UNIVERSIDADE FEDERAL DE LAVRAS

# CADEIA DE MARKOV COM ESTADOS LATENTES COM APLICAÇÃO EM ANÁLISES DE SEQÜÊNCIAS DE DNA

DEIVE CIRO DE OLIVEIRA

2005

**DEIVE CIRO DE OLIVEIRA**

# CADEIA DE MARKOV COM ESTADOS LATENTES COM APLICAÇÃO EM ANÁLISES DE SEQÜÊNCIAS DE DNA

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, Área de Concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Prof. Dr. Lucas Monteiro Chaves
(Orientador)

Prof. Dra. Cibele Queiroz da Silva
(Co-orientadora)

LAVRAS
MINAS GERAIS – BRASIL
2005

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da Biblioteca Central da UFLA**

Oliveira, Deive Ciro de
Cadeia de Markov com estados latentes com aplicação em análises de seqüências de DNA / Deive Ciro de Oliveira. -- Lavras : UFLA, 2005.
180 p. : il.

Orientador: Lucas Monteiro Chaves.
Dissertação (Mestrado) – UFLA.
Bibliografia.

1. Modelo de Markov. 2. Análise sequencial. 3. Estatistica. I. Universidade Federal de Lavras. II. Título.

CDD-519.233
-519.5

**DEIVE CIRO DE OLIVEIRA**

# CADEIA DE MARKOV COM ESTADOS LATENTES COM APLICAÇÃO EM ANÁLISES DE SEQÜÊNCIAS DE DNA

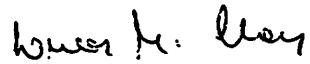
Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, Área de Concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

APROVADA em 28 de fevereiro de 2005

Prof. Dra. Cibele Queiroz da Silva - UFMG

Prof. Dra. Thelma Sáfadi - UFLA

Prof. Dr. Lucas Monteiro Chaves



LAVRAS
MINAS GERAIS – BRASIL
2005

À

família e amigos,

**OFEREÇO**

Meus pais, minha irmã e Carol,

**DEDICO**

## AGRADECIMENTOS

A meu pai Adelson Francisco de Oliveira e minha mãe Maria de Lourdes Oliveira pelo apoio e paciência; a minha irmã Dili Luiza de Oliveira e minha namorada Carolline Augusta Milani Baroni pelo afeto e compreensão.

Aos orientadores Lucas Monteiro Chaves, Cibele Queiroz da Silva, pela oportunidade de trabalhar em conjunto, paciência e ampla liberdade.

Aos familiares e amigos de maneira geral pelo afeto e carinho.

Aos professores que participaram de toda minha trajetória acadêmica superior.

## SUMÁRIO

## RESUMO

OLIVEIRA, Deive Ciro de. **Cadeia de Markov com Estados Latentes com aplicação em análises de seqüências de DNA.** LAVRAS: UFLA, 2004, 180 p. (Dissertação – Mestrado em Agronomia/Estatística e Experimentação Agropecuária)[1]

A teoria das Cadeias de Markov com estados latentes, *HMM's* (*hidden markov models*), é aplicada ao problema de discriminação de regiões homogêneas em seqüências de *DNA*. Uma abordagem didática é apresentada e os métodos de uma extensão do modelo são deduzidos. O algoritmo *EM* (Expectation-Maximization) é utilizado para obtenção de estimativas de máxima verossimilhança. Um software específico foi desenvolvido e aplicado na seqüência de *DNA* completa do vírus *bacteriófago lambda*, e em fragmentos do genoma das bactérias *Xylella fastidiosa*, *Escherichia coli*, *Streptococcus pneumoniae*, *Xanthomonas axonopodis* pv. citri. Os resultados obtidos são comparados com os trabalhos de Churchill (1989) e da Silva (2003).


[1] **Comitê Orientador:** Prof. Dr. Lucas Monteiro Chaves – UFLA (Orientador), Profª Drª Cibele Queiroz da Silva – UFMG e Profª. Drª. Thelma Sáfadi - UFLA.

## ABSTRACT

OLIVEIRA, Deive Ciro de. **Hidden Markov Chain with application in DNA sequence analysis.** LAVRAS: UFLA, 2004, 180 p. (Dissertação – Mestrado em Agronomia/Estatística e Experimentação Agropecuária[2])

*Hidden Markov models* (*HMM's*) are applied to locate segments with homogeneous *C+G* proportions (*DNA sequence segmentation*). The *HMM's* are showed in a didactic way and the methods of an extended model are obtained. The *EM* algorithm is used to obtain maximum-likehood estimators of *HMM's*. A software that implement some *HMM*'s methods was developed and applied to sequences fragments from *Xylella fastidiosa*, *Escherichia coli*, *Streptococcus pneumoniae*, *Xanthomonas axonopodis* pv. citri, and the complete genome of bacteriophage *lambda*. The results are discussed and compared with the research of Churchill (1989) and da Silva (2003).


[2] **Guidance Committee:** Prof. Dr. Lucas Monteiro Chaves – UFLA (Major Professor), Profª Drª Cibele Queiroz da Silva – UFMG e Profª Drª Thelma Sáfadi - UFLA.

# 1. INTRODUÇÃO

Nas últimas décadas, o desenvolvimento de novas tecnologias, em diversas áreas do conhecimento, proporcionou a obtenção de dados biológicos que, até então, eram impraticáveis de se estudar. Um exemplo disso foi o desenvolvimento do sequenciamento do código completo de *DNA* de um organismo. Isso só foi possível devido ao desenvolvimento de tecnologias na área de Bioquímica, Engenharia, Computação (Técnicas de sequenciamento) e Estatística (Modelagem).

É crescente a massa de dados biológicos tratada pela Biologia, em particular a Biologia Molecular. A dimensão destas massas de dados demanda a utilização de técnicas computacionais, bem como modelos teóricos para obtenção de informações pertinentes sobre estes dados. A área que realiza a implementação de ferramentas computacionais para análise e armazenamento de dados biológicos é denominada *Bioinformática*. A *Bioinformática* funciona como uma interface entre a área Biológica (Bioquímica, Biologia Molecular) e Exatas (Ciência da Computação, Engenharia, Matemática e Estatística).

Esta dissertação tem como objetivo o estudo, de maneira didática, de um modelo probabilístico, denominado *Cadeia de Markov com Estados Latentes (ou Ocultos)*, utilizado em uma série de aplicações ligadas à *Bioinformática*. Foi desenvolvido, nesta dissertação, um software na linguagem *Delphi*, que possibilita a manipulação e estimação dos parâmetros associados ao modelo em questão. Além disso, discutimos algumas aplicações das Cadeias de Markov com Estados Latentes. A maioria dos exemplos está relacionada a um tipo de aplicação específica associada à investigação de regiões funcionais em seqüências de *DNA*.

Esta dissertação está organizada como a seguir:

- Capítulo 1: Alguns conceitos de Biologia Molecular, particularizando a estrutura DNA e enfatizando o problema de *segmentação de DNA*. Além disso, realizamos uma breve introdução do tema *Modelos Markovianos com Estados Latentes*, apresentando uma primeira revisão de literatura sobre o assunto.

- Capítulo 2: Apresentação das Cadeias de Markov com Estados Latentes de maneira didática, mostrando os três problemas básicos associados ao modelo bem como os algoritmos que tratam tais problemas. São apresentados os métodos de normalização essenciais para aplicação do modelo à seqüência de grande dimensão. Além disso, são apresentados os algoritmos para uma extensão do modelo em questão.

- Capítulo 3: Aplicações das Cadeias de Markov com Estados Latentes não restritas à Bioinformática e a modelagem ao problema de segmentação de DNA.

- Capítulo 4: Apresentação de alguns resultados obtidos na aplicação das Cadeias de Markov com Estados Latentes ao problema de segmentação de *DNA*, utilizando seqüências reais de dados (segmentos das seqüências de *DNA* de organismos).

- Capítulo 5: Conclusões obtidas neste trabalho, especificando a aplicação do modelo ao problema de segmentação de DNA. Além disso, são apresentados possíveis trabalhos futuros.

Inicialmente vamos apresentar conceitos básicos da Biologia Molecular, os quais serão fundamentais para o entendimento do problema de segmentação de *DNA*.

## 1.1 Alguns conceitos de Biologia Molecular

Todos os conceitos aqui apresentados são baseados em Meidanis *et al* (1994), Kreuzer (2002), Metzler (1977), Lewin (1994) e Durbin et al (1998).

Desde os tempos antigos, os homens notaram que as características genéticas de um indivíduo eram similares às de seus pais. Gregor Mendel, ao realizar estudos com ervilhas, constatou que algumas características de uma planta eram passadas para gerações futuras. Para explicar suas hipóteses, Mendel propôs a existência de determinados fatores que regulavam estas características. Estes fatores foram denominados *Genes*. No entanto, restava ainda determinar a composição estrutural dos *Genes*, além de obter a forma (regras) com que os mesmos regulavam as características de um organismo.

Em meados do século XIX, uma série de trabalhos científicos mostrou que os genes eram compostos pelo chamado *Ácido Desoxirribonucléico (DNA)*. Mais tarde, foi comprovado que a maioria dos seres vivos tem o *DNA* como material genético. Isso porque alguns organismos possuem como material genético o *RNA*, que é uma estrutura similar ao *DNA*. A forma de regulação pela qual o *DNA* transmite as características genéticas foi descoberta no início do século XX. Tal forma consiste basicamente em associar um gene específico (seqüência de *DNA*) a uma determinada enzima. Enzimas são proteínas particulares, que têm a função de atuar no metabolismo dos organismos. Assim, os genes particularizavam as enzimas que, por sua vez, tornavam únicos os indivíduos.

O *DNA* é composto de uma fita dupla de formato helicoidal (vide figura 1).

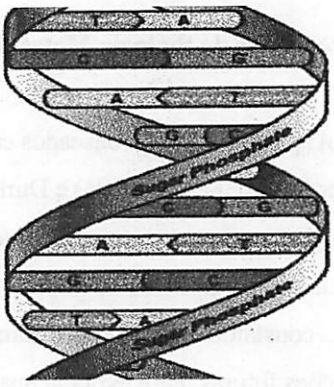


FIGURA 1. Representação do Ácido Desoxirribonucléico (*DNA*).

Cada uma das fitas é composta pela interligação de vários nucleotídeos (polinucleotídeos). O nucleotídeo tem, em sua composição estrutural, uma desoxirribose ligando-se a um fosfato e uma das quatro bases nitrogenadas (Adenina, Timina, Guanina e Citosina respectivamente representadas por *A*, *T*, *G*, *C*). Segundo as interligações entre os nucleotídeos, é possível distinguir as direções das seqüências (ligações de Carbono). Cada uma das fitas do *DNA* é interligada de maneira que, cada uma das bases em uma fita, associada a um respectivo nucleotídeo, corresponde a uma base complementar específica na outra fita. No caso do *DNA*, uma vez observada a ocorrência de uma base *A*, será observada, na outra fita, a ocorrência de *T*, assim como, se observada *G* em uma das fitas, a outra fita apresentará a base *C* na posição respectiva. Então, o conhecimento da seqüência de bases de uma das fitas do *DNA* implica no conhecimento total da estrutura.

Uma determinada proteína produzida em um organismo está associada a um gene específico. O gene é composto por um segmento de *DNA*. Um segmento de *DNA* pode ser definido pela seqüência de bases em uma das fitas presentes no *DNA*, ou simplesmente, por uma seqüência de caracteres tendo como alfabeto o conjunto $\{A, C, G, T\}$. Vale ressaltar que nem todo segmento do *DNA* é expresso (codifica) como uma proteína. Em determinados organismos grande parte da seqüência de *DNA* não possui funcionalidade conhecida. Em resumo, toda informação genética da maioria dos organismos, bem como a regulação de seu metabolismo, pode ser codificada em seqüências de caracteres. A informação genética necessária para a realização de todas as funções vitais associadas a um determinado organismo é denominada *Genoma*. O *Genoma* é composto de um certo número de cromossomos, que podem variar de acordo com a espécie. Os cromossomos são moléculas de *DNA* de grande dimensão.

Através da automatização do processo químico de sequenciamento de *DNA* descrito por Meidanis et al (1994) foi possível obter, em larga escala, as seqüências de bases nitrogenadas associadas a uma determinada molécula de *DNA*. A utilização desta técnica produziu um grande número de seqüências associadas aos organismos. Houve a necessidade de armazenar estas seqüências e analisá-las. A natureza dos dados (seqüências de caracteres) permite a aplicação de métodos de análise em texto (Busca e Comparação) (Durbin et al, 1998). Métodos associados a Reconhecimento de Padrões também são amplamente utilizados, possibilitando a aplicação de métodos probabilísticos.

Além da implementação de métodos para a análise de problemas ligados a este tipo peculiar de dados (seqüências de caracteres), existe a questão de armazenamento. Vários Bancos de Dados disponibilizam ferramentas de análise e consultas de seqüências de *DNA*, *RNA*, aminoácidos e proteínas, entre outras, através da *Internet*. Em particular, podemos citar o Genbank (*Genbank*), do qual algumas das seqüências de *DNA* utilizadas neste trabalho foram obtidas.

As técnicas de sequenciamento de *DNA* possibilitam a obtenção do código de DNA. Porém, a informação de que determinado trecho da seqüência é expresso (responsável pela codificação de uma proteína) como uma proteína qualquer não é conhecida. Em muitos casos, algumas proteínas codificadas pela seqüência de *DNA* são, isoladamente ou em conjunto, responsáveis por uma determinada funcionalidade de um organismo. Localizar estas regiões funcionais é um tipo de problema de análise de *DNA*. Alguns autores denominam este tipo de análise como *Segmentação de DNA* (Boys, 2004). Outros denominam este problema como *Análise de seqüências de DNA heterogêneas* (Churchill, 1992). No capítulo 3, apresentaremos o problema de maneira mais detalhada.

A seguir, iremos apresentar uma breve revisão de literatura sobre Cadeias de Markov com Estados Latentes, apresentando textos que tratam de aplicações diversas, enfatizando o problema de segmentação de *DNA*. Ilustraremos alguns artigos sobre o problema de ponto de mudança em variáveis aleatórias, particularizando o caso Bernoulli, que está correlacionado ao problema de segmentação de *DNA*.

## 1.2 Breve Revisão de Literatura

As Cadeias de Markov com Estados Latentes são modelos probabilísticos baseados em Cadeias de Markov e fazem parte da teoria de Processos Estocásticos. Basicamente estes modelos apresentam grande utilidade em problemas nos quais a intenção é modelar e emular sinais ou séries de resultados. Inicialmente, as Cadeias de Markov com Estados Latentes foram propostas por Baum (1966), tendo sua formalização complementada por uma série de artigos posteriores, publicados até meados da década de 70.

A partir da proposição do modelo inicial (Baum, 1966), diversos problemas foram modelados, utilizando as Cadeias de Markov com Estados

Latentes. Os Modelos Markovianos com Estados Latentes têm sido utilizados em áreas tão diversas quanto Climatologia (Hughes et al, 1999), Econometria (Ryden et al, 1998), Reconhecimento de Texto (Vlontzos et al, 1992), Processamento de Imagens (Aas et al, 1999) e Reconhecimento de Fala (Rabiner, 1989). No âmbito da *Bioinformática*, as Cadeias de Markov com Estados Latentes são aplicadas em uma série de problemas, por exemplo, o problema do *Alinhamento Múltiplo* (Gusfield, 1997) e o problema de *detecção de regiões homogêneas em seqüências de DNA* (Segmentação de *DNA*) (Churchill, 1989), os quais vão ser tratados no capítulo 3.

Existem vários textos, em formato de tutoriais, sobre o assunto. Um deles é devido a Rabiner (1989), onde o autor descreve os *Modelos Markovianos com Estados Latentes* de forma clara e sucinta. Este texto, apesar de ser um tutorial voltado à aplicação em Reconhecimento de Fala, é excelente para uma visão inicial destas cadeias. Rabiner (1989) apresenta a descrição do modelo associado a dados discretos e dados contínuos, a descrição dos três problemas básicos das *Cadeias de Markov com estados Latentes* (Problemas envolvendo cálculo de probabilidades e estimação) e os respectivos algoritmos de solução que serão abordados no capítulo 2. Neste texto apresentam-se ainda técnicas para contornar problemas de natureza numérica na realização de inferências sobre o modelo.

da Silva (2002), apresenta outro Tutorial, em português, sobre as Cadeias de Markov com Estados Latentes no qual são apresentados conceitos básicos sobre Probabilidades, Cadeias de Markov, algoritmo *EM* e por fim as Cadeias de Markov com estados Latentes. Apresentam-se ainda os algoritmos associados aos três problemas básicos do modelo. Ao final, citam-se algumas aplicações.

A aplicação básica desta dissertação está voltada para o problema de segmentação de *DNA*. O problema é caracterizado pela busca de distintos

segmentos de *DNA* homogêneos, que são responsáveis por diferentes funções de regulação celular de um determinado organismo. Este problema pode ser modelado utilizando métodos para a estimação de múltiplos pontos de mudança em seqüências de variáveis aleatórias (Hinkley et al, 1970) (Hinkley, 1970) (Smith, 1975). Contudo, este tipo de abordagem se mostra limitada à aplicação em problemas de *Segmentação de DNA*, devido às dependências entre as bases encontradas nos segmentos (Mais detalhes no capítulo 3).

Inicialmente, Churchill (1989) aplicou um modelo de *Cadeias de Markov com estados Latentes* ao problema de segmentação. No texto (Churchill, 1989) realiza-se a apresentação teórica do modelo, aplicando-o a algumas seqüências de dados reais (seqüências de *DNA*). Dentre estas seqüências, cita-se a seqüência do bacteriófago *Lambda* (Genbank), a qual será utilizada nesta dissertação. No trabalho de Churchill (1992), a temática de segmentação de *DNA* usando este tipo de modelo também é abordada. Alguns trabalhos posteriores seguiram este tipo de aplicação. No trabalho de Boys et al. (2000), os autores utilizam o mesmo modelo utilizado por Churchill (1989), com a diferença de que a abordagem adotada foi a *Bayesiana*. Boys et al (2004) estende o trabalho de Boys (2000), permitindo a escolha de modelos mais adequados, utilizando Inferência Bayesiana. Este tipo de procedimento não será abordado nesta dissertação. No trabalho de da Silva (2003), é aplicado um modelo idêntico ao formulado por Churchill (1989) em uma seqüência de *DNA* do organismo *Xylella fastidiosa*. O problema de segmentação de *DNA* será amplamente abordado no capítulo 3.

No capítulo seguinte, apresentaremos os *Modelos de Markov com Estados Latentes* de maneira formal. Apresentaremos modelos mais simples e alguns exemplos. Ainda apresentaremos os três problemas básicos associados ao modelo e a forma de obtenção dos algoritmos para resolução destes problemas. São obtidos os métodos que tratam de um Modelo de Markov com estados

latentes com uma estrutura de parâmetros diferenciada do modelo originalmente apresentado na literatura. Tal modelagem representa uma contribuição inovadora desta dissertação.

# 2. MODELOS MARKOVIANOS COM ESTADOS LATENTES

Os Modelos Markovianos com Estados Latentes, como visto no capítulo inicial, são estruturas probabilísticas amplamente utilizadas em uma série de aplicações ligadas a diversas áreas do conhecimento. Estes modelos são referenciados na literatura como *HMM* (*Hidden Markov Models*) e, por simplicidade, utilizaremos esta denominação em toda a dissertação.

No presente capítulo, o objetivo principal é apresentar, de forma didática, as definições básicas necessárias para o entendimento dos *HMM's*. De imediato, faz-se necessária a apresentação de modelos teóricos gerais que fornecem as estruturas matemáticas e probabilísticas requeridas para a descrição dos *HMM's*. Tais modelos são denominados processos estocásticos.

## 2.1 Processos Estocásticos

Os processos estocásticos são famílias arbitrárias de variáveis aleatórias $X_t$ indexadas por $t$ onde $t \in T$ (sendo $T$ um conjunto qualquer, geralmente $\mathbb{N}_+$). O índice $t$ pode ser visto como um indexador de tempo ou espaço. Os resultados assumidos por $X_t$ (contra-domínio de $X_t$) são denotados como conjunto dos estados do processo. Os processos estocásticos são definidos de maneira estritamente formal através do Teorema de Kolmogorov (Teorema Fundamental dos Processos Estocásticos) apresentado abaixo (Biswas, 1995).

**Definição 2.1** Se a função distribuição de probabilidade conjunta das variáveis aleatórias $X_{t_1}, X_{t_2}, \ldots X_{t_n}$ é conhecida para todo $n$ enumerável positivo, e para todo conjunto de valores $t_1, t_2, \ldots t_n$ onde $t_k$ qualquer pertence a

um conjunto $T$, podemos denotar o conjunto destas variáveis, $\{X_t\}$, por *Processo Estocástico*.

**Exemplo 2.1** ***Processo Estocástico.*** Um conjunto de variáveis aleatórias independentes da forma $X_t$, que possuem uma distribuição conjunta normal multivariada, onde $t \in \mathbb{Z}_+$.

Existem vários tipos de processos estocásticos. Como exemplos, temos os Processos de Poisson, Gaussianos, de Martingale (Biswas, 1995), e os *HMM's*. Vamos nos ater, inicialmente, a um processo estocástico particular denominado *Cadeia de Markov (Markov's Chain)*. Tal processo será de suma importância, pois os *HMM's* são generalizações das *Cadeias de Markov*.

## 2.2 Cadeia de Markov

Considere um sistema físico qualquer. Se as velocidades e posições das partículas são conhecidas, a cada momento, é possível estabelecer sua evolução. Entretanto, para obtermos informações sobre o tempo ou estado futuro, necessitamos apenas da configuração atual do sistema. Qualquer informação adicional sobre as posições e velocidades passadas das partículas torna-se irrelevante. De maneira análoga, a propriedade que caracteriza a *Cadeia de Markov* é que a ocorrência de eventos futuros depende, exclusivamente, dos eventos atuais. Uma Cadeia de Markov é definida como a seguir (Breiman,1969):

**Definição 2.2** Considere um espaço de eventos $\Omega$. O conjunto de variáveis aleatórias $X_1, X_2, \ldots$ onde $X_t \in \Omega$ para qualquer $t$ positivo e enumerável, é uma *Cadeia de Markov de ordem* 1 se:

$$P(X_t = x_t \mid X_1 = x_1, X_2 = x_2, \ldots X_{t-1} = x_{t-1}) =$$
$$P(X_t = x_t \mid X_{t-1} = x_{t-1}).$$

Observe que, em uma *Cadeia de Markov de ordem* 1, a probabilidade da ocorrência de $X_t$ depende exclusivamente da ocorrência da variável $X_{t-1}$. O conjunto de valores assumidos pelas variáveis $X_t$ é comumente denotado por estados. Então, de maneira equivalente, para uma *Cadeia de Markov de ordem* 1, podemos dizer que o estado que a cadeia assume em um instante futuro depende apenas do estado atual. A definição 2.2. pode ser generalizada, de forma que a dependência entre os estados seja ampliada.

**Definição 2.3** Considere um espaço de eventos $\Omega$. O conjunto de variáveis aleatórias $X_1, X_2, \ldots$, onde $X_t \in \Omega$ para qualquer $t$ positivo e enumerável, é uma *Cadeia de Markov de ordem r (ou de r-ésima ordem)*, se:

$$P(X_t = x_t \mid X_1 = x_1, X_2 = x_2, \ldots X_{t-1} = x_{t-1}) =$$
$$P(X_t = x_t \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2} \ldots X_{t-r} = x_{t-r}).$$

Pode-se observar que, nas definições 2.2 e 2.3, as variáveis com índice $t$ inicial (em 2.2 $t = 1$ e em 2.3 $t < r+1$) não estão contempladas nas relações de dependência. De modo a possibilitar o cálculo da distribuição conjunta dos

$X_t's$, faz-se necessária a definição de *Probabilidades Iniciais* da Cadeia na definição 2.4 (Breiman, 1969).

**Definição 2.4** Dada uma *Cadeia de Markov* de 1ª ordem, denotamos a *probabilidade inicial* de $X_1$ (ou densidade inicial de $X_1$) por $P(X_1 = x_1)$ (ou $f_{x_1}(x)$).

De forma similar, podemos generalizar a definição 2.4 para Cadeias de ordem superiores. A seguir, apresentaremos alguns exemplos de *Cadeia de Markov* tanto de natureza discreta quanto contínua.

**Exemplo 2.2** ***Cadeia de Markov com espaço de estados contínuos.*** Um alpinista vai realizar uma escalada de uma montanha de altitude $N$. É natural pensarmos que o desenvolvimento da escalada (acréscimo ou decréscimo de altitude, por exemplo) no dia subseqüente dependerá do ponto atingido no dia atual. Seja a variável aleatória $X_t$, designando a altitude atingida pelo alpinista no *t-ésimo* dia de escalada. A situação explicitada pode ser modelada como uma *Cadeia de Markov*, de 1ª ordem, de natureza contínua. O contra domínio de $X_t$ está contido no intervalo $[0, N]$. O valor aleatório de $X_t$ pode ser uma função da proximidade do cume da montanha. Um exemplo desta função é descrito abaixo:

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}) = \begin{cases} \dfrac{N - x_t}{N}, \text{ se } x_t > x_{t-1}; \\ \dfrac{x_t}{N}, \text{ se } x_t \leq x_{t-1}; \\ 0, \text{ em outro caso.} \end{cases}$$

Em vários casos as *Cadeias de Markov* possuem um conjunto de estados enumerável. Para trabalhar com este tipo de Cadeia, torna-se importante definir os chamados *Vetores de Probabilidades* ou *Estocásticos*.

**Definição 2.5** Um vetor de probabilidades $P = (p_1, p_2 ..., p_n) \in (0,1)^n$, satisfaz:

$$\text{I)} \quad \sum_{i=1}^{n} p_i = 1;$$

$$\text{II)} \quad p_i \geq 0, \text{ sendo que } 1 \leq i \leq n.$$

Uma *Matriz de Probabilidades* é tal, que suas linhas são formadas por *vetores de probabilidades*. Assim, podemos descrever uma definição de *Cadeia de Markov* em que a variável aleatória $X_t$ é discreta (uma *Cadeia de Markov* que assume um número de estados enumerável) (Boldrini et al, 1984).

**Definição 2.6** Uma *Cadeia de Markov* é um processo aleatório que pode assumir estados $s_1, s_2, ...s_r$, de tal modo que a probabilidade de transição do estado $s_i$ para $s_j$ é $a_{ij}$, onde $1 \leq i \leq r$ e $1 \leq j \leq r$.

A estrutura utilizada para armazenar todas as probabilidades de transição é uma matriz estocástica:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1r} \\ a_{21} & a_{22} & \ldots & a_{2r} \\ & & & \\ a_{r1} & a_{r2} & \ldots & a_{rr} \end{bmatrix}.$$

No caso de uma Cadeia de 1ª ordem, a distribuição inicial $P(S_1 = s_i) = \pi_i$, onde $1 \leq i \leq r$, é representada por um vetor de probabilidades.

$$\pi = [\pi_1 \quad \ldots \quad \pi_r].$$

A estrutura de parâmetros de uma Cadeia de Markov de 1ª ordem é composta pela matriz de transições $A$ e pelo vetor de distribuição inicial $\pi$.

Um exemplo clássico das Cadeias de Markov é o *Passeio Aleatório*. Segundo Breiman (1969), tal processo é descrito como no exemplo a seguir.

**Exemplo 2.3** ***Passeio Aleatório Unidimensional.*** Considere o movimento de uma partícula ao longo de um eixo horizontal através de posições inteiras $(\ldots, -k, 1-k, \ldots -1, 0, 1, \ldots, k-1, k, \ldots)$. Se a partícula estiver na posição $k$, num dado instante, a probabilidade que ela se mova para a posição $k+1$ no próximo instante é $p$ e a probabilidade de que ela se mova para a posição $k-1$ é $q = 1-p$. Quaisquer outros movimentos têm probabilidade zero de ocorrência. A situação em questão pode ser representada por uma *Cadeia de Markov* com número infinito de estados correspondendo às posições assumidas pela partícula através do tempo.

O passeio aleatório pode ser restrito, criando-se uma ou duas barreiras ao longo das posições assumidas pela partícula. Um caso particular do *Passeio Aleatório* é conhecido como "Problema da Ruína do Jogador" (*Gambler's Ruin Problem*).

**Exemplo 2.4** ***Ruína do Jogador.*** Considere um jogo disputado por dois jogadores (João e José), regido por um evento aleatório (por exemplo: o

lançamento de dados, de moedas, etc.) em um instante $t$. A cada momento, o jogador poderá apostar uma unidade monetária que possuir, podendo ganhar com uma probabilidade $p$ ou perder com probabilidade $1-p$. No momento inicial, $(t=1)$ João possui $k_1$ unidades monetárias e José possui $k_2$. A variação da quantidade de dinheiro que um jogador possui ao longo do tempo pode ser modelada como uma Cadeia de Markov discreta finita. Seja $S_t$ uma variável aleatória que designa a quantidade monetária que José possui no momento $t$. Esta variável pode assumir valores inteiros no intervalo $[0, k_1+k_2]$, ou seja, a Cadeia de Markov possui $k_1+k_2$ estados. Observe que, se ao longo do jogo, José atingir os estados 0 (José perdeu todo o dinheiro) ou $k_1+k_2$ (João perdeu todo o dinheiro), não será possível realizar uma transição para um estado distinto do atual. Tais estados são denominados *estados absorventes*.

**Exemplo 2.5** ***Modelo de Ehrenfest.*** (da Silva ,2002) Seja um total de $2n$ bolas numeradas. Suponha que as bolas são distribuídas aleatoriamente entre duas urnas, $I$ e $II$. No momento da distribuição, a urna $I$ contém $k$ bolas e a urna $II$, contém $2n-k$ bolas. Considere o experimento que consiste em sortear um número inteiro no intervalo $[0, 2n]$. A bola cujo número for sorteado deverá ser retirada de sua urna original e colocada na outra urna. Repetindo-se o experimento, tendo a repetição indexada por $t$, e considerando-se a variável aleatória $X_t$ como sendo o número de bolas na urna $I$ no momento $t$, podemos modelar este problema com uma *Cadeia de Markov*. Os estados são os possíveis números de bolas que a urna $I$ possui. Assim, esta cadeia possui $2n$ estados. Observe que, a partir de um estado $x$, só é possível realizar transições para os estados $x+1$ ou $x-1$. As probabilidades condicionais são dadas por:

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}) = \begin{cases} \frac{x_{t-1}}{2n}, \text{ se } x_t = x_{t-1} + 1; \\ 1 - \frac{x_{t-1}}{2n}, \text{ se } x_t = x_{t-1} - 1; \\ 0, \text{ em outro caso .} \end{cases}$$

O exemplo 2.5 é associado a um modelo famoso em Estatística Mecânica, denominado Modelo de *Ehrenfest*. Este modelo explica as trocas de calor em um sistema termodinâmico.

Em *Cadeias de Markov* cujo contra-domínio de $X_t$ (conjunto de estados) é um espaço discreto finito, podemos utilizar estruturas gráficas para a visualização do modelo. Estas ferramentas são denominadas *grafos* (Santos, 1995). Um grafo é definido por um conjunto de *nós*, representando os estados de um sistema, e um conjunto de *arestas* que estabelecem as interligações ou transições entre os nós. As arestas podem assumir valores (ou pesos) que dependem da modelagem específica. A Cadeia de Markov pode ser vista como um grafo cujas distribuições de probabilidade condicionais (vide definições 2.2 e 2.3) são os pesos das arestas oriundas de um estado.

**Exemplo 2.6** ***Clima de uma Região.*** (Ross, 1984) Suponha que, em uma dada região do planeta, as condições climáticas do dia seguinte dependam apenas das condições climáticas do clima do dia atual. Considere que a condição do tempo seja classificada em um "dia chuvoso" ou "dia não chuvoso". Considere ainda que, se chover no dia atual, a probabilidade de chover no dia seguinte seja $\alpha$. Além disso, se não chover no dia atual, a probabilidade de chuva no dia seguinte é $\beta$. Suponha que, no tempo $t = 1$, a probabilidade de um dia de chuva seja $\theta$. Denotando o estado "dia chuvoso" por 1 (estado $S_t = 1$) e

o estado "dia não chuvoso" por 0 (estado $S_t = 0$), temos uma *Cadeia de Markov* associada, em que as probabilidades são:

$$P(S_t = s_t \mid S_{t-1} = s_{t-1}) = \begin{cases} \alpha, \text{ se } s_t = 1 \text{ e } s_{t-1} = 1; \\ 1-\alpha, \text{ se } s_t = 0 \text{ e } s_{t-1} = 1; \\ \beta, \text{ se } s_t = 1 \text{ e } s_{t-1} = 0; \\ 1-\beta, \text{ se } s_t = 0 \text{ e } s_{t-1} = 0; \\ 0, \text{ em outro caso.} \end{cases}$$

A distribuição inicial da Cadeia é dada por:

$$P(S_1 = s_1) = \begin{cases} \theta, \text{ se } s_1 = 1; \\ 1-\theta, \text{ se } s_1 = 0; \\ 0, \text{ em outro caso.} \end{cases}$$

Podemos utilizar uma matriz e um vetor estocásticos para representação da cadeia. No contexto de *Cadeias de Markov*, a matriz estocástica é denominada *Matriz de Transições*. A matriz de transições e o vetor de distribuição inicial associados ao exemplo 2.6 são os seguintes:

$$A = \begin{bmatrix} 1-\beta & \beta \\ 1-\alpha & \alpha \end{bmatrix}, \qquad \pi = \begin{bmatrix} 1-\theta & \theta \end{bmatrix}.$$

A representação, em forma de grafo, da *Cadeia de Markov* associada no exemplo 2.6 é ilustrada através da figura 2.

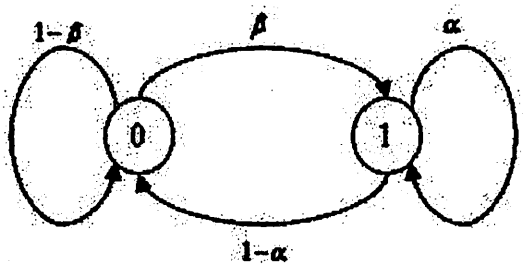



FIGURA 2. Grafo representando a Matriz de Transições do exemplo 2.6.

Em uma *Cadeia de Markov*, uma questão natural é o cálculo da probabilidade de ocorrência de uma seqüência de estados qualquer $S_1 = s_1, S_2 = s_2, ... S_T = s_T$. A *distribuição conjunta*, ou *verossimilhança* de uma seqüência de estados, considerando uma Cadeia de Markov de ordem 1, com Matriz de Transições $A$ e vetor de probabilidades iniciais $\pi$, é dada por:

$$P(S_1 = s_1, S_2 = s_2, ..., S_T = s_T) = P(S_1 = s_1) \times \prod_{t=2}^{T} P(S_t = s_t \mid S_{t-1} = s_{t-1})$$

$$= \pi_{s_1} \times \prod_{t=2}^{T} a_{s_{t-1} s_t}.$$

**Exemplo 2.7** ***Cálculo de verossimilhança.*** Considerando novamente o exemplo 2.6, se fixarmos $\theta = 0.6$, $\beta = 0.2$ e $\alpha = 0.3$, a probabilidade da ocorrência da seqüência de estados $S_1 = 0, S_2 = 1, S_3 = 0, S_4 = 0$, é apresentada abaixo:

$$P(S_1 = 0, S_2 = 1, S_3 = 0, S_4 = 0) = \pi_1 \times a_{12} \times a_{21} \times a_{22}$$
$$= (1-\theta) \times \beta \times (1-\alpha) \times (1-\beta)$$
$$= 0.4 \times 0.2 \times 0.7 \times 0.8 = 0.0448.$$

Outro aspecto interessante para análise é o *tempo de permanência* da Cadeia de Markov em um estado. A natureza deste comportamento é aleatória, sendo que a função distribuição de probabilidade que regula este evento é geométrica (Churchill, 1989). Seja uma Cadeia de Markov com matriz $A$ de transições entre os estados e vetor $\pi$ de distribuição inicial. Suponha que o sistema esteja em um estado $s$ em um tempo $t$ qualquer. Sendo $X$ a variável aleatória, que representa o tempo em que a Cadeia permanece no estado $s$, então $X$ segue distribuição geométrica com parâmetro $\lambda = a_{ss}$. Assim, $E[X] = \frac{a_{ss}}{1-a_{ss}}$ e $VAR[X] = \frac{a_{ss}}{(1-a_{ss})^2}$. Esta propriedade permite obter o tempo (em escala contínua) de permanência da cadeia em um estado qualquer.

**Exemplo 2.8** ***Tempo de Permanência em um estado markoviano.*** Considerando novamente o exemplo 2.6, suponha que em um dia $t$ qualquer não choveu na região. Fixando-se $\beta = 0.2$, a probabilidade da região ficar sem chuva, durante uma semana, é:

$$P(S_{t+1} = 0, S_{t+2} = 0, \ldots S_{t+7} = 0 \mid S_t = 0) = (a_{11})^7 (1 - a_{11})$$
$$= (1-\beta)^7 (\beta)$$
$$= (0.8)^7 (0.2) = 0.04194304.$$

Dado que não choveu no dia $t$, o número esperado de dias sem chuva e sua respectiva variabilidade (variância e desvio padrão) na região são dados por:

$$E[N^{\circ} \text{ de dias sem chuva}] = \frac{(a_{11})}{(1-a_{11})} = \frac{0.8}{0.2} = 4;$$

$$VAR[N^{\circ} \text{ de dias sem chuva}] = \frac{(a_{11})}{(1-a_{11})^2} = \frac{0.8}{0.04} = 20;$$

$$DP[N^{\circ} \text{ de dias sem chuva}] = \sqrt{\frac{(a_{11})}{(1-a_{11})^2}} = \sqrt{\frac{0.8}{0.04}} = 4.4721.$$

De acordo com os valores assumidos pela matriz de transições associada a uma Cadeia de Markov, podemos dizer que o processo é *estável*, à medida que ocorrem poucas transições entre estados distintos. Esta situação se acentua quando as diagonais da matriz de transições possuem valores próximos de um.

Na próxima sessão, introduziremos os *HMM's*.

## 2.3 Modelos Markovianos com Estados Latentes (*HMM's*)

Um *HMM* é um processo duplamente estocástico composto por uma Cadeia de Markov latente (não observável) $S_1, S_2, \ldots$, mas que se manifesta através de outro processo estocástico $O_1, O_2, \ldots$ (observável) (da Silva, 2002). Para visualizar o *HMM* de maneira intuitiva, relataremos um experimento envolvendo amostragem de bolas em urnas. Este exemplo vai elucidar os conceitos de *processo observável* e *processo latente*, fundamentais para o entendimento dos *HMM's*.

**Exemplo 2.9** ***Amostragem de bolas em Urnas.*** (da Silva, 2002) Considere $r$ urnas. Em cada urna existem bolas de $n$ cores distintas {branco, preto, azul...}, com proporções específicas para cada urna. Suponha que urnas e bolas são escolhidas de acordo com o experimento descrito a seguir:

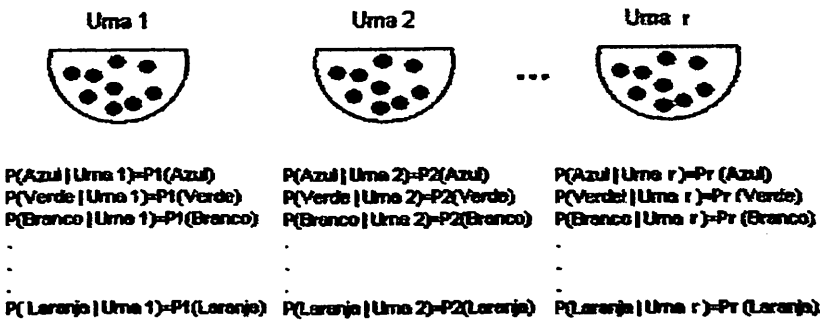



FIGURA 3. Modelo de Bolas e Urnas.

