CADEIAS DE MARKOV OCULTAS COMO METODOLOGIA NO ESTUDO DE IMPACTOS ANTROPOGÊNICOS NO COMPORTAMENTO DE CETÁCEOS

Francisco M. C. de Medeiros, Paulo S. Lucio, Diana G. Lunardi, André G. C. Pereira
Universidade Federal do Rio Grande do Norte, CCET.
CEP 59072-970, Natal - R.N. - Brasil
e-mail: moises@ccet.ufrn.br, pslucio@ccet.ufrn.br, lunardi.diana@gmail.com, andre@ccet.ufrn.br

Resumo. Neste trabalho estudamos os modelos de Markov ocultos em espaço de estados discretos. Apresentamos os algoritmos *forward* e *backward* para determinar a probabilidade da sequência observada e, em seguida, estimamos os parâmetros do modelo via algoritmo EM. O objetivo do uso de cadeias de Markov ocultas neste estudo é estimar os parâmetros do modelo via simulação, para isso utilizamos os dados de comportamento do boto-cinza, *Sotalia guianensis*, uma espécie comum no litoral do Brasil. As observações foram realizadas a partir de um mirante localizado a 25m acima do nível do mar, na enseada do Madeiro, RN, entre dezembro de 2007 e fevereiro de 2009 e entre 6h e 16h. O estado comportamental predominante da agregação focal foi registrado a cada 2 minutos e definido como mutuamente exclusivo. Foram observadas 658 agregações de golfinhos em 176 dias de amostragem, totalizando 1.223h de esforço.

Palavras Chave. Cadeia de Markov. Cadeia de Markov oculta. Algoritmo EM.

Abstract. In this work we study the Hidden Markov Models with finite state space. We studied the forward and backward algorithms in order to calculate the probability of a given observed sequence, next we use the EM algorithm to estimate the model parameters. The purpose of the use of Hidden Markov Chains in this study is to estimate the models parameters by simulation, we used in the behavior of Guiana dolphin, Sotalia guianensis, a species common on the coast of Brazil. Systematic observations of Guiana dolphin were made from an observatory located in cliffs about 25m above sea level at Madeiro Bay, from December 2007 to February 2009 and between 6h and 16h The predominant behavioral state of the focal aggregation was recorded every 2 minutes and defined as mutually exclusive. We observed 658 aggregations of dolphins in 176 days of sampling, totalizing 1.223h of effort.

Key Words. Markov chain. Hidden Markov models. EM algorithm.

1 Introdução

Efeitos de processos antrópicos são aqueles derivados de atividades humanas, em oposição a aqueles que ocorrem em ambientes naturais sem influência humana. A análise de dados em sequência ou em cadeia que incluam levantamentos comportamentais ao longo do tempo definido por propriedades ou estados, são muito comuns em Ciências Biológicas (e.g., Nowacek, 2002; Lusseau, 2003; Williams et al., 2006). As técnicas usuais para examiná-los devem fazer parte dos conceitos fundamentais de Processos Estocásticos. A espécie em estudo é o boto-cinza, Sotalia guianensis, um golfinho que frequenta as enseadas da região de Pipa no Rio Grande do Norte (RN), principalmente para forrageio (busca por alimento) e alimentação.

O objetivo do uso de cadeias de Markov ocultas neste estudo é avaliar as estimativas de máxima verossimilhança do modelo via simulação. Cinco estados comportamentais foram avaliados: Forrageio - FO, Alimentação - FE, Deslocamento - TR, Socialização - SO e Repouso - RE. Os registros comportamentais tiveram duração de 2 minutos, ou seja, em 10 minutos poderíamos ter registrado a seguinte sequência: FO-FO-SO-FO-FE. Foram observados 658 grupos de golfinhos em 176 dias de amostragem quando foram registradas um total de 21.658 transições entre estados comportamentais na presença de embarcação (impacto).

2 Cadeias de Markov Ocultas

Os Modelos de Markov Ocultos (HMMs) têm se tornado, ao passar dos anos, uma importante ferramenta na modelagem de seqüências de variáveis aleatórias fracamente dependentes. Estes modelos estocásticos estão sendo aplicados em várias áreas do conhecimento, servindo de modelagem para problemas de processamento de voz (Rabiner, 1989), neurofisiologia (Fredkin-Rice, 1992), biologia (Leroux-Puterman, 1992), processamento de imagens (Huang et al., 2008), entre outras. Esta calsse de processos estocáticos é baseada em uma cadeia de Markov $\{X_t\}_{t\in\mathbb{N}}$ não observável que descreve a evolução dos estados do processo. Considerando, então, a importância e utilidade dos HMMs em várias áreas de pesquisa, nosso principal objetivo é estudar esses modelos.

