \prod_{\mu=1}^P p(\mathbf{x}^\mu|\mathbf{w}). \quad (9.2)$$

Em um problema prático, no entanto, não se tem acesso ao professor e, desta forma, construir-se-á um algoritmo de aprendizado que possa estimá-lo a partir dos dados e com isso se tornará possível realizar predições em novos exemplos. Para isso, interessa a distribuição $p(\mathbf{w}|\mathcal{D})$, chamada de *distribuição a posteriori* e que pode ser lida como “*probabilidade de que a função estocástica possua o vetor professor \mathbf{w} dado que se dispõe do conjunto de dados \mathcal{D}* ”. Em outras palavras o posteriori quantifica o conhecimento a respeito de \mathbf{w} após a observação de \mathcal{D} . Por meio do teorema de Bayes, pode-se escrever a distribuição posteriori como

$$p(\mathbf{w}|\mathcal{D}) = \frac{p_0(\mathbf{w})p(\mathcal{D}|\mathbf{w})}{p(\mathcal{D})} \quad (9.3)$$

onde $p_0(\mathbf{w})$, chamado de *distribuição a priori*, contém toda informação prévia disponível a respeito do vetor professor e $p(\mathcal{D})$, obtido pela normalização, é a *evidência* do modelo em questão. Considerando a verossimilhança (9.2), o posteriori também pode ser escrito na forma (que será útil quando se introduzir o algoritmo PE):

$$p(\mathbf{w}|\mathcal{D}) = \frac{1}{p(\mathcal{D})} \prod_{\mu=0}^P F_\mu(\mathbf{w}), \quad (9.4)$$

onde se define

$$\begin{cases} F_0(\mathbf{w}) \equiv p_0(\mathbf{w}) \\ F_\mu(\mathbf{w}) \equiv p(\mathbf{x}^\mu|\mathbf{w}) \quad \text{se } \mu \neq 0 \end{cases}$$

Um esquema de todos os parâmetros e distribuições bayesianos envolvidos está apresentado na tabela (9.3).

Os algoritmos de aprendizado, dos quais se tratará aqui, pertencem a duas categorias: os algoritmos *on-line* e *off-line* [54]. Os algoritmos *off-line* são caracterizados pela apresentação de todos os P dados disponíveis de uma só vez em cada passo de atualização. No entanto, em algumas situações, tem-se apenas acesso momentâneo

Conjunto de Treinamento	$\mathcal{D} = \{\mathbf{x}^\mu\}_{\mu=1..P}$
Verossimilhança	$p(x^\mu \mathbf{w}) \equiv F_\mu(\mathbf{w})$
Verossimilhança do conjunto de dados	$p(\mathcal{D} \mathbf{w})$
Evidência	$p(\mathcal{D})$
Priori	$p_0(\mathbf{w})$
Posteriori	$p(\mathbf{w} \mathcal{D})$

Tabela 9.3: Resumo de notações.

a um certo dado. Essa situação leva aos algoritmos *on-line*, que usam cada dado disponível uma única vez para a atualização. Geralmente, os algoritmos *off-line* são bastante eficientes, mas pagam o preço por serem mais lentos. Os algoritmos *on-line*, por sua vez, apresentam resultados bastante satisfatórios em comparação com os algoritmos *off-line*.

A questão que surge no cenário *on-line* é: como se muda o conhecimento sobre o professor após a observação de um novo dado \mathbf{x}^{t+1} ? Para responder a essa pergunta suponha-se, primeiramente, que se pretenda aproximar o posteriori $p(\mathbf{w}|\mathcal{D}_t)$, onde \mathcal{D}_t é o conjunto de todos os dados obtidos até a apresentação do t -ésimo dado, pela aproximação $p(\mathbf{w}|\mathbf{A}_t)$, onde \mathbf{A}_t é o conjunto de parâmetros pertencente ao espaço \mathcal{A} , que caracteriza essa distribuição aproximativa (por exemplo, os primeiros momentos de \mathbf{w} se $p(\mathbf{w}|\mathbf{A})$ é uma distribuição gaussiana). Nesse instante, observa-se o dado \mathbf{x}^{t+1} , dando origem à nova distribuição a posteriori

$$p(\mathbf{w}|\mathbf{A}_t, \mathbf{x}^{t+1}) = \frac{p(\mathbf{w}|\mathbf{A}_t)p(\mathbf{x}^{t+1}|\mathbf{w})}{\int d\mathbf{w}p(\mathbf{w}|\mathbf{A}_t)p(\mathbf{x}^{t+1}|\mathbf{w})}, \quad (9.5)$$

que considera como distribuição a priori a própria distribuição posteriori anterior $p(\mathbf{w}|\mathbf{A}_t)$, uma vez que esta resume todo o conhecimento prévio disponível a respeito do professor. A distribuição posteriori atualizada não pertence necessariamente ao espaço \mathcal{A} e, portanto, deve-se escolher uma distribuição aproximativa que pertença a \mathcal{A} e que mais se aproxime da posteriori original, indo de-para:

$$p(\mathbf{w}|\mathbf{A}_t, \mathbf{x}^{t+1}) \rightarrow p(\mathbf{w}|\mathbf{A}_{t+1}). \quad (9.6)$$

Isso significa que se deve escolher uma distribuição aproximativa que minimize a perda de informação causada pela substituição da distribuição original.

Uma forma de se determinar essa distribuição aproximativa (e que se adotará aqui) é escolhendo a distribuição que minimiza a divergência Kullback-Leibler (K-L)

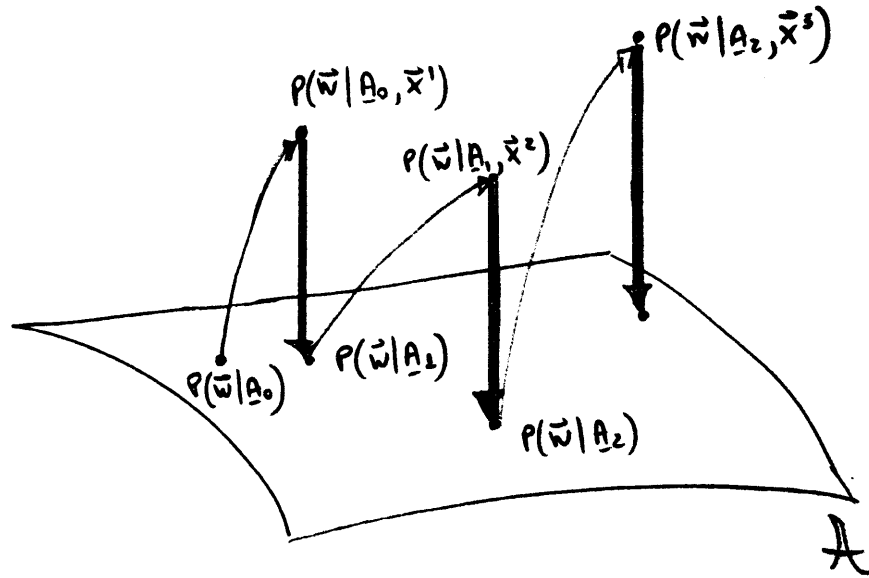


Figura 9.1: Esquema para a atualização da distribuição posteriori via algoritmo bayesiano *on-line*. A posteriori inicia com $p(\mathbf{w}|\mathbf{A}_0)$, escolhido de forma aleatória mas com o vínculo de pertencer ao espaço \mathcal{A} ; a apresentação do dado \mathbf{x}^1 leva à posteriori $p(\mathbf{w}|\mathbf{A}_0, \mathbf{x}^1)$, que não necessariamente pertence ao espaço \mathcal{A} . Em seguida, projeta-se essa distribuição no espaço \mathcal{A} para se obter a posteriori aproximada $p(\mathbf{w}|\mathbf{A}_1)$. Esse processo é feito iterativamente com todos os P dados disponíveis.

¹ com a distribuição original $p(\mathbf{w}|\mathbf{A}_t, \mathbf{x}^{t+1})$. Esse processo, quando repetido a cada novo dado adquirido, recebe o nome de *algoritmo bayesiano on-line* [54] e pode ser visualizado na figura (9.1). Para que esse algoritmo seja consistente, é necessário que a distribuição inicial $p(\mathbf{w}|\mathbf{A}_0)$ não exclua nenhum dos valores de \mathbf{w} pelo qual $p(\mathbf{w})$ seja não nula. Essa condição é satisfeita pela escolha $p(\mathbf{w}|\mathbf{A}_0) = p(\mathbf{w})$.

Aprendizado Supervisionado

Nesse trabalho, será abordada uma classe de problemas de aprendizado chamada de *Aprendizado Supervisionado*. Nesse tipo de aprendizado o μ -ésimo dado disponível é composto por um vetor de entrada $\mathbf{s}^\mu \in \mathbb{R}^N$ (o *input*) e uma etiqueta τ^μ associada a este vetor (o *output*). Usando a representação introduzida anteriormente, temos $\mathbf{x}^\mu = (\tau^\mu, \mathbf{s}^\mu)^T$ e o conjunto de treinamento $\mathcal{D} = \{(\tau^\mu, \mathbf{s}^\mu)^T\}$.

O Aprendizado Supervisionado abrange dois tipos de problemas: o problema de

¹Detalhes são apresentados no apêndice B.

classificação e o de regressão. Em um problema típico de classificação, o parâmetro τ assume valores inteiros, enquanto que, em um problema de regressão, τ assume valores contínuos. Como exemplo canônico para o aprendizado supervisionado, pode-se citar o caso em que uma rede neural (ou *perceptron*) professor com parâmetros \mathbf{w} classifica os vetores de entrada em dois grupos, representados pelo *output* binário $\tau = \pm 1$ [54]. Esse *perceptron* pode, por exemplo, responder aos *input*'s de acordo com a regra

$$\tau = \text{sign} \left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}} \right). \quad (9.7)$$

Esse *output* pode ser, por exemplo, o diagnóstico médico sobre a presença ou não de câncer em um paciente, mediante os resultados de imagens de raio-X. Nesse caso, o vetor \mathbf{s} representa as intensidades dos *pixels* na imagem e o *perceptron* (com configurações internas \mathbf{w}) representa a máquina que fará o processamento da imagem (\mathbf{s}) e o consequente diagnóstico. Esse diagnóstico será $+1$, presença de câncer, ou -1 , paciente saudável.

Assumindo que o *input* \mathbf{s} é uniforme e independente de \mathbf{w} , i.e. $p(\mathbf{s}|\mathbf{w}) = p(\mathbf{s}) = \text{cte}$, pode-se escrever $p((\tau, \mathbf{s})|\mathbf{w}) = p(\tau|\mathbf{s}, \mathbf{w})p(\mathbf{s}) \propto p(\tau|\mathbf{s}, \mathbf{w})$. Considerando a regra (9.7), pode-se escrever a verossimilhança

$$p(\tau|\mathbf{w}, \mathbf{s}) = \Theta \left(\tau \frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}} \right) \quad (9.8)$$

que significa “*probabilidade de que o perceptron professor com parâmetros \mathbf{w} retorne o output τ dado que se aplica o input \mathbf{s}* ”. A função $\Theta(\dots)$ é a *função de Heaviside*². Pela estrutura dessa verossimilhança, pode-se notar que o *perceptron* professor possui probabilidade 1 de gerar o mesmo sinal que o do campo $h = \frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}$, recuperando a regra (9.7). Nesse tipo de problema, deseja-se inferir os parâmetros ocultos \mathbf{w} por meio do cálculo da posteriori, mediante o conjunto de treinamento \mathcal{D} .

Previsibilidade

Uma vez computado a distribuição posteriori de um dado problema, pode-se utilizá-la, por exemplo, para computar a chamada *Probabilidade de Predição* ou *Previsibilidade*. A previsibilidade é denotada por $p(\tau|\mathbf{s}, \mathcal{D})$, pois representa a probabilidade que o professor gere a saída τ dado o *input* \mathbf{s} e que tenhamos disponível um conjunto

²Algumas funções matemáticas utilizadas ao longo do texto principal estão definidas no apêndice B.

de treinamento \mathcal{D} . Ela é computada por meio do valor esperado da verossimilhança em relação à distribuição posteriori, i.e.

$$p(\tau|\mathbf{s}, \mathcal{D}) = \int d\mathbf{w} p(\tau|\mathbf{s}, \mathbf{w}) p(\mathbf{w}|\mathcal{D}) \equiv \langle p(\tau|\mathbf{s}, \mathbf{w}) \rangle. \quad (9.9)$$

Aqui, introduz-se a notação $\langle \dots \rangle$ para o valor esperado em relação à distribuição posteriori.

Ao longo deste trabalho, lidaremos com uma classe de modelo denominado na literatura de redes neurais artificiais por *perceptron linear* [59] [54]. Nesse tipo de modelo, a verossimilhança depende do *input* e do vetor professor apenas por meio do campo $h = \frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}$, e portanto podemos reescrever a verossimilhança como $p(\tau|\mathbf{w}, \mathbf{s}) = p(\tau|h)$. Dessa forma, para essa classe de modelos, a previsibilidade (9.9) pode ser reescrita por intermédio de uma mudança de variável $\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}} \rightarrow h$, como

$$p(\tau|\mathbf{s}, \mathcal{D}) = \int dh p(\tau|h) p(h|\mathbf{s}, \mathcal{D}) = \langle p(\tau|h) \rangle. \quad (9.10)$$

Para o caso particular de uma distribuição posteriori Gaussiana, a atualização *on-line* da estimativa do vetor-professor $\langle \mathbf{w} \rangle$ pode ser computada a partir da previsibilidade (9.10). O trabalho [62] parte da propriedade

$$\langle z f(z) \rangle = \langle f'(z) \rangle \langle z^2 \rangle, \quad (9.11)$$

de uma função $f(z)$ bem comportada, para mostrar que

$$\langle w_i \rangle^{new} = \langle w_i \rangle + \sum_j M_{ij} \frac{\partial}{\partial \langle w_j \rangle} \log \langle p(\tau|\mathbf{s}, \mathbf{w}) \rangle \quad (9.12)$$

$$M_{ij}^{new} = M_{ij} + \sum_{k,l} M_{ik} M_{lj} \frac{\partial^2}{\partial \langle w_k \rangle \partial \langle w_l \rangle} \log \langle p(\tau|\mathbf{s}, \mathbf{w}) \rangle \quad (9.13)$$

onde

$$M_{ij} = \langle w_i w_j \rangle - \langle w_i \rangle \langle w_j \rangle \quad (9.14)$$

é o elemento da matriz covariância \mathbf{M} e $\langle \dots \rangle^{new}$ a média sobre a distribuição posteriori atualizada. Escrevendo

$$\frac{\partial(\dots)}{\partial \langle \mathbf{w} \rangle} = \frac{\partial(\dots)}{\partial \langle h \rangle} \frac{\partial \langle h \rangle}{\partial \langle \mathbf{w} \rangle} = \frac{1}{\sqrt{N}} \mathbf{s} \frac{\partial(\dots)}{\partial \langle h \rangle} \quad (9.15)$$

as equações (9.12) e (9.13) podem ser reescritas como

$$\langle \mathbf{w} \rangle^{new} = \langle \mathbf{w} \rangle + \frac{1}{\sqrt{N}} (\mathbf{M} \mathbf{s}) \frac{\partial}{\partial \langle h \rangle} \log \langle p(\tau|h) \rangle \quad (9.16)$$

$$\mathbf{M}^{new} = \mathbf{M} + \frac{1}{N} (\mathbf{M} \mathbf{s} \mathbf{s}^T \mathbf{M}) \frac{\partial^2}{\partial \langle h \rangle^2} \log \langle p(\tau|h) \rangle. \quad (9.17)$$

Esse resultado mostra que se considerarmos uma posteriori gaussiana, pode-se então computar as atualizações por meio da derivada da previsibilidade.

9.2 Propagação de Expectativas

Em grande parte dos problemas práticos, em especial nos problemas que serão abordados neste trabalho, tanto a marginalização $p(\mathcal{D})$ quanto a posteriori $p(\mathbf{w}|\mathcal{D})$ são intratáveis e, dessa forma, algum tipo de aproximação é necessário. O Algoritmo Propagação de Expectativas (PE) vai nessa direção, utilizando da minimização da divergência Kullback-Leibler (KL) para ir de uma distribuição intratável para uma *distribuição aproximativa* tratável. Aqui distribuição intratável é aquela em que a computação de seus valores esperados não pode ser realizada, seja pela alta dimensionalidade das variáveis ocultas ou pela alta complexidade dessa distribuição.