(1) Escolhe-se uma urna inicial por um processo aleatório qualquer. (2) Da urna escolhida sorteia-se uma bola com reposição, anotando-se sua cor. (3) Novamente escolhe-se uma urna baseando-se em um processo aleatório condicionado a urna anteriormente escolhida. (4) Retira-se a bola e registra-se sua cor. Repete-se o experimento a partir de (2) indefinidamente.

Vamos denotar por $S_t$ e $O_t$, respectivamente, as variáveis aleatórias associadas à urna sorteada e à cor da bola retirada no *t-ésimo* sorteio. Suponha que exista independência entre as bolas sorteadas, ou seja, que $O_1, O_2 \ldots$ são independentes, dado o conhecimento das urnas $S_1, S_2 \ldots$ sorteadas. Pela descrição do experimento, a única informação registrada é a cor da bola sorteada a cada momento, ou seja, as realizações de $O_1, O_2 \ldots$. Embora a urna $s_t$ escolhida no momento $t$ não seja observada, ela influi na probabilidade da cor da bola observada $o_t$.

No exemplo 2.9, existe uma Cadeia de Markov não observável no experimento. Esta cadeia está associada à escolha aleatória, a cada momento, da urna da qual se realizará o sorteio da bola. Os estados associados a esta cadeia

são as possíveis urnas (urna 1, urna 2, ... urna $r$). O evento observável do experimento (cor da bola) é dependente da urna sorteada. Isso porque as urnas podem conter diferentes proporções de cores de bolas. A cor da bola sorteada a cada momento corresponde à variável observável. Modificando o exemplo 2.9, podemos obter a situação em que as observações têm dependências entre si, dado o conhecimento dos estados.

Embora o exemplo 2.9 trate as seqüências $S_1, S_2, \ldots$ e $O_1, O_2, \ldots$ como sendo de natureza discreta, é possível generalizar este conceito. Tanto $S_1, S_2, \ldots$ quanto $O_1, O_2, \ldots$ podem ser seqüências de variáveis aleatórias contínuas ou discretas. Isso dá origem a *HMM's* discretos ou contínuos. Podem existir combinações dos dois tipos de variáveis aleatórias em que um dos processos pode ser discreto e o outro contínuo. Como a aplicação estudada neste trabalho trata de *HMM's* discretos, vamos restringir nossas definições posteriores a estes modelos em particular.

Para a representação de um *HMM* discreto, temos os parâmetros associados à Cadeia de Markov Latente discreta $S_1, S_2, \ldots$. São eles, a matriz de transições $A$, entre os estados latentes, e o vetor de distribuição inicial $\pi$. Além dos parâmetros $A$ e $\pi$, uma estrutura associada ao processo estocástico observável $O_1, O_2, \ldots$ deve estar presente. Tal estrutura incorpora probabilidades da observação $O_t$ dado um estado latente $S_t$, onde $O_t$ tem contradomínio enumerável e $t \in \mathbb{Z}_+$. Vamos denotá-la por $B$. Dependendo da relação de dependência entre as observações, existem variantes da estrutura de $B$. Com isso, podemos especificar a definição de *HMM's* em que os dois processos estocásticos (processo observável e latentes) são discretos.

**Definição 2.7** Seja um *HMM* definido por dois processos estocásticos $S_1, S_2, \ldots$ e $O_1, O_2, \ldots$, em que as variáveis aleatórias $O_t \mid S_t$ para todo $t$ são independentes entre si. As variáveis do tipo $S_t$ e $O_t$ têm contradomínios, $\Omega_S$ e $\Omega_O$, respectivamente discretos e $t \in \mathbb{Z}_+$. Um *HMM* é descrito por um modelo $\lambda$ tal que $\lambda = (\pi, A, B)$ (***HMM's* com $O_t$'s condicionalmente independente**), em que $\pi$ é o vetor de distribuição inicial, $A$ é a matriz de transições entre os estados latentes $S_t$ e $B$ é a matriz de distribuições condicionais de $O_t$'s dado um estado latente $S_t$, no qual:

$$\pi_{s_1} = P(S_1 = s_1);$$
$$a_{s_{t-1}s_t} = P(S_t = s_t \mid S_{t-1} = s_{t-1});$$
$$b_{s_t o_t} = P(O_t = o_t \mid S_t = s_t).$$

Existem *HMM's* em que se consideram, adicionalmente, a dependência markoviana de ordem 1 entre as variáveis $O_t \mid S_t$ para todo $t$ (***HMM's* com $O_t$'s condicionalmente dependente**). Neste caso, o modelo possui as mesmas estruturas, $\pi$, $A$ e $B$, e uma estrutura adicional $\pi o$ no qual:

$$\pi_{s_1} = P(S_1 = s_1);$$
$$\pi o_{s_1 o_1} = P(O_1 = o_1 \mid S_1 = s_1);$$
$$a_{s_{t-1}s_t} = P(S_t = s_t \mid S_{t-1} = s_{t-1});$$
$$b^{s_t}_{o_{t-1}o_t} = P(O_t = o_t \mid S_t = s_t, O_{t-1} = o_{t-1}).$$

Para efeito de notação, vamos denotar $r = \#\Omega_s$ e $n = \#\Omega_o$, ou em outras palavras, $r$ será o número de estados possíveis e $n$ o número de observações possíveis. Em muitas aplicações as distribuições iniciais são desconsideradas sendo o modelo composto somente pelas matrizes $A$ e $B$ associadas, respectivamente, à transição dos estados e a emissão de observações.

O número de parâmetros associados aos *HMM's* é uma informação interessante de se abordar. É desejável trabalhar com modelos mais simples (com menos parâmetros). Observe que num *HMM*, com ambos, o espaço de estados e o espaço de observações discretos, e que prevê independência entre as variáveis $O_t's$ dado o conhecimento dos estados, o número de parâmetros é $(r)(r-1)+(r-1)(n)+(r-1)$, correspondente a todos os elementos desconhecidos dos parâmetros $\pi, A$ e $B$. No modelo mais complexo que prevê dependência de primeira ordem entre as observações dado o conhecimento dos estados, o número de parâmetros é $(r)(r-1)+(r-1)(n)+(r-1)+(r)(n-1)(n)$.

A sequir, discutiremos alguns exemplos simples de *HMM's*, que servirão para fixar os conceitos discutidos até o momento.

**Exemplo 2.10 Lançamento de Moedas.** (da Silva, 2002) Uma pessoa tem em mãos três moedas. A moeda 1 é honesta. A moeda 2 é viciada e tem probabilidade $\frac{1}{3}$ do resultado "cara". A moeda 3 também é viciada, mas com probabilidade de "cara" igual a $\frac{3}{4}$. Considere um experimento que consiste nos seguintes passos:

1. Escolha, aleatoriamente, uma moeda no tempo 1, sendo $S_1$ a variável aleatória associada à moeda escolhida, seguindo a seguinte distribuição de probabilidade:

$$P(S_1 = 1) = \frac{1}{2},$$

$$P(S_1 = 2) = \frac{1}{4},$$

$$P(S_1 = 3) = \frac{1}{4}.$$

2. Lance a moeda escolhida e denote por $O_t$ a variável aleatória associada ao resultado da moeda, sendo $t$ a repetição do lançamento. Anote o resultado $o_t$.
3. Novamente escolha aleatoriamente uma moeda. Denote $S_t$ à variável aleatória associada à moeda escolhida no tempo $t$.

O processo aleatório de escolha da moeda é dependente da moeda escolhida anteriormente, segundo as seguintes probabilidades:

$$P(S_t = 1 | S_{t-1} = 1) = \frac{1}{2}, P(S_t = 2 | S_{t-1} = 1) = \frac{1}{4}, P(S_t = 3 | S_{t-1} = 1) = \frac{1}{4};$$

$$P(S_t = 1 | S_{t-1} = 2) = \frac{1}{3}, P(S_t = 2 | S_{t-1} = 2) = \frac{1}{3}, P(S_t = 3 | S_{t-1} = 2) = \frac{1}{3};$$

$$P(S_t = 1 | S_{t-1} = 3) = \frac{3}{4}, P(S_t = 2 | S_{t-1} = 3) = \frac{1}{8}, P(S_t = 3 | S_{t-1} = 3) = \frac{1}{8}.$$

Este experimento pode ser descrito como um *HMM* em que existem 3 estados latentes associados às moedas distintas. A única informação fornecida é o resultado do lançamento ("cara" ou "coroa"). Em nenhum momento a

informação sobre qual moeda foi lançada é conhecida. Neste exemplo, o *HMM* em questão é representado por $\lambda = (\pi, A, B)$, no qual:

$$\pi = [\frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{4}], A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}, B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}.$$

Inicialmente, o modelo possui 3 moedas, mas não há como saber se as moedas são todas lançadas, duas moedas são lançadas ou se uma moeda apenas é lançadas. Esta é uma informação extremamente importante, pois a partir da mesma, podemos construir *HMM*'s com um número maior ou menor de parâmetros (estados). Obviamente buscamos os modelos mais parcimoniosos. No caso do exemplo 2.10, o modelo possui 11 parâmetros.

**Exemplo 2.11 Tiro-ao-Alvo.** Dois atiradores participam de um experimento baseado em tiro-ao-alvo. O resultado obtido do disparo no alvo é classificado como acerto (1) ou erro (2). Os dois atiradores realizarão tiros no mesmo alvo. Um dos atiradores (atirador 1) possui extrema perícia, sendo de 0.8 sua probabilidade de acertar o alvo. O outro atirador (atirador 2) tem menos técnica que seu oponente, possuindo probabilidade 0.3 de acerto no alvo. O atirador 1, com mais técnica, propõe um esquema de rodízio entre os atiradores. Se o atirador 1 realizou o último disparo, então sua probabilidade de atirar novamente é 0.4. Se o atirador 2 realizou o último disparo, então sua probabilidade de atirar novamente é 0.1. No início do experimento, a probabilidade do atirador 1 efetuar o primeiro disparo é de 0.3. A probabilidade

do atirador 2 efetuar o primeiro disparo é 0.7. Denotemos por $O_t$ e $S_t$, respectivamente, o resultado obtido no *t-ésimo* tiro e o atirador que efetua o disparo no tempo $t$. Realizado o experimento, suponha que a seqüência de acertos e erros é anotada. Porém a informação sobre qual atirador realizou o disparo, a cada momento, é negligenciada. Este experimento pode ser modelado como um *HMM* com estados $S_t$ (atiradores) e observações $O_t$ (resultado dos tiros) discretas. Observe que cada estado possui uma característica particular na emissão de observações, uma vez que cada atirador tem uma perícia particular. Assim, os parâmetros $\pi$, $A$ e $B$ associados a este *HMM* são:

$$\pi = [0.3 \quad 0.7], A = \begin{bmatrix} 0.4 & 0.6 \\ 0.9 & 0.1 \end{bmatrix}, B = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}.$$

A função de verossimilhança associada a um *HMM* é baseada no princípio de dados aumentados. Este princípio é aplicado se alguma parte das informações é não observável. No caso do *HMM*, os estados não são observáveis. Dado $T$ realizações de $S_t$ e $T$ realizações independentes das variáveis $O_t \mid S_t$, a função de verossimilhança pode ser apresentada da seguinte forma:

$$\begin{aligned} & P(O_1 = o_1, \ldots, O_T = o_T, S_1 = s_1, \ldots, S_T = s_T \mid \pi, A, B) \\ & = P(S_1 = s_1) \times \left[ \prod_{t=1}^{T} P(O_t = o_t \mid S_t = s_t) \right] \times \left[ \prod_{t=2}^{T} P(S_t = s_t \mid S_{t-1} = s_{t-1}) \right] \\ & = \pi_{s_1} \times \left[ \prod_{t=1}^{T} b_{s_t o_t} \right] \times \left[ \prod_{t=2}^{T} a_{s_{t-1} s_t} \right]. \end{aligned}$$

Para um *HMM* que considera dependência de primeira ordem entre $O_t$'s dado o conhecimento dos estados, a função de verossimilhança considerando os dados aumentados é:

$$P(O_1=o_1,\ldots,O_T=o_T,S_1=s_1,\ldots,S_T=s_T \mid \pi,\pi o,A,B)$$
$$=P(S_1=s_1)\times P(O_1=o_1 \mid S_1=s_1)\times\left[\prod_{t=2}^{T}P(O_t=o_t \mid S_t=s_t,O_{t-1}=o_{t-1})\right]\times$$
$$\left[\prod_{t=2}^{T}P(S_t=s_t \mid S_{t-1}=s_{t-1})\right]$$
$$=\pi_{s_1}\times\pi o_{s_1 o_1}\times\left[\prod_{t=2}^{T}b^{o_t}_{s_t o_t}\right]\times\left[\prod_{t=2}^{T}a_{s_{t-1}s_t}\right].$$

Existem três problemas básicos relativos aos *HMM's*. O objetivo da próxima seção é mostrar os métodos de resolução destes três problemas.

## 2.4 Os três problemas básicos dos *HMM's*

É natural, quando se utiliza um modelo probabilístico qualquer, calcular a probabilidade associada a um dado evento. Outra questão importante é a obtenção de inferências sobre os parâmetros associados ao modelo. Os três problemas associados aos *HMM's* não são exceção e têm, como cerne, o cálculo de probabilidades e a realização de inferência estatística.

Dado um *HMM* composto pelos processos estocásticos $S_t$ (latente) e $O_t$ (observável) onde $t \in \mathbb{Z}_+$ parametrizado por $\lambda=(\pi,A,B)$ ou $\lambda=(\pi,\pi o,A,B)$ apresentam-se as seguintes questões:

1. Problema 1: Seja uma seqüência de realizações ou observações $O = \{O_1 = o_1, O_2 = o_2, ..., O_T = o_T\}$. Dado o modelo $\lambda$, qual a probabilidade da ocorrência da seqüência em questão, Isto é, qual é $P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T \mid \lambda)$?

2. Problema 2: Dada uma seqüência de realizações $O = \{O_1 = o_1, O_2 = o_2, ... O_T = o_T\}$, qual é a seqüência de realizações $S = \{S_1 = s_1, S_2 = s_2, ..., S_T = s_T\}$ que maximiza a probabilidade conjunta de $O$ e $S$? Isso é, qual a seqüência de estados $S$ que maximiza $P(S, O \mid \lambda)$?

3. Problema 3: Dada uma seqüência de realizações ou observações $O = \{O_1 = o_1, O_2 = o_2, ... O_T = o_T\}$, como podemos obter estimativas dos parâmetros do modelo $\lambda$?

### 2.4.1 Resolvendo o Problema 1

A obtenção de $P(O \mid \lambda)$ (verossimilhança) é uma informação bastante importante. Um exemplo de sua utilidade está na obtenção de testes de adequação de modelos (Teste de Razão de Verossimilhança) (Mood, 1963). Para encontrar $P(O \mid \lambda)$, devemos utilizar o Teorema da Probabilidade Total (Mood, 1963). Seja $\Omega_{ST} = \Omega_s \times \Omega_s \times ... \times \Omega_s$ (Produto Cartesiano de $T$ espaços unidimensionais $\Omega_s$) o conjunto de todas as seqüências de estados possíveis de tamanho $T$, podemos obter $P(O \mid \lambda)$ como:

$$P(O|\lambda) = \sum_{s \in \Omega_{ST}} P(O,S|\lambda) = \sum_{s \in \Omega_{ST}} P(O|S,\lambda)P(S|\lambda).$$

(1)

Desenvolvendo o primeiro dos termos internos do somatório da expressão (1) e supondo independência condicional entre $O_t's$ dados os $S_t's$, temos:

$$\begin{aligned}
&P(O|S,\lambda) = P(O_1 = o_1, O_2 = o_2, ..., O_T = o_T | S_1 = s_1, S_2 = s_2, ..., S_T = s_T, \lambda) \\
&= P(O_1 = o_1 | S_1 = s_1, \lambda) \times P(O_2 = o_2 | S_2 = s_2, \lambda) \times ... \times P(O_T = o_T | S_T = s_T, \lambda) \\
&= \prod_{t=1}^{T} P(O_t = o_t | S_t = s_t, \lambda) = \prod_{t=1}^{T} b_{s_t o_t}.
\end{aligned}$$

(2)

Em (1) resta-nos desenvolver o termo $P(S|\lambda)$. Este termo está associado aos estados da Cadeia de Markov Latente. Supondo dependência markoviana de ordem 1 entre os estados, temos:

$$\begin{aligned}
&P(S|\lambda) = P(S_1 = s_1, S_2 = s_2, ..., S_T = s_T | \lambda) \\
&= P(S_1 = s_1 | \lambda) \times P(S_2 = s_2 | S_1 = s_1, \lambda) \times ... \times P(S_T = s_T | S_{T-1} = s_{T-1}, \lambda) \\
&= P(S_1 = s_1 | \lambda) \times \left[ \prod_{t=2}^{T} P(S_t = s_t | S_{t-1} = s_{t-1}, \lambda) \right] \\
&= \pi_{s_1} \times \left[ \prod_{t=2}^{T} a_{s_{t-1} s_t} \right].
\end{aligned}$$

(3)

Portanto, de (2) e (3) temos:

$$P(O|\lambda)=\sum_{s\in\Omega_{ST}}\left[\pi_{s_1}\times\left(\prod_{t=2}^{T}a_{s_{t-1}s_t}\right)\times\left(\prod_{t=1}^{T}b_{s_t o_t}\right)\right] \tag{4}$$

É importante observar que o espaço $\Omega_{ST}$ representa o conjunto de todas as seqüências possíveis de estados presentes no modelo. Então, para calcular $P(O|\lambda)$, devemos obter todas as seqüências $S$ de estados possíveis, obtendo para cada seqüência, em particular, a respectiva verossimilhança.

**Exemplo 2.12** ***Cálculo de*** $P(O|\lambda)$ ***com base na equação (4).*** Suponha que no experimento que envolvia os atiradores, no exemplo 2.11, a seqüência observada foi $O=\{O_1=1,O_2=2\}$ com *acerto=1* e *erro=2*. Este *HMM* tem parâmetros:

$$\pi=[0.3 \quad 0.7],\ A=\begin{matrix}1\\2\end{matrix}\begin{bmatrix}0.4 & 0.6\\ 0.9 & 0.1\end{bmatrix},\ B=\begin{matrix}1\\2\end{matrix}\begin{bmatrix}0.8 & 0.2\\ 0.3 & 0.7\end{bmatrix}.$$

Devemos obter todas as seqüências $S$ de estados possíveis associadas à realização de $O$. Temos 2 atiradores, isto é, 2 estados. Dado que $O$ representa uma seqüência de 2 tiros, devemos buscar todas as seqüências de dois atiradores possíveis. Este conjunto de seqüências contém todos os possíveis arranjos, de tamanho $T=2$, dos atiradores:

$$\Omega_{ST}=\{(s_1=1,s_2=1),(s_1=1,s_2=2),(s_1=2,s_2=1),(s_1=2,s_2=2)\}.$$

O passo seguinte é realizar o cálculo da verossimilhança conjunta para cada uma das seqüências de estados possíveis. Temos assim as seguintes quantidades:

$$P(O=o,S=s\mid\lambda)=\pi_{s_1}\times\prod_{t=2}^{T}a_{s_{t-1}s_t}\times\prod_{t=1}^{T}b_{s_to_t};$$

Para s $=\{s_1=1,s_2=1\}$ e $o=\{o_1=1,o_2=2\}$,
$P(O_1=1,O_2=2,S_1=1,S_2=1\mid\lambda)=\pi_1\times a_{11}\times b_{11}\times b_{12}$
$=0.3\times0.4\times0.8\times0.2=0.0192.$

Para $s=\{s_1=1,s_2=2\}$ e $o=\{o_1=1,o_2=2\}$,
$P(O_1=1,O_2=2,S_1=1,S_2=2\mid\lambda)=\pi_1\times a_{12}\times b_{11}\times b_{22}$
$=0.3\times0.6\times0.8\times0.7=0.1008.$

Para $s=\{s_1=2,s_2=1\}$ e $o=\{o_1=1,o_2=2\}$,
$P(O_1=1,O_2=2,S_1=2,S_2=1\mid\lambda)=\pi_2\times a_{21}\times b_{21}\times b_{12},$
$=0.7\times0.9\times0.3\times0.2=0.0378.$

Para $s=\{s_1=2,s_2=2\}$e $o=\{o_1=1,o_2=2\}$,
$P(O_1=1,O_2=2,S_1=2,S_2=2\mid\lambda)=\pi_2\times a_{22}\times b_{21}\times b_{22},$
$=0.7\times0.1\times0.3\times0.7=0.0147.$

O procedimento seguinte consiste em adicionarmos as verossimilhanças parciais obtidas de todas as seqüências de estados possíveis presentes em $\Omega_{ST}$. Assim, temos que a verossimilhança da seqüência de observações $o=\{o_1=1,o_2=2\}$ é:

$$P(O=o\mid\lambda)=\sum_{s\in\Omega_{ST}}(\pi_{s_1}\times\prod_{t=2}^{T}a_{s_{t-1}s_t}\times\prod_{t=1}^{T}b_{s_to_t})$$
$$=0.0192+0.1008+0.0378+0.0147$$
$$=0.1725.$$

Visto que em trabalhos reais nos quais utilizam-se longas seqüências de observações, o cálculo de $P(O\mid\lambda)$ feito, diretamente a partir da expressão (1), torna-se impraticável. Consideremos as operações de somas e multiplicações envolvidas na obtenção de $P(O\mid\lambda)$. No cálculo de $P(O,S\mid\lambda)$, para uma dada seqüência $s\in\Omega_{ST}$, são necessárias $2T-1$ operações de multiplicação. Temos o conjunto $\Omega_{ST}$ composto por $r^T$ seqüências de tamanho $T$ em que $r$ é o número de estados. Como devemos realizar, para cada elemento de $\Omega_{ST}$, o cálculo de $P(O,S\mid\lambda)$, temos um total parcial de $(r^T)\times(2T-1)$ operações. Observe que também precisamos realizar $r^T-1$ somas para obter $P(O\mid\lambda)$. Isto totaliza um número de operações de:

$$\text{nº de operaçoes} = (r^T)\times(2T-1)+(r^T-1)=2Tr^T-1.$$

O número de operações total é de natureza exponencial em função do tamanho da seqüência de observações e polinomial no número de estados. Isso não é desejável, pois, à medida que o tamanho da seqüência de observações aumenta, o custo computacional aumenta de forma explosiva. Para o exemplo 2.11, tendo $T=2$ e $r=2$, temos 15 operações envolvidas. Entretanto, é possível efetuar o cálculo de forma mais eficiente.Observe, por exemplo, as operações $\pi_1\times b_{11}$ e $\pi_1\times b_{21}$. Neste exemplo simples, cada uma destas operações é, de forma

desnecessária, realizada 2 vezes. Existem 2 algoritmos que eliminam este tipo de desperdício no cálculo de $P(O|\lambda)$. Estes métodos, denominados *forward* e *backward*, serão vistos na seguinte seção.

#### 2.4.1.1 Procedimento *Forward*

O método *forward* é baseado em princípios de Programação Dinâmica. A idéia é tentar eliminar operações redundantes na solução de um problema. O método *forward*, desenvolvido através do uso de operações recorrentes, torna possível o cálculo de $P(O|\lambda)$, com redução significativa do custo computacional.

Para entendermos o método *forward*, é necessário desenvolvermos de maneira algébrica, a expressão $P(O|\lambda)$. Desta forma, obteremos um método computacional no qual as operações não sejam repetidas desnecessariamente.

Vamos apresentar, nas duas subseções seguintes, os modelos que consideram, respectivamente, independência e dependência condicional entre as observações dado o conhecimento dos estados.

##### 2.4.1.1.1 Procedimento *Forward* (*HMM* com $O_t | S$ condicionalmente independentes)

Desenvolvendo $P(O|\lambda)$ em que $T$ é o tamanho da seqüência $O$, temos:

$$P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T \mid \lambda)$$
$$= \sum_{s_T \in \Omega_s} P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T, S_T = s_T \mid \lambda).$$

(5)

Supondo independência condicional dos $O_t's$, dados os $S_t's$, e dependência markoviana de primeira ordem entre os $S_t's$, o termo geral no somatório em (5) pode ser expresso por:

$$\begin{aligned}
&P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T, S_T = s_T \mid \lambda) \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(O_1 = o_1, O_2 = o_2, ..., O_T = o_T, S_{T-1} = s_{T-1}, S_T = s_T \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(S_T = s_T, O_T = o_T \mid O_1 = o_1, O_2 = o_2, ..., O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 ..., O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(S_T = s_T, O_T = o_T \mid S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 ..., O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(O_T = o_T \mid S_T = s_T, S_{T-1} = s_{T-1}, \lambda) \times P(S_T = s_T \mid S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 ..., O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(O_T = o_T \mid S_T = s_T, \lambda) \times P(S_T = s_T \mid S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 ..., O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= P(O_T = o_T \mid S_T = s_T, \lambda) \times \sum_{s_{T-1} \in \Omega_s} \left[ P(S_T = s_T \mid S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 ..., O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right].
\end{aligned}$$

(6)

Para $2 \le t \le T$, (6) se reduz a:

$$P(O_1 = o_1, O_2 = o_2 \ldots, O_t = o_t, S_t = s_t \mid \lambda)$$
$$= P(O_t = o_t \mid S_t = s_t, \lambda) \times \sum_{s_{t-1} \in \Omega_s} \left[ P(S_t = s_t \mid S_{t-1} = s_{t-1}, \lambda) \times \right.$$
$$\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{t-1} = o_{t-1}, S_{t-1} = s_{t-1} \mid \lambda) \right]. \tag{7}$$

Denotando:

$$\alpha_{(t,s_t)} = P(O_1 = o_1, O_2 = o_2 \ldots, O_t = o_t, S_t = s_t \mid \lambda). \tag{8}$$

Podemos reescrever (7) como:

$$\alpha_{(t,s_t)} = \sum_{s_{t-1} \in \Omega_s} b_{s_t o_t} \times a_{s_{t-1} s_t} \times \alpha_{(t-1, s_{t-1})}$$
$$= b_{s_t o_t} \times \sum_{s_{t-1} \in \Omega_s} \left[ a_{s_{t-1} s_t} \times \alpha_{(t-1, s_{t-1})} \right]. \tag{9}$$

Para $t = 1$, temos:

$$\alpha_{(1,s_1)} = b_{s_1 o_1} \times \pi_{s_1}. \tag{10}$$

Logo o valor de $P(O \mid \lambda)$ é:

$$P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T \mid \lambda)$$
$$= \sum_{s_T \in \Omega_s} P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T, S_T = s_T \mid \lambda)$$
$$= \sum_{s_T \in \Omega_s} \alpha_{(T,s_T)}. \qquad (11)$$

Os elementos de $\alpha_{(t,s_t)}$ podem ser dispostos em uma matriz que possui $T$ linhas e $r$ colunas.

$$\alpha = \begin{bmatrix} \alpha_{(1,1)} & \cdots & \alpha_{(1,r)} \\ \vdots & \ddots & \vdots \\ \alpha_{(T,1)} & \cdots & \alpha_{(T,r)} \end{bmatrix}$$

No trabalho de Rabiner(1989) $\alpha_{(t,s_t)}$ é chamada de variável *forward*.

A obtenção de $P(O \mid \lambda)$ se resume ao cálculo dos valores dos termos $\alpha_{(t,s_t)}$ da referida matriz e, posteriormente, da soma dos elementos da última linha. Observe que existem dependências entre os valores dos $\alpha_{(t,s_t)}$'s. A construção de cada elemento $\alpha_{(t,s_t)}$ é realizada em função dos elementos $\alpha_{(t-1,s_t)}$ da linha anterior. A figura 4, adaptada de Rabiner (1989), ilustra este aspecto.

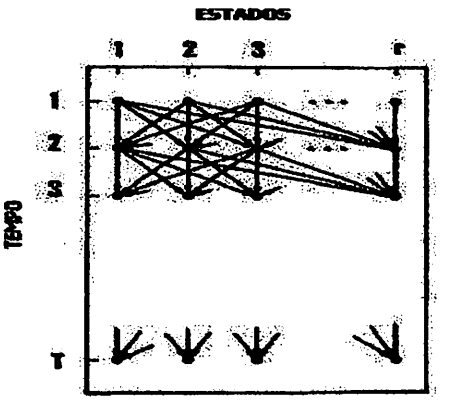



FIGURA 4. Dependência de $\alpha_{(t,s_t)}$.

Desta forma, o algoritmo para o cálculo de $P(O|\lambda)$, para uma dada seqüência de observações $O=\{O_1=o_1, O_2=o_2, ...O_T=o_T\}$, pode ser descrito através dos seguintes passos:

1. Faça $\alpha_{(1,s_1)}=b_{s_1 o_1}\times\pi_{s_1}$ para todo $s_1\in\Omega_s$.
2. Faça $\alpha_{(t,s_t)}=b_{s_t o_t}\times\sum_{s_{t-1}\in\Omega_s}\left[a_{s_{t-1}s_t}\times\alpha_{(t-1,s_{t-1})}\right]$ para todo $s_t\in\Omega_s$ e para a seqüência de observações $O_1=o_1, O_2=o_2, ..., O_T=o_T$.
3. Obtenha $P(O|\lambda)=\sum_{s_T\in\Omega_s}\alpha_{(T,s_T)}$.

**Exemplo 2.13** ***Método forward (HMM com $O_t|S$ independentes).*** Utilizando o experimento do exemplo 2.11, envolvendo os atiradores, vamos calcular $P(O|\lambda)$, utilizando o método *forward*. Dado que a seqüência de observações ou tiros foi $O=\{O_1=1, O_2=2\}$, temos:

$$\alpha_{(1,1)} = b_{11} \times \pi_1 = 0.8 \times 0.3 = 0.24,$$
$$\alpha_{(1,2)} = b_{21} \times \pi_2 = 0.3 \times 0.7 = 0.21,$$
$$\alpha_{(2,1)} = b_{12} \times \sum_{i=1}^{2} (a_{i1} \times \alpha_{(1,i)}) = 0.2 \times (0.24 \times 0.4 + 0.21 \times 0.9) = 0.057,$$
$$\alpha_{(2,2)} = b_{22} \times \sum_{i=1}^{2} (a_{i2} \times \alpha_{(1,i)}) = 0.7 \times (0.24 \times 0.6 + 0.21 \times 0.1) = 0.1155.$$

Assim, a matriz associada ao método *forward* é:

$$\alpha = \begin{bmatrix} 0.2400 & 0.2100 \\ 0.0570 & 0.1155 \end{bmatrix}.$$

Como descrito na expressão (5), dado o modelo $\lambda$ em estudo, $r = 2$ e $T = 2$, a probabilidade da seqüência observada é dada por:

$$P(O \mid \lambda) = \sum_{i=1}^{r} \alpha_{(T,i)} = \alpha_{(2,1)} + \alpha_{(2,2)} = (0.057 + 0.1155) = 0.1725.$$

Utilizando o procedimento *forward*, é possível diminuir o número de operações (somas e multiplicações) necessárias no cálculo de $P(O \mid \lambda)$. Observe que o cálculo de $P(O \mid \lambda)$ envolve a construção da matriz $\alpha$ e o cálculo da soma da última linha. Para cada um dos elementos da 1ª linha de $\alpha$, o custo envolvido é de uma operação, totalizando $r$ operações. Para cada uma das células restantes de $\alpha$, são necessárias $r-1$ somas e $r$ multiplicações. Como a matriz $\alpha$ possui $(r \times T) - r$ elementos, excluindo-se a 1ª linha, o custo total na obtenção dos elementos de $\alpha$ é de $((r \times T) - r) \times (r - 1 + r)$ operações. Resta

contabilizar o custo do somatório para encontrar $P(O \mid \lambda)$. Neste somatório são realizadas mais $r-1$ somas. Contabilizando todas as operações, temos:

$$\begin{aligned} &\text{nº de operaçoes} = ((r-1)+r)\times(r\times(T-1))+r+(r-1) \\ &= 2r^2(T-1)-r(T-3)-1. \end{aligned}$$

Desta forma, o número de operações necessárias no cálculo de $P(O \mid \lambda)$, utilizando o método *forward*, é da ordem de $r^2T$ operações contra $2Tr^T$ operações requeridas pelo cálculo direto (equação 4). No exemplo 2.11 em que $T=2$ e $r=2$, necessitamos de 11 operações, número menor que as 15 operações realizadas pelo *cálculo direto*.

Quando $T$ cresce, a diferença entre os dois métodos (*foward* e cálculo direto) se acentua ainda mais. Como exemplo, considere uma aplicação em que $T=10$ e $r=2$. O número de operações necessárias para execução do método *forward* é 66, enquanto o cálculo direto necessita de 20479 operações.

#### 2.4.1.1.2 Procedimento *Forward* (*HMM* com $O_t \mid S$ condicionalmente dependentes)

Utilizando o mesmo princípio do método *forward* aplicado na subseção anterior, podemos obter $P(O \mid \lambda)$ para um modelo que prevê dependência de primeira ordem entre as observações $O_t$'s dados $S_t$'s. Lembrando que:

$$\begin{aligned} &P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2 \ldots, O_T = o_T \mid \lambda) \\ &= \sum_{s_T \in \Omega_s} P(O_1 = o_1, O_2 = o_2 \ldots, O_T = o_T, S_T = s_T \mid \lambda) \end{aligned} \qquad (12)$$

nosso objetivo é desenvolver o termo geral do somatório em (12), com a finalidade de encontrar uma relação de recorrência.

Desenvolvendo o termo geral, respeitando as restrições impostas neste modelo, temos que:

$$
\begin{aligned}
&P(O_1 = o_1, O_2 = o_2 \ldots, O_T = o_T, S_T = s_T \mid \lambda) \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T, S_{T-1} = s_{T-1}, S_T = s_T \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(S_T = s_T, O_T = o_T \mid O_1 = o_1, O_2 = o_2, \ldots, O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(S_T = s_T, O_T = o_T \mid S_{T-1} = s_{T-1}, O_{T-1} = o_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(O_T = o_T \mid S_T = s_T, S_{T-1} = s_{T-1}, O_{T-1} = o_{T-1}, \lambda) \times \right. \\
&P(S_T = s_T \mid S_{T-1} = s_{T-1}, O_{T-1} = o_{T-1}, \lambda) \times \\
&\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= \sum_{s_{T-1} \in \Omega_s} \left[ P(O_T = o_T \mid S_T = s_T, O_{T-1} = o_{T-1}, \lambda) \times P(S_T = s_T \mid S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right] \\
&= P(O_T = o_T \mid S_T = s_T, O_{T-1} = o_{T-1}, \lambda) \times \sum_{s_{T-1} \in \Omega_s} \left[ P(S_T = s_T \mid S_{T-1} = s_{T-1}, \lambda) \times \right. \\
&\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{T-1} = o_{T-1}, S_{T-1} = s_{T-1} \mid \lambda) \right]. \qquad (13)
\end{aligned}
$$

Para $2 \leq t \leq T$, generalizando (13), temos:

$$P(O_1 = o_1, O_2 = o_2 \ldots, O_t = o_t, S_t = s_t \mid \lambda)$$
$$= P(O_t = o_t \mid S_t = s_t, O_{t-1} = o_{t-1}, \lambda) \times \sum_{s_{t-1} \in \Omega_s} \left[ P(S_t = s_t \mid S_{t-1} = s_{t-1}, \lambda) \times \right.$$
$$\left. P(O_1 = o_1, O_2 = o_2 \ldots, O_{t-1} = o_{t-1}, S_{t-1} = s_{t-1} \mid \lambda) \right]. \tag{14}$$

Considerando as estruturas paramétricas nas quais a dependência de primeira ordem entre as observações $O_t's$ é considerada, dados $S_t's$, e nos utilizando da variável $\alpha_{d_1} = P(O_1 = o_1, O_2 = o_2 \ldots, O_t = o_t, S_t = s_t \mid \lambda)$, podemos escrever a expressão (14) como:

$$\alpha_{d_1(t,s_t)} = b^{s_t}_{o_{t-1}o_t} \times \sum_{s_{t-1} \in \Omega_s} (a_{s_{t-1}s_t} \times \alpha_{d_1(t-1,s_{t-1})}). \tag{15}$$

Quando $t = 1$, temos:

$$\alpha_{d_1(1,s_1)} = \pi o_{s_1 o_1} \times \pi_{s_1}. \tag{16}$$

Quando a matriz $\alpha$ for totalmente construída, é possível obter $P(O \mid \lambda)$. Para isso basta obter o somatório da última linha da matriz $\alpha_{d_1}$ :

$$P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T \mid \lambda)$$
$$= \sum_{s_T \in \Omega_s} P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T, S_T = s_T \mid \lambda)$$
$$= \sum_{s_T \in \Omega_s} \alpha_{d_1(T, s_T)}.$$

(17)

**Exemplo 2.14** ***Método forward (HMM com dependência de primeira ordem nas observações dados os estados).*** Dado uma seqüência de observações $O = \{o_1 = 1, o_2 = 2\}$ e utilizando um *HMM* que possui dependência de 1ª ordem entre $O_t$'$s$ ,dados $S_t$'$s$ , com $\lambda = (\pi, \pi o, A, B)$:

$$\pi = [0.5 \quad 0.5], \pi o = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$
$$A = \begin{bmatrix} 0.4 & 0.6 \\ 0.9 & 0.1 \end{bmatrix}, B^1 = \begin{bmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{bmatrix}, B^2 = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}.$$

Podemos calcular $P(O \mid \lambda)$ utilizando o método *forward*. Temos os seguintes valores atribuídos a matriz $\alpha_{d_1}$ :

$$\alpha_{d_1(1,1)} = \pi o_{11} \times \pi_1 = 0.5 \times 0.5 = 0.25,$$
$$\alpha_{d_1(1,2)} = \pi o_{21} \times \pi_2 = 0.5 \times 0.5 = 0.25,$$
$$\alpha_{d_1(2,1)} = b^1_{12} \times \sum_{i=1}^{2} (a_{i1} \times \alpha_{d_1(1,i)}) = 0.7 \times (0.25 \times 0.4 + 0.25 \times 0.9) = 0.2275 \text{ e}$$
$$\alpha_{d_1(2,2)} = b^2_{12} \times \sum_{i=1}^{2} (a_{i2} \times \alpha_{d_1(1,i)}) = 0.6 \times (0.25 \times 0.6 + 0.25 \times 0.1) = 0.105.$$

Assim, a matriz associada ao método *forward* é dada a seguir:

$$\alpha_{d_1} = \begin{bmatrix} 0.2500 & 0.2500 \\ 0.2275 & 0.1050 \end{bmatrix}$$

Como visto na expressão (5), dado $r = 2$ e $T = 2$, o valor da probabilidade $P(O \mid \lambda)$ é:

$$\begin{aligned} P(O \mid \lambda) &= \sum_{i=1}^{r} \alpha_{d_1(T,i)} \\ &= \alpha_{d_1(2,1)} + \alpha_{d_1(2,2)} \\ &= (0.2275 + 0.105) \\ &= 0.3325. \end{aligned}$$

O método *forward*, aplicado a modelos que consideram dependências entre as observações dados os estados, possui a mesma ordem de complexidade computacional que o método aplicado a um *HMM* que prevê independência entre as observações.

Existe outro procedimento para o cálculo de $P(O \mid \lambda)$. Este procedimento denominado *backward*, em conjunto com o método *forward*, será de grande importância na resolução do **problema 3**.

#### 2.4.1.2 Procedimento *Backward*

O objetivo do procedimento *backward* é realizar o cálculo de $P(O \mid \lambda)$ eliminando operações redundantes. Para isso, faz-se necessária a busca de uma relação de recorrência. Vamos trabalhar nas próximas duas subseções,

respectivamente, com modelos que prevêem independência e dependência condicional de ordem 1 entre as observações $O_1, O_2, ..., O_t, ...$ dados $S_1, S_2, ..., S_t, ...$ .