Para termos idéia do desenvolvimento da teoria inferencial em HMMs, ela foi publicada pela primeira vez por Baum-Petrie (1966) no caso em que o processo observável $\{Y_t\}_{t\in\mathbb{N}}$ tinha espaço de estados finito. Neste trabalho, mostram a consistência e a normalidade assintótica dos estimadores de máxima verossimilhança (EMVs). Em Petrie (1969), as condições de consistência dos EMVs são enfraquecidas. Lindgren (1978) constroe estimadores consistentes e assintoticamente normais para os parâmetros, determinando as densidades condicionais de Y_t dado X_t, mas não considera a estimação das probabilidades de transição. Mais tarde, Leroux (1992) prova a consistência dos EMVs para HMMs gerais sob condições fracas. Em 1994, Rydén propõe uma nova classe de estimadores e prova a consistência e a normalidade assintótica sob algumas condições de regularidade. Depois, Bickel-Ritov (1996) constroem a normalidade assintótica local dos EMVs e em Bickel et al. (1998) a normalidade assintótica desses estimadores é provada.

Uma importante bibliografia no estudo inferencial em HMMs é o livro de Cappé et al. (2005). Trata-se de um estudo completo de inferência em HMMs gerais.

Considere um experimento aleatório $\mathcal{E}$. Cada realização desse experimento gera uma seqüência de observações $\{y_1, ..., y_T\}$, que é considerada como uma realização de comprimento T de algum processo aleatório $\{Y_t : t \in \mathbb{N}\}$ com espaço de estados finito S. O processo $\{Y_t\}_{t\in\mathbb{N}}$ é gerado por dois mecanismos probabilísticos: em primeiro lugar, uma cadeia de Markov homogênea não observável $\{X_t : t \in \mathbb{N}\}$ com m estados que ocorrem em qualquer instante de tempo t e, em segundo lugar, um conjunto de distribuições de probabilidades, uma para cada estado, que produzem as observações a partir de um conjunto finito de T possibilidades.

Definição 1. *Uma cadeia de Markov oculta (HMM) a tempo discreto é um processo estocástico duplo* $\{(X_t, Y_t)\}_{t\in\mathbb{N}}$ *tal que:*

1. $\{X_t\}_{t\in\mathbb{N}}$ *é uma cadeia de Markov homogênea não observável, de espaço de estados finito* $E = \{1, 2, \ldots, m\}$.

2. $\{Y_t\}_{t\in\mathbb{N}}$ *é uma seqüência de variáveis aleatórias condicionalmente independentes com a distribuição de* Y_k *dependendo somente de* X_k*, ou seja*

$$\begin{aligned} P(\boldsymbol{Y} \mid \boldsymbol{X}) &= P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T \mid X_1 = x_1, X_2 = x_2, \ldots, X_T = x_T) \\ &= \prod_{l=1}^{T} P(Y_l = y_l \mid X_l = x_l). \end{aligned}$$

O termo HMM é motivado por assumirmos que $\{X_t : t \in \mathbb{N}\}$ é uma cadeia de Markov não observável, ou seja, não sabemos de qual estado o processo observado foi gerado.

Para caracterizarmos uma cadeia de Markov basta que conheçamos a distribuição inicial e a matriz de probabilidades de transição do processo. Já nos HMMs além da distribuição inicial (denotada por δ) definida por,

$$\delta_i = P(X_1 = i), \qquad \forall\, i \in E,$$

e da matriz de probabilidades de transição da cadeia de Markov (denotada por Γ)

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{bmatrix},$$

sendo

$$\gamma_{ji} = P(X_{t+1} = i \,|\, X_t = j), \qquad \forall\, t \in \mathbb{N} \quad \text{e} \quad i,\, j \in E,$$

é necessário que conheçamos uma matriz de probabilidades $T \times m$ (denotada por Π)

$$\Pi = \begin{bmatrix} \pi_{y_{1,1}} & \pi_{y_{1,2}} & \cdots & \pi_{y_{1,m}} \\ \pi_{y_{2,1}} & \pi_{y_{2,2}} & \cdots & \pi_{y_{2,m}} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{y_{T,1}} & \pi_{y_{T,2}} & \cdots & \pi_{y_{T,m}} \end{bmatrix},$$

sendo

$$\pi_{y_t,j} = P(Y_t = y_t \,|\, X_t = j),$$

com

$$\sum_{t=1}^{T} \pi_{y_t,j} = 1, \qquad \forall\, j \in E.$$

Para calcularmos a probabilidade da sequência observado e obtermos os estimadores de máxima verossimilhança do modelo, precisaremos de uma série de propriedades dos modelos de Markov ocultos. Todas estas propriedades referem-se ao cenário da definição, isto é, o processo $\{X_t\}_{t\in\mathbb{N}}$ é uma cadeia de Markov homogênea com espaço de estados finito E e $\{Y_t : t \in \mathbb{N}\}$ tem, para todo T, a seguinte propriedade: condicionado a $\mathbf{X} = \{X_t : t = 1, \ldots, T\}$, as variáveis aleatórias $Y_1, \ldots, Y_T$ são mutuamente independentes e a distribuição condicional de Y_t só depende de X_t e não das variáveis X_k, $k \neq t$.