A distribuição aproximativa é geralmente escrita pelo produto

$$q(\mathbf{w}) = \prod_{\mu} g_{\mu}(\mathbf{w}), \quad (9.18)$$

onde o termo $g_{\mu}(\mathbf{w})$ pertencente à família de distribuições tratáveis [48] [49]. Dessa forma, se a função que se deseja aproximar é a posteriori com a forma apresentada em (9.4), o termo $g_{\mu}(\mathbf{w})$ corresponde (ou está associado) ao termo $F_{\mu}(\mathbf{w})$ da distribuição original.

Para se obter um algoritmo prático, vai-se restringir g_{μ} à *família exponencial* (Veja apêndice D). Com esse argumento, a distribuição aproximativa $q(\mathbf{w})$ também pertence a essa família, podendo ser descrita por um conjunto finito de “estatísticas suficientes” [49, 55]. Por exemplo, se os termos $g_{\mu}(\mathbf{w})$ são gaussianos, então a distribuição $q(\mathbf{w})$ também será gaussiana, descrita por apenas duas estatísticas suficientes: sua *média* e *variância*.

A ideia principal do PE é escolher os fatores $\{g_{\mu}(\mathbf{w})\}$ de tal forma que a distribuição aproximativa $q(\mathbf{w})$ seja tão próxima quanto possível da distribuição original $p(\mathbf{w}|\mathcal{D})$. Em outras palavras, será minimizada a perda de informação resultante

da substituição da distribuição verdadeira pela aproximada. Uma forma que se pode imaginar num primeiro momento para a minimização dessa perda de informação é escolher $q(\mathbf{w})$ de tal forma que haja a minimização da divergência KL $KL\left(p(\mathbf{w}|\mathcal{D})\parallel q(\mathbf{w})\right)$ entre a distribuição original e a aproximativa. No entanto, essa minimização envolve integrais sobre a distribuição original, ou seja, continua-se com o mesmo problema inicial (intratável!) [49]. O que se poderia fazer, então, seria minimizar a divergência KL entre os pares $F_\mu(\mathbf{w})$ e $g_\mu(\mathbf{w})$ em vez de minimizar a divergência entre $p(\mathbf{w}|\mathcal{D})$ e $q(\mathbf{w})$. Essa proposta foi introduzida por Thomas Minka em [52] e caracteriza o algoritmo PE ³.

O algoritmo PE começa pela inicialização dos fatores $\{g_\mu(\mathbf{w})\}$ que são atualizados iterativamente, um a cada passo. Suponha que se escolha o termo $g_\nu(\mathbf{w})$ para a atualização. Para isso, retira-se esse termo da distribuição aproximativa por meio da divisão

$$\frac{q(\mathbf{w})}{g_\nu(\mathbf{w})} = \prod_{\mu \neq \nu} g_\mu(\mathbf{w}). \quad (9.19)$$

Essa divisão é combinada com o fator original $F_\nu(\mathbf{w})$ dando origem à distribuição

$$Q_\nu(\mathbf{w}) = \frac{1}{Z_{Q_\nu}} \frac{q(\mathbf{w})}{g_\nu(\mathbf{w})} F_\nu(\mathbf{w}), \quad (9.20)$$

onde Z_{Q_ν} é uma constante de normalização. Em suma, a distribuição $Q_\nu(\mathbf{w})$ é construída, retirando-se o ν -ésimo termo da distribuição aproximativa e introduzindo em seu lugar a verossimilhança original $F_\nu(\mathbf{w})$.

A forma revisada do ν -ésimo termo da distribuição aproximativa, que se chamará aqui $g_\nu^{new}(\mathbf{w})$, é aquele que faz com que a “nova” distribuição aproximativa

$$q^{new}(\mathbf{w}) = g_\nu^{new}(\mathbf{w}) \prod_{\mu \neq \nu} g_\mu(\mathbf{w}) \quad (9.21)$$

seja tão próxima quanto possível da aproximação $Q_\nu(\mathbf{w})$. Esse novo termo é encontrado minimizando a divergência

$$KL\left(Q_\nu(\mathbf{w})\parallel q^{new}(\mathbf{w})\right) = KL\left(\frac{1}{Z_{Q_\nu}} \frac{q(\mathbf{w})}{g_\nu(\mathbf{w})} F_\nu(\mathbf{w})\parallel q^{new}(\mathbf{w})\right). \quad (9.22)$$

Essa minimização é relativamente fácil de ser obtida, uma vez que a distribuição aproximativa pertence à família exponencial. Nesse caso, de acordo com os resulta-

³Em muitos trabalhos da literatura, o algoritmo PE é chamado também de *Algoritmo de Minka*.

dos apresentados no apêndice D, a distribuição $q^{new}(\mathbf{w})$ que estamos interessados é aquela que possui as mesmas estatísticas suficientes que a distribuição $Q_\nu(\mathbf{w})$. Isso significa que

$$\langle \mathbf{u}(\mathbf{w}) \rangle_{q^{new}} = \langle \mathbf{u}(\mathbf{w}) \rangle_{Q_\mu}, \quad (9.23)$$

onde $\langle \dots \rangle_{q^{new}}$ e $\langle \dots \rangle_{Q_\mu}$ são os valores esperados em relação às duas distribuições em questão e $\mathbf{u}(\mathbf{w})$ uma função qualquer do vetor professor \mathbf{w} . Por exemplo, se a distribuição aproximativa é uma gaussiana, então a divergência (9.22) será minimizada se essa distribuição possui média e covariância idênticas às da distribuição $Q_\mu(\mathbf{w})$. Uma vez computado $q^{new}(\mathbf{w})$ pode-se determinar o fator atualizado $g_\nu^{new}(\mathbf{w})$ através da divisão

$$g_\nu^{new}(\mathbf{w}) = K \frac{q^{new}(\mathbf{w})}{\prod_{\mu \neq \nu} g_\mu} = K \frac{q^{new}(\mathbf{w})}{\left(\frac{q(\mathbf{w})}{g_\nu(\mathbf{w})} \right)}, \quad (9.24)$$

onde

$$K = Z_{Q_\nu}. \quad (9.25)$$

A igualdade (9.25) pode ser facilmente demonstrada, se considerarmos a propriedade (9.23), que permite escrever

$$\int d\mathbf{w} q^{new}(\mathbf{w}) = \int d\mathbf{w} Q_\nu(\mathbf{w}). \quad (9.26)$$

Seguindo as distribuições (9.20) e (9.24) conclui-se que

$$K = \int d\mathbf{w} g_\nu^{new}(\mathbf{w}) \frac{q(\mathbf{w})}{g_\nu(\mathbf{w})} = \int d\mathbf{w} \frac{q(\mathbf{w})}{g_\nu(\mathbf{w})} F_\nu(\mathbf{w}) = Z_{Q_\mu}. \quad (9.27)$$

Esse processo de atualização é repetido até que a convergência seja atingida. A ordem de apresentação dos termos para atualização não é relevante [49], o que permite que cada fator possa ser atualizado uma única vez (algoritmo *on-line*) ou mais de uma vez (algoritmo *off-line*). Um esquema prático do algoritmo PE pode ser visto a seguir.

Algoritmo Propagação de Expectativas

1. Inicializar todos os fatores $\{g_\nu(\mathbf{w})\}$;

2. Inicializar a distribuição aproximativa pelo produto $q(\mathbf{w}) = \prod_{\mu} g_{\mu}(\mathbf{w})$;
3. Até a convergência, repetir:
 - (a) Escolher um fator $g_{\nu}(\mathbf{w})$ para a atualização;
 - (b) Retirar este fator de $q(\mathbf{w})$ pela divisão $\frac{q(\mathbf{w})}{g_{\nu}(\mathbf{w})}$.
 - (c) Obter a distribuição mais realística $Q_{\nu}(\mathbf{w}) = \frac{1}{Z_{Q_{\nu}}} \frac{q(\mathbf{w})}{g_{\nu}(\mathbf{w})} F_{\nu}(\mathbf{w})$;
 - (d) A nova distribuição aproximativa $q^{new}(\mathbf{w})$ será aquela que possui os mesmos momentos que $Q_{\nu}(\mathbf{w})$;
 - (e) Computar o novo fator $g_{\nu}^{new}(\mathbf{w}) = K \frac{q^{new}(\mathbf{w})}{(g_{\nu}(\mathbf{w}))}$;
4. A aproximação para a distribuição posteriori original será

$$p(\mathbf{w}|\mathcal{D}) \approx \prod_{\mu} g_{\mu}(\mathbf{w}) \quad (9.28)$$

10

ALGORITMO DE APRENDIZADO VIA PROPAGAÇÃO DE EXPECTATIVAS

Será apresentada neste capítulo a formulação geral do aprendizado supervisionado obtido a partir do algoritmo Propagação de Expectativas. As aplicações desse algoritmo desenvolvido serão apresentadas nos próximos capítulos, onde se analisará o caso de um *Perceptron* Binário e um *Perceptron* Linear.

O capítulo está organizado da seguinte forma. O desenvolvimento da teoria inicia a partir da seção (10.1) onde se introduz as formas da distribuição aproximativa e da verossimilhança. O algoritmo encontrado para atualização dos parâmetros da distribuição aproximativa é descrito na seção (10.2). Na seção (10.3) será apresentado o cálculo da função de partição do algoritmo PE. Finalmente, na seção (10.4), será apresentada uma lista com todos os resultados obtidos e o consequente algoritmo de aprendizado via PE.

10.1 Desenvolvimento da Teoria

Consideraremos um *perceptron* professor que emite um *output* τ , dado um *input* \mathbf{s} , de acordo com uma certa função composta por N parâmetros ocultos guardados no vetor professor \mathbf{w} . Ao final de P *input*'s independentes aplicados ao professor teremos o conjunto de treinamento: $\mathcal{D} = \{(\tau^\mu, \mathbf{s}^\mu)^T\}_{\mu=1\dots P}$. Interessa aqui construir um algoritmo que possibilite estimar esses N parâmetros a partir dos dados \mathcal{D} .

Consideraremos também que a geração de \mathbf{s} é independente de \mathbf{w} (o que nos permite escrever $p((\tau, \mathbf{s})|\mathbf{w}) = p(\tau|\mathbf{w}, \mathbf{s})$). Além disso, a dependência da verossimilhança com os vetores \mathbf{w} e \mathbf{s} será por meio do campo $h = \frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}$ com a forma

$$p(\tau|\mathbf{w}, \mathbf{s}) = F\left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}\right). \quad (10.1)$$

Este tipo de verossimilhança (com dependência exclusiva com o campo h) caracteriza o *perceptron simples*[59]. A função $F(\dots)$ pode ser, por exemplo, a função de *Heaviside* para um problema de classificação ou uma gaussiana para um problema de regressão. Considerando ainda o conjunto de treinamento \mathcal{D} e a verossimilhança (10.1), pode-se escrever a distribuição a posteriori como

$$p(\mathbf{w}|\mathcal{D}) = \frac{1}{p(\mathcal{D})} p_0(\mathbf{w}) \prod_{\mu=1}^P F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right). \quad (10.2)$$

onde o a priori $p_0(\mathbf{w})$ pode ser, por exemplo, a gaussiana

$$p_0(\mathbf{w}) = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{\mathbf{w} \cdot \mathbf{w}}{2}} \quad (10.3)$$

ou a distribuição binária (ou de *Bernoulli*)

$$p_0(\mathbf{w}) = \prod_{i=1}^N \left[\frac{1}{2} \delta(w_i - 1) + \frac{1}{2} \delta(w_i + 1) \right]. \quad (10.4)$$

A dificuldade em lidar com o posteriori (10.2) se deve ao fato de que a função $F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right)$ é normalmente intratável, com interação entre os componentes $\{w_i\}$. Dessa forma algum tipo de aproximação pode ser bastante útil para a solução do problema. O que iremos realizar neste trabalho será substituir a verossimilhança $F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right)$ por uma distribuição aproximativa não interagente $g_\mu(\mathbf{w}) = \prod_{i=1}^N g_{i\mu}(w_i)$. Nesse sentido, a distribuição posteriori $p(\mathbf{w}|\mathcal{D})$, que se chamará de posteriori original, será substituída pela distribuição aproximativa não interagente

$$q(\mathbf{w}) = p_0(\mathbf{w}) \prod_{\mu=1}^P g_\mu(\mathbf{w}), \quad (10.5)$$

com o mesmo a priori da distribuição original. Consideraremos que a distribuição aproximativa para a verossimilhança pertence à família exponencial com a forma gaussiana

$$g_\mu(\mathbf{w}) \propto e^{-\frac{1}{2} \sum_{i=1}^N a_{i\mu} w_i^2 + \sum_{i=1}^N b_{i\mu} w_i} \quad (10.6)$$

construída a partir do conjunto de parâmetros $\{a_{i\mu}\}$ e $\{b_{i\mu}\}$. Essa forma gaussiana de $g_\mu(\mathbf{w})$ implica que a distribuição $q(\mathbf{w})$ também seja uma gaussiana com forma

$$q(\mathbf{w}) \propto p_0(\mathbf{w}) e^{-\frac{1}{2} \sum_{i=1}^N A_i w_i^2 + \sum_{i=1}^N B_i w_i} \quad (10.7)$$

onde são definidos os conjuntos de parâmetros

$$A_i \equiv \sum_{\mu=1}^P a_{i\mu}. \quad (10.8)$$

e

$$B_i \equiv \sum_{\mu=1}^P b_{i\mu}. \quad (10.9)$$

O que se irá descrever nas próximas seções é a elaboração de um algoritmo baseado nas idéias do PE para computar os componentes das matrizes $\mathbf{A} = \{a_{i\mu}\}$ e $\mathbf{B} = \{b_{i\mu}\}$, ambas $N \times P$. Uma vez computadas essas matrizes, pode-se determinar a distribuição aproximativa $q(\mathbf{w})$ por meio das distribuições (10.6) e (10.7) e, em seguida, estimar o vetor professor \mathbf{w} .

Estimando o vetor professor \mathbf{w}

Antes de apresentarmos a descrição do algoritmo de aprendizado, apresentar-se-á o modo como se pode estimar o vetor professor \mathbf{w} em termos dos componentes das matrizes \mathbf{A} e \mathbf{B} . O valor estimado $\langle \mathbf{w} \rangle_q$ do vetor professor é a sua média sobre a distribuição $q(\mathbf{w})$, que pela relação (10.7) será

$$\langle \mathbf{w} \rangle_q \propto \int \mathbf{w} p_0(\mathbf{w}) e^{-\frac{1}{2} \sum_{i=1}^N A_i w_i^2 + \sum_{i=1}^N B_i w_i} d\mathbf{w}. \quad (10.10)$$

Consequentemente, o valor estimado do i -ésimo parâmetro será

$$\langle w_i \rangle_q = \frac{\int w_i p_0(w_i) e^{-\frac{1}{2} A_i w_i^2 + B_i w_i} dw_i}{\int p_0(w_i) e^{-\frac{1}{2} A_i w_i^2 + B_i w_i} dw_i}. \quad (10.11)$$

Para o caso do a priori gaussiano (10.3), por exemplo, essa média será

$$\langle w_i \rangle_q = \frac{B_i}{1 + A_i}. \quad (10.12)$$

Por outro lado, com o a priori binário (10.4) a média será

$$\langle w_i \rangle_q = \tanh(B_i). \quad (10.13)$$

Note-se que nesse último caso somente o termo linear do expoente de $q(\mathbf{w})$ será necessário para estimar as variáveis ocultas.

Propagação de Expectativas

Seguiremos nesta seção as ideias introduzidas na seção (9.2) para realizar atualizações nos termos $\{g_\mu(\mathbf{w})\}$ via PE. Primeiramente, precisa-se encontrar a nova distribuição $q^{new}(\mathbf{w})$, restrita à família exponencial e com a mesma forma de (10.7), que seja tão próxima quanto possível da aproximação

$$Q_\mu(\mathbf{w}) = \frac{1}{Z_{Q_\mu}} \frac{q(\mathbf{w})}{g_\mu(\mathbf{w})} F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right), \quad (10.14)$$

onde Z_{Q_μ} é encontrado pela normalização. Como já se discutiu anteriormente, “tão próximo quanto possível” significa minimização de perda de informações pela troca de $Q_\mu(\mathbf{w})$ por $q^{new}(\mathbf{w})$. Isso pode ser obtido pela minimização da divergência

$$\text{KL}(Q_\mu(\mathbf{w}) || q^{new}(\mathbf{w})), \quad (10.15)$$

que caracteriza uma forma alternativa de aproximação tipo campo médio [63]. Essa minimização é relativamente fácil de ser computada, uma vez que a nova distribuição aproximativa pertence à família exponencial. Sendo assim, os resultados do apêndice D dizem que a versão atualizada $q^{new}(\mathbf{w})$ que minimiza (10.15) é aquela que obedece à igualdade

$$\langle \mathbf{u}(\mathbf{w}) \rangle_{q^{new}} = \langle \mathbf{u}(\mathbf{w}) \rangle_{Q_\mu}, \quad (10.16)$$

onde $\mathbf{u}(\mathbf{w})$ é uma função qualquer do vetor professor. Em outras palavras, a nova distribuição aproximativa é obtida pelo cálculo das “estatísticas suficientes” da distribuição mais realística $Q_\mu(\mathbf{w})$. Por exemplo, voltando ao *perceptron* binário, pode-se tomar o resultado (10.13) e determinar o conjunto atualizado dos parâmetros $\{B_i\}$ por meio da média de $\{w_i\}$ sobre a distribuição $Q_\mu(\mathbf{w})$, i.e. $B_i^{new} = \tanh^{-1}(\langle w_i \rangle_{Q_\mu})$.