### 2.4.1.2.1 Procedimento *Backward* (*HMM* com $O_t \mid S$ condicionalmente independentes)

Considerando o *HMM* mais simples, que prevê independência entre as observações, $O_1, O_2, ...O_t, ...$ dados $S_1, S_2, ...S_t, ...$, a expressão $P(O \mid \lambda)$ pode ser escrita como:

$$
\begin{aligned}
&P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T \mid \lambda) \\
&= \sum_{s_1 \in \Omega_s} \left[ P(O_1 = o_1, O_2 = o_2 ..., O_T = o_T, S_1 = s_1 \mid \lambda) \right] \\
&= \sum_{s_1 \in \Omega_s} \left[ P(O_1 = o_1 \mid O_2 = o_2 ..., O_T = o_T, S_1 = s_1, \lambda) \times \right. \\
&\left. P(O_2 = o_2 ..., O_T = o_T, S_1 = s_1 \mid \lambda) \right] \\
&= \sum_{s_1 \in \Omega_s} \left[ P(O_1 = o_1 \mid S_1 = s_1, \lambda) \times \right. \\
&\left. P(O_2 = o_2 ..., O_T = o_T \mid S_1 = s_1, \lambda) \times P(S_1 = s_1 \mid \lambda) \right].
\end{aligned}
\tag{18}
$$

Assim, devemos trabalhar nesta expressão para encontrar uma relação recorrente, possibilitando o cálculo da probabilidade de maneira eficiente. Desenvolvendo o termo $P(O_2 = o_2 ..., O_T = o_T \mid S_1 = s_1, \lambda)$ presente no somatório em (18):

$$
\begin{aligned}
&P(O_2 = o_2, O_3 = o_3, \ldots, O_T = o_T \mid S_1 = s_1, \lambda) \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_2 = o_2, O_2 = o_2, \ldots, O_T = o_T, S_2 = s_2 \mid S_1 = s_1, \lambda) \right] \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_2 = o_2 \mid O_3 = o_3, \ldots, O_T = o_T, S_2 = s_2, S_1 = s_1, \lambda) \times \right. \\
&\left. P(O_3 = o_3 \ldots, O_T = o_T, S_2 = s_2 \mid S_1 = s_1, \lambda) \right] \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_2 = o_2 \mid S_2 = s_2, \lambda) \times \right. \\
&\left. P(O_3 = o_3 \ldots, O_T = o_T \mid S_2 = s_2, S_1 = s_1, \lambda) \times P(S_2 = s_2 \mid S_1 = s_1, \lambda) \right] \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_2 = o_2 \mid S_2 = s_2, \lambda) \times \right. \\
&\left. P(O_3 = o_3, \ldots, O_T = o_T \mid S_2 = s_2, \lambda) \times P(S_2 = s_2 \mid S_1 = s_1, \lambda) \right].
\end{aligned}
\tag{19}
$$

Observe que podemos generalizar o resultado obtido em (19) de modo a estabelecer uma relação de recorrência. Desta forma, para $1 \leq t \leq T-1$, podemos escrever:

$$
\begin{aligned}
&P(O_{t+1} = o_{t+1}, O_{t+2} = o_{t+2}, \ldots, O_T = o_T \mid S_t = s_t, \lambda) = \\
&\sum_{s_{t+1} \in \Omega_s} \left[ P(O_{t+1} = o_{t+1} \mid S_{t+1} = s_{t+1}, \lambda) \times \right. \\
&\left. P(O_{t+2} = o_{t+2}, \ldots, O_T = o_T \mid S_{t+1} = s_{t+1}, \lambda) \times P(S_{t+1} = s_{t+1} \mid S_t = s_t, \lambda) \right].
\end{aligned}
\tag{20}
$$

Denotando:

$$
\beta_{(t,s_t)} = P(O_{t+1} = o_{t+1}, O_{t+2} = o_{t+2} \ldots, O_T = o_T \mid S_t = s_t, \lambda),
\tag{21}
$$

é possível reescrever a expressão (20), utilizando os parâmetros associados ao modelo.

$$\beta_{(t,s_t)} = \sum_{s_{t+1}\in\Omega_s} b_{o_{t+1}s_{t+1}} \times \beta_{(t+1,s_{t+1})} \times a_{s_t s_{t+1}}. \tag{22}$$

No trabalho de Rabiner (1989), $\beta$ é chamada de variável *backward*. Observe que $\beta$ pode ser vista como uma matriz de probabilidades condicionais. Esta matriz possui, por definição, $T-1$ linhas e $r$ colunas. Entretanto, como $\alpha$ e $\beta$ vão ser utilizadas na resolução do Problema 3, por aspectos computacionais vamos criar uma linha adicional na matriz $\beta$, tendo $s_T \in \Omega_S$ ,definida por:

$$\beta_{(T,s_T)} = 1. \tag{23}$$

Em (23) $\beta_{(T,s_t)} = 1$, de modo a garantir que $P(S_t = s_t, O \mid \lambda) = \alpha_{(t,s_t)} \times \beta_{(t,s_t)}$ se verifique para todo $s_t \in \Omega_S$ e $1 \le t \le T$. Na resolução do problema 3 será importante obter a distribuição $P(S_t = s_t, O \mid \lambda)$.

$$\beta = \begin{bmatrix} \beta_{(1,1)} & \cdots & \beta_{(1,r)} \\ \vdots & \ddots & \vdots \\ \beta_{(T,1)} & \cdots & \beta_{(T,r)} \end{bmatrix}$$

Na obtenção de $P(O \mid \lambda)$ consideremos, novamente, a equação (18). Basta reescrevê-la observando os parâmetros associados ao *HMM* com independência condicional nas observações dados os estados e as probabilidades associadas à matriz $\beta$ :

$$P(O \mid \lambda) = \sum_{s_1 \in \Omega_s} [P(O_1 = o_1 \mid S_1 = s_1, \lambda) \times$$
$$P(O_2 = o_2 \ldots, O_T = o_T \mid S_1 = s_1, \lambda) \times P(S_1 = s_1 \mid \lambda)]$$
$$= \sum_{s_1 \in \Omega_s} \left[ b_{o_1 s_1} \times \beta_{(1,s_1)} \times \pi_{s_1} \right].$$

(24)

Após todo o desenvolvimento anterior, tendo como resultados principais as equações (22), (23) e (24), o cálculo de $P(O \mid \lambda)$ se reduz a obter a matriz $\beta$. Isso é realizado através de operações recorrentes, observando as dependências entre os elementos de $\beta$. O método possibilita a construção da matriz $\beta$, inicialmente, obtendo os valores dos elementos da ultima linha, e, progressivamente, obtendo os valores dos elementos das linhas antecessoras. A figura 5 ilustra a dependência na construção do elementos da matriz $\beta$.

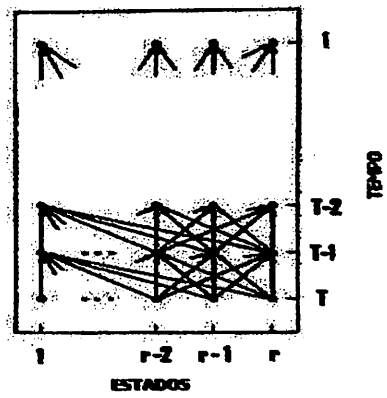



FIGURA 5. Dependência de $\beta_{(t,s_t)}$.

Esquematicamente, $P(O|\lambda)$ é obtida através dos seguintes passos:

1. Faça $\beta_{(T,s_T)}=1$ para todo $s_1 \in \Omega_s$.
2. Faça $\beta_{(t,s_t)} = \sum_{s_{t+1}\in\Omega_s} b_{o_{t+1}s_{t+1}} \times \beta_{(t+1,s_{t+1})} \times a_{s_t s_{t+1}}$, para todo $s_t \in \Omega_s$ e para a seqüência de observações $O_1 = o_1, O_2 = o_2, \ldots, O_T = o_T$.
3. Obtenha $P(O|\lambda) = \sum_{s_1\in\Omega_s} \left[ b_{o_1 s_1} \times \beta_{(1,s_1)} \times \pi_{s_1} \right]$.

**Exemplo 2.15** ***Método backward (HMM com $O_t$'s dados $S_t$'s condicionalmente independentes).*** Utilizando o experimento do exemplo 2.11, envolvendo os atiradores, vamos calcular $P(O|\lambda)$, utilizando o método *backward*. Dado que a seqüência de observações (tiros) foi $O=\{O_1=1, O_2=2\}$, temos:

$$\beta_{(2,1)} = 1,$$

$$\beta_{(2,2)} = 1,$$

$$\beta_{(1,1)} = \sum_{i=1}^{2} b_{2i} \times \beta_{(2,i)} \times a_{1i} = (0.2 \times 1 \times 0.4 + 0.7 \times 1 \times 0.6) = 0.5 \text{ e}$$

$$\beta_{(1,2)} = \sum_{i=1}^{2} b_{2i} \times \beta_{(2,i)} \times a_{2i} = (0.2 \times 1 \times 0.9 + 0.7 \times 1 \times 0.1) = 0.25,$$

Assim, a matriz *backward* é igual a:

$$\beta = \begin{bmatrix} 1.00 & 1.00 \\ 0.50 & 0.25 \end{bmatrix}.$$

Com base na expressão (24):

$$\begin{aligned} P(O \mid \lambda) &= \sum_{i=1}^{2} \left[ b_{1i} \times \beta_{(1,i)} \times \pi_i \right] \\ &= \left[ b_{11} \times \beta_{(1,1)} \times \pi_1 \right] + \left[ b_{12} \times \beta_{(1,2)} \times \pi_2 \right] \\ &= \left[ 0.8 \times 0.5 \times 0.3 \right] + \left[ 0.3 \times 0.25 \times 0.7 \right] \\ &= 0.1725 \end{aligned}$$

O custo computacional associado ao método *backward* é de ordem quadrática, pois não há custo associado ao cálculo dos $r$ elementos da última linha da matriz *Backward*, uma vez que todos os elementos desta linha são iguais à unidade. Já, para cada um dos $(T-1) \times r$ elementos restantes de $\beta$, são necessárias $3r-1$ operações. Isso totaliza $((T-1) \times r) \times (3r-1)$ para construção completa da matriz *backward*. Resta o cálculo final para obtenção de

$P(O \mid \lambda)$, que necessita de $3r-1$ operações. Assim, o número total de operações é:

$$\begin{aligned} \text{nº de operaçoes} &= ((T-1)\times r)\times(3r-1)+(3r-1) \\ &= 3r^2\times(T+1)+(T-4)\times(r)-1. \end{aligned}$$

O número de operações necessárias no método *backward* é um pouco maior do que no método *forward*. No entanto, em termos de complexidade, os dois métodos são equivalentes.

Na próxima seção, vamos mostrar o desenvolvimento do método *backward* aplicado ao *HMM*, considerando dependência de 1ª ordem entre as observações dados os estados.

#### 2.4.1.2.2 Procedimento *Backward* (*HMM* com $O_t \mid S$ condicionalmente dependentes)

Inicialmente vamos desenvolver a expressão de $P(O \mid \lambda)$ em termos de "*variáveis backward*", considerando o modelo com $O_t \mid S$ dependentes:

$$
\begin{aligned}
&P(O_1 = o_1, O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid \lambda) = \\
&\sum_{s_1 \in \Omega_s} \left[ P(O_1 = o_1, O_2 = o_2, O_2 = o_2, ..., O_T = o_T, S_1 = s_1 \mid \lambda) \right] \\
&= \sum_{s_1 \in \Omega_s} \left[ P(O_1 = o_1, O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid S_1 = s_1, \lambda) \times \right. \\
&\left. P(S_1 = s_1 \mid \lambda) \right] \\
&= \sum_{s_1 \in \Omega_s} \left[ P(O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid O_1 = o_1, S_1 = s_1, \lambda) \times \right. \\
&\left. P(O_1 = o_1 \mid S_1 = s_1, \lambda) \times P(S_1 = s_1, \lambda) \right].
\end{aligned}
\tag{25}
$$

Desenvolvendo $P(O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid O_1 = o_1, S_1 = s_1, \lambda)$, observando as restrições do *HMM* em questão, temos:

$$
\begin{aligned}
&P(O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid O_1 = o_1, S_1 = s_1, \lambda) \\
&= \sum_{s_2 \in \Omega_s} P(O_2 = o_2, O_3 = o_3, ..., O_T = o_T, S_2 = s_2 \mid O_1 = o_1, S_1 = s_1, \lambda) \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid S_2 = s_2, O_1 = o_1, S_1 = s_1, \lambda) \times \right. \\
&\left. P(S_2 = s_2 \mid O_1 = o_1, S_1 = s_1, \lambda) \right] \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid S_2 = s_2, O_1 = o_1, \lambda) \times \right. \\
&\left. P(S_2 = s_2 \mid S_1 = s_1, \lambda) \right] \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_3 = o_3, ..., O_T = o_T \mid O_2 = o_2, S_2 = s_2, O_1 = o_1, \lambda) \times \right. \\
&\left. P(O_2 = o_2 \mid S_2 = s_2, O_1 = o_1, \lambda) \times P(S_2 = s_2 \mid S_1 = s_1, \lambda) \right] \\
&= \sum_{s_2 \in \Omega_s} \left[ P(O_3 = o_3, ..., O_T = o_T \mid O_2 = o_2, S_2 = s_2, \lambda) \times \right. \\
&\left. P(O_2 = o_2 \mid S_2 = s_2, O_1 = o_1, \lambda) \times P(S_2 = s_2 \mid S_1 = s_1, \lambda) \right].
\end{aligned}
\tag{26}
$$

Podemos estabelecer uma generalização para a equação (26):

$$P(O_{t+1}=o_{t+1},O_{t+2}=o_{t+2},\ldots,O_T=o_T \mid O_t=o_t,S_t=s_t,\lambda)$$
$$=\sum_{s_2\in\Omega_s}\left[P(O_{t+2}=o_{t+2},\ldots,O_T=o_T \mid O_{t+1}=o_{t+1},S_{t+1}=s_{t+1},\lambda)\times\right.$$
$$\left.P(O_{t+1}=o_{t+1} \mid S_t=s_{t+1},O_t=o_t,\lambda)\times P(S_{t+1}=s_{t+1} \mid S_t=s_t,\lambda)\right]. \quad (27)$$

Denotando:

$$\beta_{d_1(t,s_t)}=P(O_{t+1}=o_{t+1},O_{t+2}=o_{t+2},\ldots,O_T=o_T \mid O_t=o_t,S_t=s_t,\lambda), \quad (28)$$

podemos reescrever a equação (27):

$$\beta_{d_1(t,s_t)}=\sum_{s_{t+1}\in\Omega_S} b_{o_t o_{t+1}}^{s_{t+1}}\times\beta_{d_1(t+1,s_{t+1})}\times a_{s_t s_{t+1}},\ 1\le t\le T-1. \quad (29)$$

Note que a variável $\beta_{d_1}$ não é definida quando $t=T$. Porém, para facilitar a utilização do método na resolução do Problema 3, para todo $s_T\in\Omega_S$, definimos:

$$\beta_{d1(T,s_T)}=1. \quad (30)$$

Reescrevendo a equação (26) em função da variável $\beta_{d_1}$ e dos parâmetros associados ao modelo, é possível obter $P(O \mid \lambda)$:

$$P(O \mid \lambda) = P(O_1 = o_1, O_2 = o_2, O_3 = o_3, ..., O_T = o_T \mid \lambda)$$
$$= \sum_{s_2 \in \Omega_s} \left[ \beta_{d_1(1,s_1)} \times \pi o_{s_1 o_1} \times \pi_{s_1} \right]. \tag{31}$$

Com os resultados das equações (29), (30) e (31) podemos obter um método recursivo para o cálculo de $P(O \mid \lambda)$, que diminui o número de operações necessárias. A variável $\beta_{d_1}$ apresenta uma estrutura matricial com $T$ linhas e $r$ colunas. A construção da matriz $\beta_{d_1}$ é realizada iniciando-se pelos valores da última linha e, progressivamente, avaliando os elementos das linhas anteriores, respeitando-se a relação de recorrência obtida em (29). Os passos necessários para o cálculo de $P(O \mid \lambda)$ são dados a seguir:

1. Faça $\beta_{d1(T,s_T)} = 1$ para todo $s_1 \in \Omega_s$.
2. Faça $\beta_{d1(t,s_t)} = \sum_{s_{t+1} \in \Omega s} b^{s_{t+1}}_{o_t o_{t+1}} \times \beta_{d1(t+1,s_{t+1})} \times a_{s_t s_{t+1}}$ para todo $s_t \in \Omega_s$ e para a seqüência de observações $O_1 = o_1, O_2 = o_2, ..., O_T = o_T$.
3. Obtenha $P(O \mid \lambda) = \sum_{s_1 \in \Omega_s} \left[ \beta_{d1(1,s_1)} \times \pi o_{o_1 s_1} \times \pi_{s_1} \right]$.

**Exemplo 2.14** ***Método Backward (HMM com dependência de primeira ordem nas observações dados os estados).*** Dada a seqüência de observações $O = \{o_1 = 1, o_2 = 2\}$ e tendo o *HMM* a seguir, definido da seção 2.3:

$$\pi = [0.5 \quad 0.5],\ \pi o = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$

$$A = \begin{bmatrix} 0.4 & 0.6 \\ 0.9 & 0.1 \end{bmatrix},\ B^1 = \begin{bmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{bmatrix},\ B^2 = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}.$$

Podemos calcular $P(O \mid \lambda)$, utilizando o método *Backward*. Os valores da matriz $\beta_{d_1}$ são:

$$\beta_{d_1(2,2)} = 1,$$

$$\beta_{d_1(2,1)} = 1,$$

$$\beta_{d_1(1,1)} = \sum_{i=1}^{2} [b_{12}^i \times \beta_{d_1(2,i)} \times a_{(1,i)}] = (0.7 \times 1 \times 0.4) + (0.6 \times 1 \times 0.6) = 0.64 \text{ e}$$

$$\beta_{d_1(1,2)} = \sum_{i=1}^{2} [b_{12}^i \times \beta_{d_1(2,i)} \times a_{(2,i)}] = (0.7 \times 1 \times 0.9) + (0.6 \times 1 \times 0.1) = 0.69.$$

Assim, a matriz $\beta_{d_1}$ é igual a:

$$\beta_{d_1} = \begin{bmatrix} 1.00 & 1.00 \\ 0.64 & 0.69 \end{bmatrix}.$$

Com base na expressão (31), o valor da probabilidade associada ao *HMM* em questão é:

$$P(O|\lambda) = \sum_{i=1}^{2}\left[\beta_{d_1(1,i)} \times \pi o_{i1} \times \pi_i\right]$$
$$= (\beta_{d_1(1,1)} \times \pi o_{11} \times \pi_1) + (\beta_{d_1(1,2)} \times \pi o_{21} \times \pi_2)$$
$$= (0.64 \times 0.5 \times 0.5) + (0.69 \times 0.5 \times 0.5)$$
$$= (0.16 + 0.1725)$$
$$= 0.3325.$$

O método *backward*, aplicado a um *HMM* com dependência de ordem 1 entre as observações, tem ordem de complexidade computacional idêntica ao método *backward* aplicado ao modelo com observações independentes dados os estados. Portanto, seu custo é:

$$\text{nº de operaçoes} = ((T-1) \times r) \times (3r-1) + (3r-1)$$
$$= 3r^2 \times (T+1) + (T-4) \times (r) - 1.$$

### 2.4.2 Resolvendo o Problema 2

O segundo problema associado aos *HMM's* é a obtenção da seqüência mais provável (ótima) de estados $S_1, S_2, ..., S_T$, dada uma seqüência de observações $O_1, O_2, ..., O_T$. Entretanto, vários fatores devem ser considerados para se definir uma *seqüência ótima* de estados. Tal seqüência, de tamanho $T$, será denominada $S^{otima}$, onde $S^{otima} \in \Omega_{ST}$.

Uma estratégia possível para encontrar $S^{otima}$ é obter uma seqüência em que os estados são individualmente os mais prováveis. Para tanto, devemos obter a distribuição da variável aleatória $S_t \mid O, \lambda$, com $1 \leq t \leq T$. Porém, se o

*HMM* possui alguma transição entre os estados cuja probabilidade é zero, podemos obter uma seqüência $S^{otima}$ inválida (que não pode ocorrer).

Um outro critério, mais utilizado, busca obter uma seqüência $S^{otima}$ tal que $S^{otima}$ maximize a verossimilhança conjunta $P(S,O|\lambda)$. Considerando a função $\arg(f(x)) = x$, devemos buscar $S^{otima}$ tal que:

$$S^{otima} = \arg \max_{S \in \Omega_{ST}} P(S,O|\lambda). \tag{32}$$

É possível encontrar $S^{otima}$ através da abordagem do problema em termos de grafos. Deste modo, é possível aplicar um Algoritmo de busca em grafos. Este método é denominado *algoritmo de Viterbi*.

Antes de descrevermos como funciona o algoritmo de Viterbi, é necessário contextualizar o Problema 2 (Vide seção 2.4), utilizando algumas definições de Teoria dos Grafos. Na apresentação do algoritmo de Viterbi nas duas subsecções a seguir, consideraremos *HMM*'s, que prevêem, respectivamente, independência e dependência condicional de 1ª ordem entre as observações dados os estados.

### 2.4.2.1 *Algoritmo de Viterbi* (*HMM* com $O_t | S$ condicionalmente independentes)

Como descrito pela expressão (32), no intuito de encontrar a seqüência $S^{otima}$ de estados, necessitamos descrever a verossimilhança $P(S,O|\lambda)$. Devido a sua simplicidade, trabalharemos com a log-verossimilhança, obtida aplicando-se o logaritmo natural à verossimilhança associada ao modelo.

Considerando uma seqüência de observações de tamanho $T$ e o modelo com observações $O_t \mid S$ condicionalmente independentes, temos:

$$\ln(P(O,S \mid \lambda)) = \ln(P(O_1 = o_1,...,O_T = o_T, S_1 = s_1,...,S_T = s_T \mid \pi, A, B))$$
$$= \ln(P(S_1 = s_1 \mid \lambda) \times \left[\prod_{t=1}^{T} P(O_t = o_t \mid S_t = s_t, \lambda)\right] \times \left[\prod_{t=2}^{T} P(S_t = s_t \mid S_{t-1} = s_{t-1}, \lambda)\right])$$
$$= \ln(\pi_{s_1} \times \left[\prod_{t=1}^{T} b_{s_t o_t}\right] \times \left[\prod_{t=2}^{T} a_{s_{t-1} s_t}\right])$$
$$= \ln(\pi_{s_1} \times b_{s_1 o_1}) + \sum_{t=2}^{T} \ln(b_{s_t o_t} \times a_{s_{t-1} s_t}).$$

(33)

O problema inicial de encontramos $S^{otima}$ equivale a encontrarmos a seqüência $S$ que minimiza $-\ln(P(O,S \mid \lambda))$. Portanto, considere:

$$S^{otima} = \arg \max_{S \in \Omega_{ST}} \{P(S,O \mid \lambda)\} = \arg \min_{S \in \Omega_{ST}} \{-\ln(P(S,O \mid \lambda))\}.$$

(34)

Utilizando a modelagem através de grafos de uma Cadeia de Markov, é possível estabelecer um *caminho* ou *trajetória* neste grafo, relacionado a uma seqüência de estados $S = \{S_1 = s_1, S_2 = s_2,..., S_T = s_T\}$. A idéia é associar um *custo* para cada transição entre estados. Desta forma, o custo total associado a uma trajetória será constituído pela soma dos custos das transições entre os estados, adicionado de um custo inicial associado a cada estado. Define-se o *custo inicial*, *CI*, por:

$$CI(s_1) = -\ln(\pi_{s_1} b_{s_1 o_1}).$$

(35)

O *Custo de transição, CT*, entre os estados $s_{t-1}$ e $s_t$, onde $2 \leq t \leq T$, é:

$$CT(s_{t-1}, s_t) = -\ln(a_{s_{t-1}s_t} \times b_{s_t o_t}).$$

(36)

O custo de transição entre estados $s_{t-1}$ e $s_t$, $CT(s_{t-1}, s_t)$,é uma função da observação no instante $t$ ($o_t$). Isso implica que os custos de transições variam à medida que $o_t$ se modifica. Assim, o *Custo Total*, $C_{Total}$, associado a uma seqüência de estados $S$, dado uma seqüência de observações $O$ qualquer, é dado por:

$$C_{Total}(S) = CI(s_1) + \sum_{i=2}^{T} CT(s_{i-1}, s_i).$$

(37)

Observe que o Custo Total de uma seqüência $S$ é igual a log-verossimilhança conjunta, $-\ln(P(S,O|\lambda))$. Assim, para obtermos $S^{otima}$, devemos obter a seqüência de estados, isto é, o caminho no grafo associado à Cadeia de Markov Latente de custo mínimo.

Vamos definir ainda *Custo Acumulado Mínimo* (*CAM*), necessário para implementação do algoritmo de Viterbi. $CAM(t, s_t)$ representa o "custo

acumulado mínimo para que se atinja o estado $s_t$ no tempo $t$, partindo do estado $s_{t-1}$". $CAM(t,s_t)$ pode ser expressa através da relação de recorrência.

$$CAM(t,s_t)=\begin{cases} CI(s_1),\ se\ t=1, \\ \min\limits_{s_{t-1}\in\Omega_s}\{CAM(t-1,s_{t-1})+CT(s_{t-1},s_t)\},\ se\ t>1. \end{cases} \tag{38}$$

Obtendo todos os estados $S_t$, a cada índice $t$, que proporcionaram um custo mínimo, quando $t=T$, estaremos obtendo $S^{otima}$.

O algoritmo de Viterbi é um método baseado em programação dinâmica que busca a seqüência ou o caminho de tamanho $T$ de custo mínimo e funciona da seguinte maneira:

(1) Inicialmente, com $t=1$, temos os Custos Iniciais associados a cada um dos estados. Devemos então obter o Custo Acumulado Mínimo, $CAM(t,s_t)$, associado a cada um dos estados quando $t=1$. O Custo Acumulado Mínimo será o próprio Custo Inicial quando $t=1$.

(2) Quando $t>1$, obtemos o custo acumulado mínimo, $CAM(t,s_t)$, de cada um dos estados da cadeia; além disso, é necessário registrar o estado que no instante $t-1$ proporcionou o custo acumulado mínimo no instante $t$.

Do ponto de vista computacional, para realizar o registro do estado, que no instante $t-1$, tem o menor custo correspondente a uma transição deste, para o estado $S_t$, *no tempo* $t$, fazemos uso de $EMCAM(t,s_t)$, definida a seguir:

$$EMCAM(t,s_t)=\begin{cases} 0 \text{ se } t=1; \\ \arg \min\limits_{s_{t-1}\in\Omega_s}\{CAM(t-1,s_{t-1})+CT(s_{t-1},s_t)\} \text{ se } t>1. \end{cases} \quad (39)$$

As matrizes $EMCAM$ e $CAM$, definidas pelas expressões (38) e (39), podem ser vistas como matrizes com $T$ linhas e $r$ colunas cada, utilizadas no armazenamento de, respectivamente, estados (valores inteiros no intervalo $[1,r]$), e valores de log-verossimilhança. É possível obter, recursivamente, a seqüência de estados ótima, $S^{otima}$, construindo $CAM$ e $EMCAM$. Isso pode ser feito, definindo-se:

$$S_t^{otima}=\begin{cases} \min\limits_{s_t\in\Omega_s}\{CAM(t,s_t)\} \text{ se } t=T; \\ EMCAM(t+1,s_{t+1}^{otima}) \text{ se } t<T. \end{cases} \quad (40)$$

O custo da seqüência ótima de estados será o menor valor da última linha de $CAM$.

Os passos para obtenção da seqüência ótima de estados $S^{otima}$, associada a um conjunto de observações $O_1,O_2,...,O_T$, e um $HMM$ sob pressuposto de que $O_t \mid S$ são independentes, são dados a seguir:

1. Faça $CAM(1,s_1)=CI(s_1)$ para todo $s_1\in\Omega_s$; Faça $EMCAM(1,s_1)=0$ para todo $s_1\in\Omega_s$;

2. Faça $CAM(t,s_t)=\min_{s_{t-1}\in\Omega_s}\{CAM(s_{t-1})+CT(s_{t-1},s_t)\}$ para todo $s_t\in\Omega_s$ e para todo $1<t\leq T$. Faça $EMCAM(t,s_t)=\arg\min_{s_{t-1}\in\Omega_s}\{CAM(t-1,s_{t-1})+CT(s_{t-1},s_t)\}$ para todo $s_t\in\Omega_s$ e para todo $1<t\leq T$.

3. Obtenha $S_T^{otima}=\min_{s_t\in\Omega_S}\{CAM(t,s_t)\}$; Obtenha $S_t^{otima}=EMCAM(t+1,s_{t+1}^{otima})$ para todo $1\leq t<T$.

Para a obtenção do valor da verossimilhança conjunta $P(S^{otima},O\mid\lambda)$, basta obter o valor $\exp(-\min_{s_{t-1}\in\Omega_s}\{CAM(T,s_T)\})$.

O exemplo 2.15 a seguir ilustra o funcionamento do algoritmo de Viterbi em um caso particular em que os custos de transição entre os estados não variam com $t$.

**Exemplo 2.15** ***Algoritmo de Viterbi – 1 (HMM com independência entre observações dados os estados).*** Considere que, no exemplo 2.11, foi observada a seqüência $O=\{o_1=1,o_2=1\}$. Através do algoritmo de Viterbi é possível descrever qual a seqüência de atiradores mais provável, $S_1,S_2$. No caso específico em que $O=\{O_1=O_2=...=O_T\}$, considerando o *HMM* do exemplo 2.11, os custos iniciais e das transições são:

$$CI(1) = -\ln(\pi_1 \times b_{11}) = -\ln(0.3 \times 0.8) = 1.4271;$$
$$CI(2) = -\ln(\pi_2 \times b_{21}) = -\ln(0.7 \times 0.3) = 1.5606;$$
$$CT(1,1) = -\ln(a_{11} \times b_{11}) = -\ln(0.4 \times 0.8) = 1.1394;$$
$$CT(1,2) = -\ln(a_{12} \times b_{21}) = -\ln(0.6 \times 0.6) = 1.0217;$$
$$CT(2,1) = -\ln(a_{21} \times b_{11}) = -\ln(0.9 \times 0.8) = 0.3285;$$
$$CT(2,2) = -\ln(a_{22} \times b_{21}) = -\ln(0.1 \times 0.6) = 2.8234.$$

A figura 6, a seguir, ilustra os custos de transição entre os estados, em termos de um grafo:

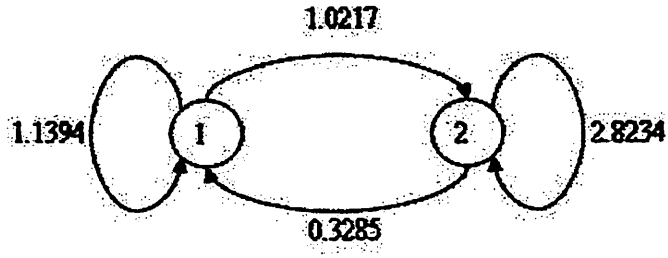



FIGURA 6. Grafo com custos (pesos) das transições entre os estados.

Com base nas quantidades calculadas na página anterior, as matrizes $CAM(t,s_t)$ e $EMCAM(t,s_t)$ são dadas por:

$$CAM = \begin{bmatrix} 1.4271 & 1.5606 \\ 1.8891 & 2.4488 \end{bmatrix},$$
$$EMCAM = \begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix},$$

pois observe que os valores da 1ª linha de $CAM$ são os custos iniciais de cada estado. Seguindo a definição apresentada em (38), podemos verificar que $CAM(2,1)$ e $CAM(2,2)$ são resultantes de:

$$
\begin{aligned}
CAM(2,1) &= \min_{s_{t-1}\in\Omega_s}\{CAM(1,s_{t-1})+CT(s_{t-1},1)\} \\
&= \min\{CAM(1,1)+CT(1,1), CAM(1,2)+CT(2,1)\} \\
&= \min\{(1.4271+1.1394),(1.5606+0.3285)\} \\
&= \min\{2.5665, 1.8891\} \\
&= 1.8891;
\end{aligned}
$$

$$
\begin{aligned}
CAM(2,2) &= \min_{s_{t-1}\in\Omega_s}\{CAM(1,s_{t-1})+CT(s_{t-1},2)\} \\
&= \min\{CAM(1,1)+CT(1,2), CAM(1,2)+CT(2,2)\} \\
&= \min\{(1.4271+1.0217),(1.5606+2.8234)\} \\
&= \min\{2.4488, 4.3840\} \\
&= 2.4488.
\end{aligned}
$$

A obtenção de $EMCAM$, ilustrada a seguir, é regida por (39):

$$
\begin{aligned}
EMCAM(2,1) &= \arg \min_{s_{t-1}\in\Omega_s} \{CAM(1,s_{t-1}) + CT(s_{t-1},1)\} \\
&= \arg \min_{s_{t-1}\in\Omega_s} \{(CAM(1,1) + CT(1,1)), (CAM(1,2) + CT(2,1))\} \\
&= \arg \ (CAM(1,2) + CT(2,1)) \\
&= 2;
\end{aligned}
$$

$$
\begin{aligned}
EMCAM(2,2) &= \arg \min_{s_{t-1}\in\Omega_s} \{CAM(1,s_{t-1}) + CT(s_{t-1},2)\} \\
&= \arg \min_{s_{t-1}\in\Omega_s} \{(CAM(1,1) + CT(1,2)), (CAM(1,2) + CT(2,2))\} \\
&= \arg \ (CAM(1,1) + CT(1,2)) \\
&= 1.
\end{aligned}
$$

O caminho ou seqüência ótima $S^{otima}$ é $\{S_1 = 2, S_2 = 1\}$. Essa seqüência é obtida a partir da expressão (40), dado que $CAM$ e $EMCAM$ já foram calculados. Assim, $S^{otima} = \{S_1^{otima} = 2, S_2^{otima} = 1\}$ foi obtida a partir de:

$$
\begin{aligned}
S_2^{otima} &= \arg \min_{s_2\in\Omega_S} \{CAM(2,s_2)\} \\
&= \arg \ \min\{CAM(2,1), CAM(2,2)\} \\
&= \arg \ \min\{1.8891, 3.1419\} \\
&= 1, \\
S_1^{otima} &= EMCAM(2, s_2^{otima}) \\
&= EMCAM(2,1) \\
&= 2.
\end{aligned}
$$

O valor da log-verossimilhança $-\ln(P(S^{otima}, O \mid \lambda))$ é $1.8891$, representando o custo acumulado mínimo no tempo $T$. Assim, o valor de $P(S^{otima}, O \mid \lambda)$ é $\exp(-1.8891) = 0.1512$.

No exemplo 2.16, é ilustrada a situação na qual os custos de transição mudam em função de $t$.

**Exemplo 2.16** ***Algoritmo de Viterbi – 2 (HMM com independência entre observações).*** Suponha que, no exemplo 2.11, envolvendo os atiradores, foi observada a seqüência de tiros $O = \{o_1 = 1, o_2 = 2, o_3 = 2\}$. Neste exemplo, os custos de transição entre os estados se modificam de acordo com o instante $t$. Isso é decorrente do fato da seqüência de observações não ser composta de apenas um único símbolo. Aplicando o algoritmo de Viterbi, é possível obter qual a seqüência $S_1, S_2, S_3$ ótima de atiradores. Os custos Iniciais e de Transição no instante $t$, supondo o *HMM* do exemplo 2.11, são:

Instante $t = 1$:

$CI(1) = -\ln(\pi_1 b_{11}) = -\ln(0.3 \times 0.8) = 1.4271;$

$CI(2) = -\ln(\pi_2 b_{21}) = -\ln(0.7 \times 0.3) = 1.5606;$

$CT(1,1) = -\ln(a_{11} \times b_{12}) = -\ln(0.4 \times 0.2) = 2.5257;$

$CT(1,2) = -\ln(a_{12} \times b_{22}) = -\ln(0.6 \times 0.4) = 1.4271;$

$CT(2,1) = -\ln(a_{21} \times b_{12}) = -\ln(0.9 \times 0.2) = 1.7147;$

$CT(2,2) = -\ln(a_{22} \times b_{22}) = -\ln(0.1 \times 0.4) = 3.2188;$

instante $t = 2$:

$CT(1,1) = -\ln(a_{11} \times b_{11}) = -\ln(0.4 \times 0.8) = 1.1394;$

$CT(1,2) = -\ln(a_{12} \times b_{21}) = -\ln(0.6 \times 0.6) = 1.0217;$

$CT(2,1) = -\ln(a_{21} \times b_{11}) = -\ln(0.9 \times 0.8) = 0.3285;$

$CT(2,2) = -\ln(a_{22} \times b_{21}) = -\ln(0.1 \times 0.6) = 2.8234.$

As Figuras 7 e 8, a seguir, ilustram, em forma de grafo, os custos de transição dos estados, para cada valor de $t$:

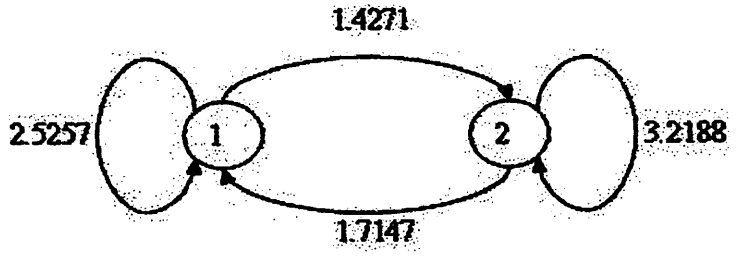



FIGURA 7. Grafo com custos (pesos) das transições entre os estados no instante $t=1$.

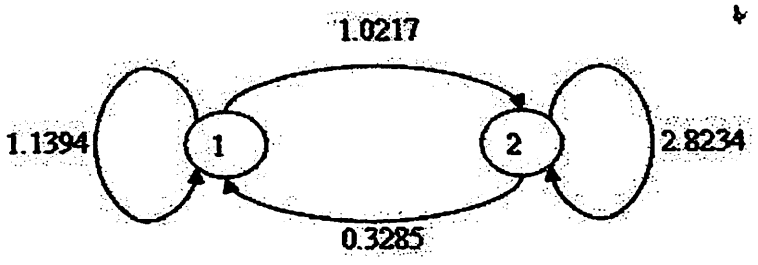



FIGURA 8. Grafo com custos (pesos) das transições entre os estados no instante $t=2$.

Com base nos custos iniciais e de transições entre os estados, temos $CAM(t,s_t)$ e $EMCAM(t,s_t)$ dadas por:

$$CAM = \begin{bmatrix} 1.4271 & 1.5606 \\ 3.2754 & 2.8542 \\ 3.1827 & 4.2971 \end{bmatrix};$$

$$EMCAM = \begin{bmatrix} 0 & 0 \\ 2 & 1 \\ 2 & 1 \end{bmatrix}.$$

A construção $CAM$ e $EMCAM$ é realizada de forma equivalente ao exemplo 2.15, observando (38) e (39), atentando que devemos utilizar os custos de transição condicionados ao índice $t$.