Por simplicidade e para não carregar a notação, o evento $\{X_t = x_t\}$ será denotado por X_t. No entanto, quando necessário, usaremos a notação $\{X_t = x_t\}$. De maneira análoga, o evento $\{Y_t = y_t\}$ será denotado por Y_t e quando necessário faremos uso da notação $\{Y_t = y_t\}$.

Para $t = 1, 2, \ldots, T$:

$$P(Y_1, \ldots, Y_T \,|\, X_t) = P(Y_1, \ldots, Y_t \,|\, X_t) P(Y_{t+1}, \ldots, Y_T \,|\, X_t). \tag{1}$$

Quando, $t = T$ usamos a seguinte convenção $P(Y_{t+1}, \ldots, Y_T \,|\, X_t) = 1$.

Para $t = 1, 2, \ldots, T - 1$:

$$P(Y_1, \ldots, Y_T \,|\, X_t, X_{t+1}) = P(Y_1, \ldots, Y_t \,|\, X_t) P(Y_{t+1}, \ldots, Y_T \,|\, X_{t+1}). \tag{2}$$

Para $1 \leq t \leq l \leq T$:

$$P(Y_l, \dots, Y_T \mid X_t, \dots, X_l) = P(Y_l, \dots, Y_T \mid X_l). \tag{3}$$

Finalmente, para $t = 1, 2, \dots, T$:

$$P(Y_t, \dots, Y_T \mid X_t) = P(Y_t \mid X_t) P(Y_{t+1}, \dots, Y_T \mid X_t). \tag{4}$$

2.1 Cálculo da Probabilidade da Sequência de Observações

Os resultados seguintes nos fornecem duas maneiras de determinarmos a probabilidade da sequência observada. O primeiro consiste no cálculo das probabilidades diretas e o segundo no cálculo das probabilidades reversas. Ambos conhecidos na literatura como algoritmo *forward* e *backward*, respectivamente. A função de verossimilhança dos dados observados é usualmente avaliada por esses procedimentos.

Definição 2. *Seja* $\mathbf{Y} = (Y_1, \dots, Y_t, \dots, Y_T)$ *a sequência de observações do modelo* $\lambda = (\delta, \Gamma, \Pi)$. *Para todo* $t \in \{1, \dots, T\}$ *e* $i \in E = \{1, \dots, m\}$ *definimos,*

$$\alpha_t(i) = P(Y_1 = y_1, \dots, Y_t = y_t, X_t = i)$$

e

$$\beta_t(i) = P(Y_{t+1} = y_{t+1}, \dots, Y_T = y_T \mid X_t = i).$$

Quando $t = T$ convencionamos que $\beta_T(i) = 1$ para todo $i \in E$.

Proposição 1. *Seja* $\mathbf{Y} = (Y_1, \dots, Y_T)$ *a sequência de observações do modelo* $\lambda = (\delta, \Gamma, \Pi)$ *então*

$$P(\mathbf{Y} \mid \lambda) = \sum_{i=1}^{m} \alpha_t(i) \beta_t(i), \qquad 1 \leq t \leq T. \tag{5}$$

Demonstração. Da Definição 2 e da Equação 1, para $t = 1, 2, \dots, T$ e $i \in E$ temos,

$$\begin{aligned}
\alpha_t(i)\beta_t(i) &= P(Y_1, \dots, Y_t, X_t = i) P(Y_{t+1}, \dots, Y_T \mid X_t = i) \\
&= P(Y_1, \dots, Y_t \mid X_t = i) P(X_t = i) P(Y_{t+1}, \dots, Y_T \mid X_t = i) \\
&= P(Y_1, \dots, Y_t \mid X_t = i) P(Y_{t+1}, \dots, Y_T \mid X_t = i) P(X_t = i) \\
&= P(Y_1, \dots, Y_T \mid X_t = i) P(X_t = i) \\
&= P(Y_1, \dots, Y_T, X_t = i).
\end{aligned}$$

Então

$$\begin{aligned}
P(\mathbf{Y} \mid \lambda) &= P(Y_1 = y_1, \dots, Y_T = y_T) \\
&= \sum_{i=1}^{m} P(Y_1, \dots, Y_T, X_t = i) \\
&= \sum_{i=1}^{m} \alpha_t(i) \beta_t(i).
\end{aligned}$$

Esse resultado nos mostra uma maneira eficiente de calcularmos a probabilidade da sequência de observações, para isso basta que as probabilidades diretas e as probabilidades reversas, em um determinado instante de tempo t, sejam conhecidas. Como a Equação (5) é válida para todo $t \in \{1, \dots, T\}$, em particular, para $t = T$ temos,

$$P(\mathbf{Y} \mid \lambda) = \sum_{i=1}^{m} \alpha_T(i)\beta_T(i) = \sum_{i=1}^{m} \alpha_T(i),$$

uma vez que $\beta_T(i) = 1$ para todo $i \in E$.