O que precisa ser feito então para a determinação da nova distribuição aproximativa é o cálculo da média $\langle \mathbf{u}(\mathbf{w}) \rangle_{Q_\mu}$, o que será feito na próxima seção. Com esse cálculo, será possível contruir um algoritmo prático para a atualização dos parâmetros das matrizes \mathbf{A} e \mathbf{B} .

10.2 Atualização dos elementos das matrizes A e B

Para a construção de um algoritmo para atualização dos parâmetros da distribuição aproximativa, precisaremos, primeiramente, calcular a média

$$\langle \mathbf{u}(\mathbf{w}) \rangle_{Q_\mu} \propto \int d\mathbf{w} \mathbf{u}(\mathbf{w}) q^{\setminus \mu}(\mathbf{w}) F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right) \quad (10.17)$$

onde introduzimos na definição (10.14) a distribuição normalizada $q^{\setminus \mu}(\mathbf{w})$ gerada pela distribuição $q(\mathbf{w})$ sem seu μ -ésimo termo, calculado por

$$q^{\setminus \mu}(\mathbf{w}) = C_\mu \frac{q(\mathbf{w})}{g_\mu(\mathbf{w})} \propto e^{-\frac{1}{2} \sum_{i=1}^N (A_i - a_{i\mu}) w_i^2 + \sum_{i=1}^N (B_i - b_{i\mu}) w_i}, \quad (10.18)$$

onde C_μ é uma constante de normalização. Note-se que $q^{\setminus \mu}(\mathbf{w})$, uma distribuição com cavidade [51], possui uma forma não-interativa $q^{\setminus \mu}(\mathbf{w}) = \prod_{i=1}^N q_i^{\setminus \mu}(w_i)$.

Como a distribuição aproximativa é uma gaussiana, o cálculo da média (10.17) é transferida para o cálculo do n -ésimo momento de suas componentes, i.e. o cálculo da média

$$\langle w_i^n \rangle_{Q_\mu} \propto \int d\mathbf{w} w_i^n \left[\prod_{j=1}^N q_j^{\setminus \mu}(w_j) \right] F\left(\frac{1}{\sqrt{N}} \sum_{j=1}^N w_j s_j^\mu\right). \quad (10.19)$$

Para se ser capaz de realizar esse cálculo, precisamos introduzir o *argumento de cavidade* [54, 59, 51]. Primeiramente, consideraremos o *campo com cavidade* $h_{\mu \setminus i}$ que nada mais é que o campo gerado pelo exemplo \mathbf{s}^μ no professor sem a influência do seu componente i . Dessa forma, a verossimilhança $F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right)$ pode ser reescrita como:

$$F\left(\frac{1}{\sqrt{N}} \sum_{j=1}^N w_j s_j^\mu\right) = \int dh_{\mu \setminus i} \delta\left(h_{\mu \setminus i} - \frac{1}{\sqrt{N}} \sum_{j \neq i} w_j s_j^\mu\right) F\left(h_{\mu \setminus i} + \frac{w_i s_i^\mu}{\sqrt{N}}\right), \quad (10.20)$$

onde $\delta(\dots)$ é a função delta de Dirac [54].

Inserindo a relação (10.20) na (10.19), assumindo N grande e o *Teorema do Limite Central*, apresentado no apêndice C, chega-se a

$$\langle w_i^n \rangle_{Q_\mu} \propto \int dw_i w_i^n q_i^{\setminus \mu}(w_i) \left\langle F\left(h_{\mu \setminus i} + \frac{w_i s_i^\mu}{\sqrt{N}}\right) \right\rangle_{h_{\mu \setminus i}}, \quad (10.21)$$

onde $\langle \dots \rangle_{h_{\mu \setminus i}}$ representa uma média sobre uma distribuição gaussiana do campo $h_{\mu \setminus i}$, com média e variância

$$\bar{h}_{\mu \setminus i} = \frac{1}{\sqrt{N}} \sum_{j \neq i} s_j^\mu \langle w_j \rangle_{q_j^\mu} \quad (10.22)$$

e

$$\sigma_{\mu \setminus i}^2 = \frac{1}{N} \sum_{j \neq i} (s_j^\mu)^2 \left(\langle w_j^2 \rangle_{q_j^\mu} - \langle w_j \rangle_{q_j^\mu}^2 \right), \quad (10.23)$$

respectivamente. Note-se que em (10.22) e (10.23) precisa-se computar o primeiro e segundo momento das componentes de \mathbf{w} sobre a distribuição $q^\mu(\mathbf{w})$, dada por (10.18).

Para minimizar a divergência (10.15), é necessário que $\langle w_i^n \rangle_{q_i^{new}} = \langle w_i^n \rangle_{Q_\mu}$, e portanto

$$\int dw_i w_i^n q_i^{new}(w_i) \propto \int dw_i w_i^n q_i^\mu(w_i) \left\langle F \left(h_{\mu \setminus i} + \frac{w_i s_i^\mu}{\sqrt{N}} \right) \right\rangle_{h_{\mu \setminus i}} \quad (10.24)$$

e conseqüentemente

$$q_i^{new}(w_i) \propto q_i^\mu(w_i) \left\langle F \left(h_{\mu \setminus i} + \frac{w_i s_i^\mu}{\sqrt{N}} \right) \right\rangle_{h_{\mu \setminus i}}. \quad (10.25)$$

Usando (10.18), o resultado acima pode ser facilmente reescrito como

$$q_i^{new}(w_i) \propto e^{-\frac{1}{2}(A_i - a_{i\mu})w_i^2 + (B_i - b_{i\mu})w_i + \ln \left\langle F \left(h_{\mu \setminus i} + \frac{w_i s_i^\mu}{\sqrt{N}} \right) \right\rangle_{h_{\mu \setminus i}}}. \quad (10.26)$$

Expandindo $\ln \left\langle F(\dots) \right\rangle_{h_{\mu \setminus i}}$ até termos $\mathcal{O}(\frac{1}{N})$, ter-se-á:

$$\begin{aligned} \ln \left\langle F \left(h_{\mu \setminus i} + \frac{w_i s_i^\mu}{\sqrt{N}} \right) \right\rangle_{h_{\mu \setminus i}} &\approx \ln \left\langle F(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}} + \frac{\left\langle F'(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}}}{\left\langle F(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}}} \frac{w_i s_i^\mu}{\sqrt{N}} \\ &+ \frac{1}{2} \left[\frac{\left\langle F''(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}}}{\left\langle F(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}}} - \frac{\left\langle F'(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}}^2}{\left\langle F(h_{\mu \setminus i}) \right\rangle_{h_{\mu \setminus i}}^2} \right] \frac{(w_i s_i^\mu)^2}{N}. \end{aligned} \quad (10.27)$$

Dado que a distribuição aproximativa atualizada pode ser escrita como

$$q_i^{new}(w_i) \propto e^{-\frac{1}{2}A_i^{new}w_i^2 + B_i^{new}w_i} \quad (10.28)$$

e recorrendo aos resultados (10.26) e (10.27), chega-se aos seguintes algoritmos PE para a atualização dos parâmetros:

$$A_i^{new} = A_i - a_{i\mu} + a_{i\mu}^{new} \quad (10.29)$$

e

$$B_i^{new} = B_i - b_{i\mu} + b_{i\mu}^{new} \quad (10.30)$$

onde

$$a_{i\mu}^{new} = -\gamma_{\mu\setminus i} \frac{(s_i^\mu)^2}{N} \quad (10.31)$$

e

$$b_{i\mu}^{new} = \eta_{\mu\setminus i} \frac{s_i^\mu}{\sqrt{N}} \quad (10.32)$$

com

$$\eta_{\mu\setminus i} \equiv \frac{\langle F'(h_{\mu\setminus i}) \rangle_{h_{\mu\setminus i}}}{\langle F(h_{\mu\setminus i}) \rangle_{h_{\mu\setminus i}}} \quad (10.33)$$

e

$$\gamma_{\mu\setminus i} \equiv \left[\frac{\langle F''(h_{\mu\setminus i}) \rangle_{h_{\mu\setminus i}}}{\langle F(h_{\mu\setminus i}) \rangle_{h_{\mu\setminus i}}} - \frac{\langle F'(h_{\mu\setminus i}) \rangle_{h_{\mu\setminus i}}^2}{\langle F(h_{\mu\setminus i}) \rangle_{h_{\mu\setminus i}}^2} \right] \quad (10.34)$$

10.3 Cálculo da Função de Partição

De acordo com Minka [53] e os trabalhos [48][41], a função de partição do algoritmo PE é dada por

$$Z_{PE} = \prod_{\mu=1}^P Z_{Q_\mu} \quad (10.35)$$

onde

$$Z_{Q\mu} = C_\mu \int d\mathbf{w} q^{\setminus\mu}(\mathbf{w}) F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right) \quad (10.36)$$

é a função de partição introduzida em (10.14) e escrita em termos da definição (10.18).

O que se precisa fazer para o cálculo de $Z_{Q\mu}$ é determinar a média da verossimilhança sobre a distribuição $q^{\setminus\mu}(\mathbf{w})$ de forma análoga ao feito na seção anterior, mas agora sem utilizar argumentos de cavidade. Realizando esse cálculo pela mudança de variável $\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}} \rightarrow h_\mu$, chega-se a

$$Z_\mu = C_\mu \langle F(h_\mu) \rangle_{h_\mu} \quad (10.37)$$

onde $\langle \dots \rangle_{h_\mu}$ representa uma média sobre uma distribuição gaussiana do campo h_μ com média e variância

$$\bar{h}_\mu = \frac{1}{\sqrt{N}} \sum_{j=1} s_j^\mu \langle w_j \rangle_{q_j^{\setminus\mu}} \quad (10.38)$$

e

$$\sigma_\mu^2 = \frac{1}{N} \sum_{j=1} (s_j^\mu)^2 \left(\langle w_j^2 \rangle_{q_j^{\setminus\mu}} - \langle w_j \rangle_{q_j^{\setminus\mu}}^2 \right), \quad (10.39)$$

respectivamente. Finalmente chega-se à função de partição

$$Z_{PE} = \prod_{\mu=1}^P \left[C_\mu \langle F(h_\mu) \rangle_{h_\mu} \right]. \quad (10.40)$$

Por meio dessa função, pode-se computar a entropia do PE. Pela definição da entropia, em termos dos estados possíveis, tem-se

$$S_{PE} = \log Z_{PE} \quad (10.41)$$

o que permite concluir por (10.40) que

$$S_{PE} = \sum_{\mu=1}^P \log \langle F(h_\mu) \rangle_{h_\mu} + \sum_{\mu=1}^P \log C_\mu. \quad (10.42)$$

10.4 Resumo geral dos resultados

Concluindo, apresenta-se nesta seção os resultados obtidos para o algoritmo de aprendizado via PE :

- Verossimilhança:

$$P(\tau|\mathbf{w}, \mathbf{s}) = F\left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}\right) \quad (10.43)$$

- Posteriori:

$$p(\mathbf{w}|\mathcal{D}) \propto p_0(\mathbf{w}) \prod_{\mu=1}^P F\left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right) \quad (10.44)$$

- Distribuição aproximativa:

$$q(\mathbf{w}) = p_0(\mathbf{w}) \prod_{\mu=1}^P q_\mu(\mathbf{w}). \quad (10.45)$$

$$q(\mathbf{w}) \propto e^{\sum_{i=1}^N (-\frac{1}{2}A_i w_i^2 + B_i w_i)} \quad (10.46)$$

$$A_i \equiv \sum_{\mu=1}^P a_{i\mu}. \quad (10.47)$$

e

$$B_i \equiv \sum_{\mu=1}^P b_{i\mu}. \quad (10.48)$$

- Atualização

$$a_{i\mu}^{new} = -\gamma_{\mu \setminus i} \frac{(s_i^\mu)^2}{N} \quad (10.49)$$

e

$$b_{i\mu}^{new} = \eta_{\mu \setminus i} \frac{s_i^\mu}{\sqrt{N}} \quad (10.50)$$

com

$$\eta_{\mu \setminus i} \equiv \frac{\langle F'(h_{\mu \setminus i}) \rangle_{h_{\mu \setminus i}}}{\langle F(h_{\mu \setminus i}) \rangle_{h_{\mu \setminus i}}} \quad (10.51)$$

$$\gamma_{\mu \setminus i} \equiv \left[\frac{\langle F''(h_{\mu \setminus i}) \rangle_{h_{\mu \setminus i}}}{\langle F(h_{\mu \setminus i}) \rangle_{h_{\mu \setminus i}}} - \frac{\langle F'(h_{\mu \setminus i}) \rangle_{h_{\mu \setminus i}}^2}{\langle F(h_{\mu \setminus i}) \rangle_{h_{\mu \setminus i}}^2} \right] \quad (10.52)$$

- Média Gaussiana

$$\langle \dots \rangle_{h_{\mu \setminus i}} = \int \frac{1}{\sqrt{2\pi\sigma_{\mu \setminus i}^2}} e^{-\frac{(h_{\mu \setminus i} - \bar{h}_{\mu \setminus i})^2}{2\sigma_{\mu \setminus i}^2}} (\dots) dh_{\mu \setminus i} \quad (10.53)$$

onde

$$\bar{h}_{\mu \setminus i} = \frac{1}{\sqrt{N}} \sum_{j \neq i} s_j^\mu \langle w_j \rangle_{q_j^\mu} \quad (10.54)$$

e

$$\sigma_{\mu \setminus i}^2 = \frac{1}{N} \sum_{j \neq i} (s_j^\mu)^2 \left(\langle w_j^2 \rangle_{q_j^\mu} - \langle w_j \rangle_{q_j^\mu}^2 \right). \quad (10.55)$$

- Função de Partição do algoritmo PE:

$$Z_{PE} = \prod_{\mu=1}^P \left[C_\mu \langle F(h_\mu) \rangle_{h_\mu} \right] \quad (10.56)$$

- Entropia do algoritmo PE:

$$S_{PE} = \sum_{\mu=1}^P \log \langle F(h_\mu) \rangle_{h_\mu} + \sum_{\mu=1}^P \log C_\mu. \quad (10.57)$$

11

APLICAÇÕES

Neste capítulo serão descritos os resultados da aplicação do algoritmo de aprendizado via inferência aproximativa em dois modelos de *perceptrons*: o *Perceptron Binário* e o *Perceptron Gaussiano*. O objetivo, claro, é inferir o *perceptron* professor por meio dos dados disponíveis. Será verificado que o desempenho do método desenvolvido é ótimo em ambos os casos que serão abordados. Será verificado também o desempenho do algoritmo em uma versão *on-line*, isto é, no caso em que cada dado disponível é apresentado uma única vez para atualização, o desempenho desse algoritmo *on-line* é equiparável a outros algoritmos presentes na literatura.

O capítulo está dividido em duas partes. A primeira, descrito na seção (11.1), dedica-se ao estudo do Perceptron Binário. Aqui será descrito o desempenho do algoritmo PE neste problema no contexto *on-line* e *off-line*. A segunda parte, descrita na seção (11.2), serão apresentados os resultados das aplicações ao *Perceptron Gaussiano*. Da mesma forma que no problema anterior, serão analisados ambos os contextos, *on-line* e *off-line*.

11.1 O Perceptron Binário

O modelo do *Perceptron Binário* consiste em um *perceptron* professor que gera, mediante o *input* \mathbf{s} , uma resposta de acordo com a regra

$$\text{sign} \left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}} \right) = \pm 1. \quad (11.1)$$

Além disso, todos os vetores envolvidos apresentam componentes binárias (± 1) e, portanto, o vetor professor deve ser construído a partir da distribuição a priori (10.4).