O caminho ou seqüência ótima, $S^{otima}$, é obtido seguindo (40), dado que $CAM$ e $EMCAM$ já foram calculados. Assim, $S^{otima} = \{S_1^{otima} = 1, S_2^{otima} = 2, S_3^{otima} = 1\}$ foi obtida a partir de:

$$\begin{aligned}
S_3^{otima} &= \arg \min_{s_3 \in \Omega_S} \{CAM(3, s_3)\} \\
&= \arg \min \{CAM(3,1), CAM(3,2)\} \\
&= \arg \min \{3.1827, 4.2971\} \\
&= 1, \\
S_2^{otima} &= EMCAM(3, s_3^{otima}) \\
&= EMCAM(2,1) \\
&= 2, \\
S_1^{otima} &= EMCAM(2, s_2^{otima}) \\
&= EMCAM(2,2) \\
&= 1.
\end{aligned}$$

O valor da log-verossimilhança conjunta $-\ln(P(S^{otima},O|\lambda))$ é $3.1827$, que é o custo acumulado mínimo. Assim, o valor de $P(S^{otima},O|\lambda)$ é $\exp(-3.1827)=0.0414$.

### 2.4.2.2 *Algoritmo de Viterbi* (*HMM* com $O_t|S$ condicionalmente dependentes)

Supondo um *HMM* que considera dependências de 1ª ordem entre as observações $O_1,O_2,...,O_t$ dados os estados $S_1,S_2,...,S_t$, também é possível construir o algoritmo de Viterbi para busca de uma seqüência ótima de estados, $S^{otima}$. Vamos considerar, de maneira equivalente ao modelo com observações independentes, que a seqüência ótima $S^{otima}$ será aquela que proporciona a maior verossimilhança $P(S,O|\lambda)$. Considerando um *HMM* $\lambda$ com dependências de 1ª ordem entre as observações dados os estados, $S^{otima}$ é obtida a partir de:

$$S^{otima}=\arg\max_{S\in\Omega_{ST}}\{P(S,O|\lambda)\}=\arg\min_{S\in\Omega_{ST}}\{-\ln(P(S,O|\lambda))\}. \quad (41)$$

É necessário obter a função log-verossimilhança respectiva ao modelo. Desta forma, considere:

$$\ln(P(O,S|\lambda)) = \ln(P(O_1 = o_1,...,O_T = o_T, S_1 = s_1,...,S_T = s_T | \pi, \pi o, A, B))$$
$$= \ln(P(S_1 = s_1) \times P(O_1 = o_1 | S_1 = s_1) \times \left[\prod_{t=2}^{T} P(O_t = o_t | S_t = s_t, O_{t-1} = o_{t-1})\right] \times$$
$$\left[\prod_{t=2}^{T} P(S_t = s_t | S_{t-1} = s_{t-1})\right])$$
$$= \ln(\pi_{s_1} \times \pi o_{s_1 o_1}) + \left[\sum_{t=2}^{T} \ln(b^{s_t}_{o_{t-1} o_t} \times a_{s_{t-1} s_t})\right].$$

Para uma dada seqüência de observações, $O = \{O_1 = o_1, O_2 = o_2,..., O_T = o_T\}$, e uma dada seqüência de estados, $S = \{S_1 = s_1, S_2 = s_2,..., S_T = s_T\}$, o valor da log-verossimilhança pode ser interpretado como o custo associado à seqüência de estados, isto é, um caminho em um grafo. Para obter $S^{otima}$, devemos encontrar a seqüência de estados de custo mínimo. Para obtê-la, é necessário estabelecermos custos iniciais e custos de transições entre os estados. Considerando um *HMM* com dependência de 1ª ordem entre as observações $O_t | S$, o *custo inicial, CI*, é dado por:

$$CI(s_1) = -\ln(\pi_{s_1} \times \pi o_{s_1 o_1}).$$

(42)

Por sua vez, o *custo de transição, CT*, entre os estado $s_{t-1}$ e $s_t$, em que $2 \le t \le T$ é:

$$CT(s_{t-1}, s_t) = -\ln(a_{s_{t-1} s_t} \times b^{s_t}_{o_{t-1} o_t}).$$

(43)

Seguindo as definições de *CAM* e *EMCAM* já estabelecidas na seção anterior (38) (39), os passos posteriores para obtenção da seqüência ótima (40), $S^{otima}$, são iguais aos passos apresentados no *HMM* que pressupõe independência condicional entre as observações $O_t$'s conhecidos os $S_t$'s.

**Exemplo 2.17** ***Algoritmo de Viterbi – 2 (HMM com dependência condicional entre observações).*** Suponha que, no exemplo 2.14, associado a um *HMM* que possui dependência de 1ª ordem entre as observações, foi observado $O = \{O_1 = 1, O_2 = 2, O_3 = 1\}$. Os custos iniciais e de transição entre os estados, no instante $t$, são:

Instante $t = 1$:

$CI(1) = -\ln(\pi_1 \times \pi o_{11}) = -\ln(0.5 \times 0.5) = 1.3863$

$CI(2) = -\ln(\pi_2 \times \pi o_{21}) = -\ln(0.5 \times 0.5) = 1.3863$

$CT(1,1) = -\ln(a_{11} \times b_{12}^1) = -\ln(0.4 \times 0.7) = 1.2730$

$CT(1,2) = -\ln(a_{12} \times b_{12}^2) = -\ln(0.6 \times 0.6) = 1.0216$

$CT(2,1) = -\ln(a_{21} \times b_{12}^1) = -\ln(0.9 \times 0.7) = 0.4620$

$CT(2,2) = -\ln(a_{22} \times b_{12}^2) = -\ln(0.1 \times 0.6) = 2.8134$

Instante $t = 2$:

$CT(1,1) = -\ln(a_{11} \times b_{22}^1) = -\ln(0.4 \times 0.1) = 3.2188$

$CT(1,2) = -\ln(a_{12} \times b_{22}^2) = -\ln(0.6 \times 0.2) = 2.1203$

$CT(2,1) = -\ln(a_{21} \times b_{22}^1) = -\ln(0.9 \times 0.1) = 2.4079$

$CT(2,2) = -\ln(a_{22} \times b_{22}^2) = -\ln(0.1 \times 0.2) = 3.9120$

As figuras 9 e 10, a seguir, ilustram em forma de grafo, os custos de Transição entre os estados através do tempo.

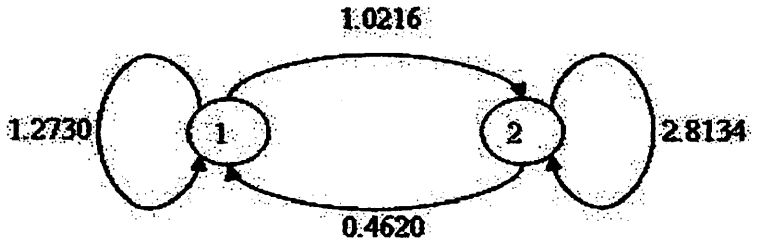



FIGURA 9. Grafo com custos (pesos) nas transições no instante $t = 1$.

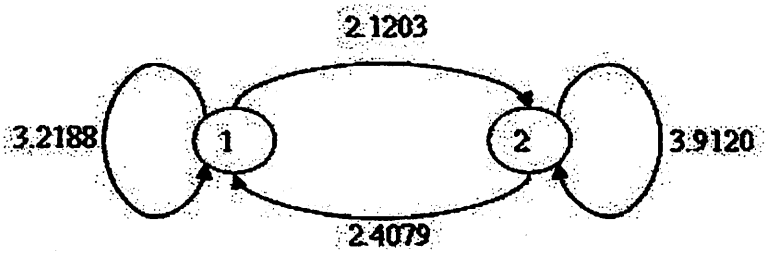



FIGURA 10. Grafo com custos (pesos) nas transições no instante $t = 2$.

Com base nos valores dos custos iniciais e de transição associados aos estados, as matrizes $CAM(t, s_t)$ e $EMCAM(t, s_t)$ obtidas da mesma maneira do exemplo 2.15, são dadas por:

$$CAM = \begin{bmatrix} 1.3863 & 1.3863 \\ 1.8473 & 2.4079 \\ 4.8158 & 3.9676 \end{bmatrix},$$

$$EMCAM = \begin{bmatrix} 0 & 0 \\ 2 & 1 \\ 2 & 1 \end{bmatrix}.$$

O caminho ou seqüência ótima, obtida de maneira similar ao exemplo 2.15, $S^{otima}$, é $\{S_1 = 1, S_2 = 2, S_3 = 2\}$. O valor da log-verossimilhança conjunta $-\ln(P(S^{otima}, O|\lambda))$ é 3.9676, que é o custo acumulado mínimo. Assim o valor de $P(S^{otima}, O|\lambda)$ é 0.0189.

O custo computacional associado à busca de uma seqüência $S^{otima}$ ótima, tanto considerando o modelo *HMM* com observações $O_t|S_t$ condicionalmente dependentes e independentes, é de ordem quadrática em função do tamanho da seqüência $O$ e linear em função do número de estados. Observe que, para obtenção de $S^{otima}$, temos que construir as matrizes $CAM$ e $EMCAM$. Observe que o custo para construção das células da 1ª linha de $CAM$ é de uma operação. Para a construção de cada uma das células fora da 1ª linha, é necessária uma busca entre as $r$ possibilidades de transição que geram o $CAM$. Isso totaliza, para cada uma das células de $CAM$, um custo de ordem $r$. Como temos $(T-1)\times r$ células em $CAM$ fora da 1ª linha, temos um custo total de $[(T-1)\times r]\times r$. Conhecidos os valores de $CAM$, a construção de cada uma das células de $EMCAM$ tem custo constante. Assim, temos:

$$\text{nº de operaçoes} = (T-1)\times r^2 + (T\times r).$$

Na próxima seção, trabalharemos os aspectos teóricos associados ao **problema 3**, ligado à inferência estatística sobre o modelo $\lambda$ associado a um *HMM*.

### 2.4.3 Resolvendo o Problema 3

O problema 3 associado aos *HMM*'s é o de natureza mais complexa dentre os problemas apresentados. O problema 3 é caracterizado pela realização da inferência estatística relativa aos valores dos parâmetros no modelo $\lambda$.

Existem vários métodos para a realização da inferência estatística. Dentre estes, podem ser citados com exemplos o Método dos Momentos, Método de Mínimos Quadrados e Método Bayesiano. Vamos utilizar o Método da Máxima Verossimilhança.

Segundo o método da máxima verossimilhança, para obtermos os estimadores dos parâmetros associados ao *HMM*, temos que obter o ponto de máximo da função de verossimilhança $L(\lambda | O) = P(O | \lambda)$ com relação à $\lambda$. No entanto, para o *HMM*, não é possível realizar a maximização direta da função de verossimilhança. Assim, torna-se necessário o uso de um método iterativo. Um dos métodos mais utilizados é denominado algoritmo *EM* (Expectation-Maximization) (Dempster, 1977). Na próxima seção, vamos apresentar o algoritmo *EM*, alguns exemplos gerais aplicados à inferência e por fim a descrição do método aplicado aos *HMM*.

#### 2.4.3.1 Algoritmo *EM*

O algoritmo *EM* é um método iterativo de maximização. Este algoritmo para maximização da verossimilhança, popularizado por Dempster et al (1977), é particularmente útil quando alguns dados são não observáveis ou perdidos. Seja $\theta$ um parâmetro associado a um espaço paramétrico $\Theta$. Seja $X \in \Omega_X$ e $Y \in \Omega_Y$, respectivamente, os dados observados e os dados latentes (não observados ou ocultos). Devemos maximizar a função de verossimilhança dos

dados incompletos $L(\theta \mid X) = P(X \mid \theta)$. Tal maximização pode ser difícil de ser feita. Em muitas situações a função de verossimilhança $l(\theta \mid X,Y) = P(X,Y \mid \theta)$ dos dados completos é mais fácil de se trabalhar. O algoritmo *EM* utiliza tal fato na maximização de $L(\theta \mid X)$. Objetivando simplicidade, vamos trabalhar com $\ln(L(\theta \mid X,Y))$. Podemos resumir o algoritmo *EM* em dois passos, que justificam seu nome:

1. Passo Esperança (*Expectation-Step*)- Neste passo devemos obter o valor esperado de $\ln(L(\theta \mid X,Y)) = \ln(P(X,Y \mid \theta))$ com relação à distribuição condicional $Y \mid X, \theta^{(m)}$, em que $\theta^{(m)}$ é o valor obtido no passo da maximização da iteração $m-1$ .
2. Passo Maximização (*Maximization-Step*)- Devemos maximizar o valor esperado de $\ln(L(\theta \mid X,Y)) = \ln(P(X,Y \mid \theta))$, encontrado no passo esperança, com relação ao parâmetro $\theta$, ou seja, devemos obter $\theta$ que maximiza $E_{X|Y,\theta^{(m)}}[\ln(l(\theta \mid X,Y))]$. Tal valor de $\theta$ será $\theta^{(m+1)}$.

No princípio do processo, $(m=1)$, devemos escolher um valor inicial para $\theta$, isto é, $\theta^{(0)}$. O procedimento é realizado iterando os dois passos acima, substituindo, ao fim do passo maximização, $\theta^{(m)}$ por $\theta^{(m+1)}$, sendo $m$ a iteração atual. Através da demonstração apresentada por da Silva (2002), é possível verificar que $P(X,Y \mid \theta^{(m+1)}) \geq P(X,Y \mid \theta^{(m)})$. A seguir vamos reproduzir esta demonstração.

Inicialmente, devemos reescrever a função de verossimilhança, dos dados originais ou observáveis, $L(\theta \mid X)$. Trabalhando com o logaritmo natural temos:

$$\begin{aligned}
&\ln(L(\theta \mid X)) = \ln(P(X \mid \theta)) \\
&= \ln\left(\frac{P(X,Y \mid \theta)P(X \mid \theta)}{P(X,Y \mid \theta)}\right) \\
&= \ln\left(\frac{P(X,Y \mid \theta)}{P(Y \mid X,\theta)}\right) \\
&= \ln\left(P(X,Y \mid \theta)\right) - \ln\left(P(Y \mid X,\theta)\right).
\end{aligned} \tag{35}$$

Utilizando-se de propriedades da Esperança Matemática, e tendo valores fixos para $X$ e $\theta^{(m)}$, podemos reescrever que:

$$\begin{aligned}
&\ln(P(X \mid \theta)) \\
&= \int_{\Omega_Y} \ln(P(X \mid \theta)) \times P(Y \mid X,\theta^{(m)})\, dY \\
&= E_{Y \mid X,\theta^{(m)}}\left[\ln(P(X \mid \theta))\right].
\end{aligned} \tag{36}$$

Utilizando os resultados (35) e (36) e propriedades de Esperança Matemática, temos que:

$$\begin{aligned}
&\ln(L(\theta \mid X)) = \ln(P(X \mid \theta)) \\
&= E_{Y \mid X,\theta^{(m)}}\left[\ln(P(X,Y \mid \theta))\right] - E_{Y \mid X,\theta^{(m)}}\left[\ln(P(Y \mid X,\theta))\right].
\end{aligned} \tag{37}$$

Em (37), a log-verossimilhança dos dados observáveis $X$ está expressa em função do valor esperado $E_{Y|X,\theta^{(m)}}\left[\ln(P(X,Y\mid\theta))\right]$ e do valor esperado $E_{Y|X,\theta^{(m)}}\left[\ln(P(Y\mid X,\theta))\right]$. Assim, maximizar $\ln(L(\theta\mid X))=\ln(P(X\mid\theta))$ implica em maximizar o lado direito da igualdade presente na expressão (37). Contudo, podemos mostrar que o termo $E_{Y|X,\theta^{(m)}}\left[\ln(P(Y\mid X,\theta))\right]$, da expressão (37), não afeta na maximização de $\ln(L(\theta\mid X))$, ou seja, basta maximizarmos $E_{Y|X,\theta^{(m)}}\left[\ln(P(Y\mid X,\theta))\right]$ para obtermos o máximo de $\ln(L(\theta\mid X))$. Para simplificar a notação, iremos denotar:

$$Q(\theta,\theta^{(m)})=E_{Y|X,\theta^{(m)}}\left[\ln(P(X,Y\mid\theta))\right];$$
$$\mathrm{H}(\theta,\theta^{(m)})=E_{Y|X,\theta^{(m)}}\left[\ln(P(Y\mid X,\theta))\right]. \qquad (38)$$

Desenvolvendo a expressão (39), utilizando-nos da desigualdade de *Jensen*, que estabelece uma relação entre $E_X\left[f(x)\right]$ e $f\left(E_X\left[x\right]\right)$ dado que $f(x)$ é côncava (Mood, 1963), vamos mostrar como $\mathrm{H}(\theta,\theta^{(m)})$ não afeta na maximização da expressão (37):

$$
\begin{aligned}
&H(\theta,\theta^{(m)})\text{-}H(\theta^{(m)},\theta^{(m)})=E_{Y|X,\theta^{(m)}}\left[\ln(P(Y\mid X,\theta))\right]\text{-}E_{Y|X,\theta^{(m)}}\left[\ln(P(Y\mid X,\theta^{(m)}))\right]\\
&=\int_{\Omega_Y}\ln(P(Y\mid X,\theta))\times P(Y\mid X,\theta^{(m)})dY-\int_{\Omega_Y}\ln(P(Y\mid X,\theta^{(m)}))\times P(Y\mid X,\theta^{(m)})dY\\
&=\int_{\Omega_Y}\left(\ln(P(Y\mid X,\theta))-\ln(P(Y\mid X,\theta^{(m)}))\right)\times P(Y\mid X,\theta^{(m)})dY\\
&=\int_{\Omega_Y}\ln\left(\frac{P(Y\mid X,\theta)}{P(Y\mid X,\theta^{(m)})}\right)\times P(Y\mid X,\theta^{(m)})dY\\
&=E_{Y|X,\theta^{(m)}}\left[\ln\left(\frac{P(Y\mid X,\theta)}{P(Y\mid X,\theta^{(m)})}\right)\right]\\
&\le\ln\left(E_{Y|X,\theta^{(m)}}\left[\frac{P(Y\mid X,\theta)}{P(Y\mid X,\theta^{(m)})}\right]\right)\\
&=\ln\left(\int_{\Omega_Y}\frac{P(Y\mid X,\theta)}{P(Y\mid X,\theta^{(m)})}\times P(Y\mid X,\theta^{(m)})dY\right)\\
&=\ln\left(\int_{\Omega_Y}(P(Y\mid X,\theta)dY\right)\\
&=\ln(1)=0.
\end{aligned}
\tag{39}
$$

Pelo resultado obtido em (39), $H(\theta^{(m)},\theta^{(m)})\text{-}H(\theta,\theta^{(m)})\ge 0$ para todo $\theta\in\Theta$. Este resultado será utilizado na fórmula (41).

Como $\ln(L(\theta\mid X,Y))=\ln(P(X,Y\mid\theta))$ é simples de se trabalhar, a obtenção de seu máximo com respeito a $\theta$ pode ser realizada através de derivação. Assim, vamos denotar:

$$
Q(\theta^{(m+1)},\theta^{(m)})=\max_{\theta\in\Theta}(Q(\theta,\theta^{(m)})).
\tag{40}
$$

Como $\theta^{(m+1)}$ é o valor que maximiza $Q(\theta,\theta^{(m)})$, $Q(\theta^{(m+1)},\theta^{(m)})-Q(\theta,\theta^{(m)})\geq 0$ para qualquer $\theta\in\Theta$. Observe que nosso objetivo é mostrar que $P(X,Y\mid\theta^{(m+1)})\geq P(X,Y\mid\theta^{(m)})$, ou, de maneira equivalente, que $P(X,Y\mid\theta^{(m+1)})-P(X,Y\mid\theta^{(m)})\geq 0$, em que $\theta^{(m+1)}$ foi obtido no passo maximização do algoritmo *EM*, a partir de um $\theta^{(m)}$, obtido na iteração anterior. Reescrevendo $P(X,Y\mid\theta^{(m+1)})-P(X,Y\mid\theta^{(m)})$, temos que:

$$\begin{aligned}
&P(X,Y\mid\theta^{(m+1)})-P(X,Y\mid\theta^{(m)})\\
&=\left[Q(\theta^{(m+1)},\theta^{(m)})-H(\theta^{(m+1)},\theta^{(m)})\right]-\left[Q(\theta^{(m)},\theta^{(m)})-H(\theta^{(m)},\theta^{(m)})\right]\\
&=\left[Q(\theta^{(m+1)},\theta^{(m)})-Q(\theta^{(m)},\theta^{(m)})\right]+\left[H(\theta^{(m)},\theta^{(m)})-H(\theta^{(m+1)},\theta^{(m)})\right]\\
&\geq\left[Q(\theta^{(m+1)},\theta^{(m)})-Q(\theta^{(m)},\theta^{(m)})\right]\geq 0.
\end{aligned}\tag{41}$$

Com o resultado (41), prova-se que $P(X,Y\mid\theta^{(m+1)})-P(X,Y\mid\theta^{(m)})\geq 0$. No trabalho de Dempster (1977), é verificado que a seqüência $\{\theta^{(1)},\theta^{(2)},\theta^{(3)}...,\theta^{(m)},...\}$ converge para um ponto de máximo associado a $L(\theta\mid X)=P(X\mid\theta)$. Entretanto, um comportamento multimodal de $P(X\mid\theta)$ pode levar o algoritmo *EM* a resultados que constituem máximos locais. A seguir, dois exemplos de aplicações do algoritmo *EM*, fora do âmbito dos *HMM*, são ilustrados.

**Exemplo 2.18** ***Algoritmo EM – 1 (Estimação de parâmetros em distribuições multinomiais).*** (Rao, 1973) Considere uma variável aleatória multidimensional $X=(X_1,X_2,X_3,X_4)$, onde $X \sim Mult(n,\frac{1}{2}+\frac{\theta}{4},\frac{1-\theta}{4},\frac{1-\theta}{4},\frac{\theta}{4})$. A função log-verossimilhança associada ao $X$ é dada por:

$$\ln(L(\theta \mid X)) = \ln\big(P(X \mid \theta)\big) = \ln\big(P(X_1=x_1, X_2=x_2, X_3=x_3, X_4=x_4 \mid \theta)\big)$$
$$= \ln\left(\frac{n!}{x_1!x_2!x_3!x_4!} \times \left(\frac{1}{2}+\frac{\theta}{4}\right)^{x_1} \times \left(\frac{1-\theta}{4}\right)^{x_2} \times \left(\frac{1-\theta}{4}\right)^{x_3} \times \left(\frac{\theta}{4}\right)^{x_4}\right)$$
$$= \ln\left(\frac{n!}{x_1!x_2!x_3!x_4!}\right) + x_1 \times \ln\left(\frac{1}{2}+\frac{\theta}{4}\right) + (x_2+x_3) \times \ln\left(\frac{1-\theta}{4}\right) + x_4 \times \ln\left(\frac{\theta}{4}\right) \tag{42}$$

O estimador $\hat{\theta}$ de $\theta$ é obtido, ao derivarmos $\ln\big(L(\theta \mid X)\big)$ com respeito a $\theta$, ou seja, $\frac{\partial \ln\big(P(X \mid \theta)\big)}{\partial \theta} = 0$. A igualdade é verificada quando:

$$(x_2+x_3-x_1-x_4)\theta^2 + (2x_2+2x_3+x_1-x_4)\theta + 2x_4 = 0 \tag{43}$$

Embora os cálculos envolvidos em (43), na obtenção de $\hat{\theta}$, sejam simples, o exemplo em estudo serve para ilustrar os passos necessários na implementação do algoritmo *EM*. Para aplicarmos o algoritmo *EM*, devemos obter uma função $P(X,Y \mid \theta)$, baseada em uma variável $Y$ latente (não observável) de tal forma que a maximização de $P(X,Y \mid \theta)$ com respeito à $\theta$ seja simples. Para isso, vamos supor que $X_1 = Y_0 + Y_1$ e $Y_i = X_i$ onde $2 \le i \le 4$. Desta forma, estamos

supondo que a variável observável $X_1$ é uma função de variáveis $Y_0$ e $Y_1$ não observáveis. Sendo $Y=(Y_0,Y_1,Y_2,Y_3,Y_4)$, então $Y \sim Mult(n,\frac{1}{2},\frac{\theta}{4},\frac{1-\theta}{4},\frac{1-\theta}{4},\frac{\theta}{4})$. Dado que conhecemos os valores das variáveis $X$ observáveis, devido à restrição $X_1=Y_0+Y_1$, somente uma das variáveis latentes $Y_0$ ou $Y_1$ não é individualmente observável. Vamos adotar que $Y_1$ é a porção não observável. Devemos obter a log-verossimilhança conjunta de $X$ e $Y$. A log-verossimilhança conjunta associada a $(X,Y)$ é dada por:

$$
\begin{aligned}
&P(Y,X\mid\theta)=P(Y\mid\theta)\times P(X\mid Y,\theta)\\
&=\left(\frac{n!}{y_0!y_1!y_2!y_3!y_4!}\times\left(\frac{1}{2}\right)^{y_0}\times\left(\frac{\theta}{4}\right)^{y_1}\times\left(\frac{1-\theta}{4}\right)^{y_2}\times\left(\frac{1-\theta}{4}\right)^{y_3}\times\left(\frac{\theta}{4}\right)^{y_4}\right)\times\\
&I_{X_1}(x_1=y_0+y_1)\times I_{X_2,X_3,X_4}(x_2=y_2,x_3=y_3,x_4=y_4)\\
&=\frac{n!}{(x_1-y_1)!y_1!x_2!x_3!x_4!}\times\left(\frac{1}{2}\right)^{x_1-y_1}\times\left(\frac{\theta}{4}\right)^{y_1}\times\left(\frac{1-\theta}{4}\right)^{x_2}\times\left(\frac{1-\theta}{4}\right)^{x_3}\times\left(\frac{\theta}{4}\right)^{x_4}\times\\
&I_{X_1}(x_1=y_0+y_1)\times I_{X_2,X_3,X_4}(x_2=y_2,x_3=y_3,x_4=y_4).
\end{aligned}
\tag{44}
$$

Sendo:

$$
I_{XY}=P(X\mid Y,\theta)=\begin{cases}1, & \text{se } X_1=y_0+y_1 \text{ e } X_2=y_2,X_3=y_3,X_4=y_4\\ 0, & \text{caso contrario;}\end{cases}
$$

Seguindo o algoritmo *EM*, o 1° passo é obter o valor esperado de $\ln(P(X,Y|\theta))$ com relação à distribuição condicional $Y|X,\theta^{(m)}$, em que $\theta^{(m)} \in \Omega_\theta$. Assim, temos que:

$$
\begin{aligned}
&E_{Y|X,\theta^{(m)}}\left[\ln(P(Y,X|\theta))\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln(P(Y|\theta)\times P(X|Y,\theta))\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln(P(Y|\theta))\right]+E_{Y|X,\theta^{(m)}}\left[\ln(P(X|Y,\theta))\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln(P(Y|\theta))\right]+\ln(P(X|Y,\theta)) \\
&=E_{Y|X,\theta^{(m)}}\left[\ln(P(Y|\theta))\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln\left(\frac{n!}{(x_1-Y_1)!Y_1!x_2!x_3!x_4!}\times\left(\frac{1}{2}\right)^{x_1-Y_1}\times\left(\frac{\theta}{4}\right)^{Y_1}\times\left(\frac{1-\theta}{4}\right)^{x_2}\times\left(\frac{1-\theta}{4}\right)^{x_3}\times\left(\frac{\theta}{4}\right)^{x_4}\right)\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln\left(\frac{n!}{(x_1-Y_1)!Y_1!x_2!x_3!x_4!}\right)+(x_1-Y_1)\times\ln\left(\frac{1}{2}\right)+(Y_1+x_4)\times\ln\left(\frac{\theta}{4}\right)+(x_2+x_3)\times\ln\left(\frac{1-\theta}{4}\right)\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln\left(\frac{n!}{(x_1-Y_1)!Y_1!x_2!x_3!x_4!}\right)\right]+E_{Y|X,\theta^{(m)}}\left[(x_1-Y_1)\times\ln\left(\frac{1}{2}\right)\right]+E_{Y|X,\theta^{(m)}}\left[(Y_1+x_4)\times\ln\left(\frac{\theta}{4}\right)\right]+ \\
&+E_{Y|X,\theta^{(m)}}\left[(x_2+x_3)\times\ln\left(\frac{1-\theta}{4}\right)\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\ln\left(\frac{n!}{(x_1-Y_1)!Y_1!x_2!x_3!x_4!}\right)\right]+\ln\left(\frac{1}{2}\right)\times\left(x_1-E_{Y|X,\theta^{(m)}}[Y_1]\right)+\ln\left(\frac{\theta}{4}\right)\times\left(x_4+E_{Y|X,\theta^{(m)}}[Y_1]\right)+ \\
&\ln\left(\frac{1-\theta}{4}\right)\times(x_2+x_3).
\end{aligned}
\tag{45}
$$

Vamos trabalhar com a maximização da expressão (45) em relação à $\theta$. Podemos assumir termos independentes de $\theta$ como constantes $C$ que podem ser descartadas. Para fins práticos temos que:

$$E_{Y|X,\theta^{(m)}}\left[\ln\left(P(Y,X\mid\theta)\right)\right]$$
$$=\ln(\theta)\times\left(E_{Y|X,\theta^{(m)}}\left[Y_1\right]+x_4\right)+\ln(1-\theta)\times(x_2+x_3)+C \tag{46}$$

Resta-nos obter a distribuição condicional de $Y\mid X,\theta^{(m)}$ a fim de obter a forma fechada da esperança $E_{Y|X,\theta^{(m)}}\left[\ln\left(P(Y,X\mid\theta)\right)\right]$, ou simplesmente obter a $E_{Y|X,\theta^{(m)}}\left[Y_1\right]$, concluindo assim o *passo esperança* do algoritmo *EM*. Desenvolvendo $P(Y\mid X,\theta^m)$, temos:

$$P(Y\mid X,\theta^{(m)})=\frac{P(Y,X\mid\theta^{(m)})}{P(X\mid\theta^{(m)})}. \tag{47}$$

Desenvolvendo $P(Y\mid X,\theta^{(m)})$, temos:

$$P(Y \mid X,\theta^{(m)}) = \frac{\dfrac{n!}{(x_1-y_1)!y_1!x_2!x_3!x_4!}}{\dfrac{n!}{x_1!x_2!x_3!x_4!}} \times$$

$$\frac{\left(\frac{1}{2}\right)^{x_1-y_1} \times \left(\frac{\theta^{(m)}}{4}\right)^{y_1} \times \left(\frac{1-\theta^{(m)}}{4}\right)^{x_2} \times \left(\frac{1-\theta^{(m)}}{4}\right)^{x_3} \times \left(\frac{\theta^{(m)}}{4}\right)^{x_4} \times I_{X|Y}}{\left(\frac{1}{2}+\frac{\theta^{(m)}}{4}\right)^{x_1} \times \left(\frac{1-\theta^{(m)}}{4}\right)^{x_2} \times \left(\frac{1-\theta^{(m)}}{4}\right)^{x_3} \times \left(\frac{\theta^{(m)}}{4}\right)^{x_4}}$$

$$= P(Y_1 \mid X_1,\theta^{(m)}) = \frac{\dfrac{1}{y_1!(x_1-y_1)!} \times \left(\frac{1}{2}\right)^{x_1-y_1} \times \left(\frac{\theta^{(m)}}{4}\right)^{y_1}}{\dfrac{1}{x_1!} \times \left(\frac{1}{2}+\frac{\theta^{(m)}}{4}\right)^{x_1-y_1+y_1}} \times I_{X|Y}$$

$$= P(Y_1 \mid X_1,\theta^{(m)}) = \frac{x_1!}{y_1!(x_1-y_1)!} \times \left(\frac{\frac{1}{2}}{\frac{1}{2}+\frac{\theta^{(m)}}{4}}\right)^{x_1-y_1} \times \left(\frac{\frac{\theta^{(m)}}{4}}{\frac{1}{2}+\frac{\theta^{(m)}}{4}}\right)^{y_1} \times I_{X|Y}$$

$$= P(Y_1 \mid X_1,\theta^{(m)}) = \frac{x_1!}{y_1!(x_1-y_1)!} \times \left(\frac{\theta^{(m)}}{2+\theta^{(m)}}\right)^{y_1} \times \left(\frac{2}{2+\theta^{(m)}}\right)^{x_1-y_1} \times I_{X|Y}. \tag{48}$$

Podemos observar que a variável aleatória $Y_1 \mid X_1,\theta^{(m)}$ possui uma distribuição de probabilidade binomial com parâmetros $X_1$ e $\dfrac{\theta^{(m)}}{2+\theta^{(m)}}$. Assim, $E_{Y|X,\theta^{(m)}}[Y_1]$ é:

$$E_{Y|X,\theta^{(m)}}[Y_1] = X_1 \times \left(\frac{\theta^{(m)}}{2+\theta^{(m)}}\right). \tag{49}$$

Substituindo (49) na expressão (46), temos que:

$$Q(\theta,\theta^{(m)}) = E_{Y|X,\theta^{(m)}}\left[\ln\left(P(Y,X\mid\theta)\right)\right]$$
$$= \ln(\theta)\times\left(x_1\times\left(\frac{\theta^{(m)}}{2+\theta^{(m)}}\right)+x_4\right)+\ln(1-\theta)\times(x_2+x_3). \tag{50}$$

Com a expressão (50), concluímos o *passo esperança* do algoritmo *EM*. Devemos obter agora a derivada de $Q(\theta,\theta^{(m)})$ com relação à $\theta$ e igualá-la a zero. Este procedimento constitui o passo maximização do algoritmo *EM*. Assim, temos:

$$\frac{\partial Q(\theta,\theta^{(m)})}{\partial\theta} = 0 \Leftrightarrow \theta^{(m+1)} = \frac{x_1\times\left(\frac{2}{2+\theta^{(m)}}\right)+x_4}{x_1\times\left(\frac{2}{2+\theta^{(m)}}\right)+x_2+x_3+x_4}. \tag{51}$$

Com a expressão (51), concluímos o *passo maximização* associado ao algoritmo *EM*. Com isso, o método para obter inferências sobre o parâmetro $\theta$ consiste em iterar a obtenção de $\theta^{(m+1)}$ (equação 51) a partir de $\theta^{(m)}$ (valor inicial), até que um critério de convergência seja alcançado.

Afim de ilustrar o processo de estimação, foram geradas 100 amostras, a partir de simulação computacional, que seguiam distribuição de probabilidade $X \sim Mult(n,\frac{1}{2}+\frac{\theta}{4},\frac{1-\theta}{4},\frac{1-\theta}{4},\frac{\theta}{4})$, com $\theta = 0.8$. Utilizam-se 5 iterações,

adotando-se como pontos iniciais $\theta^{(0)} = 0.8$, $\theta^{(0)} = 0.5$, $\theta^{(0)} = 0.3$ e $\theta^{(0)} = 0.1$ . A figura 11 ilustra a convergência do processo de estimação do parâmetro $\theta$.

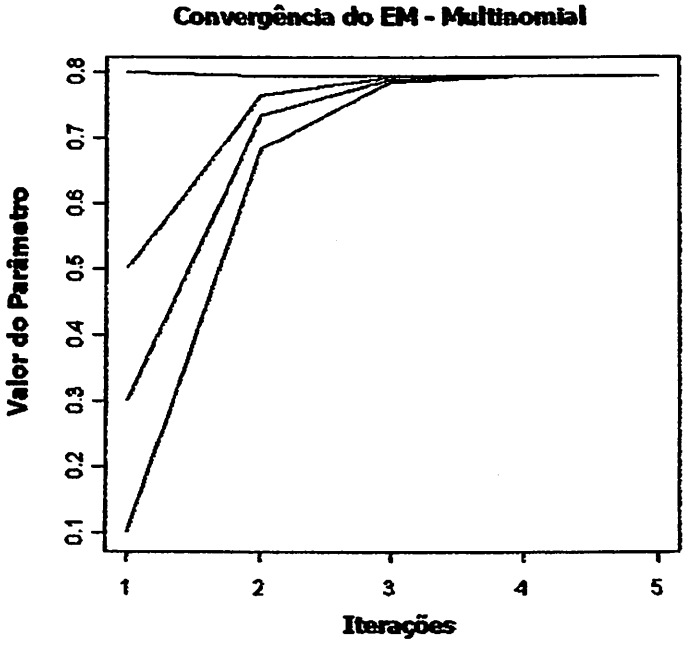



FIGURA 11. Convergência da Estimação de $\theta$ utilizando o *EM*.

Na figura 12, é apresentada a verossimilhança respectiva ao parâmetro $\theta^{(m)}$ estimado na $m$ -ésima iteração, considerando 4 pontos iniciais distintos.

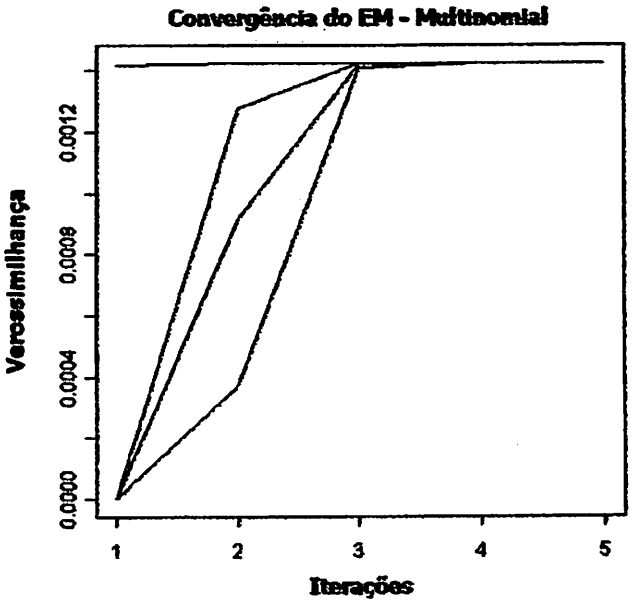



FIGURA 12. Convergência da Verossimilhança utilizando o *EM*.

Observe, na figura 12, que o comportamento da verossimilhança é sempre não decrescente para todo $\theta^{(m)}$, $m = 1,2,3,4,5$.

**Exemplo 2.19** ***Algoritmo EM – 2 (Estimação de parâmetros em uma Mistura de Normais).***(da Silva, 2002) Seja $g(x|\theta_i)$ uma função densidade de probabilidade qualquer. Suponha uma variável aleatória $X$, cuja função densidade de probabilidade, parametrizada por $\theta = (p_1, p_2, ..., p_k, \theta_1, \theta_2, ..., \theta_k)$ com $\sum_{i=1}^{k} p_i = 1$, é expressa abaixo:

$$f(x|\theta)=\sum_{i=1}^{k} p_i \times g(x|\theta_i). \qquad (52)$$

Funções densidade de probabilidade que possuem esta forma são chamadas de misturas. No exemplo, trataremos o caso de Misturas de Normais, ou seja, em que $\theta=(p_1,p_2,\ldots,p_k,\mu_1,\mu_2,\ldots,\mu_k,\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2)$ e $g(x|\theta_i)=N(\mu_i,\sigma_i^2)$. Na figura 13 é ilustrada uma mistura de duas normais $f(x|\theta)$ com $\theta=(p_1,p_2,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$ onde $p_1=p_2,\sigma_1^2=\sigma_2^2$ e $\mu_1=a,\mu_2=b$.

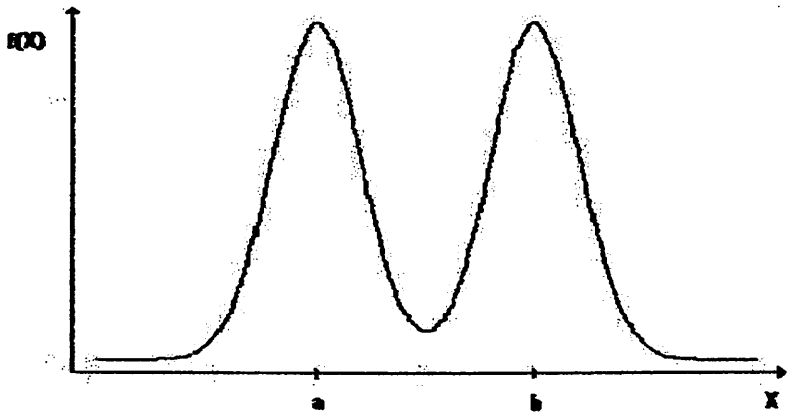



FIGURA 13. Mistura de duas Normais.