Portanto, basta determinarmos $\alpha_T(i)$, $i \in E$. O resultado seguinte nos dá uma maneira recursiva de calcularmos essas probabilidades.

Teorema 1 (Cálculo das probabilidades diretas)**.** *Seja* $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ *a sequência observável do modelo* $\lambda = (\delta, \Gamma, \Pi)$*. Então*

$$\alpha_1(i) = \pi_{y_1,i}\delta_i, \qquad 1 \leq i \leq m$$

e para $1 \leq j \leq m$ *e* $1 \leq t \leq T-1$,

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^{m} \alpha_t(i)\gamma_{ij}\right) \pi_{y_{t+1},i}.$$

A estrutura do algoritmo *forward* pode ser resumida da seguinte forma:

1° Passo: Inicialização:

$$\alpha_1(i) = \pi_{y_1,i}\delta_i, \qquad 1 \leq i \leq m.$$

2° Passo: Indução:

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^{m} \alpha_t(i)\gamma_{ij}\right) \pi_{y_{t+1},j}, \qquad 1 \leq j \leq m \quad \text{e} \quad 1 \leq t \leq T-1.$$

3° Passo: Finalização:

$$P(\mathbf{Y} \mid \lambda) = \sum_{j=1}^{m} \alpha_T(j).$$

Uma outra maneira de obtermos a probabilidade da sequência de observações é a partir do cálculo das probabilidades reversas.

Proposição 2. *Seja* $\mathbf{Y} = (Y_1, \dots, Y_T)$ *a sequência observável do modelo* $\lambda = (\delta, \Gamma, \Pi)$*. Então,*

$$P(\mathbf{Y} \mid \lambda) = \sum_{j=1}^{m} \pi_{y_1,j}\beta_1(j)\delta_j.$$

Demonstração. Fazendo $t = 1$ na Equação 4 vemos que,

$$\begin{aligned} P(\mathbf{Y} \mid \lambda) &= P(Y_1 = y_1, \dots, Y_T = y_T) = \sum_{j=1}^{m} P(Y_1, \dots, Y_T, X_1 = j) \\ &= \sum_{j=1}^{m} P(Y_1, \dots, Y_T \mid X_1 = j) P(X_1 = j) \\ &= \sum_{j=1}^{m} \underbrace{P(Y_1 \mid X_1 = j)}_{\pi_{y_1,j}} \underbrace{P(Y_2, \dots, Y_T \mid X_1 = j)}_{\beta_1(j)} \underbrace{P(X_1 = j)}_{\delta_j} \\ &= \sum_{j=1}^{m} \pi_{y_1,j}\beta_1(j)\delta_j. \end{aligned}$$

Portanto, conhecendo δ_j, $\pi_{y_1,j}$ e $\beta_1(j)$, para todo $j \in E$, determinamos $P(\mathbf{Y} \mid \lambda)$. Como δ_j e $\pi_{y_1,j}$ são dados, basta calcularmos $\beta_1(j)$ para todo $j \in E$. O Teorema 2 nos mostra uma maneira recursiva para o cálculo das probabilidades reversas.

Teorema 2 (Cálculo das probabilidades reversas)**.** *Seja* $\mathbf{Y} = (Y_1, \ldots, Y_T)$ *a sequência observável do modelo* $\lambda = (\delta, \Gamma, \Pi)$*. Então para* $1 \leq i \leq m$ *e* $1 \leq t \leq T-1$,

$$\beta_t(i) = \sum_{j=1}^{m} \pi_{y_{t+1},j} \beta_{t+1}(j) \gamma_{ij}.$$

A estrutura do algoritmo *backward* pode ser resumida da seguinte forma:

<u>1° Passo:</u> Inicialização:

$$\beta_T(i) = 1, \qquad\qquad 1 \leq i \leq m.$$

<u>2° Passo:</u> Indução:

$$\beta_t(i) = \sum_{j=1}^{m} \pi_{y_{t+1},j} \beta_{t+1}(j) \gamma_{ij}, \quad 1 \leq i \leq m \quad \text{e} \quad 1 \leq t \leq T-1.$$

<u>3° Passo:</u> Finalização:

$$P(\mathbf{Y} \mid \lambda) = \sum_{j=1}^{m} \pi_{y_1,j} \beta_1(j) \delta_j.$$

Os dois algoritmos acima nos fornecem resultados iguais. Portanto, para calcularmos a probabilidade da sequência de observações basta utilizarmos um dos procedimentos (*forward* ou *backward*). Enquanto o algoritmo *forward* calcula a $P(\mathbf{Y} \mid \lambda)$ utilizando as probabilidades diretas ($\alpha_t(i)$), o algoritmo *backward* calcula a mesma quantidade utilizando as probabilidades reversas ($\beta_t(i)$). Daí a terminologia dos algoritmos.