Inferir-se-á o vetor professor \mathbf{w} por meio dos dados disponíveis, porém não se tem acesso direto à resposta (11.1) gerada pelo professor. Tem-se acesso apenas ao

output $\tau = \pm 1$ que é idêntico à resposta do professor a menos de ruído. Isso implica que há uma probabilidade κ de que o *output* τ e a resposta do professor tenham sinais opostos e probabilidade $1 - \kappa$ de que esses tenham sinais idênticos. Usando processo análogo à (9.8), pode-se escrever a verossimilhança

$$P(\tau|\mathbf{w}, \mathbf{s}) = F\left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}\right) = \kappa + (1 - 2\kappa)\Theta\left(\tau \frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}\right). \quad (11.2)$$

Ao final de P exemplos submetidos ao professor, tem-se o conjunto de pares de dados $\mathcal{D} = \{(\mathbf{s}^\mu, \tau^\mu)\}_{\mu=1, \dots, P}$.

Graças ao a priori binário, tem-se a constante $w_i^2 = 1$, para qualquer i . Isso implica que a correção $\mathcal{O}(\frac{1}{N})$ da expansão (10.27), que está correlacionada com os componentes da matriz \mathbf{A} via atualização (10.34), não dependa de w_i . Dessa forma, a estimativa do vetor professor por meio do algoritmo PE necessita apenas dos parâmetros da matriz \mathbf{B} . Esses parâmetros, por sua vez, estão relacionados com o termo $\mathcal{O}(\frac{1}{\sqrt{N}})$ da expansão (10.27) e suas atualizações devem obedecer ao algoritmo (10.30) com

$$b_{i\mu}^{new} = \eta_{\mu\setminus i} \frac{s_i^\mu}{\sqrt{N}}, \quad (11.3)$$

onde

$$\eta_{\mu\setminus i} = \frac{\frac{(1-2\kappa)}{\sqrt{2\pi\sigma_{\mu\setminus i}^2}} e^{-\frac{(\bar{h}_{\mu\setminus i})^2}{2\sigma_{\mu\setminus i}^2}}}{\kappa + \frac{1}{2}(1-2\kappa)\text{erfc}\left(-\frac{\tau^\mu \bar{h}_{\mu\setminus i}}{\sqrt{2\sigma_{\mu\setminus i}^2}}\right)} \tau^\mu, \quad (11.4)$$

obtido pelo resultado (10.33) aplicado à verossimilhança (11.2). Para computar $\eta_{\mu\setminus i}$, por meio de (11.4) precisa-se determinar a média $\bar{h}_{\mu\setminus i}$ e a variância $\sigma_{\mu\setminus i}^2$, que são obtidas por intermédio de (10.22) e (10.23), respectivamente. Note-se que, para esse cálculo, precisa-se dos dois primeiros momentos da distribuição (10.18), que para o modelo do *perceptron* binário são

$$\langle w_i \rangle_{q_i^\mu} = \tanh(B_i - b_{i\mu}) \quad (11.5)$$

e

$$\langle w_i^2 \rangle_{q_i^\mu} = 1. \quad (11.6)$$

Após aplicar esse algoritmo iterativamente e obter convergência dos parâmetros da matriz \mathbf{B} e conseqüentemente a convergência da distribuição aproximativa, pode-se estimar o vetor professor por meio de

$$\langle w_i \rangle_q = \tanh(B_i), \quad (11.7)$$

conforme as ideias discutidas na seção (10.1).

11.1.1 Versão *on-line*

Nesta seção, apresentar-se-ão os resultados de simulações computacionais desse algoritmo PE para o caso em que cada um dos P pares de dados $\{(\tau^\mu, \mathbf{s}^\mu)\}_\mu$ é apresentado uma única vez para atualização. Isso representa a versão *on-line* do algoritmo PE para o *perceptron* binário. Para inicializar a distribuição aproximativa como $q(\mathbf{w}) = p_0(\mathbf{w})$, devem-se fazer todos os N parâmetros $\{b_{i\mu}\}$ nulos. Dessa forma, a atualização (10.30) deve obedecer à recorrência

$$B_i^{\mu+1} = B_i^\mu + \eta_{\mu \setminus i} \frac{s_i^\mu}{\sqrt{N}} \quad (11.8)$$

com $\mu = 1, \dots, P$ e $B_i^{t=0} = 0$. Por meio de (11.7) o valor estimado do vetor professor obedece à atualização

$$\langle w_i \rangle_q^{\mu+1} = \tanh \left(\tanh^{-1} \langle w_i \rangle_q^\mu + \eta_{\mu \setminus i} \frac{s_i^\mu}{\sqrt{N}} \right), \quad (11.9)$$

onde se introduz a notação $\langle \dots \rangle_q^\mu$ para o valor esperado em relação à distribuição aproximativa no instante μ .

O fato interessante é que o algoritmo *on-line* gerado pelo PE proporciona a mesma regra de atualização gerada pelo algoritmo Bayesiano *On-line*, estudado e apresentado por Whinter e Solla em [59] e [60]. Na verdade, isso não é uma coincidência, uma vez que ambos os algoritmos iniciam com mesmo a priori e utilizam da minimização da divergência KL para computar as estatísticas suficientes da distribuição aproximativa.

Parâmetros de Ordem

Alguns parâmetros importantes para a medida de desempenho do algoritmo serão introduzidos agora. Entre esses, têm-se os *Parâmetros de Ordem* Q , R e T , definidos em termos do vetor professor \mathbf{w} e sua estimativa $\langle \mathbf{w} \rangle_q$:

$$R \equiv \frac{1}{N} \langle \mathbf{w} \rangle_q \cdot \mathbf{w} \quad (11.10)$$

$$Q \equiv \frac{1}{N} \langle \mathbf{w} \rangle_q \cdot \langle \mathbf{w} \rangle_q \quad (11.11)$$

$$T \equiv \frac{1}{N} \mathbf{w} \cdot \mathbf{w}. \quad (11.12)$$

Com esses parâmetros, pode-se introduzir o *erro de generalização*

$$e_g = \kappa + (1 - 2\kappa) \frac{1}{\pi} \arccos \left(\frac{R}{\sqrt{QT}} \right) \quad (11.13)$$

que dá a probabilidade de que a estimativa esteja errada [54].

A predição teórica da evolução desses parâmetros para o algoritmo Bayesiano *On-line* foi feita nos trabalhos de Winther e Solla. Eles encontraram que

$$R = Q = \int Dz \tanh(\sqrt{A}z + A) \quad (11.14)$$

com

$$\frac{dA}{d\alpha} = \frac{1}{\pi} \frac{1}{\sqrt{1-Q}} \int Dt \frac{(1-2\kappa)^2 e^{-\frac{1}{2}Qt^2}}{\kappa + (1-2\kappa)H(-\sqrt{Q}t)} \quad (11.15)$$

onde

$$\alpha \equiv \frac{P}{N}. \quad (11.16)$$

As funções matemáticas introduzidas em (11.14) e (11.15) estão definidas no apêndice B. Estão apresentados na figura (11.1) o resultado teórico (obtido pela solução numérica de (11.14) junto com (11.15)) e resultados médios de simulações independentes em que se usa o algoritmo PE (11.9) para estimar o vetor professor. O perfeito ajuste dos dados de simulações com a curva obtida por Winther e Solla evidencia que o algoritmo PE em sua versão *on-line* é idêntico ao algoritmo estudado por eles. No entanto, o algoritmo PE é mais robusto no sentido de que ele admite a re-apresentação dos dados de modo a se tornar um algoritmo *off-line*. Esse fato será discutido com mais detalhes na próxima seção.

11.1.2 A versão *off-line*

A partir do momento em que se reapresenta um dado para uma nova atualização, deixa-se o cenário *on-line* e se entra no chamado aprendizado *off-line*. Nesse cenário,

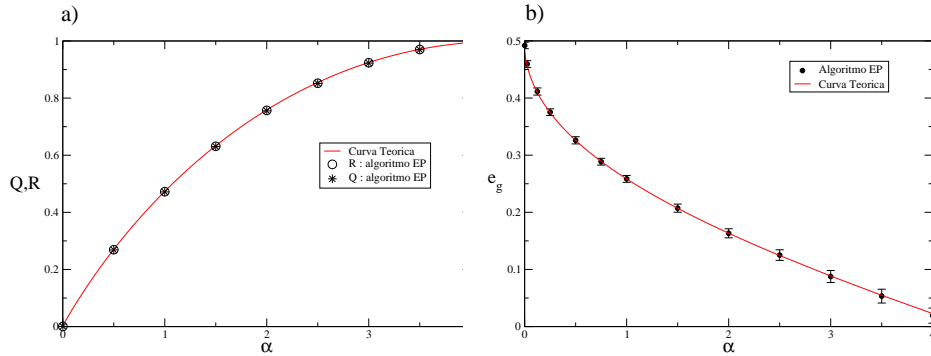


Figura 11.1: Evolução *on-line* dos parâmetros de ordem Q e R (figura a) e do erro de generalização (figura b) em função de α . Nesses gráficos, é apresentada a predição teórica do algoritmo bayesiano estudado por Winther e Solla juntamente com os resultados de simulações com o algoritmo PE. As simulações foram realizadas com $N = 1000$, $\kappa = 0$ e com médias realizadas sobre 100 experimentos numéricos independentes.

o conjunto de treinamento é reapresentado para atualização até que a convergência seja atingida.

A figura (11.2) mostra o erro de generalização do algoritmo PE em função de α . Cada ponto nesse gráfico representa o valor de convergência propiciado por esse algoritmo. Para $0 < \alpha < 1.245$, o erro de generalização do algoritmo PE é o mesmo do algoritmo ótimo, representado pela curva cheia. O algoritmo ótimo é aquele que minimiza o erro de generalização devido ao fato de ele utilizar o conhecimento disponível de uma forma ótima, absorvendo toda a informação presente nos dados. Sendo assim, esse algoritmo apresenta um erro de generalização que não pode ser ultrapassado por nenhuma outra regra de aprendizado. Portanto, no intervalo $0 < \alpha < 1.245$, o algoritmo PE *off-line* desenvolvido aqui consegue absorver toda informação contida nos dados, pois apresenta desempenho ótimo.

A curva do erro de generalização do algoritmo ótimo foi obtida em [57] por meio do cálculo de réplicas. A versão replica simétrica desse cálculo, que está apresentada com detalhes no apêndice E, conduz a um resultado correto, exceto na região $1.245 < \alpha < 1.492$, onde esse ansatz promove uma entropia negativa. Nesse mesmo intervalo, pode-se perceber que o erro de generalização do algoritmo PE é o mesmo do obtido por cálculo replicas simétricas, que são, ambos, diferentes do erro de generalização do algoritmo ótimo; nessa região o algoritmo ótimo promove generalização perfeita ($e_g = 0$). Note que o algoritmo PE também propicia uma entropia negativa neste

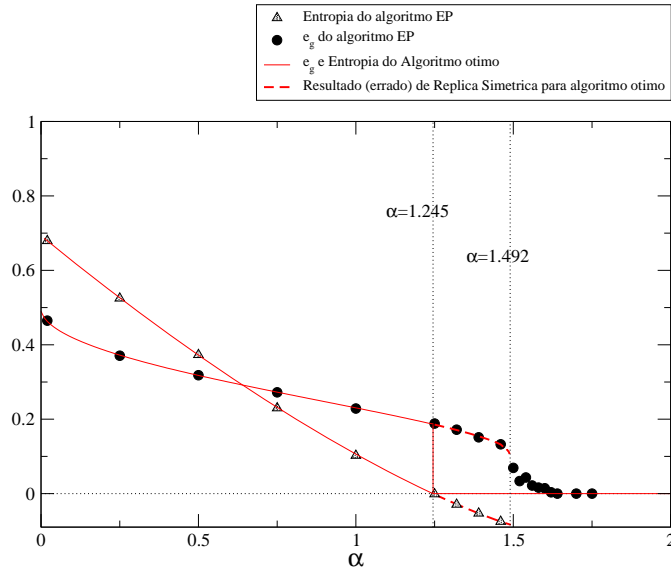


Figura 11.2: Gráfico da Entropia e do erro de generalização para a previsão de réplica simétrica para o algoritmo ótimo e os resultados do algoritmo PE. As simulações com o algoritmo PE utilizaram $N = 1000$, $\kappa = 0$ e cada ponto gerado pela média sobre 30 experimentos numéricos independentes.

intervalo. A entropia do algoritmo PE é computado aplicando-se o resultado (10.42) à verossimilhança (11.2).

O motivo dos resultados idênticos entre algoritmo PE e o cálculo de réplica simétrica, bem como a razão de eles gerarem uma entropia negativa e um erro de generalização diferente de zero nessa região de α , pode ser entendido por meio da figura (11.3). Nela, é apresentada a entropia do algoritmo ótimo, dada por (E.24), em função do *overlap* R . Para um dado α , o máximo dessa entropia corresponde ao valor mais provável de R , quando se utiliza o algoritmo ótimo. A conjuntura de réplica simétrica somada ao limite termodinâmico “concentram toda a atenção” no ponto máximo que seja diferenciável. Quando $\alpha < 1.245$ existe um único máximo diferenciável e este é também o máximo global; nesta região de α o cálculo de réplicas simétricas reproduz o resultado correto para o algoritmo ótimo, e da mesma forma o algoritmo PE, que converge para este ponto de máximo. Para $\alpha > 1.245$, o ponto máximo diferenciável se torna um máximo *local*, enquanto o ponto $R = 1$ se torna o ponto máximo *global*, porém não diferenciável. Nessa região de α o algoritmo ótimo consegue uma generalização perfeita, pois o máximo global está em $R = 1$; no entanto, o cálculo de réplica simétrica não prevê esse resultado, mas sim

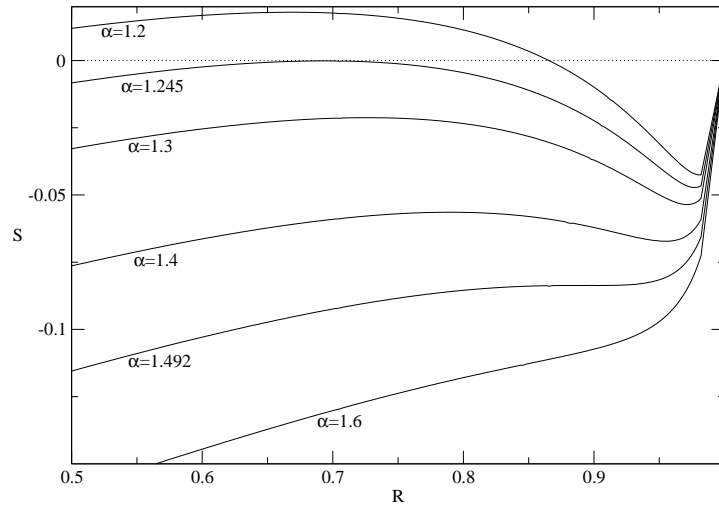


Figura 11.3: Gráfico da entropia em função do overlap R para o algoritmo ótimo. Para um dado α , o máximo dessa entropia corresponde ao valor mais provável de R .

um valor $R < 1$ dado pelo máximo local diferenciável. Nessa mesma região de α ($1.245 < \alpha < 1.492$), o algoritmo PE também não consegue atingir o resultado do algoritmo ótimo, pois ele fica preso ao mesmo máximo local que prevê o cálculo de réplica simétrica. Ainda referindo a figura (11.3), pode-se ver que, quando $\alpha = 1.492$ o máximo local desaparece, o que dá origem a um único máximo em $R = 1$. A partir desse ponto, tanto o cálculo de réplica quanto o algoritmo PE obtêm a generalização perfeita.