A função de verossimilhança para uma amostra de tamanho $n$ associada a uma mistura de normais é dada por:

$$f(x_1, x_2 ... x_n \mid \theta) = \prod_{j=1}^{n} \left[ \sum_{i=1}^{k} p_i \times N(x_j \mid \mu_i, \sigma_i^2) \right]. \tag{53}$$

Observe que em (53) não é possível precisar a partir de qual das $k$ distribuições normais $x_j$ é proveniente. Para obtermos estimadores de máxima verossimilhança, devemos obter as derivadas parciais em relação aos parâmetros em $\theta$. Não é possível obter estimadores de forma fechada para esta função de verossimilhança $f(x \mid \theta)$. Podemos então aplicar o algoritmo *EM* a fim de obter a maximização desta função. Para aplicar o *EM*, devemos considerar os parâmetros $p_k$, como pesos associados a cada uma das distribuições envolvidas na mistura. Podemos considerar que cada realização da variável aleatória $X$ é a realização, com uma determinada incerteza (ponderada pelos pesos), de uma das variáveis aleatórias cujas funções de densidade estão envolvidas na mistura. Lembrando que a aplicação do algoritmo *EM* prevê uma verossimilhança $f(x \mid \theta)$ complexa e a adoção de uma $f(x, y \mid \theta)$ que seja tratável. Suponha que $Y$ constitui a informação de qual das distribuições presentes na mistura foi obtida uma determinada realização da variável aleatória $X$, ou seja, $Y$ informa qual das $k$ distribuições normais gerou cada um dos valores de $X$. Observe que esta informação($Y$) é latente, ou não observável. Vamos considerar $Y \in \{0,1\}^k$, em que $Y_i \in \{0,1\}$ com $1 \leq i \leq k$. Dado que $y_i = 1$ para um $i = l$, $y_i = 0$ para todo $i \neq l$, ou equivalentemente, $\sum_{i=1}^{k} y_i = 1$. Note que o fato de $y_i = 1$ representa a informação que uma determinada amostra de $X$ foi gerada pela função densidade $g(x \mid \theta_i) = N(\mu_i, \sigma_i^2)$. Assim, podemos escrever a função

densidade de probabilidade conjunta $f(x,y|\theta)$ dos dados completos $(X,Y)$ como:

$$f(x,y|\theta)=\prod_{i=1}^{k}[p_i \times N(x|\mu_i,\sigma_i^2)]^{y_i}. \tag{54}$$

Tendo $n$ realizações da variável aleatória $X$, podemos escrever a log-verossimilhança associada à distribuição $f(x,y|\theta)$ como:

$$\ln(f(x,y|\theta))=\sum_{j=1}^{n}\sum_{i=1}^{k} y_{ji}\left[\ln(p_i)+\ln(N(x_j|\mu_i,\sigma_i^2))\right]. \tag{55}$$

Desenvolvendo a primeira etapa do *EM*, devemos obter a esperança da expressão (55) com respeito à distribuição da variável aleatória $Y|X,\theta^{(m)}$, na qual $\theta^{(m)}$ constitui um valor obtido na $m$-ésima iteração, restrito ao espaço paramétrico de $\theta$. Assim, temos que:

$$\begin{aligned}
&E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right] \\
&=E_{Y|X,\theta^{(m)}}\left[\sum_{j=1}^{n}\sum_{i=1}^{k} Y_{ji}\times\left(\ln(p_i)+\ln(N(x_j|\mu_i,\sigma_i^2))\right)\right] \\
&=\sum_{j=1}^{n}\sum_{i=1}^{k}\left[E_{Y|X,\theta^{(m)}}\left[Y_{ji}\right]\times\left(\ln(p_i)+\ln(N(x_j|\mu_i,\sigma_i^2))\right)\right]
\end{aligned} \tag{56}$$

Para obtermos $E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]$, basta calcularmos $E_{Y|X,\theta^{(m)}}\left[Y_{ji}\right]$. Desenvolvendo $E_{Y|X,\theta^{(m)}}\left[Y_{ji}\right]$, temos:

$$
\begin{aligned}
& E_{Y|X,\theta^{(m)}}\left[Y_{ji}\right] \\
&= \sum_{y_{ji}\in\{0,1\}} y_{ji} \times f(y_{ji} \mid x_1, x_2, \ldots, x_n, \theta^{(m)}) \\
&= \sum_{y_{ji}\in\{0,1\}} y_{ji} \times f(y_{ji} \mid x_j, \theta^{(m)}) \\
&= \sum_{y_{ji}\in\{0,1\}} y_{ji} \times f(y_{ji}, x_j \mid \theta^{(m)}) \times \frac{1}{f(x_j \mid \theta^{(m)})} \\
&= \frac{1}{f(x_j \mid \theta^{(m)})} \times \sum_{y_{ji}\in\{0,1\}} y_{ji} \times f(y_{ji}, x_j \mid \theta^{(m)}) \\
&= \frac{1}{f(x_j \mid \theta^{(m)})} \times \sum_{y_{ji}\in\{0,1\}} \left[ y_{ji} \times \left( p_i^{(m)} \times N(x_j \mid \mu_i^{(m)}, \sigma_i^{2(m)}) \right)^{y_{ji}} \right] \\
&= \frac{1}{f(x_j \mid \theta^{(l)})} \times 1 \times \left( p_i^{(m)} \times N(x_j \mid (\mu_i)^{(m)}, (\sigma_i^2)^{(m)}) \right)^1 \\
&= \frac{p_i^{(m)} \times N(x_j \mid \mu_i^{(m)}, \sigma_i^{2(m)})}{\sum_{i=1}^{k} p_i^{(m)} \times N(x_j \mid \mu_i^{(l)}, \sigma_i^{2(m)})}.
\end{aligned}
\tag{57}
$$

Vamos denotar $E_{Y|X,\theta^{(m)}}\left[Y_{ji}\right]$ por $\widehat{y}_{ji}^{(m)}$. Podemos completar o passo esperança do algoritmo *EM* obtendo a esperança associada aos dados completos $(X,Y)$. Temos então:

$$E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]$$
$$=\sum_{j=1}^{n}\sum_{i=1}^{k}\left[\widehat{y_{ji}}^{(m)}\times\left(\ln(p_i)+\ln(N(x_j|\mu_i,\sigma_i^2))\right)\right].$$
(58)

Desta forma, finalizamos o passo esperança do algoritmo *EM*. Obtida a esperança condicional da função de distribuição dos dados completos, devemos maximizá-la com relação a todos os parâmetros em $\theta$ associados à função densidade da mistura. Devemos utilizar a teoria que envolve máximos condicionados devido à restrição $\sum_{i=1}^{K} p_i = 1$. Utilizando-se da técnica "*Multiplicadores de Lagrange*" é possível maximizar $E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]$ com relação à $\theta$. Considere:

$$E_{Y|X,\theta^{(m)}}\left[\ln(f_{X,Y}(x,y|\theta))\right]$$
$$=\sum_{j=1}^{n}\sum_{i=1}^{k}\left[\widehat{y_{ji}}^{(m)}\times\left(\ln(p_i)+\ln(N(x_j|\mu_i,\sigma_i^2))\right)\right]$$
$$=\sum_{j=1}^{n}\sum_{i=1}^{k}\left[\widehat{y_{ji}}^{(m)}\times\left(\ln(p_i)+\ln(N(x_j|\mu_i,\sigma_i^2))\right)\right]+\lambda\times\left(\sum_{i=1}^{K}p_i-1\right).$$
(59)

Obtendo as derivadas parciais de $E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]$ com relação a cada $\mu_i, \sigma_i^2$ e $p_i$ com $1 \le i \le k$, temos:

$$\frac{\partial\left(E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]\right)}{\partial(\lambda)}=\sum_{i=1}^{k}p_i-1;$$

$$\frac{\partial\left(E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]\right)}{\partial(p_i)}=\frac{\widehat{y}_{ji}^{(m)}}{p_i}+\lambda;$$

$$\frac{\partial\left(E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]\right)}{\partial(\mu_i)}=\frac{1}{\sigma_i^2}\times\sum_{j=1}^{n}\left[\widehat{y}_{ji}^{(m)}\times(x_j-\mu_i)\right];$$

$$\frac{\partial\left(E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]\right)}{\partial(\sigma_i^2)}=\frac{1}{\sigma_i^4}\times\sum_{j=1}^{n}\left[\widehat{y}_{ji}^{(m)}\times\left(x_j-\mu_i\right)^2\right]-\frac{1}{\sigma_i^2}\times\sum_{j=1}^{n}\left[\widehat{y}_{ji}^{(m)}\right];\tag{60}$$

Igualando cada uma das derivadas parciais a zero, obtemos um sistema que resulta na obtenção do máximo de $E_{Y|X,\theta^{(m)}}\left[\ln(f(x,y|\theta))\right]$ com relação à $p_i$, $\mu_i$ e $\sigma_i^2$. Assim, concluindo o passo maximização do Algoritmo *EM*, temos que as estimativas de $p_i$, $\mu_i$ e $\sigma_i^2$ para a $m+1$-ésima iteração do processo são dadas por:

$$\widehat{p}_i^{(m+1)} = \frac{1}{n} \times \sum_{j=1}^{n} \widehat{y}_{ji}^{(m)}$$

$$\widehat{\mu}_i^{(m+1)} = \frac{\sum_{j=1}^{n} \left[ \widehat{y}_{ji}^{(m)} \times x_j \right]}{\sum_{j=1}^{n} \widehat{y}_{ji}^{(m)}}$$

$$\widehat{\sigma}_i^{2(m+1)} = \frac{\sum_{j=1}^{n} \left[ \widehat{y}_{ji}^{(m)} \times \left( x_j - \widehat{\mu}_i^{(m+1)} \right)^2 \right]}{\sum_{j=1}^{n} \widehat{y}_{ji}^{(m)}} \qquad (61)$$

Concluímos o passo maximização associado ao algoritmo *EM*. Devemos substituir o valor $\theta^{(m)}$ pelo $\hat{\theta}^{(m+1)}$, em que $\hat{\theta}^{(m+1)} = (\widehat{p}_1^{(m+1)}, \widehat{p}_2^{(m+1)}, \ldots, \widehat{p}_k^{(m+1)}, \widehat{\mu}_1^{(m+1)}, \widehat{\mu}_2^{(m+1)}, \ldots \widehat{\mu}_k^{(m+1)}, \widehat{\sigma}_1^{2(m+1)}, \widehat{\sigma}_2^{2(m+1)}, \ldots, \widehat{\sigma}_k^{2(m+1)})$.

Devemos repetir a obtenção da estimativa de $\hat{\theta}^{(m+1)}$ até que processo atinja a convergência, que pode ser baseada na verossimilhança ou no número de iterações.

Como exemplo, ilustramos uma simulação de uma variável aleatória $X$ que possui função densidade mistura de duas normais com $\theta = \{p_1 = p_2 = 0.5, \sigma_1^2 = \sigma_2^2 = 1, \mu_1 = 0, \mu_2 = 5\}$. Foram geradas 1000 observações de $X$. A partir de um valor inicial $\theta^{(0)} = \{p_1^{(0)} = 0.3, p_2^{(0)} = 0.7, \sigma_1^{2(0)} = 4, \sigma_2^{2(0)} = 4, \mu_1^{(0)} = 2, \mu_2^{(0)} = 3\}$ foi obtido como resultado, aplicando-se o algoritmo *EM* e adotando-se 50 iterações

como critério de convergência, $\hat{\theta} = \{\widehat{p_1} = 0.53, \widehat{p_2} = 0.47, \widehat{\sigma_1^2} = 1.06, \widehat{\sigma_2^2} = 1.01, \widehat{\mu_1} = 0.03, \widehat{\mu_2} = 5.07\}$.

### 2.4.3.2 Aplicação do algoritmo *EM* aos *HMM's*

Como vimos nos exemplos e nos tópicos anteriores, o algoritmo *EM* é um método de estimação baseado na teoria de máxima verossimilhança. A característica principal do algoritmo *EM* está na expressão de uma "forma aumentada" da verossimilhança original, que permite a obtenção de estimadores de máxima verossimilhança de forma fechada para os parâmetros de interesse. A forma aumentada da verossimilhança é obtida, ao incluirmos no problema, variáveis aleatórias adicionais (dados faltantes) que tornem simples a maximização da verossimilhança.

Ao observarmos a função de verossimilhança original dos *HMM's*, iremos constatar a impossibilidade de obtermos estimadores de forma fechada. Na aplicação do algoritmo *EM* nos *HMM's*, os dados faltantes são caracterizados pela seqüência $S_1, S_2, ..., S_t, ...$ de estados latentes.

A seguir, vamos apresentar o desenvolvimento do algoritmo *EM* aplicado a *HMM's* considerando independência e dependência condicionais entre as observações $O_1, O_2, ..., O_t, ...$ dados $S_1, S_2, ..., S_t, ...$, sendo o primeiro dos casos baseado em desenvolvimentos descritos em da Silva(2002).

#### 2.4.3.2.1 Algoritmo *EM* (*HMM* com $O_t \mid S$ condicionalmente independentes)

Considerando independência condicional entre as observações $O_1, O_2, ..., O_T$ dados $S_1, S_2, ..., S_T$, com parametrizações $\pi$, $A$ e $B$, em que $\pi$

é um vetor que descreve a distribuição inicial associada aos estados em $\Omega_S$, $A$ é a matriz de Transições da Cadeia de Markov latente e $B$ é a matriz que descreve as distribuições de probabilidade de emissão de observações em $\Omega_o$ condicionadas aos estados.

A verossimilhança associada ao modelo $\lambda$ é dada por $P(O \mid \lambda)$, enquanto a função de verossimilhança aumentada (com presença dos estados latentes) é dada por:

$$\begin{aligned} &L(\lambda \mid O,S) = P(O,S \mid \lambda) \\ &= P(O_1 = o_1,...,O_T = o_T, S_1 = s_1,...,S_T = s_T \mid \pi, A, B) \\ &= \pi_{s_1} \times \left[\prod_{t=1}^{T} b_{s_t o_t}\right] \times \left[\prod_{t=2}^{T} a_{s_{t-1} s_t}\right]. \end{aligned}$$

A maximização da função de verossimilhança dos dados aumentados permite a obtenção dos estimadores de $\lambda$.

Aplicando o primeiro passo do algoritmo *EM*, devemos obter $E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S \mid \lambda))\right]$ em que $\lambda^{(m)}$ é um valor qualquer no espaço paramétrico de $\lambda$. Desenvolvendo esta esperança, em que $\Omega_{ST}$ representa o conjunto de todas as seqüências possíveis de estados $S = (S_1 = s_1,...,S_T = s_T)$ de tamanho $T$, temos que:

$$E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S\mid\lambda))\right]$$
$$=\sum_{S\in\Omega_{ST}}\ln(P(O,S\mid\lambda))\times P(S\mid O,\lambda^{(m)})$$
$$=\frac{1}{P(O\mid\lambda^{(m)})}\sum_{S\in\Omega_{ST}}\ln(P(O,S\mid\lambda))\times P(S,O\mid\lambda^{(m)}).$$

(62)

Observe que o termo $\frac{1}{P(O\mid\lambda^{(m)})}$, presente no resultado expresso em (62), é uma constante com respeito a $S$, podendo ser desconsiderada na maximização de $E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S\mid\lambda))\right]$ com respeito a $\lambda$. Observe também que o termo $\sum_{S\in\Omega_{ST}}\ln(P(O,S\mid\lambda))\times P(S,O\mid\lambda^{(m)})$ pode ser expresso por:

$$\sum_{S\in\Omega_{ST}}\ln(P(O,S\mid\lambda))P(O,S\mid\lambda^{(m)})$$
$$=\sum_{S_1\in\Omega_S}\sum_{S_2\in\Omega_S}\cdots\sum_{S_T\in\Omega_S}\left(\ln(P(O,S_1,S_2,\ldots,S_T\mid\lambda))P(O,S_1,S_2,\ldots,S_T\mid\lambda^{(m)})\right),$$

(63)

isto é,

$$
\begin{aligned}
Q(\lambda,\lambda^{(m)}) &= \sum_{S\in\Omega_{ST}}\left[\ln(P(O,S\mid\lambda))\times P(S,O\mid\lambda^{(m)})\right] \\
&= \sum_{S\in\Omega_{ST}}\left[\left(\ln(\pi_{s_1})+\sum_{t=1}^{T}\ln(b_{s_t o_t})+\sum_{t=2}^{T}\ln(a_{s_{t-1}s_t})\right)\times P(S,O\mid\lambda^{(m)})\right] \\
&= \sum_{i\in\Omega_S}\left[\ln(\pi_{S_1=i})\times P(S_1=i,O\mid\lambda^{(m)})\right] \\
&+ \sum_{i\in\Omega_S}\left[\sum_{t=1}^{T}\left(\ln(b_{S_t=i,O_t=k})\times P(S_t=i,O\mid\lambda^{(m)})\right)\right] \\
&+ \sum_{i\in\Omega_S}\sum_{j\in\Omega_S}\left[\sum_{t=2}^{T}\left(\ln(a_{S_{t-1}=i,S_t=j})\times P(S_{t-1}=i,S_t=j,O\mid\lambda^{(m)})\right)\right].
\end{aligned}
\tag{64}
$$

Observe que na expressão dada por (64), as três partes são expressas, respectivamente, em função de $\pi$, $A$ e $B$ apenas. Como o objetivo é maximizar a expressão (63) em relação a cada um dos parâmetros presentes nas estruturas $\pi$, $A$ e $B$, basta maximizarmos cada uma das três partes de $Q$, independentemente, em relação a cada parâmetro. Definido $a_i$ como a $i$-ésima linha da matriz $A$, ou seja, $a_i = (a_{i1}, a_{i2}, ..., a_{ir})$, na qual $1 \le i \le r$, e $b_i$ $i$-ésima linha da matriz $B$, ou seja, $b_i = (b_{i1}, b_{i2}, ..., b_{in})$, na qual $1 \le i \le r$, podemos escrever as funções que compõem $Q$ como:

$$Q(\pi,\lambda^{(m)})=\sum_{i\in\Omega_S}\left[\ln(\pi_{S_1=i})\times P(S_1=i,O\mid\lambda^{(m)})\right]$$

$$Q(a_i,\lambda^{(m)})=\sum_{j\in\Omega_S}\left[\sum_{t=2}^{T}\left(\ln(a_{S_{t-1}=i,S_t=j})\times P(S_{t-1}=i,S_t=j,O\mid\lambda^{(m)})\right)\right]$$

$$Q(b_i,\lambda^{(m)})=\sum_{t=1}^{T}\left(\ln(b_{S_t=i,O_t=k})\times P(S_t=i,O\mid\lambda^{(m)})\right). \tag{65}$$

Assim, concluímos o passo esperança associado ao *EM*, sendo a esperança dos dados completos, segundo a distribuição condicional de $S\mid O,\lambda^{(m)}$, dada por:

$$E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S\mid\lambda))\right]$$
$$=\frac{1}{P(O\mid\lambda^{(m)})}\times\left[Q(\pi,\lambda^{(m)})+\sum_{i\in\Omega_S}Q(a_i,\lambda^{(m)})+\sum_{i\in\Omega_S}Q(b_i,\lambda^{(m)})\right] \tag{66}$$

Observe que cada uma das funções, $Q(\pi,\lambda^{(m)}),Q(a_i,\lambda^{(m)})$ e $Q(b_i,\lambda^{(m)})$, possui forma:

$$Q(\lambda_j)=\sum_j w_j\ln(\lambda_j). \tag{67}$$

Para maximizarmos $Q(\lambda,\lambda^{(m)})$ em relação a cada um dos parâmetros envolvidos, basta obtermos o máximo de funções $Q(\lambda_j)$, explicitadas em (67),

sujeito à restrição $\sum_j \lambda_j = 1$. Para tanto, maximizaremos $Q(\lambda, \lambda^{(m)})$ utilizando "Multiplicadores de Lagrange".

Obtendo o sistema de equações associadas à aplicação do "Multiplicador de Lagrange", temos como máximo da função condicionada $Q(\lambda_j)$ em relação a $\lambda_j$:

$$\widehat{\lambda}_j = \frac{w_j}{\sum_j w_j} \tag{68}$$

Aplicando o resultado obtido em (68), obedecendo respectivamente às restrições $\sum_{i \in \Omega_s} \pi_i = 1$, $\sum_{i \in \Omega_s} a_{ij} = 1$, $\sum_{k \in \Omega_o} b_{ik} = 1$, sendo $\pi_i \geq 0$, $a_{ij} \geq 0$, $b_{ik} \geq 0$, temos que a maximização de $Q(\pi, \lambda^{(m)}), Q(a_i, \lambda^{(m)})$ e $Q(b_i, \lambda^{(m)})$ em relação a cada $\pi_i$, $a_{ij}$ e $b_{ik}$, respectivamente implica em:

$$\pi_i^{(m+1)} = \frac{P(S_1 = i, O \mid \lambda^{(m)})}{\sum_{i \in \Omega_S} \left[ P(S_1 = i, O \mid \lambda^{(m)}) \right]}$$

$$= \frac{P(S_1 = i, O \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})} \quad \text{onde } i \in \Omega_S,$$

$$a_{ij}^{(m+1)} = \frac{\sum_{t=2}^{T} \left( P(S_{t-1} = i, S_t = j, O \mid \lambda^{(m)} \right)}{\sum_{j \in \Omega_S} \left[ \sum_{t=2}^{T} \left( P(S_{t-1} = i, S_t = j, O \mid \lambda^{(m)} \right) \right]}$$

$$= \frac{\frac{1}{P(O \mid \lambda^{(m)})} \times \sum_{t=2}^{T} \left( P(S_{t-1} = i, S_t = j \mid O, \lambda^{(m)} \right)}{\frac{1}{P(O \mid \lambda^{(m)})} \times \left[ \sum_{t=2}^{T} \left( P(S_{t-1} = i, O \mid \lambda^{(m)} \right) \right]}$$

$$= \frac{\sum_{t=2}^{T} \left( P(S_{t-1} = i, S_t = j \mid O, \lambda^{(m)} \right)}{\left[ \sum_{t=2}^{T} \left( P(S_{t-1} = i, O \mid \lambda^{(m)} \right) \right]} \quad \text{onde } i, j \in \Omega_S. \tag{69.1}$$

Para a obtenção de $b_{ik}^{(m+1)}$, devemos utilizar uma função indicadora $I(O_t, k) = 1$ se $O_t = k$ e $I(O_t, k) = 0$ se $O_t \neq k$. Então $b_{ik}^{(1)}$ é expresso por:

$$b_{jk}^{(m+1)} = \frac{\sum_{t=1}^{T}\left(P(S_t = j, O \mid \lambda^{(m)}) \times I(O_t = k)\right)}{\sum_{t=1}^{T}\left(P(S_t = j, O \mid \lambda^{(m)})\right)}$$

$$= \frac{\frac{1}{P(O \mid \lambda^{(m)})} \times \sum_{t=1}^{T}\left(P(S_t = j \mid O, \lambda^{(m)}) \times I(O_t = k)\right)}{\frac{1}{P(O \mid \lambda^{(m)})} \times \sum_{t=1}^{T}\left(P(S_t = j \mid O, \lambda^{(m)})\right)}$$

$$= \frac{\sum_{t=1}^{T}\left(P(S_t = j \mid O, \lambda^{(m)}) \times I(O_t = k)\right)}{\sum_{t=1}^{T}\left(P(S_t = j \mid O, \lambda^{(m)})\right)} \quad j \in \Omega_S,\ k \in \Omega_O. \tag{69.2}$$

Assim, concluímos o passo maximização associado ao algoritmo *EM*. Resta-nos obter os valores das probabilidades $P(S_t = i \mid O, \lambda^{(m)})$ e $P(S_{t-1} = i, S_t = j \mid O, \lambda^{(m)})$, que estão presentes nas estimativas obtidas em (69). Denotemos estas probabilidades por:

$$\gamma_{t(i)}^{(m)} = P(S_t = i \mid O, \lambda^{(m)}),$$
$$\xi_{t(i,j)}^{(m)} = P(S_{t-1} = i, S_t = j \mid O, \lambda^{(m)}). \tag{70}$$

Desenvolvendo $\gamma_{t(i)}$ e recordando as definições das variáveis *backward* e *forward*, apresentadas nas seções anteriores, temos que:

$$\gamma_{t(i)}^{(m)} = P(S_t = i \mid O, \lambda^{(m)})$$
$$= \frac{P(S_t = i, O \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$
$$= \frac{P(O_{t+1},...,O_T \mid O_1,...,O_t, S_t = i, \lambda^{(m)}) \times P(O_1,...,O_t, S_t = i \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$
$$= \frac{P(O_{t+1},...,O_T \mid S_t = i, \lambda^{(m)}) \times P(O_1,...,O_t, S_t = i \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$
$$= \frac{\beta_{(t,i)} \times \alpha_{(t,i)}}{P(O \mid \lambda^{(m)})} .$$

(71)

Reescrevendo $\xi_{t(i,j)}$ em função dos parâmetros associados ao *HMM* $\lambda^{(m)}$, temos que:

$$\xi_{t(i,j)}^{(m)} = P(S_{t-1}=i, S_t=j \mid O, \lambda^{(m)})$$

$$= \frac{P(S_{t-1}=i, S_t=j, O \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid O_1,\ldots,O_t, S_{t-1}=i, S_t=j, \lambda^{(m)}) \times$$

$$P(O_1,\ldots,O_t, S_{t-1}=i, S_t=j \mid \lambda^{(m)}) \Big]$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid S_t=j, \lambda^{(m)}) \times P(O_t \mid O_1,\ldots,O_{t-1}, S_{t-1}=i, S_t=j, \lambda^{(m)}) \times$$

$$P(O_1,\ldots,O_{t-1}, S_{t-1}=i, S_t=j \mid \lambda^{(m)}) \Big]$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid S_t=j, \lambda^{(m)}) \times P(O_t \mid S_t=j, \lambda^{(m)}) \times$$

$$P(S_t=j \mid O_1,\ldots,O_{t-1}, S_{t-1}=i, \lambda^{(m)}) \times P(O_1,\ldots,O_{t-1}, S_{t-1}=i, \lambda^{(m)}) \Big]$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid S_t=j, \lambda^{(m)}) \times P(O_t \mid S_t=j, \lambda^{(m)}) \times$$

$$P(S_t=j \mid S_{t-1}=i, \lambda^{(m)}) \times P(O_1,\ldots,O_{t-1}, S_{t-1}=i \mid \lambda^{(m)}) \Big]$$

$$= \frac{\beta_{(t,j)} \times b_{jo_t} \times a_{ij} \times \alpha_{(t-1,i)}}{P(O \mid \lambda^{(m)})}.$$

(72)

Dados $i \in \Omega_S$, $j \in \Omega_S$ e $k \in \Omega_o$, e a $m$-ésima iteração (estimativa) de $\lambda$, $\lambda^{(m)} = (\pi^{(m)}, A^{(m)}, B^{(m)})$, a $m+1$-ésima iteração de $\lambda$ é dada por:

$$\pi_i^{(m+1)} = \gamma_{1(i)}^{(m)},$$

$$a_{ij}^{(m+1)} = \frac{\sum_{t=2}^{T} \xi_{t(i,j)}^{(m)}}{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)}},$$

$$b_{ik}^{(m+1)} = \frac{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)} \times I(O_t, k)}{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)}}. \qquad (73)$$

Para a aplicação do *EM* aos *HMM's* temos, em resumo, que:

1. Estabeleça um valor inicial $\lambda^{(0)} = (\pi^{(0)}, A^{(0)}, B^{(0)})$.
2. Obtenha as variáveis $\gamma^{(m)}$ e $\xi^{(m)}$ a partir do modelo inicial $\lambda^{(m)}$, utilizando as equações (71) e (72).
3. Obtenha as novas estimativas $\lambda^{(m+1)}$ a partir de $\gamma^{(m)}, \xi^{(m)}$ e $\lambda^{(m)}$, utilizando as equações vistas em (73).
4. Substitua $\lambda^{(m)}$ por $\lambda^{(m+1)}$ e itere os passos 2, 3 e 4 até que um critério de convergência seja alcançado.

No tópico a seguir, apresentamos o desenvolvimento do algoritmo *EM* aplicado ao caso de modelos *HMM*, que pressupõem dependência condicional de ordem 1 entre as observações $O_t$'s, dados $S_t$'s.

### 2.4.3.2.1 Algoritmo *EM* (*HMM* com $O_t \mid S$ condicionalmente dependentes)

Considere o modelo $\lambda = (\pi, \pi o, A, B)$ que prevê dependência de Markov de ordem 1 entre variáveis latentes $S_t$ e dependência de ordem 1 das variáveis observáveis $O_t$ dado $S_t$. A função de verossimilhança conjunta envolvendo a seqüência de observações $O_1, O_2..., O_T$ e a seqüência de estados $S_1, S_2..., S_T$ associada ao modelo $\lambda$ é:

$$
\begin{aligned}
&L(\lambda \mid O,S) = P(O,S \mid \lambda) \\
&= P(O_1 = o_1,...,O_T = o_T, S_1 = s_1,...,S_T = s_T \mid \pi, \pi o, A, B) \\
&= P(S_1 = s_1) \times P(O_1 = o_1 \mid S_1 = s_1) \times \left[ \prod_{t=2}^{T} P(O_t = o_t \mid S_t = s_t, O_{t-1} = o_{t-1}) \right] \\
&\times \left[ \prod_{t=2}^{T} P(S_t = s_t \mid S_{t-1} = s_{t-1}) \right] \\
&= \pi_{s_1} \times \pi o_{s_1 o_1} \times \left[ \prod_{t=2}^{T} b^{s_t}_{o_t s_t} \right] \times \left[ \prod_{t=2}^{T} a_{s_{t-1} s_t} \right].
\end{aligned}
$$

De maneira similar ao caso de um *HMM* no qual se considera independência entre observações dados os estados, na estimação dos parâmetros associados ao modelo $\lambda$ em estudo, faz-se necessário o uso do algoritmo *EM*.

Para a aplicação do algoritmo *EM*, devemos obter a esperança da log-verossimilhança dos dados completos (dados observáveis e latentes) segundo a distribuição dos dados latentes (ou não observáveis) condicionados aos dados observados. Os dados observáveis são as realizações das variáveis $O_1, O_2,..., O_T$, enquanto os dados latentes são as realizações da seqüência de estados $S_1, S_2,..., S_T$.

Seja $\lambda^{(m)}$ um conjunto qualquer de valores dos parâmetros do modelo $\lambda$. Desenvolvendo $E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S|\lambda))\right]$, temos:

$$E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S|\lambda))\right] = \sum_{S\in\Omega_{ST}} \ln(P(O,S|\lambda)) \times P(S|O,\lambda^{(m)})$$

$$= \sum_{S\in\Omega_{ST}} \ln(P(O,S|\lambda)) \times \frac{P(S,O|\lambda^{(m)})}{P(O|\lambda^{(m)})}$$

$$= \frac{1}{P(O|\lambda^{(m)})} \times \sum_{S\in\Omega_{ST}} \ln(P(O,S|\lambda)) \times P(S,O|\lambda^{(m)}). \quad (74)$$

Como $\frac{1}{P(O|\lambda^{(m)})}$ é uma constante com respeito a $\lambda$, este termo pode ser desconsiderado na maximização de $E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S|\lambda))\right]$ com relação à $\lambda$. Desenvolvendo o somatório em (74), e utilizando a igualdade expressa em (63), temos:

$$E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S\mid\lambda))\right]=\sum_{S\in\Omega_{ST}}\ln(P(O,S\mid\lambda))\times P(S,O\mid\lambda^{(m)})$$

$$=\sum_{S\in\Omega_{ST}}\left[\left(\ln(\pi_{s_1=i})+\ln(\pi o_{s_1=i,o_1})+\sum_{t=2}^{T}\ln(b_{o_{t-1}o_t}^{s_t=i})+\sum_{t=2}^{T}\ln(a_{s_{t-1}=i,s_t=j})\right)\times\right.$$

$$\left.P(S,O\mid\lambda^{(m)})\right]$$

$$=\sum_{i\in\Omega_S}\left[\ln(\pi_{s_1=i})\times P(S_1=1,O\mid\lambda^{(m)})\right]$$

$$+\sum_{i\in\Omega_S}\left[\ln(\pi o_{s_1=i,o_1=k})\times P(S_1=1,O\mid\lambda^{(m)})\right]$$

$$+\sum_{i\in\Omega_S}\left[\sum_{t=2}^{T}\ln(b_{o_{t-1}=k,o_t=q}^{s_t=i})\times P(S_t=i,O\mid\lambda^{(m)})\right]$$

$$+\sum_{i\in\Omega_S}\sum_{j\in\Omega_S}\left[\sum_{t=2}^{T}\ln(a_{s_{t-1}=i,s_t=j})\times P(S_1=i,S_1=j,O\mid\lambda^{(m)})\right]. \tag{75}$$

Note que $\sum_{S\in\Omega_{ST}}\ln(P(O,S\mid\lambda))\times P(S,O\mid\lambda^{(m)})$ pode ser escrita como uma soma de funções dos parâmetros $\pi,\pi o,A,B$. Considerando $\pi o_i$ como sendo a $i$-ésima linha da matriz $\pi o$, $a_i$ como a $i$-ésima linha da matriz $A$ e $b_k^i$ a $k$-ésima linha da matriz $B^i$, sejam:

$$Q(\pi,\lambda^{(m)})=\sum_{i\in\Omega_S}\left[\ln(\pi_{s_1=i})\times P(S_1=1,O\mid\lambda^{(m)})\right],$$

$$Q(\pi o_i,\lambda^{(m)})=\sum_{i\in\Omega_S}\left[\ln(\pi o_{s_1=i,o_1=k})\times P(S_1=1,O\mid\lambda^{(m)})\right],$$

$$Q(a_i,\lambda^{(m)})=\sum_{j\in\Omega_S}\left[\sum_{t=2}^{T}\ln(a_{s_{t-1}=i,s_t=j})\times P(S_1=i,S_1=j,O\mid\lambda^{(m)})\right],$$

$$Q(b_k^i,\lambda^{(m)})=\sum_{t=2}^{T}\ln(b^{s_t=i}_{o_{t-1}=k,o_t=q})\times P(S_t=i,O\mid\lambda^{(m)}).$$

(76)

Como no modelo tratado no tópico anterior, temos que a esperança condicional da distribuição dos dados completos é:

$$E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S\mid\lambda))\right]$$
$$=\frac{1}{P(O\mid\lambda^{(m)})}\times\Big[Q(\pi,\lambda^{(m)})+Q(\pi o_i,\lambda^{(m)})+$$
$$\sum_{i\in\Omega_S}Q(a_i,\lambda^{(m)})+\sum_{i\in\Omega_S}Q(b_k^i,\lambda^{(m)})\Big].$$

(77)

Para obter a maximização da esperança condicional em (77) com respeito a $\lambda$, basta realizar a maximização individual de cada uma das funções, $Q(\pi,\lambda^{(m)})$, $Q(\pi o_i,\lambda^{(m)})$, $\sum_{i\in\Omega_S}Q(a_i,\lambda^{(m)})$ e $\sum_{i\in\Omega_S}Q(b_k^i,\lambda^{(m)})$. Cada um destes termos possui a mesma forma geral dada em (67). As restrições associadas são,

respectivamente, $\sum_{i\in\Omega_S}\pi_i=1$, $\sum_{k\in\Omega_O}\pi o_{ik}=1$, $\sum_{j\in\Omega_S}a_{ij}=1$ e $\sum_{k\in\Omega_O}b_{kq}^{i}=1$, para todo $i,j\in\Omega_S$ e para todo $k,q\in\Omega_O$ sendo $\pi_i\geq 0$, $a_{ij}\geq 0$ e $b_{kq}^{i}\geq 0$. Desta forma, $E_{S|O,\lambda^{(m)}}\left[\ln(P(O,S\mid\lambda))\right]$ é maximizada, quando:

$$\pi_i^{(m+1)}=\frac{P(S_1=i,O\mid\lambda^{(m)})}{\sum_{i\in\Omega_S}\left[P(S_1=i,O\mid\lambda^{(m)})\right]}$$
$$=\frac{P(S_1=i,O\mid\lambda^{(m)})}{P(O\mid\lambda^{(m)})},\quad i\in\Omega_S,$$

$$\pi o_{ik}^{(m+1)}=\frac{P(S_1=i,O\mid\lambda^{(m)})\times I(O_t,k)}{\sum_{i\in\Omega_S}\left[P(S_1=i,O\mid\lambda^{(m)})\right]},\quad i\in\Omega_S,\quad k\in\Omega_O,$$

$$a_{ij}^{(m+1)}=\frac{\sum_{t=2}^{T}\left(P(S_{t-1}=i,S_t=j,O\mid\lambda^{(m)})\right)}{\sum_{j\in\Omega_S}\left[\sum_{t=2}^{T}\left(P(S_{t-1}=i,S_t=j,O\mid\lambda^{(m)})\right)\right]}$$
$$=\frac{\frac{1}{P(O\mid\lambda^{(m)})}\times\sum_{t=2}^{T}\left(P(S_{t-1}=i,S_t=j\mid O,\lambda^{(m)})\right)}{\frac{1}{P(O\mid\lambda^{(m)})}\times\left[\sum_{t=2}^{T}\left(P(S_{t-1}=i,O\mid\lambda^{(m)})\right)\right]}$$
$$=\frac{\sum_{t=2}^{T}\left(P(S_{t-1}=i,S_t=j\mid O,\lambda^{(m)})\right)}{\left[\sum_{t=2}^{T}\left(P(S_{t-1}=i,O\mid\lambda^{(m)})\right)\right]}\quad i,j\in\Omega_S.$$

(78.1)

Utilizando-se da função indicadora $I(O_t,k)=1$ se $O_t=k$ e $I(O_t,k)=0$ caso contrário, temos a estimativa de $b_{kq}^{i\ (m+1)}$ dada por:

$$b_{kq}^{i\ (m+1)}=\frac{\sum_{t=2}^{T}\left(P(S_t=i,O|\lambda^{(m)})\times I(O_{t-1}=k)\times I(O_t=q)\right)}{\sum_{t=2}^{T}\left(P(S_t=i,O|\lambda^{(m)})\right)}$$

$$=\frac{\frac{1}{P(O|\lambda^{(m)})}\times\sum_{t=2}^{T}\left(P(S_t=j|O,\lambda^{(m)})\times I(O_{t-1}=k)\times I(O_t=q)\right)}{\frac{1}{P(O|\lambda^{(m)})}\times\sum_{t=2}^{T}\left(P(S_t=i|O,\lambda^{(m)})\right)}$$

$$=\frac{\sum_{t=2}^{T}\left(P(S_t=j|O,\lambda^{(m)})\times I(O_{t-1}=k)\times I(O_t=q)\right)}{\sum_{t=2}^{T}\left(P(S_t=i|O,\lambda^{(m)})\right)},\quad k,q\in\Omega_O. \tag{78.2}$$

Devemos ainda calcular as probabilidades $P(S_t=i,O|\lambda^{(m)})$ e $P(S_{t-1}=i,S_t=j,O|\lambda^{(m)})$, sendo que o modelo $\lambda^{(m)}$ pressupõe a dependência condicional entre observações $O_t$'s dados os estados $S_t$'s. Denotemos estas probabilidades por:

$$\gamma_{t(i)}^{(m)}=P(S_t=i|O,\lambda^{(m)}),$$
$$\xi_{t(i,j)}^{(m)}=P(S_{t-1}=i,S_t=j|O,\lambda^{(m)}). \tag{79}$$

Desenvolvendo $\gamma_{t(i)}^{(m)}$, temos como resultado:

$$\gamma_{t(i)}^{(m)} = P(S_t = i \mid O, \lambda^{(m)})$$
$$= \frac{P(S_t = i, O \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$
$$= \frac{P(O_{t+1},\ldots,O_T \mid O_1,\ldots,O_t, S_t = i, \lambda^{(m)}) \times P(O_1,\ldots,O_t, S_t = i \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$
$$= \frac{P(O_{t+1},\ldots,O_T \mid O_t, S_t = i, \lambda^{(m)}) \times P(O_{t+1},\ldots,O_T, S_t = i \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$
$$= \frac{\beta_{d_1(t,i)} \times \alpha_{d_1(t,i)}}{P(O \mid \lambda^{(m)})}. \qquad (80)$$

Desenvolvendo o termo $\xi_{t(i,j)}^{(m)}$, tem-se:

$$\xi_{t(i,j)}^{(m)} = P(S_{t-1}=i, S_t=j \mid O, \lambda^{(m)})$$

$$= \frac{P(S_{t-1}=i, S_t=j, O \mid \lambda^{(m)})}{P(O \mid \lambda^{(m)})}$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid O_1,\ldots,O_t, S_{t-1}=i, S_t=j, \lambda^{(m)}) \times$$

$$P(O_1,\ldots,O_t, S_{t-1}=i, S_t=j \mid \lambda^{(m)}) \Big]$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid O_t, S_t=j, \lambda^{(m)}) \times P(O_t \mid O_1,\ldots,O_{t-1}, S_{t-1}=i, S_t=j, \lambda^{(m)}) \times$$

$$P(O_1,\ldots,O_{t-1}, S_{t-1}=i, S_t=j \mid \lambda^{(m)}) \Big]$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid O_t, S_t=j, \lambda^{(m)}) \times P(O_t \mid O_{t-1}, S_t=j, \lambda^{(m)}) \times$$

$$P(S_t=j \mid O_1,\ldots,O_{t-1}, S_{t-1}=i, \lambda^{(m)}) \times P(O_1,\ldots,O_{t-1}, S_{t-1}=i, \lambda^{(m)}) \Big]$$

$$= \frac{1}{P(O \mid \lambda^{(m)})} \times \Big[ P(O_{t+1},\ldots,O_T \mid O_t, S_t=j, \lambda^{(m)}) \times P(O_t \mid O_{t-1}, S_t=j, \lambda^{(m)}) \times$$

$$P(S_t=j \mid S_{t-1}=i, \lambda^{(m)}) \times P(O_1,\ldots,O_{t-1}, S_{t-1}=i \mid \lambda^{(m)}) \Big]$$

$$= \frac{\beta_{d_1(t,j)} \times b^j_{o_{t-1}} \times a_{ij} \times \alpha_{d_1(t-1,i)}}{P(O \mid \lambda^{(m)})}.$$

(81)

Dados $i, j \in \Omega_S$ e $k, l \in \Omega_O$, e a $m$-ésima iteração (estimativa) de $\lambda$, a $m+1$-ésima iteração de $\lambda$ é dada por:

$$\pi_i^{(m+1)} = \gamma_{1(i)}^{(m)},$$

$$\pi o_{ik}^{(m+1)} = \frac{\gamma_{1(i)}^{(m)} \times I(O_1, k)}{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)}},$$

$$a_{ij}^{(m+1)} = \frac{\sum_{t=2}^{T} \xi_{t(i,j)}^{(m)}}{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)}},$$

$$b_{kl}^{i(m+1)} = \frac{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)} \times I(O_{t-1}, k) \times I(O_t, l)}{\sum_{t=2}^{T} \gamma_{t(i)}^{(m)}}. \qquad (82)$$

Assim, temos como resumo do algoritmo *EM* aplicado ao *HMM* com dependência entre observações dados os estados:

1. Estipule um valor inicial para $\lambda^{(0)} = (\pi^{(0)}, \pi o^{(0)}, A^{(0)}, B^{(0)})$
2. Obtenha $\gamma^{(m)}$ e $\xi^{(m)}$ a partir do modelo $\lambda^{(m)}$, utilizando as equações (80) e (81).
3. Obtenha a nova estimativa $\lambda^{(m+1)} = (\pi^{(m+1)}, \pi o^{(m+1)}, A^{(m+1)}, B^{(m+1)})$ a partir de $\gamma^{(m)}$ e $\xi^{(m)}$ utilizando as equações (82).
4. Substitua a $\lambda^{(m)}$ por $\lambda^{(m+1)}$ e repita os passos 2, 3 e 4 até que a convergência seja atingida.