2.2 Estimação dos parâmetros

Queremos determinar os estimadores de máxima verossimilhança de um modelo de Markov oculto a tempo discreto. Para isso, utilizamos a metodologia do algoritmo EM na obtenção dos estimadores.

O algoritmo EM (do inglês, *expectation-maximization*) é um dos métodos mais populares em problemas de otimização em que se busca um ótimo local e foi proposto por Dempster et al. (1977). Esse algoritmo consiste na construção de uma função auxiliar, conforme a Definição 3.

Definição 3. *A quantidade intermediária do EM é uma família* $\{\mathcal{Q}(\cdot\,;\,\lambda^{'})\}_{\lambda^{'} \in \Theta}$ *de funções em* λ*, indexada por* $\lambda^{'}$ *definida por,*

$$\mathcal{Q}(\lambda^{'}, \lambda) = E\left\{\left[\log P(\mathbf{Y}, \mathbf{X} \mid \lambda^{'})\right] \mid \mathbf{Y}, \lambda\right\}.$$

O algoritmo proposto por Dempster et al. (1977) consiste na construção de uma sequência de estimadores a partir de um valor inicial λ_0. Cada iteração é dividida em dois passos

Passo E: Determinar a função $\mathcal{Q}(\lambda\,;\,\lambda^{'})$;

Passo M: Escolher λ^{i+1} para ser o valor de $\lambda \in \Theta$ que maximiza $\mathcal{Q}(\lambda\,;\,\lambda^i)$.

O algoritmo EM garante dois resultados importantes. Para qualquer sequência $\{\lambda^i\}_{i\geq 0}$ tal que $\mathcal{Q}(\lambda^{i+1}\,;\,\lambda^i) \geq \mathcal{Q}(\lambda^i\,;\,\lambda^i)$, a sequência $\{\log P(\mathbf{Y},\mathbf{X}\,|\,\lambda^i)\}_{i\geq 0}$ é não decrescente, e assim, o EM é um algoritmo de otimização monótono. Segundo, se o processo de iteração parar no ponto λ^*, então $\mathcal{Q}(\lambda\,;\,\lambda^*)$ tem um máximo em λ^* (caso contrário, ainda será possível encontrar um outro λ^*), e então λ^* será o ponto de estacionariedade da função de verossimilhança.

Precisaremos definir algumas ferramentas auxiliares:

Definição 4. *Seja $\lambda = (\delta, \Gamma, \Pi)$ uma cadeia de Markov oculta. Definimos a função de verossimilhança da sequência observada como,*

$$L(\lambda; \mathbf{Y}) = P(\mathbf{Y}\,|\,\lambda) = P(Y_1 = y_1, \ldots, Y_T = y_T\,|\,\lambda).$$

Queremos determinar $\widehat{\lambda} = (\widehat{\delta}, \widehat{\Gamma}, \widehat{\Pi})$ tal que $L(\lambda; \mathbf{Y})$ seja máxima.

Definição 5. *Seja $\mathbf{Y} = (Y_1, \ldots, Y_T)$ a sequência observada do modelo $\lambda = (\delta, \Gamma, \Pi)$. Para todo $i \in E$ definimos*

$$e_t(i) = P(X_t = i\,|\,\mathbf{Y}, \lambda), \qquad 1 \leq t \leq T \tag{6}$$

e

$$a_t(jk) = P(X_t = j, X_{t+1} = k\,|\,\mathbf{Y}, \lambda), \qquad 1 \leq t \leq T-1. \tag{7}$$

A Proposição seguinte nos mostra uma maneira de calcularmos $e_t(i)$ e $a_t(jk)$ por meio das probabilidades diretas e reversas.

Proposição 3. *Considere $e_t(i)$ e $a_t(jk)$ definidos em (6) e (7). Então, para todo $i, j \in E$,*

$$e_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^m \alpha_t(i)\beta_t(i)}, \qquad 1 \leq t \leq T$$

e

$$a_t(jk) = \frac{\alpha_t(j)\gamma_{jk}\pi_{y_{t+1},k}\beta_{t+1}(k)}{\sum_{j=1}^m \sum_{k=1}^m \alpha_t(j)\gamma_{jk}\pi_{y_{t+1},k}\beta_{t+1}(k)}, \qquad 1 \leq t \leq T-1.$$

Teorema 3. *Se $\lambda = (\delta, \Gamma, \Pi)$ é uma cadeia de Markov oculta. Os estimadores de máxima verossimilhança de λ são dados por,*

$$\widehat{\delta}_i = e_1(i), \quad \widehat{\gamma}_{ij} = \frac{\displaystyle\sum_{t=1}^{T-1} a_t(ij)}{\displaystyle\sum_{t=1}^{T-1} e_t(i)} \quad e \quad \widehat{\pi}_{k,j} = \frac{\displaystyle\sum_{\{t; y_t = k\}}^{T} e_t(j)}{\displaystyle\sum_{t=1}^{T} e_t(j)}.$$

3 Aplicação

Para ilustrar o método, faremos uma aplicação usando o conjunto de dados que mencionamos na introdução, com o objetivo de avaliar as estimativas de máxima verossimilhança do modelo.