11.2 O Perceptron Gaussiano

Nesta seção, dar-se-á continuidade às aplicações do algoritmo de aprendizado PE. Considera-se agora a classe de problemas em que o professor gera uma resposta linear. Mais especificamente, considerar-se-á que o professor gera como saída, dada uma entrada \mathbf{s} , o próprio valor do campo $h = \left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}}\right)$. Esse tipo de professor é conhecido na literatura como *perceptron simples* [59]. Essa resposta do professor chega, porém, sujeita a um ruído aditivo, o que faz com que o *output* τ possua a forma

$$\tau = h + \eta, \quad (11.17)$$

onde η , o ruído, é uma variável aleatória que obedece a uma distribuição gaussiana com média nula e variância λ^2 . A forma (11.17) implica a distribuição

$$p(\tau|\mathbf{w}, \mathbf{s}, \eta) = \delta(\tau - h - \eta), \quad (11.18)$$

em que uma média sobre a distribuição de η dá origem à verossimilhança

$$p(\tau|\mathbf{w}, \mathbf{s}) = \begin{cases} \delta(\tau - h) & \text{para } \lambda^2 = 0 \text{ (ausência de ruído)} \\ \frac{1}{\sqrt{2\pi\lambda^2}} e^{-\frac{(\tau-h)^2}{2\lambda^2}} & \text{para } \lambda^2 \neq 0 \text{ (presença de ruído)}. \end{cases}$$

Para que esse modelo possa ser resolvido exatamente, considerar-se-á que o vetor professor seja gerado a partir do a priori gaussiano (10.3). Tem-se então o modelo conhecido por *Perceptron Gaussiano*.

O algoritmo PE proporciona a esse modelo uma regra de atualização bastante simples. Aplicando os resultados (10.31) e (10.32) à verossimilhança acima, os parâmetros atualizados das matrizes \mathbf{A} e \mathbf{B} obedecem às formulas

$$a_{i\mu}^{new} = \frac{(s_i^\mu)^2}{N(\lambda^2 + \sigma_{\mu|i}^2)} \quad (11.19)$$

e

$$b_{i\mu}^{new} = \frac{(\tau^\mu - \bar{h}_\mu^i) s_i^\mu}{(\lambda^2 + \sigma_{\mu|i}^2) \sqrt{N}}, \quad (11.20)$$

respectivamente.

Medida de desempenho

Introduzir-se-á aqui a forma que se considera mais conveniente para analisar o desempenho (ou acuracidade) da estimativa do professor nesse modelo. Considere-se que o erro de generalização e_g entre o vetor professor \mathbf{w} e a sua estimativa $\langle \mathbf{w} \rangle_q$ possa ser calculado pela média do quadrado da diferença entre o campo do professor e da sua estimativa sobre a distribuição dos *input's* \mathbf{s} ; isso claro, a menos do ruído. Esta ideia pode ser equacionada na forma

$$e_g = \lambda^2 + \left\langle \left(\frac{\mathbf{w} \cdot \mathbf{s}}{\sqrt{N}} - \frac{\langle \mathbf{w} \rangle_q \cdot \mathbf{s}}{\sqrt{N}} \right)^2 \right\rangle_{\mathbf{s}}, \quad (11.21)$$

onde $\langle \dots \rangle_{\mathbf{s}} = \int d\mathbf{s}(\dots)p(\mathbf{s})$ é a média sobre a distribuição $p(\mathbf{s})$ dos *input*'s. Como se está considerando que os *input*'s são independentes, i.e. $p(\mathbf{s}) = cte$, pelas definições dos parâmetros de ordem apresentadas na seção (11.1.1) chega-se a

$$e_g = \lambda^2 + T + Q - 2R. \quad (11.22)$$

11.2.1 A versão *On-line*

Iniciam-se as simulações desse modelo com $q(\mathbf{w}) = p_0(\mathbf{w})$, da mesma forma que foi feito no caso do *perceptron* binário. Essa escolha implica que todos os parâmetros das matrizes \mathbf{A} e \mathbf{B} são inicialmente nulos. Como foi visto anteriormente, quando se apresenta cada um dos dados uma única vez para atualização, tem-se a versão *on-line* do algoritmo. Com essas considerações, a μ -ésima atualização *on-line* dos parâmetros A_i e B_i será dada por

$$\begin{cases} A_i^\mu = A_i^{\mu-1} + a_{i,\mu}^{new} \\ B_i^\mu = B_i^{\mu-1} + b_{i,\mu}^{new}. \end{cases}$$

Essa atualização, juntamente com a estimativa (10.12) e apenas os termos $\mathcal{O}(\frac{1}{\sqrt{N}})$, implica o algoritmo PE

$$\langle w_i \rangle_q^\mu = \langle w_i \rangle_q^{\mu-1} + \frac{1}{\sqrt{N}} \frac{(\tau^\mu - \bar{h}_\mu^i)}{(\lambda^2 + \sigma_{\mu|i}^2)} s_j^\mu. \quad (11.23)$$

Esse algoritmo PE *on-line*, embora simples, é sub-ótimo. Esse fato pode ser verificado a partir da figura (11.4) que apresenta a evolução do erro de generalização desse algoritmo em comparação com o do Algoritmo Bayesiano *On-Line*, que é ótimo nesse contexto[61].

O Algoritmo ótimo

O Algoritmo Bayesiano *on-line*, que segue a atualização (9.16) e (9.17), é ótimo quando aplicado ao modelo do *perceptron* gaussiano. Isso se deve ao fato de que a distribuição a posteriori original $p(\mathbf{w}|A_t, \mathbf{x}^{t+1})$ não precisa ser projetada no espaço das distribuições gaussianas para se obter uma distribuição aproximativa; ela própria é uma gaussiana (pois é obtida pelo produto de gaussianas) e, portanto, já está neste espaço. Dessa forma, a nova distribuição não é uma aproximação, mas sim a

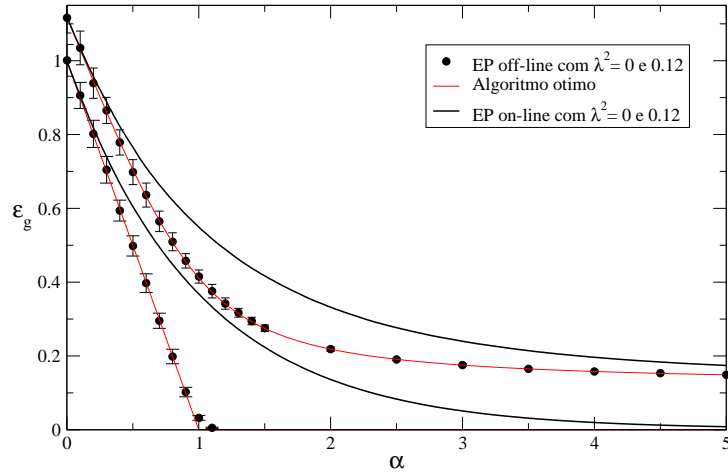


Figura 11.4: Resultados de simulações com o algoritmo PE nas versões *on-line* e *off-line* em comparação com o algoritmo ótimo. Essas simulações foram realizadas utilizando-se $N = 1000$. Na versão *on-line*, os resultados foram obtidos através da média sobre 100 experimentos numéricos independentes; as barras de erro desses experimentos são menores que a largura da curva. Na versão *off-line* as médias foram realizadas sobre 30 experimentos numéricos independentes.

própria distribuição original. Em outras palavras, não há perda de informação nessa atualização *on-line* o que faz com que o algoritmo Bayesiano seja ótimo.

No trabalho [59], é mostrado que o algoritmo bayesiano *on-line* apresenta a atualização

$$\langle \mathbf{w} \rangle^\mu = \langle \mathbf{w} \rangle^{\mu-1} + \frac{1}{\sqrt{N}} (\mathbf{M}^{\mu-1} \mathbf{s}^\mu) \frac{(\tau^\mu - \bar{h}_\mu)}{(\lambda^2 + \sigma_\mu^2)} \quad (11.24)$$

com

$$\mathbf{M}^\mu = \mathbf{M}^{\mu-1} + \frac{1}{N} (\mathbf{M}^{\mu-1} \mathbf{s}^\mu (\mathbf{s}^\mu)^T \mathbf{M}^{\mu-1}) \frac{1}{(\lambda^2 + \sigma_\mu^2)}. \quad (11.25)$$

Nesse mesmo trabalho, Solla e Winther sugerem que um algoritmo mais simples pode ser obtido a partir do algoritmo bayesiano. Isso é feito mudando a direção de atualização do algoritmo bayesiano, que é $\mathbf{M}^{\mu-1} \mathbf{s}^\mu$, para a direção do *input* \mathbf{s}^μ ; em outras palavras, pela substituição

$$\mathbf{M}^{\mu-1} \mathbf{s}^\mu \rightarrow \mathbf{s}^\mu \frac{(\mathbf{s}^\mu)^T \mathbf{M}^{\mu-1} \mathbf{s}^\mu}{\mathbf{s}^\mu \mathbf{s}^\mu} \quad (11.26)$$

nas equações de recorrência (11.24) e (11.25), que dá origem ao algoritmo sub-ótimo

$$\langle w_i \rangle^\mu = \langle w_i \rangle^{\mu-1} + \frac{1}{\sqrt{N}} \frac{(\tau^\mu - \bar{h}_\mu)}{(\lambda^2 + \sigma_\mu^2)} s_i^\mu \quad (11.27)$$

chamado por eles de *algoritmo simplificado*. Na ausência de ruído ($\lambda^2 = 0$), o erro de generalização desse algoritmo obedece ao decaimento $e_g = e^{-\alpha}$. Nota-se que o algoritmo (11.27) é idêntico ao algoritmo (11.23), obtido pelo PE a menos do termo de cavidade. Na verdade, dado que esse termo de cavidade é $\mathcal{O}(\frac{1}{N})$, pode-se considerar que, no limite termodinâmico, o algoritmo simplificado e o algoritmo PE são idênticos. A curva da evolução do erro de generalização do algoritmo simplificado sobrepõe exatamente a curva do algoritmo PE.

11.2.2 Versão *Off-line*

Embora o desempenho do algoritmo PE *on-line* seja sub-ótimo, a sua versão *off-line* apresenta o mesmo desempenho do algoritmo ótimo, como pode ser verificado pela figura (11.4). Nela estão apresentados os resultados de simulações em que o conjunto de dados \mathcal{D} é apresentado e reapresentado quantas vezes forem necessárias até que o algoritmo PE promova a convergência. Nesse problema, o número de reapresentações é, tipicamente, em torno de 5. Note-se que, na ausência de ruído, há generalização perfeita a partir de $\alpha = 1$.

12

CONCLUSÕES

Foi desenvolvido neste trabalho um algoritmo de inferência aproximativa para o aprendizado supervisionado por meio da técnica *Propagação de Expectativas*. Utilizando do Teorema do Limite Central e do Método de Cavidade, foi possível obter uma formulação geral dado a verossimilhança de uma regra que se queira inferir. O algoritmo desenvolvido possibilita estimar os componentes do vetor professor, a partir dos dados disponíveis, com custo computacional bastante razoável ($\mathcal{O}(N)$). Nos dois casos estudados (Modelo do *Perceptron* Binário e do *Perceptron* Gaussiano), este o algoritmo PE um erro de generalização ótimo.

O algoritmo PE permite ir de uma distribuição a posteriori original intratável para uma outra distribuição aproximativa pertencente a uma classe de distribuições tratáveis. Em particular, foi considerada neste trabalho uma distribuição aproximativa pertence à família exponencial. Dessa forma, esta distribuição possui componentes não iteragentes e pode ser computada a partir de estatísticas suficientes. O algoritmo PE propicia uma atualização local nessa distribuição aproximativa, de modo que seus termos sejam atualizados um de cada vez.

No caso do *Perceptron* Binário, foi verificado que a versão *on-line* do algoritmo PE apresenta uma regra de atualização idêntica ao do algoritmo Bayesiano *On-line*. Isso se deve ao fato que esses dois algoritmos utilizam da mesma filosofia de aprendizado; uma filosofia baseada na minimização da divergência KL para a determinação das estatísticas suficientes. No entanto, o algoritmo PE é mais geral, de modo que ele pode lidar com a reapresentação dos dados e assim agir no cenário *off-line*. Nesse cenário, ele é ótimo quando aplicado ao *Perceptron* Binário, exceto, porém, quando a quantidade de dados disponíveis está em um certo intervalo específico. Quando o parâmetro α (razão entre a quantidade de dados e a dimensão dos vetores) pertence ao intervalo $1.245 < \alpha < 1.492$, o algoritmo PE *off-line* ap-

resenta desempenho sub-ótimo. Isso se deve à existência de um máximo local da entropia do sistema (com valores negativos) que prende a dinâmica do algoritmo, não permitindo a generalização perfeita. O resultado do erro de generalização do algoritmo PE é o mesmo resultado obtido por meio do ansatz de réplica simétrica para a determinação do erro de generalização ótimo. O resultado é o mesmo até no intervalo sub-ótimo, onde as duas técnicas apresentam entropia negativa, mostrando a incoerência de seus resultados nesta região. Quando se trata do *perceptron* gaussiano, foi verificado que a versão *on-line* do algoritmo PE propicia um algoritmo sub-ótimo, ao contrário do algoritmo Bayesiano *on-line*, que é ótimo nesse modelo. No entanto o algoritmo PE dá origem a uma regra de atualização *on-line* bastante satisfatória se considerar o seu baixo custo computacional. De qualquer forma, a versão *off-line* do algoritmo PE apresenta desempenho ótimo, o que foi verificado com simulações usando diferentes níveis de ruído.

Em suma, o algoritmo de aprendizado PE se mostrou bastante satisfatório nos quesitos precisão, dada a obtenção de desempenhos ótimos, e eficiência computacional. A proporção do custo computacional da versão *off-line* e da versão *on-line*, que é $\mathcal{O}(P)$, onde P é o número de dados disponíveis, é apenas um fator constante. E, como o algoritmo apresenta convergência após poucos *loops* (cerca de 10), conclui-se que seu desempenho computacional é bastante razoável. No entanto, embora nos dois problemas estudados a convergência tenha sido atingida, não se pode estender esta expectativa a qualquer problema. Isto se deve, principalmente, à dinâmica de atualizações locais feitas por esse algoritmo; como cada fator é individualmente aproximado, o produto dos fatores pode apresentar uma aproximação pobre. Mas, no geral, os resultados são bastante animadores, o que permite aventurar em outros campos de aplicações, obtendo-se soluções precisas sem pagar um alto custo computacional para isso.

13

CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Os resultados apresentados nesta tese concluíram uma parte significativa de nossos estudos a respeito da Especiação Simpátrica e de Algoritmos de Inferência Aproximativa. Mas claro, tais estudos não acabam aqui, pois há um imenso leque de novas formulações e cenários a se atacar a partir das idéias introduzidas. Novos direcionamentos devem ser exploradas no sentido de se construir tanto uma descrição mais precisa e geral para a evolução de preferência sexual, quanto uma formulação mais robusta para o algoritmo de Inferência Aproximativa. Todavia, os resultados introduzidos aqui são os primeiros *insight's* sobre dois campos que prometem maiores desenvolvimentos num futuro próximo. Entre os possíveis desenvolvimentos, destacamos alguns a seguir.

A respeito do modelo biológico, seria muito interessante entender por que uma dada estratégia (como ser sexualmente seletivo ou não) pode ser (ou não) evolutivamente viável, dadas as condições do ambiente. Deve ser possível determinar analiticamente a fronteira dessas duas estratégias evolutivas e conseqüentemente a linha de transição de primeira ordem observada pelas simulações. Um cálculo nessa direção foi realizado e apresentado em [12] para o modelo de Competição com a simplificação de que o traço aparente e o ecológico são os mesmos. Foi demonstrado pelos autores que há uma “invasão” de indivíduos sexualmente seletivos, quando o chamado *Fitness Malthusiano* dos homozigotos se torna maior que o dos heterozigotos. O *Fitness Malthusiano* é construído a partir da adição de duas componentes: uma, dependente da probabilidade de sobrevivência do indivíduo e, outra, da probabilidade de que esse indivíduo seja gerado. Note-se a semelhança dessa definição

de *fitness* e a função energia livre discutida no capítulo 6, com o seu primeiro termo ligado à Energia e , o segundo, à Entropia. A fronteira entre as fases com ou sem preferência sexual acontece quando há igualdade dos *fitness Malthusiano* dos dois tipos de indivíduos, exatamente como prescreve-se pela teoria da ME das transições de fase. Acreditamos que um cálculo análogo ao realizado por esses pesquisadores pode ser realizado numa versão mais complexa, como a do modelo introduzido aqui, com indivíduos compostos por dois loci. Com esse cálculo, acredita-se ser possível a construção da função energia livre desse sistema biológico. Numa perspectiva ambiciosa, acreditamos também que será possível equacionar, a partir da energia livre, a relação da histerese com a taxa de recombinação r , evidenciada pelas simulações. Um pergunta que se gostaria de responder é se haveria uma transição de segunda ordem (transição contínua) no limite $r \rightarrow 0$. Nas simulações apresentadas na figura (6.1), verificou-se que a histerese deixa de existir nesse limite, mas não foi possível verificar se há ou não um ponto crítico r_c a partir do qual se iniciam transições contínuas.