Nos próximos tópicos, trataremos de outras questões importantes associadas aos *HMM*'s.

## 2.5 Simulando *HMM'S*

O *HMM* é um modelo probabilístico baseado na caracterização de dois tipos de seqüências de variáveis aleatórias, $S_t$ e $O_t$, onde $t = 1,..,T$. Assim, dado um modelo $\lambda$, é possível simular amostras das variáveis $S_t$ e $O_t$. Naturalmente, as realizações das variáveis aleatórias $S_t$ são não observáveis em aplicações reais. Entretanto, em dados simulados, elas são conhecidas. Para a utilização de amostras simuladas nas inferências, devemos utilizar apenas as variáveis observáveis dos dados, ou seja, as realizações dos $O_t$'s.

Nesta seção, iremos abordar apenas o caso da simulação de *HMM's* de natureza discreta. Segundo Rabiner (1989), para a simulação de uma amostra de tamanho $T$ de um *HMM* parametrizado por $\lambda$, que pressupõe independência condicional entre as observações $O_t$'s dados os estados $S_t$'s, o procedimento pode ser sumarizado nos passos seguintes:

1. A partir da distribuição inicial $P(S_1 \mid \lambda)$ dada pelo vetor $\pi$, simule uma realização de $S_1$ denotada por $s_1$.
2. A partir da distribuição $P(O_1 \mid S_1 = s_1, \lambda)$, estabelecida pelo vetor de probabilidades $b_i$ contido na matriz $B$, simule uma observação de $O_1$. Este valor será denotada por $o_1$.
3. A partir da distribuição $P(S_t \mid S_{t-1} = s_{t-1}, \lambda)$, simule uma observação de $s_t$ condicionalmente ao valor simulado $s_{t-1}$ de $S_{t-1}$. Observe que $P(S_t \mid S_{t-1} = s_{t-1}, \lambda)$ é definida por um vetor (linha que define a distribuição condicional de

$S_t \mid S_{t-1} = s_{t-1}, \lambda)$ contido na matriz de transições $A$. Esta observação será denotada por $s_t$.

4. A partir da distribuição $P(O_t \mid S_t = s_t, \lambda)$, simule uma observação de $O_t$ condicionalmente ao valor simulado $s_t$ de $S_t$. Esta observação será denotada por $o_t$.
5. Repita os passos 3 e 4 até que $t = T$.

É possível obter uma seqüência de passos para simulação de observações de um modelo *HMM*, que prevê dependência condicional entre as observações $O_t \mid S$ com pequenas alterações.

No próximo tópico, vamos trabalhar com uma esperança particular para a verificação de padrões de emissões de observações $O_t$.

## 2.6 Construção de Mapas das Probabilidades de Emissão das Observações

Em algumas aplicações é interessante obter a probabilidade da ocorrência de um dado valor de $k$ a cada instante $t$ $(O_t = k)$. Podemos, para isso, trabalhar com a média ponderada das probabilidades de emissão de um valor $k$, fazendo a ponderação da cadeia estar no estado $i$ no tempo $t$ $(S_t = i)$, dada a seqüência de observações $P(S_t \mid O, \lambda)$. Desta forma, temos a esperança de $P(O_t = k \mid S_t, \lambda)$ em que $k \in \Omega_O$, com relação à distribuição da variável aleatória $S_t \mid O, \lambda$ apresentada por Churchill (1989). Desenvolvendo então a $E_{S_t \mid O, \lambda}\left[P(O_t = k \mid S_t, \lambda)\right]$, sendo o modelo $\lambda$ associado a um *HMM* com independência entre as observações dados os estados, temos que:

$$E_{S_t|O,\lambda}\left[P(O_t=k\mid S_t,\lambda)\right]=\sum_{i\in\Omega_S}P(O_t=k\mid S_t=i,\lambda)\times P(S_t=i\mid O,\lambda)$$

$$=\sum_{i\in\Omega_S}b_{ik}\times\gamma_{t(i)}.$$

(83)

Trabalhando com as estimativas obtidas pelo algoritmo *EM*, substituímos o modelo $\lambda$ por $\hat{\lambda}$, no qual $\hat{\lambda}$ é o estimador de máxima verossimilhança.

Em algumas aplicações nas quais existem padrões distintos de emissão das observações $O_t$ ao longo do tempo $t$, a construção de um gráfico, cujos pontos são constituídos dos pares $\left(t, E_{S_t|O,\lambda}\left[P(O_t=k\mid S_t,\hat{\lambda})\right]\right)$, possibilita a localização de regiões homogêneas quanto à emissão de $k$. Casos em que o conjunto de observações $\Omega_O$ têm dois elementos (somente dois tipos de observações), através de um único gráfico, de qualquer dos dois elementos de $\Omega_O$, é possível visualizar regiões que possuem padrões homogêneos. Isso, pois $\sum_{k\in\Omega_o}E_{S_t|O,\lambda}\left[P(O_t=k\mid S_t,\hat{\lambda})\right]=1 \quad \forall \ t$. Já, quando $\Omega_O$ tem mais de dois elementos, é necessário traçar os gráficos associados a cada elemento de $\Omega_O$ para verificar regiões de homogeneidade em emissão de $O_t$'s. No trabalho de Churchill (1989), esta esperança é denotada por composição local e será esta denominação adotada no decorrer do trabalho.

## 2.7 Problemas Numéricos

Em várias aplicações, a grande dimensão da seqüência de observações $O$ pode gerar problemas de natureza numérica no cálculo de $P(O \mid \lambda)$. Isso se deve à capacidade limitada de endereçamento de valores pelos softwares ou linguagens pelas quais são realizadas as implementações dos métodos associados ao *HMM* que pode ser ilustrada na seguinte situação. Suponha que uma máquina ou determinada linguagem possa endereçar apenas valores com precisão máxima $10^{-5000}$, ou seja, valores menores do que esta precisão serão considerados zero. Agora suponha um *HMM* equiprovável, ou seja, todas as seqüências $O$ de tamanho $T$ possuem valores de probabilidades iguais. Assim, $P(O \mid \lambda) = \frac{1}{n^T}$, para qualquer seqüência $O$ de tamanho $T$. Suponha uma seqüência de $O$ com tamanho $T = 50000$ e a precisão apresentada anteriormente. Nesta máquina, a probabilidade $P(O \mid \lambda)$ iguala a zero. Este tipo de situação tornaria inviável a aplicação.

Com uma pequena manipulação algébrica no cálculo das variáveis, *forward* e *backward* ($\alpha$ , $\beta$), e com a utilização da transformação logarítmica, é possível obter o valor correto de $P(O \mid \lambda)$. Trabalhando com um *HMM* $\lambda$ que prevê independência condicional entre as observações dados os estados latentes, as definições das variáveis, *forward* e *backward* são:

$$\alpha_{(t,s_t)} = P(O_1 = o_1, O_2 = o_2 \ldots, O_t = o_t, S_t = s_t \mid \lambda)$$
$$\beta_{(t,s_t)} = P(O_{t+1} = o_{t+1}, O_{t+2} = o_{t+2} \ldots, O_T = o_T \mid S_t = s_t, \lambda). \quad (84)$$

Se observarmos a construção das matrizes *forward* e *backward*, à medida que os valores são obtidos, eles tendem a se aproximar de zero. No caso específico da

variável *forward*, à medida que os valores $\alpha_{(t,s_t)}$, da linha $t$, são construídos, em que $t$ se aproxima de $T$, eles se aproximam de zero. No caso da variável *backward*, os valores de $\beta_{(t,s_t)}$, da linha $t$, aproximam-se de zero, quando $t$ se aproxima de 1.

A estratégia a se utilizar é evitar que estes valores das matrizes $\alpha$ e $\beta$ se aproximem de zero. Uma maneira de contornar o problema é efetuar uma *normalização* na $t$-ésima linha de $\alpha$ e $\beta$ após a sua construção. Ao procedermos desta forma, estaremos dividindo cada elemento da $t$-ésima linha pela soma da mesma. Os valores $\alpha_{(t,s_t)}$ e $\beta_{(t,s_t)}$ normalizados, denotados por $\alpha^{Norm}_{(t,s_t)}$ e $\beta^{Norm}_{(t,s_t)}$, não são necessários para realizar o cálculo $P(O \mid \lambda)$. Para isso, armazenaremos a constante normalizadora de cada uma das linhas associadas às matrizes *forward* e *backward*. Vamos armazenar cada uma das constantes em um vetor $C_t$. Os valores do vetor $C_t^f$, da variável *forward*, onde $1 \le t \le T$, são:

$$C_t^f = \begin{cases} P(O_1 \mid \lambda), se\ t = 1, \\ P(O_t \mid O_1, \ldots, O_{t-1}, \lambda), se\ t > 1. \end{cases} \tag{85}$$

Os valores de $C_t^b$ para a variável *backward*, onde $1 \le t \le T-1$, são expressos exclusivamente em função do índice $t$. Tal fato será fundamental, para validar a utilização do processo de normalização.

Observe que é possível obter a log-verossimilhança $\ln(P(O \mid \lambda))$ através da soma do logaritmo de cada elemento de $C_t$. Assim, temos que:

$$\ln(P(O \mid \lambda)) = \sum_{t=2}^{T} \ln(P(O_t \mid O_{t-1}, \ldots, O_1, \lambda)) + \ln(P(O_1 \mid \lambda))$$

$$= \sum_{t=1}^{T} \ln(C_t^f).$$

(86)

As matrizes $\alpha^{Norm}$ e $\beta^{Norm}$, embora desnecessárias no cálculo de $P(O \mid \lambda)$, serão fundamentais na obtenção de estimativas do modelo $\lambda$ (algoritmo *EM*). Os valores normalizados $\alpha_{(t,s_t)}$ e $\beta_{(t,s_t)}$ são dados por:

$$\alpha_{(t,s_t)}^{Norm} = \frac{P(O_1 = o_1, \ldots, O_t = o_t, S_t = s_t \mid \lambda)}{C_t^f}$$

$$= \frac{\alpha_{(t,s_t)}}{C_t^f},$$

$$\beta_{(t,s_t)}^{Norm} = \frac{P(O_{t+1} = o_{t+1}, \ldots, O_T = o_T \mid S_t = s_t, \lambda)}{C_t^b}$$

$$= \frac{\beta_{(t,s_t)}}{C_t^b}.$$

(87)

Para realizarmos a inferência sobre o modelo $\lambda$, devemos obter com as matrizes $\alpha^{Norm}$ e $\beta^{Norm}$ as variáveis $\gamma_{t(i)} = P(S_t = i \mid O, \lambda)$ e $\xi_{t(i,j)} = P(S_{t-1} = i, S_t = j \mid O, \lambda)$ necessárias para aplicação do algoritmo *EM*.

Vamos utilizar $\alpha^{Norm}$ e $\beta^{Norm}$ em lugar de $\alpha$ e $\beta$ na equação (71). Vamos denotar esta estrutura por $\gamma_{t(i)}^{p}$:

$$\gamma_{t(i)}^{p} = \frac{\alpha_{(t,i)}^{Norm} \times \beta_{(t,i)}^{Norm}}{P(O|\lambda)}$$

$$= \frac{\frac{\alpha_{(t,s_t)}}{C_t^f} \times \frac{\beta_{(t,s_t)}}{C_t^b}}{P(O|\lambda)}$$

$$= \frac{\alpha_{(t,s_t)} \times \beta_{(t,s_t)} \times P(O|\lambda)}{C_t^f \times C_t^b}$$

$$= \frac{P(S_t = i, O|\lambda)}{C_t^f \times C_t^b}. \qquad (88)$$

Vamos obter o somatório da estrutura $\gamma_{t(i)}^{p}$ com relação ao índice $i$. Isso será necessário para obter a normalização da estrutura $\gamma_{t(i)}^{p}$. Vamos denotar esta constante por $C_t^q$:

$$C_t^q = \sum_{i \in \Omega_S} \gamma_{t(i)}^{p}$$

$$= \sum_{i \in \Omega_S} \left( \frac{P(S_t = i, O|\lambda)}{C_t^f \times C_t^b} \right)$$

$$= \frac{1}{C_t^f \times C_t^b}$$

$$\times \sum_{i \in \Omega_S} P(S_t = i, O|\lambda)$$

$$= \frac{P(O|\lambda)}{C_t^f \times C_t^b}. \qquad (89)$$

Desta forma, cada elemento normalizado da estrutura $\gamma_{t(i)}^{p}$ terá o seguinte valor:

$$\gamma_{t(i)}^{Norm} = \frac{\gamma_{t(i)}^{p}}{C_t^q}$$

$$= \frac{\frac{P(S_t = i, O \mid \lambda)}{C_t^f \times C_t^b}}{\frac{P(O \mid \lambda)}{C_t^f \times C_t^b}}$$

$$= \frac{P(S_t = i, O \mid \lambda)}{P(O \mid \lambda)}$$

$$= P(S_t = i \mid O, \lambda)$$

$$= \gamma_{t(i)}.$$

(90)

Iremos utilizar as variáveis $\alpha_{(t,i)}^{Norm}$ e $\beta_{(t,i)}^{Norm}$ em lugar de $\alpha_{(t,i)}$ e $\beta_{(t,i)}$, na equação (72), para verificar sua relação com $\xi_{t(i,j)}$. Denotemos esta estrutura por $\xi_{t(i,j)}^{p}$, na qual $1 \leq t \leq T$, $i \in \Omega_S$ e $j \in \Omega_S$. Temos, então:

$$\xi^{p}_{t(i,j)} = \frac{\beta^{Norm}_{(t,j)} \times b_{jo_t} \times a_{ij} \times \alpha^{Norm}_{(t-1,i)}}{P(O|\lambda)}$$

$$= \frac{1}{P(O|\lambda)} \times [\frac{P(O_{t+1},...,O_T | S_t = j, \lambda)}{C^b_t} \times P(O_t | O_{t-1}, S_t = j, \lambda) \times$$

$$P(S_t = j | S_{t-1} = i, \lambda) \times \frac{P(O_1,...,O_{t-1}, S_{t-1} = i | \lambda)}{C^f_{t-1}}]$$

$$= \frac{\beta_{(t,j)} \times b_{jo_t} \times a_{ij} \times \alpha_{(t-1,i)}}{P(O|\lambda) \times C^b_t \times C^f_{t-1}}$$

$$= \frac{P(S_{t-1} = i, S_t = j, O | \lambda)}{P(O|\lambda) \times C^b_t \times C^f_{t-1}}$$

$$= \frac{\xi_{t(i,j)}}{C^b_t \times C^f_{t-1}} . \qquad (91)$$

Portanto, agora iremos obter o somatório de $\xi^{p}_{t(i,j)}$ com relação à $i$ e $j$, a fim de calcular a constante normalizadora da variável $\xi^{p}_{t(i,j)}$. Iremos armazenar as constantes em um vetor $C^e_t$. Assim, temos:

$$
\begin{aligned}
C_t^e &= \sum_{i\in\Omega_S}\sum_{j\in\Omega_S}\xi_{t(i,j)}^p \\
&= \sum_{i\in\Omega_S}\sum_{j\in\Omega_S}\left(\frac{\xi_{t(i,j)}}{C_t^b \times C_{t-1}^f}\right) \\
&= \sum_{i\in\Omega_S}\sum_{j\in\Omega_S}\left(\frac{\xi_{t(i,j)}}{C_t^b \times C_{t-1}^f}\right) \\
&= \frac{1}{C_t^b \times C_{t-1}^f} \times \sum_{i\in\Omega_S}\sum_{j\in\Omega_S}\xi_{t(i,j)} \\
&= \frac{1}{C_t^b \times C_{t-1}^f}.
\end{aligned}
\tag{92}
$$

Desta forma, cada elemento normalizado da estrutura $\xi_{t(i,j)}^p$ terá o seguinte valor:

$$
\begin{aligned}
\xi_{t(i,j)}^{Norm} &= \frac{\xi_{t(i,j)}^p}{C_t^e} \\
&= \frac{\xi_{t(i,j)} \times C_t^b \times C_{t-1}^f}{C_t^b \times C_{t-1}^f} \\
&= \xi_{t(i,j)}.
\end{aligned}
\tag{93}
$$

Assim, a partir das variáveis normalizadas, é possível realizar, de forma eficiente, as inferências sobre os parâmetros do modelo $\lambda$ associado a um *HMM*.

## 2.8 Comparando *HMM*'s

Até o momento, todos os desenvolvimentos realizados têm como base a pressuposição do conhecimento do número de estados que um *HMM* possui, ou seja, o valor $r$ que corresponde ao número de elementos do conjunto $\Omega_S$. Existem alguns critérios para se escolher o número de estados que melhor descreve um dado conjunto de dados. Os critérios apresentados aqui são baseados, invariavelmente, no ajuste de modelos, considerando várias possibilidades para o número de estados.

Observe que, como estamos trabalhando com um modelo probabilístico, a verossimilhança é um bom indicativo para a determinação do número de estados. Assim, um critério, para o estabelecimento do número de estados, consiste em adotar o modelo $\lambda$, com número de estados $r$, que maximiza a verossimilhança. Isso pode ser feito de maneira formal, adotando-se o Teste de Razão de Verossimilhança (Mood, 1963). No entanto, ao adotarmos a verossimilhança isoladamente, modelos com maior número de estados serão privilegiados. Modelos com maior número de estados são mais onerosos do ponto de vista computacional, além de serem de interpretação mais complexas. Preferencialmente, devemos estabelecer uma ponderação entre a complexidade e a verossimilhança. Então, devemos obter um equilíbrio entre o custo (número de parâmetros) e o benefício (verossimilhança) que um determinado modelo oferece. Os critérios *Bayesian Information Criterion* (*BIC*) (Schwarz, 1978) e *Akaikes's Information Criterion* (*AIC*) (Sakamoto, 1986) foram desenvolvidos de modo a permitir tal equilíbrio na escolha de modelos. Seja $\lambda$ um modelo associado a um *HMM*, $K$ os graus de liberdade associados ao modelo, e $T$ o tamanho da seqüência em estudo, os critérios *AIC* e *BIC* são dados pelas expressões:

$$AIC = -2\ln(P(O \mid \hat{\lambda})) + 2K,$$
$$BIC = -2\ln(P(O \mid \hat{\lambda})) + K\ln(T). \tag{94}$$

Devemos optar pelo modelo que minimiza os valores de *AIC* e *BIC*.

Muitas vezes é interessante observar mudanças significativas dos valores de *BIC* e *AIC*, em relação a um modelo mais simples, de acordo com a modificação do modelo. Podemos, como exemplo, adotar o modelo independente e equiprovável como mais simples. Neste modelo, temos que $\ln(P(O \mid \lambda)) = \ln\left(\frac{1}{n^T}\right) = T \times \ln\left(\frac{1}{n}\right) = -T \times \ln(n)$, e o grau de liberdade associado é zero. Assim, o *BIC* associado ao modelo mais simples é dado por $2 \times T \times \ln(n)$. Podemos adotar a diferença entre os valores do *BIC*, associado a um modelo qualquer e a um modelo mais simples, denotado por $\Delta BIC$, como um critério (Churchill,1992). O valor de $\Delta BIC$ é expresso por:

$$\Delta BIC = \left[-2\ln(P(O \mid \hat{\lambda})) + K\ln(T)\right] - \left[2 \times T \times \ln(n)\right]. \tag{95}$$

No capítulo seguinte, apresentaremos algumas aplicações dos *HMM's*. A ênfase maior está no estudo de segmentação de DNA (Churchil, 1989).

# 3. APLICAÇÕES DAS CADEIAS DE MARKOV COM ESTADOS LATENTES (*HMM'S*)

Os modelos apresentados no decorrer do capítulo 2 podem ser aplicados a uma série de problemas em diversas áreas do conhecimento. No início deste capítulo, iremos abordar algumas aplicações de uso corrente na literatura. Num segundo momento, iremos abordar, especificamente, o problema da *Segmentação de DNA,* que é o foco principal deste trabalho.

## 3.1 Reconhecimento de Fala

A "*Fala*" é uma das mais importantes formas de comunicação entre os homens. A "*Fala*" poseria ser uma forma de interface muito eficaz entre homens e máquinas. No entanto, há a necessidade de sistemas que possam estabelecer esta comunicação de maneira eficiente. O *HMM*, como estrutura associada a problemas de reconhecimento de padrão, tem uma de suas aplicações ligada ao problema de *Reconhecimento de Fala*. Neste sub-tópico, objetivamos apresentar o problema de *Reconhecimento de Fala* de forma sucinta e a respectiva modelagem ao problema utilizando o *HMM* (Rabiner, 1989) (Russel, 1995).

O problema do *Reconhecimento da Fala* pode ser definido como o processo de mapeamento de uma seqüência de sinais acústicos digitalizados em uma seqüência de palavras.

A *Fala*, primordialmente, é uma seqüência de sons. Em outras palavras, a *Fala* pode ser caracterizada como um sinal analógico. Este sinal pode ser digitalizado através de uma interface de som qualquer (microfone e placas de som). Se analisarmos a *Fala*, podemos verificar que sua composição básica é feita através dos fonemas. A linguagem humana possui entre 40 e 50 fonemas. A questão básica é tentar estabelecer uma relação entre os fonemas (som) e a sua

forma textual (forma escrita). Existem questões que complicam o processo de *Reconhecimento de Fala*. Uma delas está associada a sons equivalentes que diferentes fonemas emitem. Outras questões estão associadas ao estabelecimento de intervalos entre as palavras.

Como dissemos, os sons são sinais acústicos analógicos que podem ser digitalizados. Estes sinais são caracterizados pela freqüência. Basicamente o que um conversor de sinal analógico-digital realiza, é capturar estas freqüências atingidas pelos sons emitidos na interface (microfone e placa de som). Assim, o que obtemos, ao fim do processo de conversão do sinal, é um vetor com os valores das freqüências atingidas pelos sons, capturadas pela interface (estrutura de dados digital).

O próximo passo é estabelecer um mapeamento dos fonemas e suas características em termos de freqüências. O vetor com as amostras das freqüências é mapeado em uma estrutura menor, denominada "*Frame*" (Russel, 1995). Cada *frame* possui um certo número de amostras do vetor inicial. Ao fim do processo, é obtido um vetor de *frames*, que é uma estrutura de dados digital que armazena uma seqüência de sons. A figura 14, adaptada, ilustra este processo (Russel, 1995).

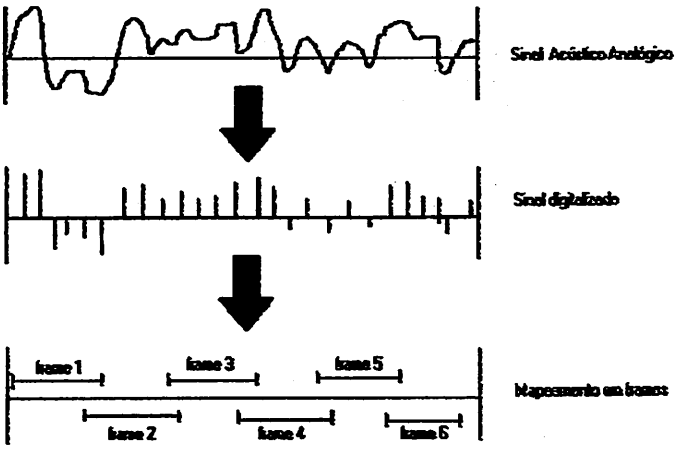



FIGURA 14. Ilustração do processo de digitalização do sinal analógico da *fala*.

Com o resultado da fase inicial, caracterizada pela digitalização do sinal acústico da fala, é necessário estabelecer a relação entre os fonemas e os frames. Como existe a presença de incerteza no processo de amostragem do sinal sonoro, uma boa estratégia é utilizarmos os métodos e modelos probabilísticos. Observe que a relação entre um *frame* e um fonema não é biunívoca. Assim, é natural estabelecermos a probabilidade de um *frame* estar associado a um fonema, isto é, a $P(frame \mid fonema)$. Observe ainda que, por construção natural da linguagem (regras de formação de palavras), existe dependência entre os fonemas consecutivos. Esta característica de dependência pode ser modelada satisfatoriamente por uma Cadeia de Markov. No entanto, a seqüência de fonemas não é a estrutura observável. A única informação conhecida é a estrutura de *frames*. Isso indica a utilização de um *HMM* para estabelecer a modelagem. O *HMM* possibilita caracterizar a dependência entre os fonemas e associá-los aos *frames*, que constituem a informação observável. Vários modelos podem ser elaborados (*HMM's*) associados a grupos de palavras similares por

formação. Observe que os fonemas constituem os estados latentes da *Cadeia de Markov*. É possível obter a seqüência de estados (fonemas) mais provável, associada a um modelo (*HMM*) e uma seqüência de *frames*. Isso é realizado através da aplicação do algoritmo de *Viterbi*. A obtenção das estimativas dos parâmetros associados ao *HMM* (*Treinamento*) pode ser realizada aplicando-se o algoritmo *EM*.

A seguir, apresentaremos outro problema muito importante, associado à análise de *DNA*, denominado *Alinhamento Múltiplo*.

## 3.2 Alinhamento Múltiplo

O problema de alinhamento múltiplo é extremamente importante na análise de *DNA*, no que diz respeito à investigação da similaridade evolucionária entre famílias de proteínas e a observação de comportamentos evolucionários. O Alinhamento Múltiplo é um método através do qual é possível estabelecer um grau de similaridade entre um conjunto de seqüências de caracteres. A idéia básica é dispor um conjunto de seqüências alinhadas. A figura 15 apresenta um alinhamento entre 4 seqüências.


FIGURA 15. Ilustração de Alinhamento Múltiplo entre as seqüências de caracteres *S1*, *S2*, *S3*, *S4*.

De maneira formal, o Alinhamento Múltiplo pode ser definido como uma função que mapeia, respectivamente, várias seqüências de caracteres sobre o mesmo alfabeto (geralmente conjunto de bases nitrogenadas $\{A,C,G,T\}$ ou conjunto de aminoácidos), em seqüências de caracteres, todas com mesmo tamanho, incluindo o caractere especial "-" sobre o alfabeto original das seqüências. Este caractere especial é denotado como "*lacuna*" (*gap*). Maiores detalhes sobre o caractere "-" serão dados adiante.

Naturalmente, existem vários Alinhamentos Múltiplos possíveis entre um conjunto de seqüências de caracteres. No entanto, existem alguns alinhamentos particulares que maximizam o número de caracteres idênticos nas colunas. Tais alinhamentos, em geral, são denominados como "*Alinhamentos Ótimos*". Estes alinhamentos podem discriminar regiões das seqüências com alta similaridade (para mais detalhes vide o trabalho de Meidanis et al (1992)) definindo um padrão denominado *domínio*.

Para compreender melhor a aplicação do *HMM* ao problema de alinhamento múltiplo, devemos apresentar o modelo "*Profile HMM*" (Krogh et al., 1994). Para termos noção da importância deste tipo de ferramenta, Eddy (1996) classifica-a como a mais popular aplicação de *HMM*'s na biologia molecular até a data de publicação do artigo. Antes de apresentar o "*Profile HMM*" vamos discutir um modelo mais simples, denominada "*Pair HMM*", da qual o "*Profile HMM*" é uma generalização.

O "*Pair HMM*" é uma modelagem estabelecida para o problema de alinhamento simples (*pairwise aligniment*), que basicamente é o Alinhamento Múltiplo entre 2 seqüências. Este modelo é um *HMM* no qual cada estado não emite uma observação a cada momento, e sim um pareamento de caracteres. Um "*Pair HMM*" possui três estados característicos denominados:

- Estado Par (*Matching State*): designa uma distribuição de probabilidade dos possíveis pareamentos de um símbolo "*x*" com "*y*", na qual *x* e *y* pertencem ao alfabeto original das seqüências.
- Estado Inserção (*Insert State*): designa uma distribuição de probabilidade dos possíveis pareamentos de um símbolo "*x*" com uma lacuna (gap) "-", na qual *x* pertence ao alfabeto original das seqüências.
- Estado Exclusão (*Delete State*): designa uma distribuição de probabilidade dos possíveis pareamentos de uma lacuna "-" com um símbolo "*x*", na qual *x* pertence ao alfabeto original das seqüências.

Embora este modelo seja baseado nos *HMM's*, algumas extensões devem ser feitas. Devemos definir os estados inicial e final para complementar o modelo. Estes estados possuem a característica de não emitirem nenhum tipo de observação, o que é uma extrapolação do modelo probabilístico *HMM*. Tais estados são denominados, na literatura, como estados silenciosos ou nulos (null or silent states) (Durbin, 1989). Com isso o modelo "*Pair HMM*" possui, de fato, três estados.

O "*Profile HMM*" é uma generalização do "*Pair HMM*". O "*Profile HMM*" é um *HMM* com três grupos (par, inserção e exclusão) de estados similares. O primeiro grupo considera os alinhamentos sem *"lacunas"* (*gaps*) e é formado pela ligação de estados do tipo par de caracteres (*matching*). O modelo é ilustrado na figura 16.


FIGURA 16. "*Profile HMM*"; Estrutura Inicial com $r$+2 estados, sendo $r$ do tipo par (*matching*).

Observe que cada estado $M_j$ do tipo "par" possui a distribuição de probabilidade de emissão (variável observável) de um determinado caractere do alfabeto das seqüências (alfabeto do *DNA*, alfabeto de aminoácidos, alfabeto de proteínas) como consenso na *j-ésima* coluna do alinhamento múltiplo.

Existe a necessidade de contemplar as situações de inserções e exclusões de caracteres na seqüência. Isso é realizado através dos estados de inserção e exclusão. Iremos trabalhar com estes estados individualmente.

Ao modelo inicial com estados "par", devemos acrescentar os estados "inserção". A cada estado "inserção", $I_j$, está associado a distribuição de probabilidade da inserção de um caractere após o pareamento na *j-ésima* coluna do alinhamento múltiplo (após a visita ao estado $M_j$). Segue na figura 17 a estrutura.

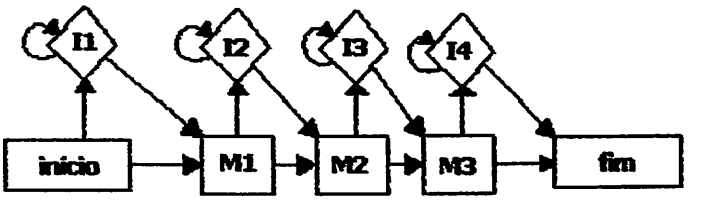



FIGURA 17. Estados Inserção (Losangos) no "*Profile HMM*" com 9 estados, sendo 3 do tipo par (*matching*) e 4 do tipo inserção.

Note que, uma vez no estado inserção $I_j$, é possível retornar ao mesmo, pois a probabilidade de transição para o mesmo estado é não nula. Na figura 17 onde há transições, existe uma probabilidade não nula associada a esta transição.

Resta adicionar os estados do tipo "exclusão". Estes estados são do tipo silenciosos ou nulos (não emitem nenhuma observação). Com a estrutura de ligações com que estes estados são inseridos no "*Profile HMM*", é possível estabelecer uma série de exclusões de caracteres após um estado $M_j$, e deslocar-se para qualquer estado com $M_k$, onde $k > j$. A estrutura é ilustrada na figura 18.

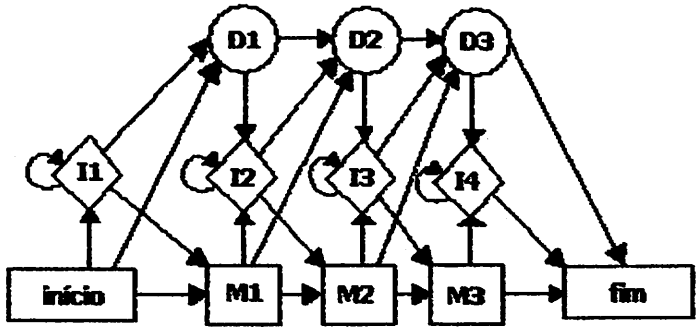



FIGURA 18. Estados de Exclusão (Círculos) no "*Profile HMM*" com 12 estados, sendo 3 do tipo par (*matching*), 4 do tipo inserção e 3 do tipo exclusão.

Os mesmos tipos de problemas encontrados em *HMM's* básicos são encontrados no "*Profile HMM*". Escolha do modelo e estimação de parâmetros são dois problemas iniciais. Basicamente a escolha do tamanho do modelo está relacionada à escolha do número de estados do tipo "par" e do tipo "inserção". Dado um certo alinhamento múltiplo, é possível obter a parametrização do "*Profile HMM*" representativa deste alinhamento (Durbin et. al, 1998).

Utilizando o método *forward*, visto no capítulo 2, é possível obter a probabilidade de uma seqüência estar associada a um alinhamento múltiplo específico. Isso é interessante, pois permite estabelecer um grau de proximidade entre a seqüência em questão e o alinhamento múltiplo (modelo).

Ainda é possível estabelecer (da Silva, 2002):

- **Estimação do "*Profile HMM*"**: tomando-se um conjunto de seqüências não alinhadas pertencentes a uma mesma família específica de proteínas ou aminoácidos ou bases (Conjunto de Treinamento), é possível estimar os parâmetros do "*Profile HMM*" usando o algoritmo *EM* (Capítulo 2).
- **Obtendo o alinhamento ótimo**: Com base no modelo estimado acima, e em uma seqüência do conjunto de treinamento, é possível obter a seqüência de estados que maximiza a probabilidade da ocorrência de tal seqüência. Isso é realizado utilizando-se o algoritmo de Viterbi (Capítulo 2).
- **Produzindo o alinhamento múltiplo**: Com base no modelo estimado e os alinhamentos ótimos associados a cada uma das seqüências do conjunto de treinamento, o alinhamento múltiplo é determinado através dos estados que são comuns em todos os alinhamentos ótimos.
- **Formando as colunas do alinhamento**: Todos os caracteres (bases, aminoácidos e proteínas) que estiverem nos alinhamentos ótimos associados a um estado "par" específico devem ser colocados na mesma coluna.

Existem variações da estrutura "*Profile HMM*" apresentada aqui. De acordo com aplicação específica, podem ser adicionados outros tipos de estados no modelo.

A seguir, apresenta-se a aplicação de segmentação de *DNA* e a utilização do modelo *HMM* original, como descrito no capítulo 2, para a localização de regiões homogêneas em seqüências de *DNA*.

## 3.3 Segmentação de *DNA*

Como vimos anteriormente no capítulo 1, as características associadas a um determinado organismo estão codificadas em seqüências de *DNA* que, por sua vez, compõem os genes. Um determinado gene está associado à síntese de uma determinada proteína. Em resumo, cada segmento de *DNA* é responsável pela codificação de uma determinada proteína, que é responsável por um tipo de funcionalidade presente no organismo. Vide o exemplo dos genes *XF1143* e *XF1144* que estão presentes na bactéria *Xylella fastidiosa*. Os genes *XF1143* e *XF1144* têm, respectivamente, 478 e 277 bases. Em conjunto com mais alguns genes, o *XF1143* e o *XF1144* são responsáveis pela síntese de *ATP*, função associada á produção de energia. Existem outros genes que codificam outras proteínas com funcionalidades distintas das dos genes *XF1143* e *XF1144*. Por exemplo, os genes *XF1151* e *XF1152* são, em conjunto com outros genes, responsáveis pela codificação de proteínas associadas aos ribossomos.

No processo de sequenciamento de *DNA* de um genoma, não é possível determinar a localização exata de um gene ou mesmo se um segmento composto de vários genes é responsável pela codificação e, por sua vez, síntese de proteínas que atuam especificamente em uma funcionalidade no organismo. É necessário um método que possa discriminar os segmentos que codificam determinadas proteínas com distintas funcionalidades. Vale ressaltar que, entre

estas regiões codificadoras, denominadas *éxons*, existem as seqüências não codificadoras chamadas *íntros*. No processo de síntese de proteínas, as regiões não codificadoras são descartadas, pois não possuem função aparente. Este processo é ilustrado na figura 19.