Na etapa de modelagem, a sequência observada foi obtida através de um estudo de simulação e a cadeia de Markov associada a essa sequência observada é o grupo impacto (presença de embarcação), compondo um processo estocástico duplo $\{X_t, Y_t\}_{t\in\mathbb{N}}$. Verificamos que a cadeia de Markov $\{X_t\}_{t\in\mathbb{N}}$ é irredutível e aperiódica, sendo assim admite uma distribuição estacionária. Além disso, a matriz de transição P converge para uma matriz na qual cada linha é a distribuição estacionária π, que é dada por $\lim_{n\to\infty} P^n = \pi$ (Markov, 1906 in Howard, 1971). Isto significa que se simularmos ou observarmos um passeio aleatório com matriz de transição P, então a probabilidade a longo prazo de que um indivíduo que realiza este passeio esteja em um certo estado é independente do estado em que o passeio inicia, e é definida pela distribuição estacionária.

A matriz de probabilidades de transições constitui uma forma simples e prática de apresentação das Cadeias de Markov. O cálculo das probabilidades de transições baseia-se na frequência de ocorrência de transições de um estágio para outro, observadas na representação do processo considerado, isto é, na matriz de registros ou de frequências de transições. A verificação da propriedade markoviana no processo é feita geralmente por meio da matriz de transições dos parâmetros comportamentais onde os estágios envolvidos no processo correspondem ao comportamento que compõe a sequência estudada. A distribuição inicial e a matriz de probabilidades de transições do processo markoviano são, respectivamente,

$$\delta = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \frac{12039}{14803} & \frac{1794}{14803} & \frac{492}{14803} & \frac{405}{14803} & \frac{73}{14803} \\ \frac{1695}{2903} & \frac{908}{2903} & \frac{194}{2903} & \frac{99}{2903} & \frac{7}{2903} \\ \frac{509}{2173} & \frac{101}{2173} & \frac{1484}{2173} & \frac{75}{2173} & \frac{4}{2173} \\ \frac{393}{1326} & \frac{84}{1326} & \frac{101}{1326} & \frac{746}{1326} & \frac{2}{1326} \\ \frac{80}{453} & \frac{1}{453} & \frac{8}{453} & \frac{2}{453} & \frac{362}{453} \end{bmatrix}.$$

Em relação às distribuições de probabilidades do processo observado dado os estados da cadeia de Markov, no instante de tempo t, supomos ser uma distribuição Poisson, com taxas $\lambda_1 = 0.9940379$, $\lambda_2 = 1.0031604$, $\lambda_3 = 1.2825642$, $\lambda_4 = 0.9810503$ e $\lambda_5 = 0.9340430$, construídas através da distribuição estacionária do processo markoviano. Ou seja,

$$\begin{aligned} \pi_{y_{t,1}} &= P(Y_t = y_t \,|\, X_t = 1) \sim Poisson(0.9940379), \\ \pi_{y_{t,2}} &= P(Y_t = y_t \,|\, X_t = 2) \sim Poisson(1.0031604), \\ \pi_{y_{t,3}} &= P(Y_t = y_t \,|\, X_t = 3) \sim Poisson(1.2825642), \\ \pi_{y_{t,4}} &= P(Y_t = y_t \,|\, X_t = 4) \sim Poisson(0.9810503), \\ \pi_{y_{t,5}} &= P(Y_t = y_t \,|\, X_t = 5) \sim Poisson(0.9340430). \end{aligned}$$

Para estimar os parâmetros, apresentamos um processo de simulação utilizando o pacote `HiddenMarkov` implementado no *software* **R**, que gera um processo de Markov oculto sendo δ, Γ

e Π os parâmetros do modelo. Assim, os algoritmos executados fornecem estimativas para a amostra específica gerada pela função `dthmm`. Listamos, a seguir, os algoritmos utilizados neste procedimento.

```
# Seja lambda=[gama,pi,delta] um Modelo de Markov Oculto
# com Espaçoo de Estados: E.
# Estados da Cadeia: E={1, 2, 3, 3, 4, 5}.
# Matriz de Transição da Cadeia de Markov:

gama.impact=matrix(c(12039/14803,1794/14803,492/14803,405/14803,
73/14803,1695/2903,908/2903,194/2903,99/2903,7/2903,509/2173,
101/2173,1484/2173,75/2173,4/217,393/1326,84/1326,101/1326,746/1326,
2/1326,80/453,1/453,8/453,2/453,362/453), ncol=5,nrow=5,byrow=T)

# Distribuição inicial da Cadeia de Markov

delta.impact=c(1.0,0.0,0.0,0.0,0.0)

# P{obs | estado oculto = 1} ~ Poisson(0.9940379)
# P{obs | estado oculto = 2} ~ Poisson(1.0031604)
# P{obs | estado oculto = 3} ~ Poisson(1.2825642)
# P{obs | estado oculto = 4} ~ Poisson(0.9810503)
# P{obs | estado oculto = 5} ~ Poisson(0.9340430)
```