A pergunta que ainda persiste é até que ponto os resultados aqui obtidos dependem dos detalhes desse modelo. Deveriam outros tipos de modelos, tais como o Modelo de Competição estudado por Dikman e Dobeli [17], apresentar também características de histerese? Ou como as transições de fase se desenvolveriam em modelos multi-loci?

Em relação ao PE, pode-se pensar que uma forma mais geral para a distribuição aproximativa apresentada aqui pode ser o primeiro passo rumo à uma formulação mais genérica e robusta. Por exemplo, considerar uma distribuição gaussiana multivariada ao invés de uma gaussiana univariada. Claro que essa implementação implica um maior custo computacional, pois necessita a computação não só do primeiro e segundo momento da distribuição, mas também dos elementos da matriz co-variância. No entanto, esse aumento do custo de processamento seria compensado pela maior robustez do algoritmo, o que permitirá atacar problemas mais complexos como, por exemplo, os chamados *Processos Gaussianos* [49]. Métodos padrões para resolução desses problemas requerem a inversão de uma matriz $P \times P$ (onde P é o número de dados disponíveis), o que implica uma computação $\mathcal{O}(P^3)$. Dessa forma, à medida que se aumenta o número de dados disponíveis, a resolução do problema se torna inviável. O que tem sido proposto por Csató e Oppen [65] é considerar apenas um subgrupo dos dados disponíveis composto por n dados, onde $n \ll P$. Sendo assim, há a necessidade de inversão de uma matriz $n \times n$, o que gera um custo $\mathcal{O}(n^3)$,

muito menor que o custo da inversão da matriz original, mas pagando um alto preço por perder as informações contidas nos dados excluídos. O que poderia ser feito seria abstrair as informações desses dados excluídos a partir do algoritmo PE. Dessa forma, o custo computacional do algoritmo (PE mais inversão da matriz $n \times n$) seria $\mathcal{O}(cPn^3)$, onde c é o número de *loops* do algoritmo PE. Alguns resultados preliminares têm sido bastante animadores, mostrando que, se o número de dados n for escolhido de forma adequada, pode-se obter um resultado bastante acurado e com custo computacional inferior aos algoritmos existentes.

Apêndice A

GLOSSÁRIO DE TERMOS BIOLÓGICOS

Neste apêndice será apresentado um glossário de palavras e termos biológicos utilizados na tese.

- **Alelo:** Forma alternativa de um gene. Uma das alternativas de um par ou série de formas de um gene, os quais são alternativos na herança, porque estão situados no mesmo locus em cromossomos homólogos.
- **Cromossomo:** Estrutura encontrada no núcleo de todas as células que contém os genes. Os cromossomos vêm geralmente aos pares.
- **Crossover:** Troca de material genético entre os cromossomos maternos e paternos durante a meiose. Esse processo pode resultar em uma combinação única dos genes no cromossomo filho em relação aos cromossomos originais (paternal e maternal).
- **Diploide:** Que contém dois conjuntos de cromossomos homólogos e, portanto, cada locus é composto por dois genes alelos.
- **Dominante:** Refere-se ao membro de um par de alelos que é expresso no fenótipo do organismo, enquanto o outro alelo não o é, apesar dos dois alelos estarem presentes. Também refere-se ao fenótipo expresso pelo alelo dominante. Oposto a recessivo.
- **Dominância:** Interação entre alelos de um mesmo locus de tal forma que nos heterozigotos um alelo tem efeito maior do que outro alelo. Pode ser completa, quando o fenótipo do heterozigoto é o mesmo do homozigoto para o alelo dominante, ou incompleta, quando o fenótipo heterozigoto situa-se no intervalo dos fenótipos homozigotos.

- **Equilíbrio Hardy-Weinberg:** Pode-se mostrar facilmente que uma população pan-mítica na ausência de seleção natural apresenta frequências genotípicas constantes ao longo das gerações [23]; além disso essas frequências podem ser obtidas por meio das frequências alélicas pelas relações

$$n_{AA} = p^2, \quad (\text{A.1})$$

$$n_{aa} = (1 - p)^2, \quad (\text{A.2})$$

$$n_{Aa} = 2p(1 - p). \quad (\text{A.3})$$

Quando as frequências genotípicas de uma população obedecem as relações acima, diz-se que ela (a população) se encontra no equilíbrio Hardy-Weinberg (HW). Em uma população pan-mítica as proporções citadas acima serão atingidas em uma única geração.

- **Fenótipo:** São características observáveis de um organismo, em função do conjunto de genes que possui (seu genótipo). Aparência final de um indivíduo como resultado da interação de seu genótipo com um determinado ambiente abiótico. Aspecto externo do organismo, subordinado às influências do ambiente; grupos de indivíduos que, embora tendo constituição genética diferente, apresentam os mesmos caracteres exteriores. O fenótipo de um organismo se manifesta como resultado de fatores genéticos e ambientais. Assim, organismos que possuem o mesmo genótipo podem apresentar diferentes fenótipos devido a fatores ambientais. Da mesma forma, organismos com o mesmo fenótipo podem ter genótipos diferentes.
- **Fitness:** Grau de sucesso de um indivíduo em passar seus genes para a próxima geração. Apresenta duas componentes: sobrevivência (viabilidade) e sucesso reprodutivo (fecundidade).
- **Gene:** Unidade física e funcional de hereditariedade que carrega informações de uma geração para outra. É também uma toda uma sequência do DNA responsável pela síntese de um polipeptídeo funcional ou de uma molécula RNA.

- **Gameta:** célula reprodutiva haploide, como esperma ou pólen.
- **Genoma :** Material genético total de um indivíduo.
- **Genótipo:** Fórmula genética de um ou mais loci. Totalidade dos fatores genéticos que formam a constituição genética de um indivíduo. Determina o fenótipo individual. Conjunto da informação genética na massa hereditária; constituição genética, hereditária , do indivíduo.
- **Haplotipo:** Combinação particular de genes. Sinônimo de Gameta.
- **Haploide:** Célula ou organismo que contém uma cópia de cada cromossomo. Que contém apenas um genoma. Nível de ploidia de células ou organismos que contém apenas um dos cromossomos de cada par de homólogos encontrados no nível diplode ($2n$) em eucariontes superiores. Os cromossomos haploides apresentam-se sozinhos, únicos, sem “irmãos” idênticos.
- **Heterozigoto:** Indivíduo que apresenta alelos diferentes em um ou vários loci. Oposto de Homozigoto.
- **Híbrido:** Indivíduo formado pelo cruzamento entre formas diferentes, usualmente populações geneticamente diferenciadas ou espécies; ocasionalmente em genética, os descendentes de um cruzamento entre genótipos fenotipicamente distinguíveis de qualquer tipo. Um indivíduo que resulta do cruzamento entre dois genitores geneticamente diferentes. Uma descendência de progenitores homozigotos que diferem em um ou mais genes.
- **Homozigoto:** Locus, genótipo ou indivíduo que tem duas cópias do mesmo alelo para determina do gene. Que apresenta alelos iguais.
- **Isolamento reprodutivo pré-zigótico:** tipo de isolamento sexual causado por algum mecanismo que evita o cruzamento entre indivíduos de especialidades (ou espécies) diferentes e a conseqüente formação do zigoto.
- **Isolamento reprodutivo pós-zigótico:** tipo de isolamento sexual em que há o cruzamento entre especialidades (ou espécie) diferentes, no entanto não há formação ou desenvolvimento do zigoto. Ou quando quando há esta formação, o indivíduo híbrido apresenta esterilidade.
- **Locus:** Local no cromossomo onde se localiza um determinado gene.

- **Pan-mítico:** tipo de acasalamento em que os indivíduos escolhem ao acaso seu parceiro sexual.

B

ALGUMAS FUNÇÕES MATEMÁTICAS

Neste apêndice seguem algumas funções ou propriedades matemáticas utilizadas ao longo do desenvolvimento deste trabalho.

B.1 Divergência Kullback-Leibler

A divergência Kullback-Leibler (KL) entre duas distribuições $p(\mathbf{x})$ e $q(\mathbf{x})$ é definida como

$$KL(p||q) = \int p(\mathbf{x}) \log \left(\frac{q(\mathbf{x})}{p(\mathbf{x})} \right) d\mathbf{x} \quad (\text{B.1})$$

Esta divergência é não simétrica, uma vez que $KL(p||q) \neq KL(q||p)$. Além disso ela satisfaz $KL(p||q) \geq 0$; a igualdade acontece somente se $q(\mathbf{x}) = p(\mathbf{x})$.

Esta divergência nos dá o valor médio da quantidade de informação necessária para especificar o valor de \mathbf{x} como resultado do uso de uma distribuição aproximativa $q(\mathbf{x})$ ao invés da distribuição verdadeira $p(\mathbf{x})$ [49].

B.2 Integrais Gaussianas

Algumas notações a respeito de integrais gaussianas são muito praticas para a apresentação mais coesa de alguns resultados. Ao longo deste trabalho usamos as seguintes abreviações:

$$Dt := dt \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \quad (\text{B.2})$$

e

$$H(x) := \int_x^\infty Dt = \int_x^\infty \frac{dt}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}. \quad (\text{B.3})$$

Esta última se relaciona com a chamada *função erro complementar*, definida como

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt, \quad (\text{B.4})$$

pela forma

$$H(x) = \frac{1}{2} \operatorname{erfc} \left(\frac{x}{\sqrt{2}} \right). \quad (\text{B.5})$$

B.3 A função de Heaviside

A função de Heaviside, também conhecida por *função degrau*, é definida como

$$\Theta(x) = \begin{cases} 0 & \text{para } x \leq 0 \\ 1 & \text{para } x > 0 \end{cases}.$$

B.4 A função característica

A função característica $g(k)$ de uma distribuição $p(x)$ é definida pela transformada de Fourier

$$g(k) = \int p(x) e^{ikx} dx = \langle e^{ikx} \rangle \quad (\text{B.6})$$

onde os braket's denotam a média com respeito à distribuição em questão. Utilizando expansões em séries de Taylor da função e^{ikx} podemos re-escrever a função característica como

$$g(k) = e^{\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n}. \quad (\text{B.7})$$

onde $\{\kappa_n\}$ são os cumulantes da distribuição $p(x)$ e são diretamente relacionados com os momentos desta distribuição. Esta relação segue a regra

$$\begin{cases} \kappa_1 = \langle x \rangle \\ \kappa_2 = \langle x^2 \rangle - \langle x \rangle^2 \\ \kappa_3 = \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3 \\ \dots \end{cases}$$

A distribuição $p(x)$ pode também ser obtida a partir de $g(k)$ pela transformada inversa de Fourier

$$p(x) = \frac{1}{\sqrt{2\pi}} \int e^{-ikx} g(k) dk. \quad (\text{B.8})$$

C

TEOREMA DO LIMITE CENTRAL

Considere um conjunto de N variáveis microscópicas $\{\xi_i\}$ que obedecem à distribuição

$$p(\boldsymbol{\xi}) = \prod_{i=1}^N p(\xi_i). \quad (\text{C.1})$$

onde $p(\xi) = p(\xi_i)$, para qualquer i , é uma distribuição com média a e variância b . Este conjunto de variáveis microscópicas dá origem a uma variável macroscópica h definida por

$$h = \frac{1}{\sqrt{Nb}} \left(\sum_{i=1}^N \xi_i - Na \right). \quad (\text{C.2})$$

O *Teorema do Limite Central* afirma que, dada a distribuição (C.1) para as variáveis microscópicas, então no limite termodinâmico ($N \rightarrow \infty$) a variável macroscópica possui a distribuição gaussiana

$$p(h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h^2}. \quad (\text{C.3})$$

Prova

Considere $g(k)$ a *função característica* da distribuição $p(h)$ (veja apêndice B), definida pela transformada de Fourier

$$g(k) = \int p(h) e^{ikh} dh. \quad (\text{C.4})$$

Uma vez que a variável aleatória h pode ser escrita em termos de um conjunto de N variáveis $\{\xi_i\}$ pela relação (C.2), a função $g(k)$ pode ser reescrita como

$$\begin{aligned}
g(k) &= \int d\boldsymbol{\xi} p(\boldsymbol{\xi}) e^{i\frac{k}{\sqrt{Nb}}(\sum_{i=1}^N \xi_i - Na)} \\
&= \prod_{i=1}^N \left[\int p(\xi_i) e^{i\frac{k}{\sqrt{Nb}} \xi_i} d\xi_i \right] e^{-i\frac{k}{\sqrt{Nb}} Na} \\
&= \left[g_\xi \left(\frac{k}{\sqrt{Nb}} \right) \right]^N e^{-i\frac{k}{\sqrt{Nb}} Na}. \tag{C.5}
\end{aligned}$$

Note que na segunda linha de (C.5) há fatorização em i , o que nos permite substituir ξ_i por ξ . A terceira linha é obtida introduzindo a função característica da distribuição $p(\xi)$, dada por

$$\begin{aligned}
g_\xi \left(\frac{k}{\sqrt{Nb}} \right) &= \int d\xi p(\xi) e^{i\frac{k}{\sqrt{Nb}} \xi} \\
&= e^{\sum_{n=1}^{\infty} \frac{\left(i\frac{k}{\sqrt{Nb}} \right)^n}{n!} \kappa_n} \tag{C.6}
\end{aligned}$$

onde introduzimos a representação em termos dos cumulantes $\{\kappa_n\}$ da distribuição $p(\xi)$ (veja apêndice B.4).

No limite termodinâmico $N \rightarrow \infty$ a função $g_\xi \left(\frac{k}{\sqrt{Nb}} \right)$ pode ser aproximada pelos dois primeiros termos da série (C.6):

$$g_\xi \left(\frac{k}{\sqrt{Nb}} \right) \approx e^{i\frac{k}{\sqrt{Nb}} \kappa_1 - \frac{1}{2} \frac{k^2}{Nb} \kappa_2}. \tag{C.7}$$

Como a distribuição $p(\xi)$ apresenta

$$\kappa_1 = \langle \xi \rangle_\xi = a \tag{C.8}$$

e

$$\kappa_2 = \langle \xi^2 \rangle_\xi - \langle \xi \rangle_\xi^2 = b \tag{C.9}$$

onde $\langle \dots \rangle_\xi$ é uma média sobre esta distribuição, chegamos a

$$\lim_{N \rightarrow \infty} \left[g_\xi \left(\frac{k}{\sqrt{Nb}} \right) e^{-i\frac{k}{\sqrt{Nb}} a} \right]^N = e^{-\frac{k^2}{2}} \tag{C.10}$$

Desta forma, voltando à relação (C.5) concluímos que

$$g(k) = e^{-\frac{k^2}{2}}. \tag{C.11}$$

Finalmente podemos determinar $p(h)$ através da transformada de Fourier inversa

$$p(h) = \frac{1}{\sqrt{2\pi}} \int e^{-ikh} g(k) dk \quad (\text{C.12})$$

que resulta

$$p(h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h^2}. \quad (\text{C.13})$$

Está mostrado o teorema!

D

FAMÍLIA EXPONENCIAL

Uma distribuição é dita pertencente à *Família Exponencial* quando pode ser escrita na forma

$$q(\mathbf{w}|\boldsymbol{\eta}) = h(\mathbf{w})g(\boldsymbol{\eta})e^{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{w})} \quad (\text{D.1})$$

onde $\boldsymbol{\eta}$ é o vetor que guarda os chamados *parâmetros naturais*, enquanto $h(\mathbf{w})$ e $g(\boldsymbol{\eta})$ são duas funções quaisquer. A normalização da distribuição acima implica que

$$g(\boldsymbol{\eta}) \int d\mathbf{w} h(\mathbf{w}) e^{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{w})} = 1. \quad (\text{D.2})$$

Como exemplo de distribuição desta família podemos citar a *distribuição Multinomial*, definida por $q(\mathbf{w}|\boldsymbol{\mu}) = \prod_k \mu_k^{w_k}$ e que pode ser re-escrita na forma

$$q(\mathbf{w}|\boldsymbol{\mu}) = e^{\sum_k w_k \log \mu_k}. \quad (\text{D.3})$$

Introduzindo $\eta_k \equiv \log \mu_k$ temos

$$q(\mathbf{w}|\boldsymbol{\eta}) = e^{\boldsymbol{\eta}^T \mathbf{w}}, \quad (\text{D.4})$$

que pertence à família exponencial pois obedece a forma (D.1) com funções

$$\begin{cases} \mathbf{u}(\mathbf{w}) = \mathbf{w} \\ h(\mathbf{w}) = 1 \\ g(\boldsymbol{\eta}) = 1 \end{cases}$$

D.1 Propriedade dos momentos

Uma importante propriedade das distribuições desta família consiste no fato de que pode-se sempre determinar seus momentos por simples diferenciação da função $g(\boldsymbol{\eta})$.