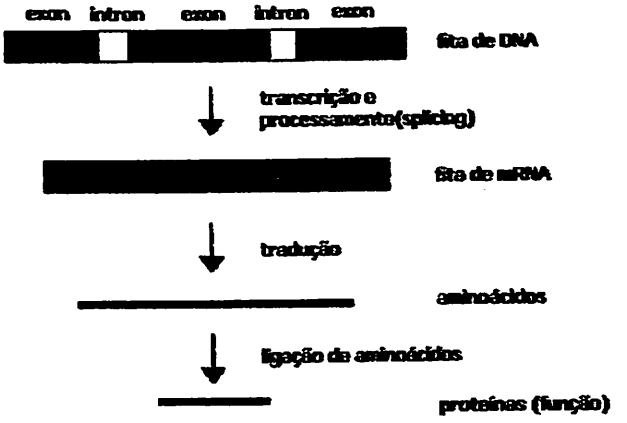



FIGURA 19. Etapas do processo de síntese de proteínas. Discriminação de regiões codificadoras e não codificadoras.

Alguns estudos sobre os segmentos distintos de *DNA* indicaram que as freqüências de bases (nucleotídeos) poderiam auxiliar na identificação e distinção entre os segmentos com diferentes funcionalidades. Existem segmentos nos quais a proporção da ocorrência de uma determinada base, ou de um determinado grupo de bases é similar. Estes segmentos são denotados como *regiões homogêneas* (*isochores*). Variações nas proporções das freqüências de *C+G* comumente refletem distinções funcionais ou estruturais entre os segmentos. A localização de tais regiões homogêneas em conteúdo de *C+G* é denotada por segmentação de *DNA*.

Na segmentação de *DNA*, uma estratégia inicial a ser adotada é dividir a seqüência de *DNA* original em $k$ segmentos de tamanhos iguais. Este processo é ilustrado na figura 20.

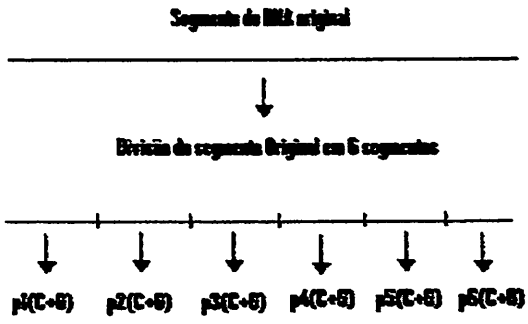


FIGURA 20. Método de caracterização de heterogeneidade baseada em segmentação onde $k = 6$.

Assim, devemos testar a hipótese de igualdade entre as proporções $p1$, $p2$, $p3, p4, p5$, $p6$ de conteúdo de *C+G* em cada segmento. Observe, no entanto, que este procedimento apresenta a limitação do conhecimento do número de fragmentos (segmentos menores). Inicialmente, é arbitrária a escolha do número $k$ de fragmentos em que será dividido o segmento original. Este método possibilita detectar heterogeneidades, caso elas existam, porém não é capaz de localizar as regiões de homogeneidade.

Outro método utilizado na caracterização da heterogeneidade na freqüência de *C+G* consiste em percorrer o segmento de *DNA*, considerando segmentos de tamanho $l$ fixo (janela) menor que o tamanho do segmento original (Staden, 1984). Devemos, inicialmente, calcular a proporção de

ocorrência de $C+G$ no segmento que se inicia na primeira base (nucleotídeo) até a $l$-ésima base. Depois devemos obter a proporção da ocorrência de $C+G$ no segmento, iniciando-se da segunda base até a $l+1$-ésima base. O processo deve ser repetido até que todo o segmento do *DNA* original seja percorrido, armazenando-se as proporções obtidas. A figura 21 ilustra este processo em uma pequena seqüência de *DNA*.

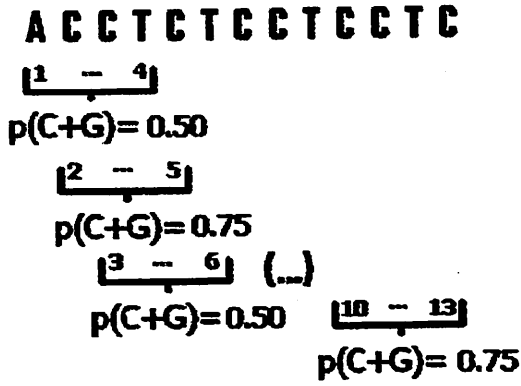



FIGURA 21. Método de caracterização de heterogeneidade baseado em verificação de proporção em janelas onde $l = 4$.

As proporções obtidas em cada janela (segmento de tamanho $l$) nos permitem verificar a ocorrência de algum comportamento sistemático na variação destas proporções ao longo da seqüência. A figura 22 ilustra um gráfico obtido a partir do procedimento de verificação de proporções através de janelas. A seqüência de observações é binária (0 ou 1).

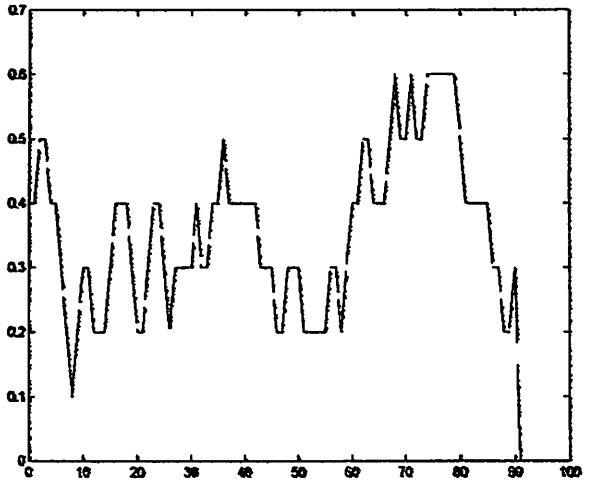



FIGURA 22. Gráfico de proporções de janelas onde $l = 10$ (comprimento da janela) e o tamanho da seqüência de observações é 100.

Embora este método seja eficiente para uma análise preliminar, ele apresenta algumas limitações. O tamanho da janela $l$ é escolhido de forma arbitrária. Além disso, este método não aponta, claramente, os pontos onde há mudança de regime na expressão de *C+G*.

Uma alternativa é tratar a busca de possíveis regiões homogêneas no *DNA* através de uma metodologia que localiza pontos de mudança de variável aleatória.

### 3.3.1 Ponto de mudança em Variáveis Aleatórias

O problema de ponto de mudança em uma seqüência de variáveis aleatórias pode ser definido da seguinte maneira (Smith, 1975). Uma dada seqüência de variáveis aleatórias $X_1, X_2, \ldots, X_n$ possui um *ponto de mudança*

em $r$ se $X_i \sim F_1(x|\theta_1)$ para $1 \leq i \leq r$ e $X_j \sim F_2(x|\theta_2)$ para $r+1 \leq j \leq n$, sendo $F_1(x|\theta_1) \neq F_2(x|\theta_2)$. As formas de $F_1$ e $F_2$ são conhecidas, porém o ponto de mudança $r$ é desconhecido. Os parâmetros $\theta_1$ e $\theta_2$ podem ser conhecidos a priori ou desconhecidos. Dada uma seqüência de realizações $x_1, x_2, ..., x_n$, o objetivo é estimar o ponto $r$, e $\theta_1$ e $\theta_2$ quando necessário. Existem alguns métodos para estimação de $r$. Como exemplo, temos métodos de máxima verossimilhança (Hinkley, 1970) e métodos bayesianos (Smith, 1975).

Observe que o problema pode ser generalizado para o caso no qual existem múltiplos pontos de mudança, $r_1$, $r_2$, ... $r_k$. Assim, torna-se possível estabelecer a modelagem do problema de segmentação de *DNA* utilizando os métodos de inferência sobre pontos de mudança em variáveis aleatórias. Devemos então considerar cada seqüência de *DNA* $(Base_1, Base_2, ...Base_n)$ como a realização de $n$ experimentos de *Bernoulli* $(X_1, X_2, ..., X_n)$ em que um "sucesso" está relacionado à ocorrência de $C$ ou $G$ na seqüência. A pressuposição básica é a de que $X_1, X_2, ..., X_n$ são variáveis aleatórias independentes..

Esta abordagem não é adequada quando tratamos de seqüências de *DNA*, pois sabidamente há dependências entre as bases (nucleotídeos).

Com a existência das dependências entre as bases ao longo da seqüência de *DNA*, a modelagem via *HMM's* torna-se útil, pois, como visto no capítulo 2, este modelo pressupõe a dependência entre as variáveis aleatórias explicitadas no problema.

### 3.3.2 Aplicando *HMM's* ao problema de Segmentação de *DNA*

O *HMM* pode ser utilizado na localização de regiões homogêneas em termos de emissão de *C+G*, de modo a se encontrar regiões que são funcional ou estruturalmente distintas. O trabalho de Churchill (1989) foi pioneiro neste tipo de aplicação e em problemas ligados à biologia computacional (Eddy, 1998).

Este tipo de modelagem estabelece os estados da cadeia latente como as distintas regiões funcionais presentes na seqüência. Cada região funcional tem uma proporção de *C+G* particular. Com o *HMM* é possível encontrar a probabilidade de cada base estar associada a um estado específico (região funcional) (equação 79) e observar o comportamento da proporção de *C+G* ao longo da seqüência (composição local).

Para realização da análise, adotanto-se o *HMM*, inicialmente, a seqüência de *DNA* é transformada em uma série binária, na qual 0 corresponde à ocorrência de *A* ou *T* e 1 corresponde à ocorrência de *C* ou *G*. O número de *estados* (regiões de homogeneidade na expressão de *C+G*) é definido utilizando-se critérios para adequação de modelos visto no final do capítulo 2. Cada região com um distinto padrão de emissão de *C +G*, é considerada como um estado distinto.

Considerando uma dada seqüência de *DNA*, o processo para a investigação das regiões homogêneas utilizando *HMM's*, deve ser realizado da seguinte forma. Devemos estimar cada um dos modelos competidores (*HMM's* com 1, 2, ... *r* estados), utilizando-se o método da máxima verossimilhança (algoritmo *EM*). Pressupõe-se que o modelo associado à seqüência de *DNA* não seja volátil, isto é, a esperança do tempo de permanência no estado é um valor alto. Assim, para aplicação do algoritmo *EM*, é natural escolhermos um modelo inicial na qual a probabilidade de permanência no estado seja alta (Matriz *A* com valores nas diagonais próximos de 1). Após obter estimativas dos

parâmetros associados aos modelos competidores, devemos escolher, segundo um critério estabelecido, (*AIC*, *BIC*, $\Delta BIC$) o modelo que é mais adequado (Vide capítulo 2). Para visualizar as regiões com homogeneidade na emissão *C+G*, basta traçarmos o gráfico da *composição local* e identificarmos os trechos onde os valores são homogêneos (apresentam pouca variabilidade).

No capítulo 4, apresenta-se a descrição dos resultados obtidos na análise de dados reais, com o auxílio de um software desenvolvido especialmente para a implementação dos métodos associados aos *HMM*'s, ligados ao problema de segmentação de *DNA*.

# 4. APLICAÇÕES A DADOS REAIS

Neste capítulo, ilustramos o problema da segmentação do *DNA* com base em dados reais de vários organismos. As seqüências foram analisadas com o auxílio do software *SIMHMM*, que foi desenvolvido durante este trabalho, com o objetivo de implementar as técnicas descritas no capítulo 2. Uma descrição sucinta da utilização deste software, que será disponibilizado gratuitamente, pode ser encontrada nos anexos.

Para realização das análises, foram utilizadas frações (segmentos) dos genomas de 5 organismos específicos. Tais seqüências foram obtidas a partir do *Genbank* (Genbank).

De forma sumária, o resultado básico que será apresentado é o gráfico da composição local de $C+G$ na seqüência de *DNA*. Os valores da composição local, para cada ponto da seqüência, podem ser descritos em um gráfico que permite a discriminação visual de regiões homogêneas em conteúdo de $C+G$. O modelo *HMM* que se mostrar mais adequado, segundo os critérios apresentados no capítulo 2 (*BIC* e *AIC*), será aquele a partir do qual os gráficos da composição local serão apresentados. Quando dois modelos com números de estados distintos forem considerados como mais adequados pelos critérios, adotamos o modelo com maior número de estados. Os valores de $\Delta BIC$ serão apresentados com objetivo de ilustração, porém não serão utilizados na análise. Todos os modelos utilizados nestas aplicações são os *HMM's*, que consideram as observações, dado os estados, condicionalmente independentes.

## 4.1 Bacteriófago *lambda*

No trabalho inicial de Churchill (1989), os dados referentes ao genoma do bacteriófago *Lambda* foram analisados. Com o objetivo de comparar os resultados obtidos, uma das seqüências escolhidas, neste trabalho foi a do Bacteriófago *Lambda*. A seqüência deste vírus foi obtida mediante consulta ao *Genbank* e está catalogada sob o código NC001416. Um bacteriófago é um tipo de vírus que ataca bactérias. O genoma associado a este bacteriófago é composto de 48514 pares de base. A obtenção da seqüência de *DNA* deste vírus (sequenciamento) foi realizada por Sanger (1982). Vale ressaltar que a seqüência de *DNA* deste bacteriófago é circular, ou seja, sua seqüência não possui início ou fim.

Os resultados obtidos por Churchill (1989) apontam um modelo com 3 estados com dependência de 1ª ordem entre as observações como o mais adequado (critério *BIC*). Analisando o gráfico da *composição local* de Churchil (1989), associado a um modelo com 4 estados, a seqüência do bacteriófago *lambda* possui dois trechos de características particulares em conteúdo de *C+G*. O trecho inicial (primeiras 23000 bases) apresenta comportamento homogêneo em relação à freqüência de *C+G*, sendo a mesma em torno de 0.55. Isso sugere que este trecho codifica proteínas com estrutura ou função similares (permanência do processo em um estado). O segundo trecho da seqüência (a partir da base 23000) sugere uma característica altamente heterogênea em relação à freqüência de *C+G*. Esta heterogeneidade, segundo o autor, sugere vários segmentos funcionais ou estruturais de características distintas.

Para realizarmos a análise da seqüência, esta foi modificada tal que toda a ocorrência de *C* ou *G* foi substituída por 1 e toda a ocorrência de A ou T foi substituída por 0. Primeiramente, foram obtidos os modelos ajustados

(estimados). O método utilizado para obtenção a de inferências sobre o modelo foi o da máxima verossimilhança, com a aplicação do algoritmo *EM*.

Para realizar a análise comparativa, foram considerados, inicialmente, 4 modelos competidores. Todos estes modelos possuem espaço de observações restrito a 0 e 1 ou $\Omega_O = \{0,1\}$. Estas observações correspondem à presença de *C* ou *G* (1) e a presença de *A* ou *T* (0). Os modelos se distinguem em número de estados que são, respectivamente, 2, 3, 4 ou 5, isto é, $\Omega_S = \{1,2\}$, $\Omega_S = \{1,2,3\}$, $\Omega_S = \{1,2,3,4\}$ e $\Omega_S = \{1,2,3,4,5\}$.

Como o algoritmo *EM* pode conduzir a máximos locais, foram escolhidos alguns pontos iniciais distintos para verificar a convergência. Foi estabelecido como critério de convergência que o módulo da diferença entre a log-verossimilhança do modelo ajustado (obtido após uma iteração do algoritmo *EM*) e a log-verossimilhança do modelo atual fosse menor que $10^{-5}$.

Obtidos os estimadores de máxima verossimilhança dos modelos (*HMM's*) com 2,3,4 e 5 estados, para escolher o modelo mais adequado, foram utilizados os critérios *BIC* e *AIC*, descritos no capítulo 2. Na tabela 1, apresentam-se os valores dos graus de liberdade (Número de parâmetros desconhecidos associados ao modelo), *BIC*, $\Delta BIC$ e *AIC* (vide sessão 2.8, expressão (97)).

TABELA 1. Valores de Graus de Liberdade, *BIC*, $\Delta BIC$ e *AIC* associados à seqüência do bacteriófago *Lambda*.

| *N° de estados* | *Grau de liberdade* | *BIC* | $\Delta BIC$ | *AIC* |
|---|---|---|---|---|
| 2 | 5 | 66456.144 | -799 | 66412.197 |
| 3 | 12 | 66353.272 | -901 | 66256.589 |
| 4 | 21 | 66409.225 | -845 | 66242.227 |
| 5 | 32 | 66502.854 | -752 | 66247.962 |

Pelos critérios *BIC* e *AIC*, devemos escolher os modelos que minimizam os valores dos mesmos. Os modelos escolhidos pelos dois critérios foram distintos. O *BIC* considerou o modelo com 3 estados como mais adequado, enquanto o *AIC* considerou o modelo com 4 estados. Embora sejam modelos distintos, os gráficos da *composição local* de *C+G* nos dois modelos sugerem (figuras 23 e 24) um perfil similar ao encontrado por Churchill (1989).

Os modelos ajustados (estimados pelo método *EM*) com 3 e 4 estados são apresentados abaixo.

$$\hat{\pi} = [1 \quad 0 \quad 0],$$

$$\hat{A} = \begin{bmatrix} 9.9904\times10^{-1} & 9.7106\times10^{-5} & 8.6283\times10^{-4} \\ 7.7961\times10^{-5} & 9.9992\times10^{-1} & 1.9397\times10^{-92} \\ 1.6179\times10^{-3} & 3.9654\times10^{-24} & 9.9838\times10^{-1} \end{bmatrix},$$

$$\hat{B} = \begin{bmatrix} 5.1867\times10^{-1} & 4.8132\times10^{-1} \\ 4.3013\times10^{-1} & 5.6987\times10^{-1} \\ 6.3735\times10^{-1} & 3.6265\times10^{-1} \end{bmatrix},$$

$$\hat{\pi} = [0 \quad 1 \quad 0 \quad 0],$$

$$\hat{A} = \begin{bmatrix} 9.9781\times10^{1} & 9.2565\times10^{15} & 2.1927\times10^{3} & 8.7577\times10^{142} \\ 1.1269\times10^{4} & 9.9951\times10^{1} & 2.9167\times10^{4} & 8.4536\times10^{5} \\ 8.6509\times10^{4} & 4.1408\times10^{4} & 9.9872\times10^{1} & 9.8067\times10^{46} \\ 1.7008\times10^{303} & 5.2317\times10^{5} & 1.1010\times10^{167} & 9.9995\times10^{1} \end{bmatrix},$$

$$\hat{B} = \begin{bmatrix} 6.6781\times10^{-1} & 3.3219\times10^{-1} \\ 5.0384\times10^{-1} & 4.9616\times10^{-1} \\ 5.6905\times10^{-1} & 4.3095\times10^{-1} \\ 4.2868\times10^{-1} & 5.7132\times10^{-1} \end{bmatrix}.$$

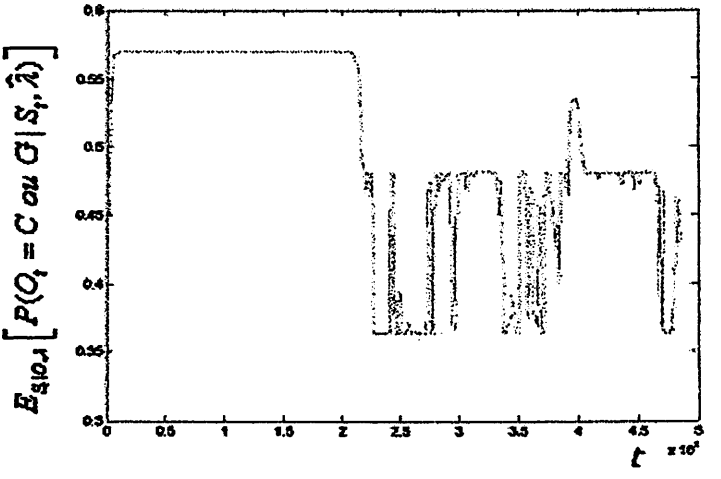



FIGURA 23. Gráfico da *composição local* de *C*+*G* em cada posição da seqüência do bacteriófago *lambda*, sob o modelo com 3 estados.

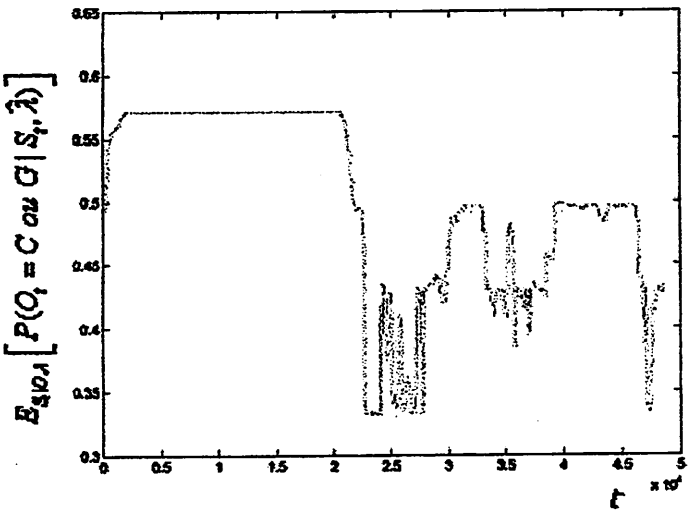



FIGURA 24. Gráfico da *composição local* de *C+G* em cada posição da seqüência do bacteriófago *lambda*, sob o modelo com 4 estados.

Em ambos os modelos, os gráficos da *composição local* se apresentam homogêneos até, aproximadamente, a base 23000, ou seja, os valores da *composição local* não apresentam variação significativa no trecho da seqüência que vai da base 1 até a 23000. Isso sugere a presença de uma região funcionalmente ou estruturalmente similar neste segmento. O segundo trecho partindo, aproximadamente, da base 23000 até o fim da seqüência apresenta uma variação maior dos valores da composição local de *C+G*, que sugere uma heterogeneidade em termos de funcionalidades associadas a este segmento. Vale ressaltar que no caso do *HMM* com 3 estados, embora distinto estruturalmente e em número de estados, o gráfico das esperanças aqui apresentado sugere o mesmo perfil encontrado por Churchill (1989).

O gráfico das esperanças condicionais de freqüência de *C+G*, obtido adotando-se o modelo com 4 estados (modelo mais adequado, segundo o critério

*AIC*), é similar ao gráfico apresentado por Churchill (1989). Independente do critério, os modelos apresentaram-se coerentes com o trabalho presente na literatura.

## 4.2 *Xylella fastidiosa*

A *Xylella fastidiosa* é uma bactéria patógena que ataca culturas cítricas no estado de São Paulo. Como forte região exportadora, a mesma sofre impactos econômicos (queda de produtividade) em virtude desta doença. Pela importância das culturas cítricas, a *FAPESP* (Fundação de Amparo a Pesquisa do Estado de São Paulo) iniciou um projeto de sequenciamento da *Xylella Fastidiosa*. Este projeto foi concluído em 2000, obtendo-se o código completo desta bactéria e os mapas funcionais associados ao seu genoma (Simpson et al, 2000).

No trabalho de da Silva (2003), foi feita uma análise similar à realizada por Churchill (1989), tendo como alvo um fragmento da seqüência da *Xylella fastidiosa*. da Silva (2003) apresenta a análise sobre os éxons da seqüência de *DNA* associado a um segmento correspondendo do gene *XF1141* até o gene *XF1196*. Como visto no capítulo 3, existem trechos na seqüência de *DNA* que não são expressos (não codificam nenhum tipo de proteína), não possuindo nenhuma funcionalidade aparente, sendo estes denominados *íntrons*. No trabalho de da Silva (2003), considera-se a seqüência composta pela concatenação de *éxons* (regiões com expressão) a partir do gene *XF1141* até o gene *XF1196*.

Nesta dissertação, procuramos analisar a mesma seqüência, considerando íntrons e éxons presentes dentre os genes *XF1141* até o gene *XF1196*. No total, este segmento possui 50716 bases. Para fim de verificação, os resultados obtidos foram comparados com os mapas de funcionalidades disponíveis no *Genbank*. A seqüência foi obtida a partir do *Genbank* sob código AE003849.

Inicialmente, foi realizada a modificação usual na seqüência (substituição de *C* e *G* por 1 e *A* e *T* por 0) como na análise do bacteriófago *Lambda*. Foram obtidos os modelos ajustados pelo método da máxima verossimilhança, com 2, 3, 4 e 5 estados. Na tabela 2, apresentam-se os valores das funções *AIC*, *BIC*, $\Delta BIC$ e os graus de liberdade associados a cada modelo.

TABELA 2. Valores de Graus de Liberdade, *BIC* e *AIC* associados à seqüência da bactéria *Xylella fastidiosa*.

| *Nº de estados* | *Grau de liberdade* | *BIC* | $\Delta BIC$ | *AIC* |
|---|---|---|---|---|
| 2 | 5 | 68640.542 | -1667 | 68596.442 |
| 3 | 12 | 68614.233 | -1693 | 68517.215 |
| 4 | 21 | 68694.287 | -1613 | 68526.711 |
| 5 | 32 | 68788.308 | -1519 | 68532.534 |

O critérios *AIC* e *BIC* consideram o modelo com 3 estados como mais adequado aos dados. Sendo assim, foi obtido o gráfico da *composição local* de *C+G*, associada ao modelo com 3 estados. A partir deste, é possível observar segmentos homogêneos na presença de *C+G* homogêneas. O gráfico é apresentado na figura 25.

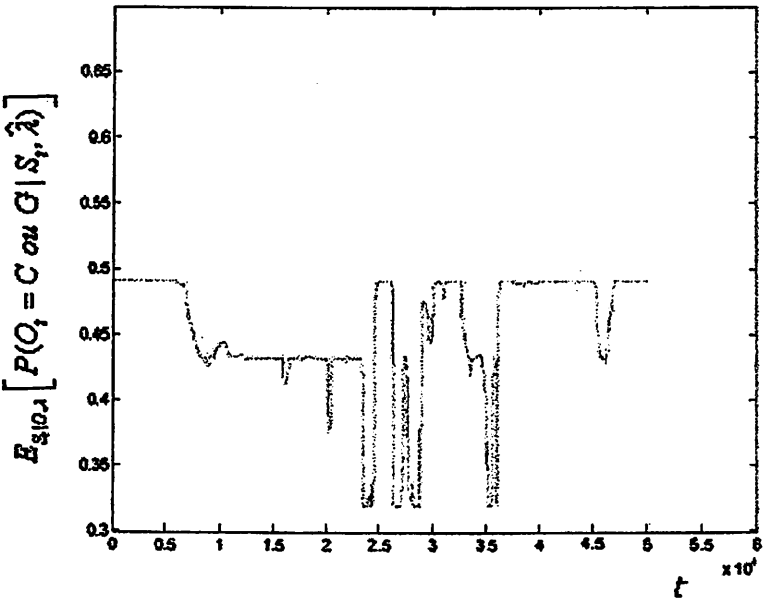



FIGURA 25. Gráfico da *composição local* de *C+G* em cada posição da seqüência da *Xylella fastidiosa*, sob o modelo com 3 estados.

Analisando-se o comportamento do gráfico da composição local associada ao modelo com 3 estados, é possível discriminar 4 trechos característicos quanto ao comportamento de homogeneidade ou heterogeneidade de *C+G*. Um trecho inicial que vai da primeira base até, aproximadamente, a base 7000, apresenta uma característica bem homogênea (pouca variação) em relação ao conteúdo de *C+G*, sendo a mesma em torno de 0.48. É possível observar ainda um segundo trecho com característica homogênea em termos do conteúdo de *C+G*. Este trecho se inicia, aproximadamente, a partir da base 9000 até a base 22000, e os valores da *composição local* variam em torno de 0.42. Embora exista a presença de homogeneidade neste trecho, ela não é tão evidente quanto no primeiro, devido à presença de variações muito bruscas, no entanto insignificantes em relação ao tamanho da seqüência. O terceiro trecho apresenta uma característica

heterogênea (variação alta nos valores da *composição local* de C+G) em relação ao conteúdo de *C+G*. Este trecho vai, aproximadamente, da base na posição 24000 até a base na posição 36000. No quarto e último trecho é possível verificar uma característica homogênea, embora exista uma variação considerável por volta da base 46000. Os valores da *composição local* neste trecho são próximos de 0.48.

Todas estas observações sugerem que o segmento analisado é expresso como 4 trechos com características distintas. O primeiro dos trechos está localizado entre as bases 1 e 7000 (permanência no estado 3) e o segundo está localizado da base 7000 até 23000 (permanência no estado 2). Um terceiro trecho possui uma característica altamente heterogênea em termos de expressão, agregando pequenas regiões de expressão (relativo ao tamanho da seqüência). Este terceiro trecho está localizado, aproximadamente, a partir da base 23000 até a base 36000. A análise sugere a presença de outra região funcional a partir da base 36000 (permanência no estado 3).

Se observarmos a estrutura funcional deste segmento (genes *XF1141* até *XF1196*), segundo os mapas funcionais e classificação de proteínas similares funcional ou estruturalmente (figura 26), podemos notar que o *HMM* proporcionou boa discriminação das distintas regiões funcionais. Na figura 26, segue uma ilustração gráfica, adaptada do *Genbank*, da localização das regiões funcionais no segmento analisado.

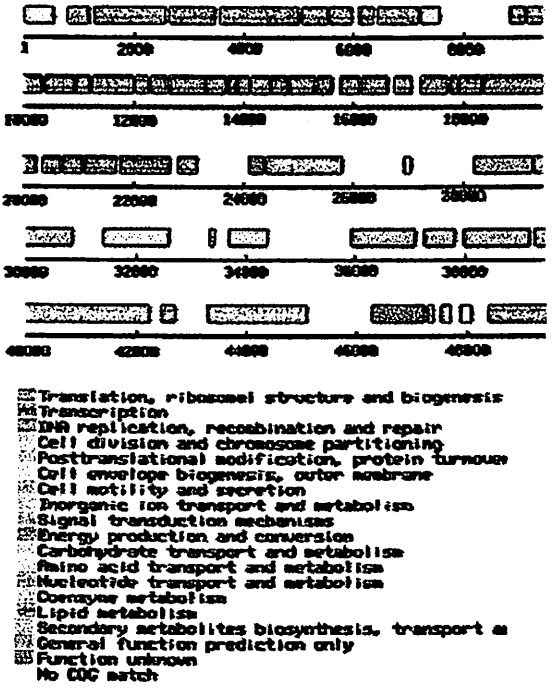



FIGURA 26. Regiões de expressão do gene *XF1141* até o gene *XF1196* da *Xylella fastidiosa* e tabela de classificação funcional.

Na figura 26, os segmentos com reticulados coloridos apresentam *éxons*. As regiões que não possuem reticulados apresentam *introns*. Os trechos da seqüência que contêm reticulados com cores distintas apresentam expressão de proteínas de características funcionais ou estruturais distintas. Analisando o mapa, podemos observar uma região inicial partindo das primeiras bases até aproximadamente a base 7000, que é responsável pela codificação de proteínas associadas à produção de energia. Podemos observar ainda um segundo segmento, partindo da base 9000, aproximadamente, até a base 23000, que é responsável pela codificação de proteínas associadas aos ribossomos. Um terceiro trecho, partindo aproximadamente de 24000 até 36000, apresenta vários

segmentos não codificadores, além de alguns pequenos segmentos codificadores. O trecho final, que parte da base 36000 até a aproximadamente 45000, apresenta uma região responsável por funções (similares) associadas ao metabolismo.

Os resultados obtidos foram coerentes com os resultados encontrados pelo trabalho de da Silva (2003). O trabalho de da Silva (2003) discrimina 4 trechos do segmento analisado. Os dois primeiros trechos (1 a 7000 e 7000 a 21000) são largamente homogêneos. Por possuir poucas regiões não codificadoras, os dois primeiros trechos são caracterizados de maneira similar ao presente trabalho. A partir do 3° trecho (22000 a 40000), a seqüência possui muitos trechos não codificadores. Este fato, no entanto, não interferiu na análise realizada neste trabalho.

Portanto a caracterização das regiões funcionais estabelecidas pela utilização do modelo *HMM* nesta seqüência apresentou resultados bastante satisfatórios.

### 4.3 *Xanthomonas axonopodis* pv. citri

O gênero *Xanthomonas* é um importante grupo de bactérias fitopatogênicas. Sua importância se deve ao fato de causar doenças em culturas de relevância econômica em todo o mundo. O *Xanthomonas axonopodis* pv.citri causa o "*Cancro Cítrico*" (*Citrus Canker*), doença que ataca culturas de frutas cítricas, causando perda significativa de produção e, consequentemente, perdas de ordem financeira (Silva et al, 2002).

Um segmento do genoma desta bactéria foi obtido a partir de consulta no *Genbank* sob catalogação NC003919. Esta seqüência inclui bases do gene *XAC0965* até o gene *XAC1014* e é composta de 49611 pares de bases.

Os modelos competidores são os mesmos utilizados nas seqüências anteriores, possuindo 2, 3, 4 e 5 estados. O espaço de observações se limita ao

conjunto 0 e 1 associados, respectivamente, à presença de *A* ou *T* e à presença de *C* ou *G*.

Os resultados dos critérios de adequação *BIC*, $\Delta BIC$ e *AIC*, associados aos modelos ajustados, são apresentados na tabela 3.

TABELA 3. Valores de Graus de Liberdade, *BIC* e *AIC* associados á seqüência da bactéria *Xanthomonas axonopodis* pv. citri.

| *Nº de estados* | *Grau de liberdade* | *BIC* | $\Delta BIC$ | *AIC* |
|---|---|---|---|---|
| 2 | 5 | 65571.414 | -3204 | 65527.354 |
| 3 | 12 | 65701.336 | -3074 | 65523.592 |
| 4 | 21 | 65620.524 | -3155 | 65533.908 |
| 5 | 32 | 65807.290 | -2968 | 65551.742 |

O modelo com 2 estados foi escolhido como o mais adequado pelo critério *BIC*. O critério *AIC* indica o modelo com 3 estados como o mais adequado.

Embora os *HMM's* sejam distintos (2 e 3 estados), os gráficos associados às esperanças condicionais da freqüência de *C+G* são similares, indicando um mesmo perfil. O gráfico associado ao modelo com 3 estados é apresentado na figura 27.

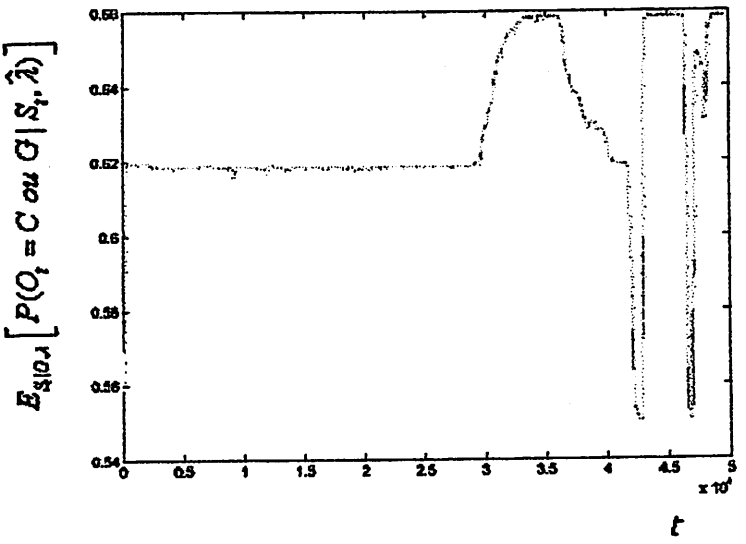



FIGURA 27. Gráfico da *composição local* de *C+G* em cada posição da seqüência do *Xanthomonas Axonopodis* associado ao modelo com 3 estados.

Observando o gráfico da *composição local* associado ao modelo com 3 estados, podemos notar um comportamento altamente homogêneo nas 30000 primeiras posições da seqüência. Este comportamento sugere a presença de uma única região funcional presente neste trecho da seqüência. No segundo trecho da seqüência, após a posição 30000, existe uma variação maior que no primeiro trecho (maior heterogeneidade) com relação aos valores da composição local. Isso sugere a possível presença de várias funcionalidades distintas em termos de produção de proteínas similares.

A figura 28 apresenta uma ilustração da divisão funcional do segmento analisado.

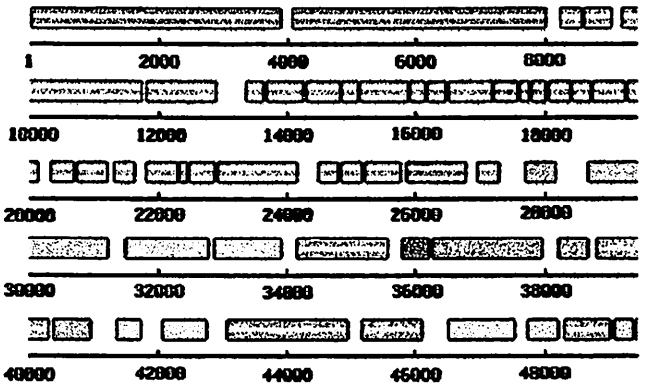



Figura 28. Regiões de expressão do gene *XAC0965* até o gene *XAC1014* da bactéria *Xanthomonas axonopodis*.

Observe que nas primeiras 30000 posições existe a codificação de proteínas que atuam na tradução e transcrição, processos ligados à síntese de proteínas. Nas 20000 posições seguintes, existem várias funcionalidades associadas ao transporte de íons, metabolismo, além de regiões sem funcionalidade conhecida.

Com relação à seqüência analisada, os resultados obtidos a partir da utilização dos *HMM's* fornecem boas descrições gráficas das diferenças funcionais da seqüência em estudo. A técnica conseguiu discriminar de maneira satisfatória distintas regiões funcionais presentes neste segmento.

### 4.4 *Streptococcus pneumoniae*

O *Streptococcus Pneumoniae* é um importante patógeno humano. Esta bactéria causa a maioria das infecções respiratórias agudas e otites (Infecções de ouvido). Cerca de 3 milhões de mortes por ano em crianças são causadas por

pneumonia e outras doenças correlacionadas a esta bactéria, evidenciando a importância da mesma (Tetelin et al, 2001).

O segmento obtido a partir do *Genbank*, sob catalogação NC003098, é correspondente à seqüência de bases a partir do gene *SPR0584* até o gene *SPR0641*. Este segmento possui um total de 49105 pares de bases. O segmento escolhido é marcadamente heterogêneo em termos de funcionalidades. Assim, o resultado da análise, utilizando o *HMM*, deve refletir este comportamento.

Como nos segmentos analisados anteriormente, foram obtidos, para os 4 modelos competidores (modelos com 2, 3, 4 e 5 estados), os ajustes através do método da máxima verossimilhança, os valores dos critérios, *BIC*, $\Delta BIC$, *AIC* e os graus de liberdade (tabela 4).

TABELA 4. Valores de Graus de Liberdade, *BIC* e *AIC* associados á seqüência da bactéria *Streptococcus pneumoniae*.

| *N° de estados* | *Grau de liberdade* | *BIC* | $\Delta BIC$ | *AIC* |
|---|---|---|---|---|
| 2 | 5 | 65401.579 | -2672 | 65357.571 |
| 3 | 12 | 65435.606 | -2638 | 65338.787 |
| 4 | 21 | 65509.551 | -2564 | 65342.318 |
| 5 | 32 | 65623.531 | -2450 | 65368.281 |

Observe que, segundo o critério *BIC*, o modelo mais adequado (modelo com valor mínimo de *BIC*) é o *HMM* com 2 estados. Porém, segundo o critério *AIC*, o modelo considerado mais adequado é o *HMM* com 3 estados. Embora os dois modelos sejam diferentes, os gráficos da composição local de *C+G*, nestes dois modelos, são bem similares e representam, de maneira satisfatória, a conhecida heterogeneidade funcional do segmento.