A função `dthmm` é utilizada para gerar a cadeia de Markov oculta com os valores iniciais δ, Γ e os parâmetros $\lambda_1 = 0.9940379$, $\lambda_2 = 1.0031604$, $\lambda_3 = 1.2825642$, $\lambda_4 = 0.9810503$ e $\lambda_5 = 0.9340430$.

```
x <- dthmm(NULL, gama.impact, delta.impact, "pois", list(lambda=c(0.9940379,
1.0031604, 1.2825642, 0.9810503, 0.9340430)), discrete = TRUE)
x <- simulate(x, nsim=500)
```

Sequência Observada

```
2 1 0 1 1 0 0 1 1 1 3 1 2 1 2 0 1 2 2 1 1 2 1 0 0 1 1 3 1 0 0 0 0 3 2 1 0 3 1 0
1 1 1 1 1 4 0 2 0 1 0 2 0 0 1 0 3 0 0 1 1 1 1 0 3 0 1 0 3 1 3 1 1 2 1 0 0 3 2 0
2 3 1 0 0 1 1 0 0 0 1 0 1 2 1 2 3 1 1 0 1 4 1 2 1 1 0 1 3 0 2 0 2 1 0 0 1 0 1 2
1 1 0 2 0 0 1 2 1 1 1 1 3 1 0 0 0 0 0 1 0 1 0 1 0 2 0 2 0 2 1 1 3 0 1 2 4 1 0 1
1 3 2 0 2 2 0 0 2 0 1 0 0 2 1 2 4 1 3 1 2 0 2 2 0 0 0 3 0 3 0 1 2 0 3 1 2 1 3 0
1 2 0 2 4 1 1 1 0 0 1 1 0 1 0 1 2 1 1 1 3 2 0 1 1 2 2 1 0 5 1 0 2 2 1 2 1 1 3 1
3 0 0 2 2 1 2 1 0 0 0 1 1 3 0 2 1 0 0 3 1 2 0 1 1 1 0 2 0 0 2 1 0 0 1 2 0 0 0 1
0 3 1 2 3 1 0 0 0 0 1 1 2 1 0 0 0 1 0 1 1 1 1 4 2 3 0 0 2 3 0 1 2 2 3 2 0 1 1 1
1 0 1 1 1 2 2 1 0 1 1 0 2 1 2 1 1 1 1 2 2 1 2 0 0 0 0 0 3 1 0 1 0 1 0 3 1 0 0 0
1 0 1 1 0 0 0 0 1 2 0 1 3 2 1 3 2 1 1 0 0 1 1 2 1 0 2 1 0 0 0 0 2 1 0 2 1 0 1 0
1 0 0 2 1 1 2 2 0 1 0 1 1 3 0 1 1 1 1 0 3 2 1 2 1 3 4 0 2 1 1 2 3 1 1 0 0 0 2 1
1 2 1 1 2 3 2 1 1 0 3 0 2 1 0 1 0 1 2 1 2 2 0 0 0 0 3 1 1 1 0 1 1 2 2 4 0 2 1 1
0 0 1 0 0 1 4 2 1 0 1 0 0 2 1 0 0 0 1 2
```

Por fim, utilizamos a função `BaumWelch` para obter as estimativas de máxima verossimilhança do modelo:

```
y <- BaumWelch(x)
$Pi
          [,1]        [,2]       [,3]        [,4]        [,5]
[1,] 0.8104595 0.121861897 0.03525423 0.027651704 0.004772718
[2,] 0.5794153 0.312636933 0.07133772 0.034292995 0.002317089
[3,] 0.2185138 0.043275766 0.68907843 0.032476824 0.016655135
[4,] 0.2927618 0.062661877 0.08148905 0.561646860 0.001440398
[5,] 0.1815126 0.002261272 0.01976586 0.004578427 0.791881876

$delta
[1] 1 0 0 0 0

$distn
[1] "pois"

$pm
$pm$lambda
[1] 1.093374 1.028008 1.086819 1.057946 1.096534
```

Notamos que as estimativas obtidas pelo processo de simulação são próximas dos valores iniciais dos parâmetros como mostram as Tabelas 1 e 2.