Para demonstrar esta propriedade tomemos o gradiente da normalização (D.2) em relação aos parâmetros $\boldsymbol{\eta}$, que nos leva a

$$\begin{aligned} \nabla_{\boldsymbol{\eta}} g(\boldsymbol{\eta}) \int h(\mathbf{w}) e^{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{w})} d\mathbf{w} + g(\boldsymbol{\eta}) \int h(\mathbf{w}) e^{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{w})} \mathbf{u}(\mathbf{w}) d\mathbf{w} &= 0 \\ -\frac{\nabla_{\boldsymbol{\eta}} g(\boldsymbol{\eta})}{g(\boldsymbol{\eta})} &= \int h(\mathbf{w}) e^{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{w})} \mathbf{u}(\mathbf{w}) d\mathbf{w}. \end{aligned} \quad (\text{D.5})$$

Introduzindo a notação $\langle \dots \rangle_q = \int d\mathbf{w}(\dots) q(\mathbf{w}|\boldsymbol{\eta})$ chegamos ao resultado

$$-\nabla_{\boldsymbol{\eta}} \ln g(\boldsymbol{\eta}) = \langle \mathbf{u}(\mathbf{w}) \rangle_q, \quad (\text{D.6})$$

que significa que podemos expressar a média de $\mathbf{u}(\mathbf{w})$ apenas em termos da derivada de $g(\boldsymbol{\eta})$.

Com cálculo análogo podemos expressar também a covariância de $\mathbf{u}(\mathbf{w})$ em termos da segunda derivada de $g(\boldsymbol{\eta})$ pela forma:

$$\nabla_{\boldsymbol{\eta}}^2 \ln g(\boldsymbol{\eta}) = \langle \mathbf{u}^2(\mathbf{w}) \rangle_q - \langle \mathbf{u}(\mathbf{w}) \rangle_q^2 \quad (\text{D.7})$$

De forma similar podemos determinar os demais momentos por simples diferenciação.

D.2 Propriedade da Minimização da divergência KL

Uma propriedade bastante útil das distribuições da família exponencial diz respeito à minimização da divergência KL. Considere que desejamos encontrar os parâmetros naturais de uma distribuição $q(\mathbf{w}|\boldsymbol{\eta})$ da família exponencial que minimize a divergência

$$KL(p(\mathbf{w})||q(\mathbf{w}|\boldsymbol{\eta})) = \int d\mathbf{w} p(\mathbf{w}) \ln \frac{p(\mathbf{w})}{q(\mathbf{w}|\boldsymbol{\eta})}, \quad (\text{D.8})$$

onde $p(\mathbf{w})$ é uma distribuição qualquer.

Para o cálculo desta minimização vamos aplicar o gradiente de $\boldsymbol{\eta}$ nos dois lados de (D.8) e fazer $\nabla_{\boldsymbol{\eta}} KL(p, q) = 0$. Isto nos permite escrever

$$-\nabla_{\boldsymbol{\eta}} \ln g(\boldsymbol{\eta}) = \langle \mathbf{u}(\mathbf{w}) \rangle_p \quad (\text{D.9})$$

onde introduzimos a notação $\langle \dots \rangle_p = \int d\mathbf{w}(\dots)p(\mathbf{w})$. O resultado acima junto com (D.6) nos levam a

$$\langle \mathbf{u}(\mathbf{w}) \rangle_q = \langle \mathbf{u}(\mathbf{w}) \rangle_p. \quad (\text{D.10})$$

Isto significa que a distribuição $q(\mathbf{w}|\boldsymbol{\eta})$ (da família exponencial) que minimiza a divergência KL com uma outra distribuição $p(\mathbf{w})$ (qualquer), é aquela que apresenta os mesmos momentos que a distribuição $p(\mathbf{w})$.

Apêndice E

ERRO DE GENERALIZAÇÃO ÓTIMO PARA O PERCEPTRON BINÁRIO

Na figura (11.2) foi apresentado o erro de generalização ótimo para o modelo do *Perceptron* Binário. Esse erro é obtido a partir do volume de soluções possíveis ao problema, que por sua vez pode ser calculado usando o método de Réplicas. Vai-se realizar este cálculo neste apêndice, mas restringindo à versão Réplica Simétrica. Na verdade, a solução completa para o volume de soluções, ou Espaço de Versões, só deve ser obtida pela quebra de simetria de Réplicas. No entanto, a versão tratada aqui, além de ser consideravelmente mais fácil de ser realizada, é suficiente para os propósitos deste trabalho.

E.1 Espaço de Versões

Ao longo do cenário Bayesiano abordado no texto principal desta tese, lidou-se com estimativas para o vetor professor. Essas estimativas, ou médias, justificam a notação com os *brackets* $\langle \dots \rangle_q$. Será adotada a partir de agora, porém, a notação simplificada $\mathbf{J} \equiv \langle \mathbf{w} \rangle_q$. Essa introdução se justifica pelo fato que será tratado neste apêndice com “candidatos” à solução do problema, e não mais com estimativas. Por exemplo, quando se consideram apenas vetores binários, o vetor \mathbf{J} , chamado de *Vetor Estudante* na literatura de Redes Neurais, é um dentre 2^N candidatos à solução do problema.

À medida que vão se apresentando novos dados, mais e mais candidatos vão sendo excluídos por se mostrarem incompatíveis com as informações disponíveis. O que se espera é que, a partir de uma quantidade suficiente de informação, apenas

um candidato sobreviva, sendo este idêntico ao vetor professor: tem-se a solução do problema. Considerando um modelo sem ruído, um estudante candidato deve ter as mesmas respostas que o professor em todos os exemplos apresentados. Para ajudar a compreender esse processo, considere-se a função

$$\chi(\mathbf{J}) = \prod_{\mu=1}^P \theta \left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}} \frac{\mathbf{J} \cdot \mathbf{s}^\mu}{\sqrt{N}} \right), \quad (\text{E.1})$$

que assume “1”, se o estudante candidato \mathbf{J} classifica todos os P exemplos disponíveis exatamente como professor e, “0”, se ele diferir em pelo menos um exemplo. Essa função nos diz, então, se o estudante \mathbf{J} continua ($\chi(\mathbf{J}) = 1$) ou não ($\chi(\mathbf{J}) = 0$) a ser compatível com o vetor professor, após a apresentação dos dados.

Pretende-se, então, calcular o número de estudantes compatíveis, à medida que se vão apresentando novos exemplos. Isso possibilitará a determinação do erro de generalização ótima para o caso do *perceptron* binário em função do número de exemplos apresentados. Após a apresentação de P exemplos, o número de estudantes compatíveis pode ser escrito como

$$\Omega(\{\mathbf{s}^\mu\}, \mathbf{w}) = \sum_{\mathbf{J}} \prod_{\mu=1}^P \theta \left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}} \frac{\mathbf{J} \cdot \mathbf{s}^\mu}{\sqrt{N}} \right). \quad (\text{E.2})$$

O conjunto desses estudantes compatíveis é chamado de *espaço de versões* [54]. Uma vez que todos os elementos desse conjunto são equiprováveis a serem a solução, ao escolher um deles ao acaso, a probabilidade de que este seja idêntico ao professor será $\frac{1}{\Omega}$. Essa descrição é chamada na literatura de *Aprendizado de Gibbs* [56] e dá o limite inferior do erro de generalização. Resumindo, será possível a partir do cálculo do espaço de versões, determinar o erro de generalização ótimo.

Para a determinação desse número de candidatos, parece então razoável estudar a média configuracional

$$\langle\langle \Omega(\{\mathbf{S}^\mu\}, \mathbf{w}) \rangle\rangle_{\mathbf{s}} = \int p(\mathbf{s}^\mu) \Omega(\{\mathbf{s}^\mu\}, \mathbf{w}). \quad (\text{E.3})$$

O espaço de versões (E.2) está relacionado com o produto de P distribuições independentes (veja (E.2)). É conhecido na literatura [54] que o produto de distribuições independentes geram uma distribuição com uma longa calda, o que implica que o valor mais provável dessa distribuição seja diferente do seu valor médio. Estamos interessados no valor médio do espaço de versões, mas com os procedimentos que

serão tomados, como, por exemplo, aproximações em torno do valor mais provável, chegar-se-á à um resultado não desejado. Por outro lado, se se considerar a função $\log \Omega(\{\mathbf{s}^\mu\}, \mathbf{w})$, o produto de (E.2) se transforma numa somatória de distribuições independentes e, de acordo com o Teorema Central do Limite (ver apêndice C), resulta numa distribuição gussiana. Agora sim, ter-se-á uma distribuição cujo valor mais provável e o valor médio coincidem. É então sensato calcular a média configuracional desse logaritmo, que dá origem à *Entropia Quenched* definida por

$$S = \langle \langle \log \Omega(\{\mathbf{s}^\mu\}, \mathbf{w}) \rangle \rangle_{\mathbf{s}}, \quad (\text{E.4})$$

e depois extrair do resultado as quantidades de interesse.

E.2 Truque de Réplicas

A média (E.4) pode ser resolvida utilizando o chamado *truque de réplicas* [58], que parte da identidade

$$\log Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n}. \quad (\text{E.5})$$

Por meio dessa identidade, a entropia *quenched* toma a forma

$$\langle \langle \log \Omega \rangle \rangle_{\mathbf{s}} = \lim_{n \rightarrow 0} \frac{\langle \langle \Omega^n \rangle \rangle_{\mathbf{s}} - 1}{n}, \quad (\text{E.6})$$

onde, para fins de simplificação, utiliza-se $\Omega \equiv \Omega(\{\mathbf{s}^\mu\}, \mathbf{w})$. O problema foi então transferido para o cálculo da média $\langle \langle \Omega^n \rangle \rangle_{\mathbf{s}}$. Porém, para n real, computar esta média é tão complexo quanto computar a média original. No entanto, para o caso em que n é inteiro, essa média pode ser realizada. Assuma-se essa hipótese e, depois que se obter um resultado dependente de n , faz-se a continuação analítica $n \rightarrow 0$. Pode-se então escrever

$$\begin{aligned} \langle \langle \Omega^n \rangle \rangle_{\mathbf{s}} &= \left\langle \left\langle \left[\sum_{\mathbf{J}} \prod_{\mu=1}^P \theta \left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu \mathbf{J} \cdot \mathbf{s}^\mu}{\sqrt{N}} \right) \right]^n \right\rangle \right\rangle_{\mathbf{s}} \\ &= \left\langle \left\langle \int \prod_{a=1}^n \sum_{\mathbf{J}^a} \prod_{\mu=1}^P \prod_{a=1}^n \theta \left(\frac{\mathbf{w} \cdot \mathbf{s}^\mu \mathbf{J}^a \cdot \mathbf{s}^\mu}{\sqrt{N}} \right) \right\rangle \right\rangle_{\mathbf{s}} \end{aligned} \quad (\text{E.7})$$

que tem a aparência de uma combinação de n diferentes cópias, ou réplicas, do modelo original com a realização do mesmo conjunto de exemplos.

Introduzindo as variáveis

$$h_\mu^a = \frac{\mathbf{J}^a \cdot \mathbf{s}^\mu}{\sqrt{N}} \quad (\text{E.8})$$

e

$$u_\mu = \frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}} \quad (\text{E.9})$$

na equação (E.7), temos

$$\begin{aligned} \langle\langle \Omega^n \rangle\rangle_{\mathbf{s}} &= \prod_{a=1}^n \sum_{\mathbf{J}^a} \prod_{a,\mu} dh_\mu^a \int \prod_u du_\mu \prod_{a,\mu} \Theta(u_\mu h_\mu^a) \\ &\times \left\langle\left\langle \prod_{a,\mu} \delta\left(h_\mu^a - \frac{\mathbf{J}^a \cdot \mathbf{s}^\mu}{\sqrt{N}}\right) \prod_\mu \delta\left(u_\mu - \frac{\mathbf{w} \cdot \mathbf{s}^\mu}{\sqrt{N}}\right) \right\rangle\right\rangle_{\mathbf{s}}. \end{aligned} \quad (\text{E.10})$$

As variáveis (E.8) e (E.9) representam os campos no estudante e no professor, respectivamente. Considerando ainda a representação Fourier da função Delta de Dirac e o limite $N \rightarrow \infty$, chega-se a

$$\begin{aligned} \langle\langle \Omega^n \rangle\rangle_{\mathbf{s}} &= \prod_{a=1}^n \sum_{\mathbf{J}^a} \int \prod_{a,\mu} \frac{dh_\mu^a \hat{h}_\mu^a}{2\pi} \int \prod_\mu \frac{du_\mu \hat{u}_\mu}{2\pi} \prod_{a,\mu} \Theta(u_\mu h_\mu^a) \\ &\times \exp \left\{ i \sum_{\mu,a} h_\mu^a \hat{h}_\mu^a + i \sum_\mu u_\mu \hat{u}_\mu - \frac{1}{2} \sum_{\mu,a} (\hat{h}_\mu^a)^2 + \right. \\ &\quad \left. - \sum_\mu \sum_{a>b} \hat{h}_\mu^a \hat{h}_\mu^b \frac{\mathbf{J}^a \cdot \mathbf{J}^b}{N} - \sum_{\mu,a} \hat{h}_\mu^a \hat{u}_\mu \frac{\mathbf{J}^a \cdot \mathbf{w}}{N} - \frac{1}{2} \sum_\mu \hat{u}_\mu^2 \right\} \end{aligned} \quad (\text{E.11})$$

onde \hat{h}_μ^a e \hat{u}_μ são as variáveis conjugadas de h_μ^a e u_μ , respectivamente.

Introduzir-se-á agora em (E.11) as variáveis auxiliares

$$q^{ab} = \frac{\mathbf{J}^a \cdot \mathbf{J}^b}{N}, \quad (\text{E.12})$$

o *overlap* entre duas réplicas, e

$$R^a = \frac{\mathbf{J}^a \cdot \mathbf{w}}{N}, \quad (\text{E.13})$$

o *overlap* entre uma réplica e o professor. Obtém-se com isso

$$\begin{aligned}
\langle\langle\Omega^n\rangle\rangle_s &= \int \prod_{a<b} \frac{dq^{ab}d\hat{q}^{ab}}{\frac{2\pi}{N}} \int \prod_a \frac{dR^a d\hat{R}^a}{\frac{2\pi}{N}} \exp \left\{ iN \sum_{a,b} q^{ab} \hat{q}^{ab} + iN \sum_a R^a \hat{R}^a \right\} \\
&\times \prod_{a=1}^n \sum_{\{J_j^a=\pm 1\}} \exp \left\{ -i \sum_{a<b} \hat{q}^{ab} \sum_j J_i^a J_i^b - i \sum_a \hat{R}^a \sum_j J_j^a \right\} \\
&\times \int \prod_\mu Du_\mu \int \prod_{\mu,a} \frac{dh_\mu^a d\hat{h}_\mu^a}{2\pi} \prod_{\mu,a} \Theta(u_\mu h_\mu^a) \\
&\times \exp \left\{ -\frac{1}{2} \sum_a (1 - (R^a)^2) \sum_\mu (h_\mu^a)^2 + \frac{1}{2} \sum_\mu \sum_{a \neq b} \hat{h}_\mu^a \hat{h}_\mu^b (R^a R^b - q^{ab}) + \right. \\
&\quad \left. + i \sum_{\mu,a} \hat{h}_\mu^a h_\mu^a - i u_\mu \sum_a \hat{h}_\mu^a R^a \right\} \tag{E.14}
\end{aligned}$$

Aqui foram introduzidas as variáveis \hat{q}^{ab} e \hat{R}^a conjugadas as variáveis q^{ab} e R^a , respectivamente. Foi aberta também a notação $\sum_{\mathbf{J}^a} = \sum_{\{J_j^a=\pm 1\}}$.