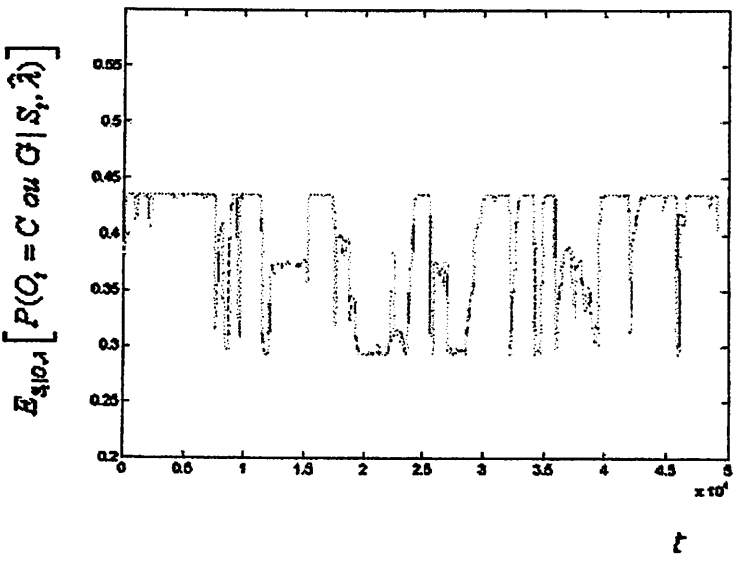



FIGURA 29. Gráfico da *composição local* de *C+G* em cada posição da seqüência do *Streptococcus pneumoniae*, sob o modelo com 3 estados.

Podemos observar, através da figura 29, a característica altamente heterogênea quanto composição local de *C+G* ao longo da seqüência. Isso sugere que este fragmento específico do genoma do *Streptococcus* possui vários segmentos pequenos (proporcionalmente ao tamanho da seqüência), que codificam várias proteínas de natureza estrutural ou funcional distintas.

Observe o gráfico ilustrativo, na figura 30, adaptado do *Genbank*, sobre regiões de expressão associadas a este segmento.

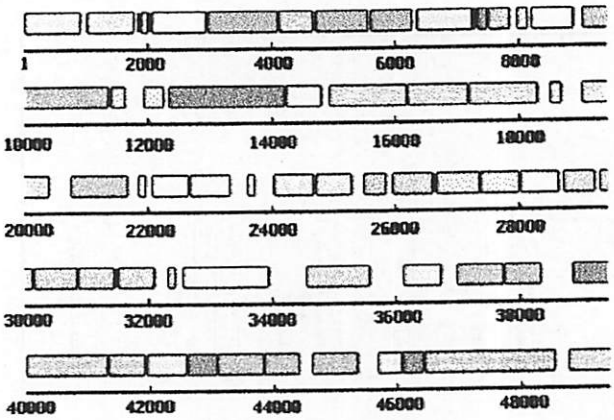



FIGURA 30. Regiões de expressão do gene *SPR0584* até o gene *SPR0681* do *Streptococcus pneumoniae*.

Podemos notar que não existem segmentos de grande dimensão (maiores que 5000) que codificam um determinado tipo de proteína similar. O que se apresenta no segmento analisado é a presença de várias regiões com expressão de proteínas com funções distintas.

Desta maneira, o resultado da análise obtida através do *HMM* mostrou-se satisfatório neste segmento da bactéria *Streptococcus pneumoniae*, refletindo a estrutura funcional heterogênea da mesma em todo o segmento.

### 4.5 *Escherichia coli*

A *Escherichia coli* é um importante patógeno humano. Esta bactéria foi descoberta em 1885 e, desde então, recebeu várias denominações. Entre estas, as *Bacillus Coli commune*, *Bacillus Coli*, *Bacterium Coli*. Entre os tipos de infecções causadas por esta bactéria, podemos citar colites, infecções urinárias e

gastrointestinais, entre outras (Santos, 2003), evidenciando a importância desta bactéria.

A seqüência de bases da bactéria *Escherichia coli* analisada foi obtida a partir do *Genbank* sob código NC004431. Esta subseqüência inclui bases do gene *C4208* até o gene *C4253*, totalizando 44600 bases.

Como nas seqüências anteriores, os modelos competidores adotados são os *HMM* 's contendo 2, 3, 4 e 5 estados. Os valores dos critérios *BIC*, $\Delta BIC$ e *AIC*, e graus de liberdade associados aos modelos estão apresentados na tabela 5.

TABELA 5. Valores de Graus de Liberdade, *BIC* e *AIC* associados á seqüência da bactéria *Escherichia coli*.

| *Nº de estados* | *Grau de liberdade* | *BIC* | $\Delta BIC$ | *AIC* |
|---|---|---|---|---|
| 2 | 5 | 61200.477 | -8114 | 61156.950 |
| 3 | 12 | 61202.015 | -8113 | 61106.255 |
| 4 | 21 | 61253.340 | -8061 | 61087.935 |
| 5 | 32 | 61357.238 | -7957 | 61104.779 |

Segundo o critério *BIC*, o *HMM* com o menor número de estados é o que melhor descreve os dados. Já o critério *AIC* aponta o modelo com 4 estados como o mais adequado.

O gráfico das esperanças condicionais do conteúdo de *C+G*, associado ao modelo com 4 estados, é apresentado na figura 31.

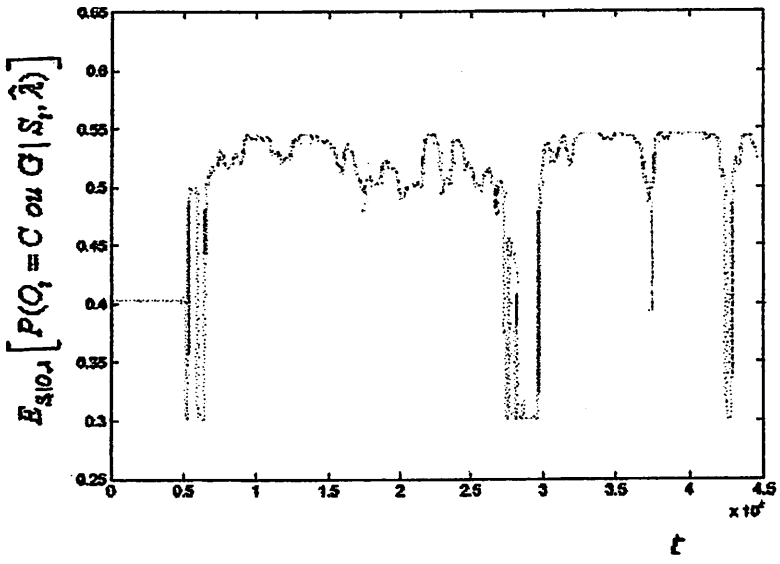



FIGURA 31. Gráfico da composição local de *C+G* em cada posição da seqüência do *Escherichia coli*.

Podemos notar, pelo comportamento do gráfico, um trecho inicial com alta homogeneidade com relação à composição local. Este trecho vai até a posição 5000. Isso sugere a presença de uma região funcional característica neste trecho. Todo o restante da seqüência (posição 5000 até posição 44600) não apresenta em nenhum trecho um comportamento homogêneo claro (posição 30000 até 40000 apresenta um indício de homogeneidade). Este resultado indica a presença de um emaranhado de funcionalidades associadas a este segmento.

Se observarmos a estrutura funcional real da seqüência analisada, apresentada na figura 32, constataremos que o *HMM* não apresentou um resultado tão satisfatório quanto nas outras seqüências, deixando de evidenciar algumas regiões com características homogêneas em relação à freqüência de *C+G*.

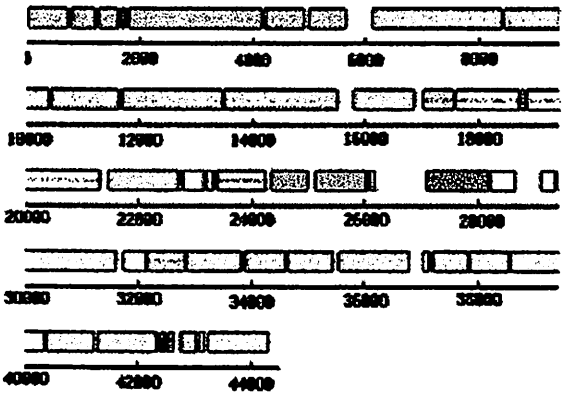


FIGURA 32. Gráfico das esperanças condicionais de probabilidades de presença de *C+G* em cada posição da seqüência do *Escherichia Coli*.

O primeiro trecho, composto pelas 5000 primeiras bases, está associado à codificação de proteínas com função ligada à secreção celular. Nesta região, o *HMM* conseguiu um bom resultado. No entanto, na região seguinte, partindo da base 6000 até a base 15000, existe uma região funcional associada a transporte de carboidratos e metabolismo, que não foi evidenciada pelo gráfico da composição local (Trecho bastante heterogêneo). O trecho do segmento de *DNA* que inicia aproximadamente na base 16000 até a base 32000 é heterogêneo em relação a características funcionais, o que fica evidenciado pela análise, utilizando o *HMM*. Da posição 33000 da seqüência até 37000, existe a presença de uma região funcional ligada ao transporte de aminoácidos que fica moderadamente evidenciada no gráfico das esperanças condicionais. A mesma situação acontece no segmento da posição 37000 até 44000 no qual ocorre uma região funcional ligada ao transporte e metabolismo de carboidratos.

Vale ressaltar que o modelo com maior número de estados (5 estados), embora não escolhido pelos critérios adotados, conseguiu capturar mais claramente as regiões funcionais do segmento analisado.

# 5. CONCLUSÕES

Nesta dissertação, podemos sumarizar as conclusões como:

- Os *HMM's* se mostraram satisfatórios, considerando o problema de *segmentação de DNA*. Os resultados das análises da seqüência do bacteriófago *lambda* foram similares aos obtidos por Churchill (1989). Nas seqüências do *Xanthomonas axonopodis* pv. citri, *Xylella fastidiosa*, *Escherichia* coli, o *HMM* conseguiu explicitar os segmentos de características homogêneas. Na seqüência do *Streptococcus pneumoniae*, o *HMM* conseguiu descrever a característica heterogênea do segmento escolhido.
- O algoritmo *EM* e os métodos para evitar problemas numéricos se mostraram eficazes para estimação dos parâmetros associados aos *HMM's*.
- O software desenvolvido que implementa métodos dos *HMM's* se mostrou eficiente na aplicação a dados reais.
- Os *HMM's* são bons modelos para problemas associados ao reconhecimento de padrão, particularmente onde existe alguma relação de dependência entre os dados.

O autor lista como possíveis trabalhos futuros:

- A utilização de Inferência Bayesiana no processo de estimação dos modelos *HMM's*.
- A busca pelos métodos associados aos 3 problemas básicos do *HMM's* em que se considera a generalização da ordem da Cadeia de Markov e a

generalização para ordem de dependência entre as variáveis observáveis dado o conhecimento dos estados.

- Implementação dos métodos associados ao *HMM* que prevê dependência de 1ª- ordem entre as observações (variáveis observáveis) dado o conhecimento dos estados.

# 6 REFERÊNCIAS BIBLIOGRÁFICAS


AAS, K.; EIKVIL, L. Text page recognition using grey - level features and hidden Markov models. **Pattern recognition,** Oxford, v. 29, n. 6, p. 977-985, June 1996.

BAUM, L. E.; PETRIE, T. Statistic inference for probabilistic functions of finite state Markov chain. **Annals of Mathematical Statistics,** Baltimore, v. 37, n. 6, p. 1554-1563, Dec. 1966.

BISWAS, S. **Applied stochastic processes:** a biostatistical and population oriented approach. New York: Jonh Wiley & Sons, 1995. 427 p.

BOLDRINI, J. L. **Álgebra linear.** 3. ed. São Paulo: Harper e Row, 1984. 411 p.

BOYS, R. J.; HENDERSON, D. A. A Bayesian approach to DNA sequence segmentation. **Biometrics,** Oxford, v. 60, n. 3, p. 573-581, Sept. 2004.

BOYS, R. J.; HENDERSON, D. A.; WILKINSON, D. J. Detecting homogeneous segments in DNA sequences by using hidden Markov models. **Journal of the Royal Statistical Society Serie C - Applied Statistics,** London, v. 49, n. 2, p. 269-285, 2000.

BREIMAN, L. **Probability and stochastic processes with a view toward application.** Boston: Houghton Mifflin Company, 1969. 324 p.

CHURCHILL, G. Hidden Markov Chain and the analysis of genome structure. **Computer Chemistry,** Oxford, v. 16, n. 2, p. 107-115, 1992.

CHURCHILL, G. Stochastic models for Heterogeneous DNA sequences. **Bulletin of Mathematical Biology,** Oxford, v. 51, n. 1, p. 79-94, 1989.

da SILVA, A. C. R.; FERRO, J. A.; REINACH, F. C.; FARAH, C. S.; FURLAN, L. R.; QUAGGIO, R. B.; MONTEIRO-VITORELLO, C. B.; VAN SLUYS, M. A.; ALMEIDA, N. F.; ALVES, L. M. C.; do AMARAL, A. M.; BERTOLINI, M. C.; CAMARGO, L. E. A.; CAMAROTTE, G.; CANNAVAN, F.; CARDOZO, J.; CHAMBERGO, F.; CIAPINA, L. P.; CICARELLI, R. M. B.; COUTINHO, L. L.; CURSINO-SANTOS, J. R. ; EL-DORRY, H.; FARIA, J. B.; FERREIRA, A. J. S.; FERREIRA, R. C. C.; FERRO, M. I. T.; FORMIGHIERI, E. F.; FRANCO, M. C.; GREGGIO, C. C.; GRUBER, A.; KATSUYAMA, A. M.; KISHI, L. T.; LEITE, R. P.; LEMOS, E. G. M.;

LEMOS, M. V. F.; LOCALI, E. C.; MACHADO, M. A.; MADEIRA, A. M. B. N.; MARTINEZ-ROSSI, N. M.; MARTINS, E. C.; MEIDANIS, J.; MENCK, C. F. M.; MIYAKI, C. Y.; MOON, D. H.; MOREIRA, L. M.; NOVO, M. T. M.; OKURA, V. K.; OLIVEIRA, M. C.; OLIVEIRA, V. R.; PEREIRA, H. A.; ROSSI, A. ; SENA, J. A. D.; SILVA, C.; de SOUZA, R. F.; SPINOLA, L. A. F.; TAKITA, M. A.; TAMURA, R. E.; TEIXEIRA, E. C.; TEZZA, R. I. D.; TRINDADE dos SANTOS, M.; TRUFFI, D.; TSAI, S. M.; WHITE, F. F.; SETUBAL, J. C.; KITAJIMA, J. P.; Comparison of genomes of two Xanthomonas pathogens with differing host specificities. **Nature,** v. 417, n. 6887, p. 459-463, May 2002.

da SILVA, C. Q. Hidden Markov models applied to a subsequence of the Xylella fastidiosa genome. **Genetics and Molecular Biology,** Ribeirão Preto, v. 2, n. 4, p. 529-535, Dec. 2003.

da SILVA, C. Q. Tutorial sobre modelos markovianos com estados latentes. in: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 25., 2002, Nova Friburgo. 79 p. Apostila.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likehood from incomplete data via the EM algorithm. **Journal of Royal Statistical Society Serie B - Methodological,** London, v. 39, n. 1, p. 1-38, 1977.

DURBIN, R.; EDDY, S.; KROGH, A.; MITCHISON, G.; **Biological Sequence analysis:** probabilistic models o proteins and nucleic acids. Cambridge: Cambridge University Press, 1999. 368 p.

EDDY, S. R. Hidden Markov models. **Current opinion in Structurel Biology,** London, v. 6, n. 3, p. 361-365, June 1996.

EDDY, S. R. Profile hidden Markov models. **Bioinformatics,** Oxford, v. 14, n. 9, p. 755-763, 1998.

GUSFIELD, D. **Algorithms on strings, trees, and sequences**. New York: Cambridge University Press, 1997. 534 p.

HINKLEY, D. V. Inference about the change point in a sequence of random variables. **Biometrika,** London, v. 57, n. 1, p. 1-17, Apr. 1970.

HINKLEY, D. V.; HINKLEY, E. A. Inference about the chage-point in a sequence of binomial variables. **Biometrika,** London, v. 57, n. 3, p. 477-488, Dec. 1970.

HUGHES, J. P.; GUTTORP, P.; CHARLES, S. P. A non-homogeneous hidden Markov model for precipitation occurrence. **Journal of the Royal Statistics Serie C - Applied Statistics**, London, v. 48, n. 1. p. 15-31, 1999.

KREUZER, H.; MASSEY, A. **Engenharia genética e biotecnologia**. Tradução de Ana Cristina de Oliveira Dias. 2. ed. Porto Alegre: Artmed, 2002. 434 p. Título Original: Recombinant DNA and biotechnology.

KROGH, A.; BROWN, M.; MIAN, I. S.; SJOLANDER, K.; HASSLER, D.; Hidden Markov Models in Computational Biology: application to protein modeling. **Journal of Molecular Biology**, London, v. 235, n. 5, p. 1501-1531, Feb. 1994.

LEWIN, B. **Genes V**. Oxford: Oxfor University Press, 1994. 1272 p. (International Student Edition).

MEIDANIS, J.; SETUBAL, J. C. **Uma introdução á Biologia Computacional.** Recife: UFPE-DI, 1994. 131 p. Apostila.

METZLER, D. E; **Biochemistry:** the chemical reactions of living cells. New York: Academic, 1977. 1129 p.

MOOD, A.; GRAYBILL, F. A.; BOES, D. C.; **Introduction to the Theory of Statistics**. 3. ed. Tokio: Mcgraw-hill Kogakusha, 1963. 564 p. (International Student Edition).

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **Genbank**. Banco de dados de seqüências de DNA. disponível em: <http://www. ncbi. nlm. nih. gov>. Acesso em: 10 ago. 2004.

RABINER, L. A tutorial on hidden Markov models and selected applications in speech recognition. **Proceedings of the IEEE**, New York, v. 77, n. 2, p. 257-286, Fev. 1989.

RAO, C. R. **Linear statistical inference and its applications**. New York: Jonh Wiley, 1973. 625 p.

ROSS, S. M. **A first course in probability**. New York: Macmilan Publishing Company, 1988. 420 p.

RUSSEL, S. J. **Atificial intelligence:** a modern approach. New Jersey: Prentice Hall, 1995. 932 p. (Prentice Hall Series in Artificial Intelligence)

RYDEN, T.; TERASVIRTA, T.; ASBRINK, S. Stylized Facts of Daily Return Series and the Hidden Markov Model. **Journal of applied econometrics,** Sussex, v. 13, n. 3, p. 217-230, May 1998.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike information criterion Statistics.** D. Reidel Publishing Company, 1986.

SANGER, F.; COULSON, A. R.; HONG, G. F.; HILL, D. F.; PETERSEN G. B. Nucleic sequence of Bacteriophage $\lambda$ DNA. **Journal of Molecular Biology,** London, v. 162, n. 4, p. 729-773, 1982.

SANTO, E. **Estudo dos fatores de virulência de Escherichia coli, isolada de infecção urinária em humanos.** 2003. 135 p. Dissertação (Mestrado) – Universidade Estadual Paulista. Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal.

SANTOS, J. P. O.; MELLO, M. P.; MURARI, I. T. C. **Introdução a análise combinatória.** Campinas: UNICAMP, 1995. 295 p.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics,** v. 6, p. 461-464, 1978.

SIMPSON, A. J. G.; REINACH, F. C.; ARRUDA, P.; ABREU, F. A.; ACENCIO, M.; ALVARENGA, R.; ALVES, L. M. C.; ARAYA, J. E.; BAIA, G. S.; BAPTISTA,C. S.; BARROS, M. H.; BONACCORSI, E. D.; BORDIN, S.; BOVE, J. M.; BRIONES, M. R. S.; BUENO, M. R. P.; CAMARGO, A. A.; CAMARGO, L. E. A.; CARRARO, D. M.; CARRER, H.; COLAUTO, N. B.; COLOMBO, C.; COSTA, F. F.; COSTA, M. C. R.; COSTA-NETO, C. M.; COUTINHO, L. L.; CRISTOFANI, M.; DIAS-NETO, E.; DOCENA, C.; EL-DORRY, H.; FACINCANI, A. P.; FERREIRA, A. J. S.; FERREIRA, V. C. A.; FERRO, J. A.; FRAGA, J. S.; FRANCA, S. C.; FRANCO, M. C.; FROHME, M.; FURLAN, L. R.; GARNIER, M.; GOLDMAN, G. H.; GOLDMAN, M. H. S.; GOMES, S. L.; GRUBER, A.; HO, P. L.; HOHEISEL, J. D.; JUNQUEIRA, M. L.; KEMPER, E. L.; KITAJIMA, J. P.; KRIEGER, J. E.; KURAMAE, E. E.; LAIGRET, F.; LAMBAIS,M. R.; LEITE, L. C. C.; LEMOS, E. G. M.; LEMOS, M. V. F.; LOPES, S. A.; LOPES, C. R.; MACHADO, J. A.; MACHADO, M. A.; MADEIRA, A. M. B. N.; MADEIRA, H. M. F.; MARINO, C. L.; MARQUES, M. V.; MARTINS, E. A. L.; MARTINS, E. M. F.; MATSUKUMA, A. Y.; MENCK, C. F. M.; MIRACCA, E. C.; MIYAKI, C. Y.;

MONTEIRO-VITORELLO, C. B.; MOON, D. H.; NAGAI, M. A.; NASCIMENTO, A. L. T. O.; NETTO, L. E. S.; NHANI JR.; A.; NOBREGA, F. G.; NUNES, L. R.; OLIVEIRA, M. A.; de OLIVEIRA, M. C.; de OLIVEIRA, R. C.; PALMIERI, D. A.; PARIS, A.; PEIXOTO, B. R.; PEREIRA, G. A. G.; PEREIRA JR.; H. A.; PESQUERO, J. B.; QUAGGIO, R. B.; ROBERTO, P. G.; RODRIGUES, V.; de M. ROSA, A. J.; de ROSA JR.;V. E.; de SA, R. G.; SANTELLI, R. V.; SAWASAKI, H. E.; da Silva, A. C. R.; da Silva, F. R.; da Silva, A. M.; SILVA JR.; W. A.; da SILVEIRA, J. F.; SILVESTRI, M. L. Z.; SIQUEIRA, W. J.; de SOUZA, A. A.; de SOUZA, A. P.; TERENZI, M. F.; TRUFFI, D.; TSAI, S. M.; TSUHAKO, M. H.; VALLADA, H.; VAN SLUYS, M. A.; VERJOVSKI-ALMEIDA, S.; VETTORE, A. L.; ZAGO, M. A.; ZATZ, M.; MEIDANIS, J.; SETUBAL, J. C.; The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis. **Nature**, London, v. 406, n. 6792, p. 151-157, July 2000.

SMITH, A. F. M. A bayesian approach to inference about a change-point in a sequence of random variables. **Biometrika**, London, v. 62, n. 2, p. 407-416, Aug. 1975.

STADEN, R. Graphic Methods to Determine the function of Nucleic Acid Sequence. **Nucleic Acid Research**, Oxford, v. 12, n. 1, p. 521-538, 1984.

TETTELIN, H.; NELSON, K. E.; PAULSEN, I. T.; EISEN, J. A.; READ, T. D.; PETERSON, S.; HEIDELBERG, J.; DEBOY, R. T.; HAFT, D. H.; DODSON, R. J.; DURKIN, A. S.; GWINN, M.; KOLONAY, J. F.; NELSON, W. C.; PETERSON, J. D.; UMAYAM, L. A.; WHITE, O.; SALZBERG, S. L.; LEWIS, M. R.; RADUNE, D. ; HOLTZAPPLE, E.; KHOURI, H.; WOLF, A. M.; UTTERBACK, T. R.; HANSEN, C. L.; MCDONALD, L. A.; FELDBLYUM, T. V.; ANGIUOLI, S.; DICKINSON, T.; HICKEY, E. K.; HOLT, I. E.; LOFTUS, B. J.; YANG, F.; SMITH, H. O.; CRAIG VENTER, J.; DOUGHERTY, B. A.; MORRISON, D. A.; HOLLINGSHEAD, S. K.; FRASER C. M.; Complete Genome Sequence of a virulent Isolate of Streptococcus pneumoniae. **Science**, Washington, v. 293, n. 5529, p. 498-505, July 2001.

VLONTZOS, J.; KUNG, S. Hidden Markov Models for character Recognition. **IEEE transactions on image processing**, Los Alamitos, v. 1, n. 4, p. 539-543, out. 1992.

# 7. ANEXOS

O objetivo deste texto é apresentar sucintamente o software *SIMHMM*, que implementa as funções associadas aos *HMM's*. Este software foi desenvolvido utilizando-se a linguagem visual *Delphi*.

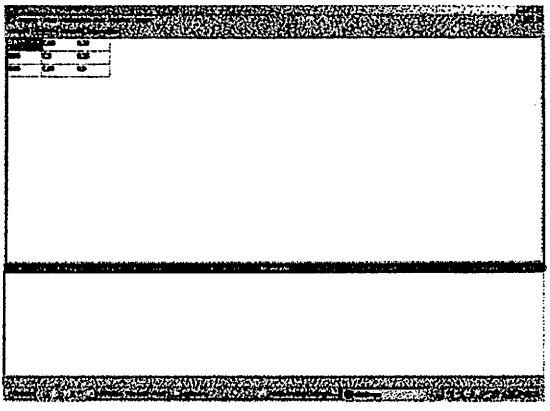


FIGURA 1A. Software *SIMHMM*

O *SIMHMM* permite manipular seqüências de observações (variáveis observáveis $O_t$), possibilitando edição (modificação) e armazenamento (salvar). Estas observações são armazenadas em arquivos texto simples (*".txt"*) (*ASCII*). O *SIMHMM* permite a manipulação de mais de um arquivo de seqüência de observações simultaneamente.

Outra funcionalidade do *SIMHMM* é permitir a manipulação de modelos *HMM* (consideram dependência markoviana de 1ª ordem entre as variáveis aleatórias latentes $S_t$ e independência entre as variáveis aleatórias $O_t \mid S_t$) em arquivos com extensão *".hmm"*. O *SIMHMM* permite trabalharmos com mais de um arquivo *HMM* simultaneamente. Estes arquivos armazenam, em código

*ASCII*, os parâmetros associados a um *HMM* ( $\lambda = (\pi, A, B)$ ), além do alfabeto ou conjunto de observações $\Omega_o$ (Figura 2A).

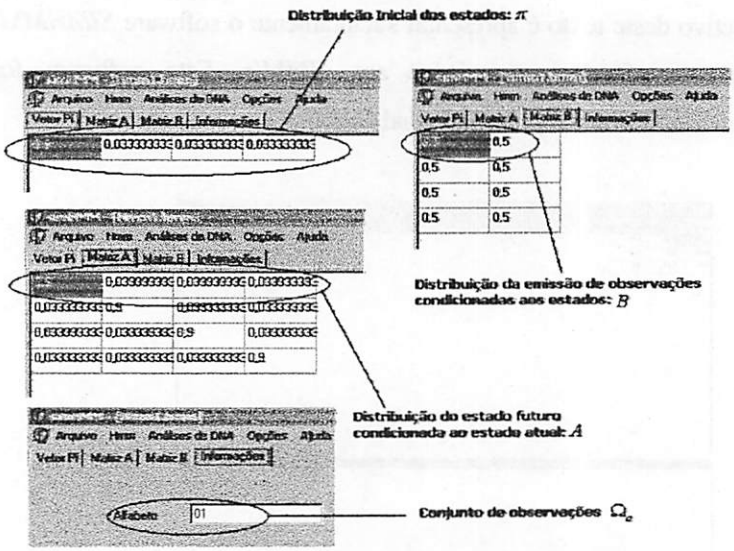



FIGURA 2A. Arquivo do tipo (*".hmm"*).

Sumarizando, temos associados aos *HMM's* dois tipos de arquivos, que são *".hmm"* e arquivos de seqüências de observações *".txt"*(*ASCII* em geral), aos quais podem ser associados as seguintes funcionalidades ilustradas na figura 3A:

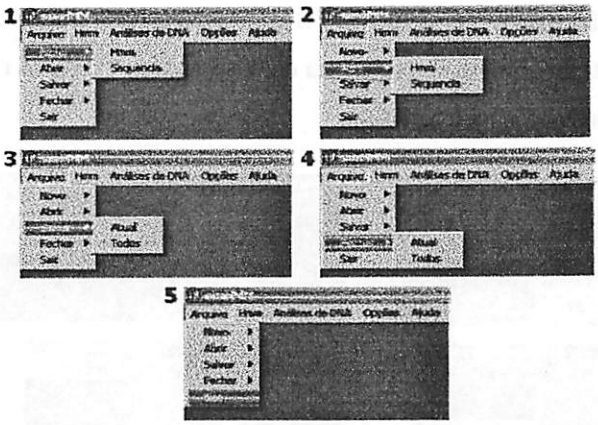



FIGURA 3A. Funcionalidades associadas aos arquivos do *SIMHMM*

- **1- Novo - Hmm**: Cria um novo arquivo *".hmm"* onde o usuário deve informar o número de estados e o alfabeto (conjunto $\Omega_o$).
- **1- Novo - Seqüência**: Cria um novo arquivo texto que armazenará uma seqüência de observações $O = \{O_1, O_2, ..., O_T\}$ associada a um *HMM*.
- **2- Abrir - Hmm**: Abre um arquivo *".hmm"* já existente onde o usuário deve informar sua localização.
- **2- Abrir - Seqüência**: Abre um arquivo texto já existente onde está armazenada uma seqüência de observações $O = \{O_1, O_2, ..., O_T\}$. O usuário deve informar a sua localização.
- **3- Salvar - Atual**: Salva o arquivo que está ativo.
- **3- Salvar - Todos**: Salva todos os arquivos abertos.
- **4- Fechar - Atual**: Fecha o arquivo que está ativo.
- **4- Fechar - Todos**: Fecha todos os arquivos abertos.
- **5- Sair**: Finaliza a execução do *SIMHMM*.

O *SIMHMM*, como dito, implementa os métodos para solução dos 3 problemas básicos dos *HMM's* (vide seção 2.1). Considerando o último arquivo "*.hmm*" ativo e o último arquivo de seqüência de observações ".txt", os métodos do *SIMHMM* são (ilustrados na figura 4A):

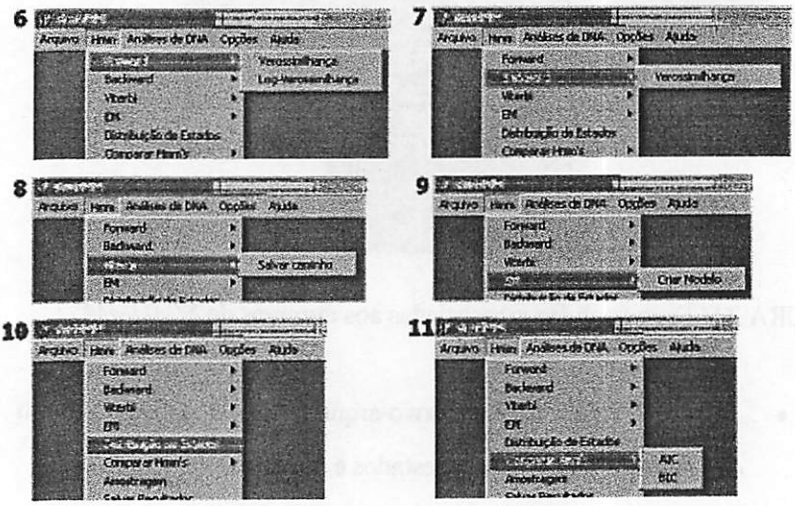



FIGURA 4A. Funcionalidades associadas aos 3 problemas básicos dos *HMM's*.

- **6- Forward - Verossimilhança**: apresenta o valor da verossimilhança na janela inferior "*Resultados*". Esta verossimilhança será associada ao último modelo *HMM* ativo (arquivo "*.hmm*") e ao último arquivo de *seqüência* ativo (arquivo *ASCII*). Para obter a verossimilhança, é implementado o método *forward* **sem utilização de variáveis normalizadas**.
- **6- Forward – log-Verossimilhança**: apresenta o valor da log-verossimilhança na janela inferior "*Resultados*". Esta log-verossimilhança será associada ao último modelo *HMM* ativo (arquivo "*.hmm*") e ao último arquivo de *seqüência* ativo

(arquivo *ASCII*). Para obter a verossimilhança, é implementado o método *forward* **com a utilização de variáveis normalizadas.**

- **7- Backward - Verossimilhança**: apresenta o valor da verossimilhança na janela inferior "*Resultados*". Esta verossimilhança será associada ao último modelo *HMM* ativo (arquivo "*.hmm*") e ao último arquivo de *seqüência* ativo (arquivo *ASCII*). Para obter a verossimilhança, é implementado o método *backward* **sem utilização de variáveis normalizadas.**
- **8- Viterbi – Salvar Caminho**: Obtém a seqüência de estados mais provável $S = \{S_1, S_2, ..., S_T\}$ a partir do último modelo (*HMM*) ativo e da última seqüência de observações ativa (*".txt"*). A seqüência de estados obtida será armazenada em um arquivo que o usuário deve indicar.
- **9- EM – Criar Modelo**: Cria um novo modelo (*HMM*) que possui o mesmo número de estados e o mesmo conjunto de observações do último modelo (*HMM*) ativo. Os parâmetros deste novo modelo são as estimativas de máxima verossimilhança obtidas através do algoritmo *EM*, tendo como as observações, o último arquivo de seqüências (*".txt"*) ativo. O modelo inicial (iteração inicial) do algoritmo *EM* contém os valores dos parâmetros do último modelo ativo.
- **10- Distribuição dos estados**: A partir do último modelo (*HMM*) ativo e do último arquivo de seqüência de observações (*".txt"*), obtém-se $\gamma_{t(i)} = P(S_t = i \mid O, \lambda)$ para todos os estados $i \in \{0,1,...,r-1\}$ e $1 \le t \le T$ em que $T$ é tamanho da seqüência de observações e $r$ o número de estados. A distribuição dos estados será armazenada em um arquivo texto a

ser indicado pelo usuário, cuja $t$-ésima linha e $i$-ésima coluna representam $\gamma_{t(i)} = P(S_t = i \mid O, \lambda)$.

- **11- Comparar *HMM's* - AIC e BIC**: Obtém o valor do *AIC* e *BIC* associado modelo com número de estados e conjunto de observações do último modelo ativo e da ultima seqüência de observações ativa. Os valores serão mostrados na janela "*Resultados*".

É possível obter amostras de observações associadas a um determinado modelo (*HMM*). As amostras devem ser salvas pelo usuário em um arquivo *ASCII* a ser indicado. É possível manipular a janela de resultados, podendo salvar (arquivo *ASCII* definido pelo usuário) ou limpar seu conteúdo. Estas funções são (ilustração na figura 5A):

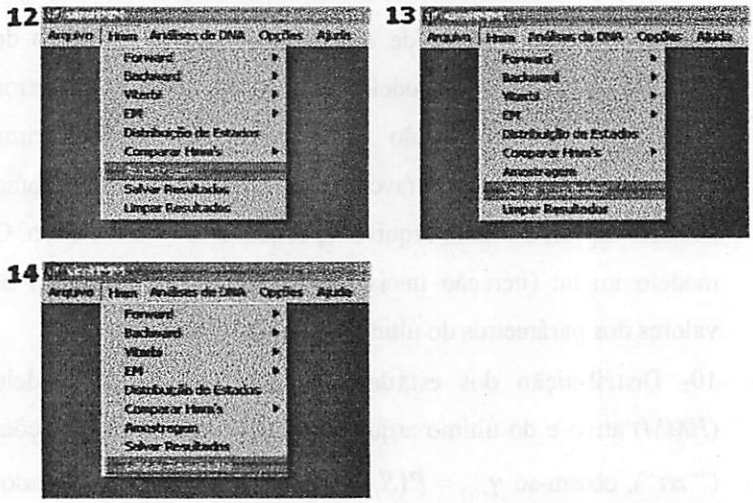



FIGURA 5A. Funcionalidades associadas à amostragem, comparação de modelos e manipulação da janela de resultados.

- **12- Amostragem**: Simula uma amostra das variáveis $O_t$ com um determinado número de observações, que é definido pelo usuário. A amostra será armazenada em um arquivo texto *ASCII*, que é determinado pelo usuário.
- **13- Salvar Resultados**: Salva o conteúdo da janela Resultados em um arquivo texto *ASCII* a ser definido pelo usuário.
- **14- Limpar Resultados**: Limpa o conteúdo da janela Resultados

Outras funcionalidades estão ligadas à análise de regiões homogêneas em seqüências de *DNA*. O *SIMHMM* possui a funcionalidade de calcular a *composição local* associada a um modelo (*HMM*) e a um determinado conjunto de dados armazenado em um arquivo (*".txt"*). Além disso, é possível obter os gráficos de janelas e de segmentos fixos ilustrados no capítulo 3. Algumas opções podem ser modificadas no *SIMHMM*. Em particular, podemos modificar as opções de convergência e número de iterações para a obtenção de estimativas através do algoritmo *EM*. Estas funcionalidades são (ilustração na figura 6A) descritas a seguir:

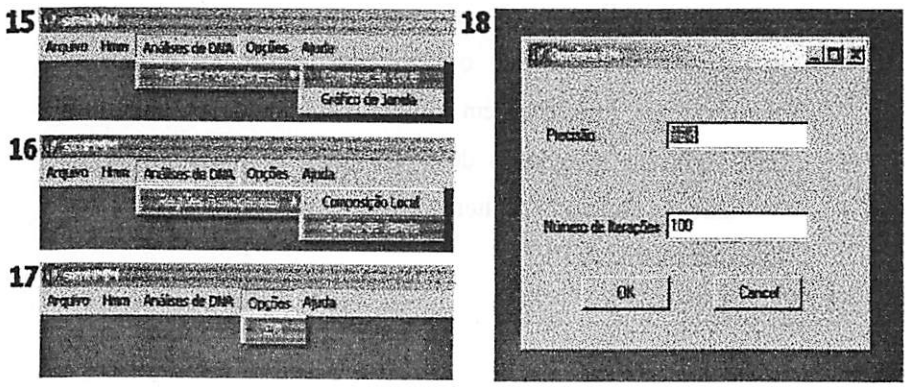


FIGURA 6A. Funcionalidades associadas à análise de seqüências de DNA e customização.

- **15- Análise de DNA – Regiões Homogêneas – Composição local**: Esta opção obtém os valores do gráfico da composição local descrita no fim do capítulo 2. Esta composição local esta associada ao último modelo (*HMM*) ativo e a última seqüência ativa (arquivo de observações *".txt"*). Os valores da composição local serão armazenados em um arquivo texto (*ASCII*), que deverá ser indicado pelo usuário.
- **16- Análise de DNA – Regiões Homogêneas – Gráfico de Janela**: Esta função obtém os valores do gráfico de janelas e de proporções por segmento de tamanhos iguais. Estes valores serão obtidos a partir do último arquivo de seqüências de observações ativo. O usuário vai indicar um arquivo texto (*ASCII*) no qual estes dados serão armazenados. Além disso, o usuário deve indicar o tamanho da janela e o número de segmentos.
- **17 e 18- Opções – EM**: Esta função permite ao usuário configurar as opções do algoritmo *EM* que dizem respeito à convergência. Estas opções são o número de iterações máximo e a precisão máxima que o algoritmo possui em termos de convergência (diferença em log-verossimilhança). O algoritmo *EM* converge quando as duas situações de convergência são verdadeiras (número de iterações mínimo e diferença de log-verossimilhança).