Tabela 1: Valores iniciais

$$\delta = [1\ 0\ 0\ 0\ 0]$$

$$\Gamma = \begin{bmatrix} 0.8132811 & 0.121191650 & 0.03323651 & 0.027359319 & 0.004931433 \\ 0.5838787 & 0.312779883 & 0.06682742 & 0.034102652 & 0.002411299 \\ 0.2342384 & 0.046479521 & 0.68292683 & 0.034514496 & 0.018433180 \\ 0.2963801 & 0.063348416 & 0.07616893 & 0.562594268 & 0.001508296 \\ 0.1766004 & 0.002207506 & 0.01766004 & 0.004415011 & 0.799116998 \end{bmatrix}$$

$$\lambda_1 = 0.9839807,\ \lambda_2 = 0.9852900,\ \lambda_3 = 1.1383732,$$
$$\lambda_4 = 1.0057043,\ \lambda_5 = 0.9372803$$

Tabela 2: Estimativas de máxima verossimilhança

$$\widehat{\delta} = [1\ 0\ 0\ 0\ 0]$$

$$\widehat{\Gamma} = \begin{bmatrix} 0.8104595 & 0.121861897 & 0.03525423 & 0.027651704 & 0.004772718 \\ 0.5794153 & 0.312636933 & 0.07133772 & 0.034292995 & 0.002317089 \\ 0.2185138 & 0.043275766 & 0.68907843 & 0.032476824 & 0.016655135 \\ 0.2927618 & 0.062661877 & 0.08148905 & 0.561646860 & 0.001440398 \\ 0.1815126 & 0.002261272 & 0.01976586 & 0.004578427 & 0.791881876 \end{bmatrix}$$

$$\widehat{\lambda}_1 = 1.093374,\ \widehat{\lambda}_2 = 1.028008,\ \widehat{\lambda}_3 = 1.086819,$$
$$\widehat{\lambda}_4 = 1.057946,\ \widehat{\lambda}_5 = 1.096534$$

4 Considerações Finais

Neste trabalho estudamos, sobre o ponto de vista inferencial, os modelos de Markov ocultos a tempo discreto e aplicamos o método a um conjunto de dados relativos ao comportamento de cetáceos. O processo observado foi obtido por simulação e a cadeia de Markov (oculta) é caracterizada pelo grupo impacto (presença de embarcação).

Como a maximização da função de verossimilhança é feita por meio de um algoritmo iterativo e depende de um valor inicial para os parâmetros, caso esses valores iniciais sejam distantes do ponto ótimo o algoritmo pode não convergir ou apresentar estimativas distorcidas.

Referências Bibliográficas

BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*. Vol. 37, p. 1554-1563, 1966.

BICKEL, P. J.; RITOV, Y. Inference in hidden Markov models I: Local asymptotic Normality in the stationary case. *Bernoulli*. Vol. 2, p. 199-228, 1996.

BICKEL, P. J.; RITOV, Y.; RYDEN, T. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics*. Vol. 26, p. 1614-1635, 1998.

CAPPE, O.; MOULINES, E.; RYDEN, T. *Inference in hidden Markov models*. 1. ed. New York: Springer, 2005. 652 p.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*. Vol. 39, p. 1-38, 1977.

FREDKIN, D. R.; RICE, J. A. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. Royal Soc. London*. Vol. 249, p. 125-132. 1992.

HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. Michigan: University of Michigan Press, 1975.

HUANG, S.; AHMADI, M.; SID-AHMED, M. A. A hidden Markov model-based character extraction method. *The Journal of the Pattern Recognition Society*. Vol. 9, p. 2890-2900, 2008.

LEROUX, B. G. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications*. Vol. 40, p. 127-143, 1992.

LEROUX, B. G.; PUTERMAN, M. L. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*. Vol. 48, p. 545-558, 1992.

LINDGREN, G. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics*. Vol. 5, p. 81-91, 1978.

LUSSEAU, D. The effects of tour boats on the behavior of bottlenose dolphins: Using Markov chains to model anthropogenic impacts. *Conservation Biology*. Vol. 17(6): p. 1785-1793, 2003.

MACDONALD, I. L.; ZUCCHINI, W. *Hidden Markov and other models for discrete-valued time series*. 1. ed. Chapman & Hall/crc. 1997. 236 p.

MARKOV, A.A. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. Reprinted in Appendix B of: (Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 2-ya seriya, tom 15, p. 135-156, 1906.), R. Howard. *Dynamic Probabilistic Systems*, volume 1: Markov Chains. John Wiley and Sons, 1971.

NOWACEK, D. P. Sequential foraging behaviour of bottlenose dolphins, Tursiops truncatus, in Sarasota Bay, *FL. Behaviour*. Vol. 139, p. 1125-1145, 2002.

PETRIE, T. Probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*. Vol. 40, p. 97-115, 1969.

RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*. Vol.77, p. 257-284, 1989.

RYDEN, T. Consistent and asymptotically normal parameter estimates for hidden Markov models. *Annals of Statistics*. Vol. 22, p. 1884-1895, 1994.

WILLIAMS, R.; LUSSEAU, D.; HAMMOND, P. S. Estimating relative energetic costs of human disturbance to killer whales (Orcinus orca). *Biological Conservation*. Vol. 133, p. 301-311, 2006.