Na segunda linha da equação (E.14), os termos fatorizam-se em j , o que dá origem a uma somatória simples sobre J^a , elevada à potência N ; ou seja, essa segunda linha pode ser reescrita pela forma

$$\prod_{a=1}^n \sum_{\{J_j^a=\pm 1\}} \exp \{ \dots \} = \prod_{a,j} \sum_{J_j^a=\pm 1} e^{-i(\sum_{b>a} \hat{q}^{ab} J_j^b + \hat{R}^a) J_j^a} \tag{E.15}$$

$$\begin{aligned}
&= \left[\sum_{\{J^a=\pm 1\}} e^{-i \sum_{b>a} \hat{q}^{ab} J^a J^b - i \sum_a \hat{R}^a J^a} \right]^N \\
&= e^{N \log \left\{ \sum_{\{J^a=\pm 1\}} e^{-i \sum_{b>a} \hat{q}^{ab} J^a J^b - i \sum_a \hat{R}^a J^a} \right\}} \tag{E.16}
\end{aligned}$$

Agora, nas linhas 3, 4 e 5 da equação (E.14), os termos u_μ , h_μ^a e \hat{h}_μ^a fatorizam-se em μ e, portanto, podem ser reduzidos à somatória sobre as variáveis u , h^a e \hat{h}^a , respectivamente, elevados à potência $P = \alpha N$.

Finalmente, pode-se escrever

$$\begin{aligned}
\langle\langle\Omega^n\rangle\rangle_s &= \int \prod_{a<b} \frac{dq^{ab}d\hat{q}^{ab}}{\frac{2\pi}{N}} \int \prod_a \frac{dR^a d\hat{R}^a}{\frac{2\pi}{N}} \\
&\times e^{N[i \sum_{a<b} q^{ab} \hat{q}^{ab} + i \sum_a R^a \hat{R}^a + G_s(\hat{q}^{ab}, \hat{R}^a) + \alpha G_E(q^{ab}, R^a)]} \tag{E.17}
\end{aligned}$$

onde foi introduzido o termo

$$G_s(\hat{q}^{ab}, \hat{R}^a) = \log \sum_{\{J^a = \pm 1\}} e^{-i \sum_{b>a} \hat{q}^{ab} J^a J^b - i \sum_a \hat{R}^a J^a} \quad (\text{E.18})$$

que está relacionado com a entropia, uma vez que ele conta o número de vetores \mathbf{J} que obedecem ao vínculo E.13 e E.12; enquanto que

$$G_E(q^{ab}, R^a) = \log \left[\int Du \int \prod_a \frac{dh^a d\hat{h}^a}{2\pi} \Theta(uh^a) \right. \\ \left. \times e^{-\frac{1}{2} \sum_a (1-(R^a)^2)(\hat{h}^a)^2 - \frac{1}{2} \sum_{a \neq b} \hat{h}^a \hat{h}^b (q^{ab} - R^a R^b) + i \sum_a h^a \hat{h}^a - iu \sum_a \hat{h}^a R^a} \right] \quad (\text{E.19})$$

é o termo relacionado com a Energia, pois especifica a função custo ou regra de aprendizado que está sendo usada. Uma característica desse termo energético é que ele é sensível apenas às variáveis macroscópicas.

Dados o limite termodinâmico e a presença de N no exponencial da expressão (E.17), o ponto de máximo dessa função dever ser o termo mais relevante. Precisa-se, dessa forma, determinar os valores de $\{q^{ab}\}$, $\{R^a\}$, etc. que maximizam a função. No entanto computar esse ponto extremo é bastante difícil, especialmente pelo fato que se deve realizar no final dos cálculos o limite $n \rightarrow 0$ [54]. No entanto, esses cálculos podem ser simplificados consideravelmente se se assumirem simetrias entre réplicas, o que será feito na próxima seção.

E.3 Ansatz Réplica Simétrica

O chamado Ansatz de Réplica Simétrica consiste na simplificação

$$-i\hat{q}^{ab} = \hat{q} \quad (\text{E.20})$$

e

$$-i\hat{R}^a = \hat{R}. \quad (\text{E.21})$$

Dessa forma, pode-se escrever a forma simplificada

$$G_S(\hat{q}^{ab}, \hat{R}^a) = -\frac{n\hat{q}}{2} + n \int Dz \log 2 \cosh \left(\sqrt{\hat{q}} z + \hat{R} \right) + \mathcal{O}(n^2) \quad (\text{E.22})$$

$$G_E = \log 2 \int Dt \int_0^\infty Du H^n \left(-\frac{\sqrt{q - R^2 t} + Ru}{\sqrt{1 - q}} \right) \quad (\text{E.23})$$

Por meio desses resultados e realizando o limite $n \rightarrow 0$, conclui-se que a entropia *quenched* pode ser computada pelo método de ponto de sela, o que permite escrevê-la pelo seu termo mais relevante, i.e os valores de q, \hat{q}, R, \hat{R} que a maximiza, ou seja

$$S = \text{extr}_{q, \hat{q}, R, \hat{R}} \left[\frac{\hat{q}}{2}(q - 1) - \hat{R}R + \int Dz \log 2 \cosh(\sqrt{\hat{q}}z + \hat{R}) + 2\alpha \int Dt H \left(-\frac{Rt}{\sqrt{q - R^2}} \right) \log H \left(-\sqrt{\frac{q}{1 - q}} t \right) \right]. \quad (\text{E.24})$$

Considerando a simetria $q = R$ e $\hat{q} = \hat{R}$ e algumas transformações de variáveis de integração, conclui-se que a entropia é maximizada, quando há obediência das relações

$$R = \int Dz \tanh(\sqrt{\hat{R}}z + \hat{R}) \quad (\text{E.25})$$

$$\hat{R} = \frac{\alpha}{\pi \sqrt{1 - R}} \int Dt \frac{e^{-\frac{Rt^2}{2}}}{H(\sqrt{R}t)} \quad (\text{E.26})$$

Essas equações devem ser resolvidas numericamente, e pelo resultado, pode-se computar o erro de generalização pela relação

$$e_g = \frac{1}{\pi} \arccos R. \quad (\text{E.27})$$

O resultado desse erro de generalização em função de α está apresentado na figura (11.2).

Apêndice F

ARTIGO PUBLICADO

BIBLIOGRAFIA

- [1] Walsh B. J. . *On Phytophagous Varieties and Phytophagous Species*. Proc. Entomol. Soc. Phila. 3:403-430 (1864).
- [2] D. Ortiz-Barrientos, B. A. Counterman and M. A. F. Noor. *The Genetics of Speciation by Reinforcement*. PLoS Biol 2(12):e416 (2005).
- [3] C. Tauber and J. Tauber. *Phenotypic Plasticity in Chrysoperla: Genetic Variation in the Sensory Mechanism and in Correlated Reproductive Traits*. Evolution, vol. 46, 1754-1773 (Dez 1992).
- [4] MAF Noor. *Absence of Species Discrimination in Drosophila pseudoobscura and D. persimilis Males*. Anim Behav 52: 1205-1210 (1996).
- [5] M. Kirkpatrick and V. Ravigne *Speciation by Natural and Sexual Selection: Models and Experiments*. American Naturalist 159, S22 (2002).
- [6] H. Levene. *Genetic Equilibrium When More than One Ecological Niche is Available* American Naturalist 87, 836, pp. 331-333 (1953).
- [7] D. Bolnick, B. Fitzpatrick. *Sympatric Speciation: Theory and Empirical Data*. Annual Review of Ecology Evolution and Systematics 38, 459-487 (2008).
- [8] J. Maynard Smith. *Sympatric Speciation*. Am. Nat. 100, 637-650 (1966).
- [9] D. Udovic. *Frequency-Dependent Selection, Disruptive Selection, and the Evolution of Reproductive Isolation*. The American Naturalist, Vol. 116, No. 5 (Nov. 1980).
- [10] E. Mayr . *Animal Species and Evolution*. Cambridge: Harvard University Press (1963).

- [11] K. Luz-Burgoa, S. Moss de Oliveira, Veit Schwammle and J. S. Sa Martins. *Thermodynamic Behavior of a Phase Transition in a Model for Sympatric Speciation*. Physical Review E 74, 021910 (2006).
- [12] PS Pennings, M Kopp, G Meszéna, U Dieckmann, J Hermisson. *An Analytically Tractable Model for Competitive Speciation*. The American Naturalist vol.171, N0.1 E44 (2008).
- [13] M. Higashi, G. Takimoto and N. Yamamura. *Sympatric Speciation by Sexual Selection*. Nature 402, 523 (1999).
- [14] M. L. Rosenzweig. *Competitive Speciation*. Biol. J. Linn. Soc (London) 10, 275-289 (1978).
- [15] B. Drossel and A. Mackane. *Competitive Speciation in Quantitative Genetic Models*. J. Theor. Biol. 204, 467-478 (2000).
- [16] A. Kondrashov and F. Kondrashov. *Interactions Among Quantitative Traits in the Course of Sympatric Speciation*. Nature 400, 351 (1999).
- [17] U. Dieckmann and M. Doebeli. *On the Origin of Species by Sympatric Speciation*. Nature 400, 354 (1999).
- [18] J. Fry. *Multilocus Models of Sympatric Speciation: Bush versus Rice Versus Felsenstein*. Evolution, 57 (8), pp. 1735-1746 (2003).
- [19] A. Kondrashov. *Multilocus Model of Sympatric Speciation. III. Computer Simulations*. Theoretical Population Biology 29, 1-15 (1986).
- [20] J. M. Yeomans. *Statistical Mechanics of Phase Transitions*. Clarendon Press, Oxford (1994).
- [21] L Kadanoff. *Statistical Physics*. World Scietific Publishing Co. Pte. Ltd. (1999).
- [22] S. Geritz and E. Kisdi. *Adaptative Dynamics in Diploid, Sexual Populations and the Evolution of Reproductive Isolation*. The Royal Society of London, 1671 (2000).
- [23] M. Ridley. *Evolution*. Third Edition, Blackwell Publishing (2004).
- [24] Paulo Otto. *Genetica de Populacoes*. <http://dreyfus.ib.usp.br/bio212/>

- [25] J. Feder et al. *Evidence for Inversion Polymorphism Related to Sympatric Host Race Formulation in the Apple Maggot Fly, *Rhagoletis pomonella**. *Genetics* 163:939-953 (March 2003).
- [26] Sergey Gavrilets and David Waxman *Sympatric Speciation by Sexual Conflict* PNAS August 6, vol. 99 no. 16 (2002).
- [27] Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. *The Genetic Architecture of Divergence Between Threespine Stickleback Species*. *Nature*. 2001 Dec 20-27;414(6866):901-5.
- [28] Schluter, D. and McPhail, J. . *Ecological Character Displacement and Speciation in Sticklebacks*. *American Naturalist* 1992. Vol. 140, pp. 85 .
- [29] Nagel L. and Schluter, D. .*Body Size, Natural Selection, and Speciation in Sticklebacks*. *Evolution*, Vol. 52, No. 1 (Feb., 1998), pp. 209-218.
- [30] Hatfield, T. and Schluter, D.. *Ecological Speciation in Sticklebacks: Environment-Dependent Hybrid Fitness*. *Evolution*, Vol. 53, No. 3 (Jun., 1999), pp. 866-873.
- [31] Kraak, S , Mundwiler, B. and Hart, P. *Increased Number of Hybrids Between Benthic and Limnetic Three-Spined Sticklebacks in Enos Lake, Canada; The Collapse of a Species Pair?* *Journal of Fish Biology* Volume 58 Issue 5 Page 1458-1464, May 2001.
- [32] Gow, J. , Peichel, C and Taylor, E. *Contrasting Hybridization Rates Between Sympatric Three-Spined Sticklebacks Highlight the Fragility of Reproductive Barriers Between Evolutionarily Young Species*. *Molecular Ecology* Volume 15 Issue 3 Page 739-752, March 2006.
- [33] Taylor, E. , Boughman, J. , Groenenboom, M. , Sniatynski, M., Schluter, D. and Gow, L. *Speciation in Reverse: Morphological and Genetic Evidence of the Collapse of a Three-Spined Stickleback (*Gasterosteus aculeatus*) Species Pair*. *Molecular Ecology* volume 15 Issue 2 Page 343-355, February 2006.
- [34] Seehausen, O. , Alphen, J. and Witte, F. . *Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection*. *Science* 19 September 1997: Vol. 277. no. 5333, pp. 1808 - 1811 .

- [35] Grant, B. R. and Grant, P. R. . *Evolution of Darwin's Finches Caused by a Rare Climatic Event*. Proceedings: Biological Sciences, Vol. 251, No. 1331 (Feb. 22, 1993), pp. 111-117.
- [36] Bolnick, D. I., and B. Fitzpatrick. *Sympatric speciation: Theory and Empirical Data*. Annual Review of Ecology Evolution and Systematics. 38:459-487 (2007)..
- [37] Doebeli, M. *Adaptive Speciation When Assortative Mating is Based on Female Preference for Male marker trait*. J. Evol. Biol. 18 (2005) 1587-1600.
- [38] M. Doebeli. *A Quantitative Genetic Competition Model for Sympatric Speciation*. Evol. Biol. 9: X93 -909 (1996).
- [39] S. Gavrillets. *Models of Speciation: What Have We Learned in 40 years?* Evolution 57:2197?215 (2003).
- [40] Gavrillets, S and Hastings, A. *Intermittency and Transient Chaos from Simple Frequency-Dependent Selection*. Proceedings: Biological Sciences, Vol. 261, No. 1361 (Aug. 22, 1995), pp. 233-238 .
- [41] U. Paquet, O. Winther and M. Opper. *Perturbation Corrections in Approximate Inference: Mixture Modelling Applications*. Journal of Machine Learning Research 1 (2007) 1-28.
- [42] R. Alamino and N. Caticha. *Bayesian Online Algorithms for Learning in Discrete Hidden Markov Models*. Discrete and Continuous Dynamical Systems - Series B 9, 1-10 (2008) | arXiv:0708.2377v1 [stat.ML].
- [43] R. Alamino. *Aprendizado em Modelos de Markov com Variáveis de Estado Escondidas*. Tese de doutorado IF-USP (2005).
- [44] M. Opper and O. Winther. *A Bayesian Approach to On-line Learning*. On-line Learning in Neural Networks. Cambridge University Press (1999).
- [45] M. Opper and O. Winther. *Expectation Consistent Approximate Inference*. Journal of Machine Learning Research, vol. 6 pp. 2177 - 2204, 2005.
- [46] M. Kuss and C. Rasmussen. *Assessing Aproximate Inference for Binary Gaussian Process Classification*. Journal of Machine Learning Research, 6:1679-1704, 2005.

- [47] M. Opper and W. Kinzel. *Statistical Mechanics of Generalization*. Physics of Neural Networks III (1994).
- [48] M. Opper. *Improving on Expectation Propagation*. Artigo em preparação.
- [49] M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [50] Z. Ghahramani and M.J. Beal. *Graphical Models and Variational Methods*. In Saad and Opper (eds) *Advanced Mean Field Method: Theory and Practice*. Neural Information Processing Series. MIT Press (2000).
- [51] M. Mezard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press (2008).
- [52] T. P. Minka. *Expectation Propagation for Approximate Bayesian Inference*. UAI 2001, pp. 362–369, 2001.
- [53] T. P. Minka. *The EP Energy Function and Minimization Schemes*.
<http://research.microsoft.com/en-us/um/people/minka/papers/ep/minka-ep-energy.pdf>
- [54] A Engel, C Van den Broeck. *Statistical Mechanics of Learning*. Cambridge University Press (2001).
- [55] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience; 2 edition (July 18, 2006).
- [56] H. Nishimori. *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press (2001).
- [57] H Seung and H Sompolinsky. *Statistical Mechanics of Learning from Examples*. Physical Review A, 45 8 6056, (1992).
- [58] M. Mézard, G. Parisi and M. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.
- [59] S. Solla and O. Winther. *Optimal Perceptron Learning: an Online Bayesian Approach*. In Proceedings of the Newton Institute Workshop on On-Line Learning, Ed. D. Saad , 379-398, Cambridge University Press (1998).

-
- [60] O. Winther and S. Solla . *Optimal Bayesian online Learning*. Theoretical Aspects of Neural Computation (TANC'97), Eds. K.-Y. M. Wong, I. King, and D.-Y. Yeung, 61-70, Springer Singapore (1998).
- [61] O. Winther. *Bayesian Mean Field Algorithms for Neural Networks and Gaussian Process*. Ph.D. Thesis University of Copenhagen, The Niels Bohr Institute (1998).
- [62] M. Opper. *On-line versus Off-line Learning from Random Examples: General Results*. Physical Review Letters, 77 22 (1996).
- [63] M. J. Wainwright and M. I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*. Foundations and Trends in Machine Learning, 1 (2008). <http://dx.doi.org/10.1561/22000000001>
- [64] J. Hertz, A. Krogh and G. Thorbergsson. *Phase Transitions in Simple Learning*. J. Phys. A: Math. Gen. 22 2133 (1989).
- [65] L. Csató and M. Opper. *Efficient Gaussian Process, Inference and Hyper-Parameter Optimisation*. Artigo em preparação.

