Modelos Ocultos de Markov

Introdução

Parte II. Modelos Ocultos de Markov

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

Agosto, 2020

Seja

$$\{S_t\} = \{S_t, t = 1, 2, \cdots\} = \{S_1, S_2, \cdots\}$$

denotando a sequência de observações,

$$\{C_t\} = \{C_t, t = 1, 2, \cdots\} = \{C_1, C_2, \cdots\}$$

uma Cadeia de Markov definida no espaço de estados $\{1, 2, \cdots, m\}$,

$$S^{(t)} = \{S_1, S_2, \cdots, S_t\} \qquad \text{e} \qquad C^{(t)} = \{C_1, C_2, \cdots, C_t\}$$

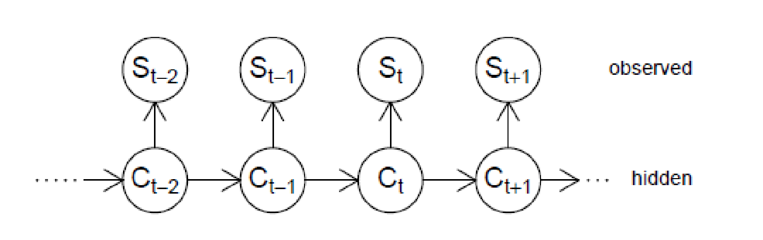
suas histórias até o tempo t, respectivamente. Considere um processo estocástico que consiste em duas partes, primeiro o processo paramêtrico subjacente, mas não observado, $\{C_t\}$, que possui a propriedade Markov

$$P(C_t \,|\, C^{(t-1)}) = P(C_t \,|\, C_{t-1}),$$

e segundo, o processo dependente do estado $\{S_t\}$, para o qual se mantêm

$$P(S_t \,|\, S^{(t-1)}, C^{(t)}) = P(S_t \,|\, C_t)\cdot$$

Então, o processo estocástico $\{S_t\}$ é chamado de Modelo Oculto de Markov de m-estados. A segunda propriedade é chamada independência condicional e significa que, no caso de C_t ser conhecido, S_t depende apenas de C_t e não de estados ou observações anteriores. Aqui e no seguinte, se não indicado de outra forma, utilizamos MacDonald & Zucchini (1997) como uma referência padrão. A estrutura básica do HMM é ilustrada na figura a seguir.



Estrutura baacute;sica de um HMM (Zucchini 38; MacDonald, 2009).

Definição II.1 (Modelo Oculto de Markov)

Um Modelo Oculto de Markov $\{S_t, C_t \,:\, t \in \mathbb{N}\}$ é um tipo particular de mistura dependente, com S_t e C_t representando representando duas séries temporais tais que:

$$P(C_t \mid C^{(t-1)}) \;=\; P(C_t \mid C_{t-1}), \qquad t = 2, 3, \cdots$$

é uma Cadeia de Markocv e

$$P(S_t \mid S^{(t-1)}, C^{(t)}) \;=\; P(S_t \mid C_t), \qquad t \in \mathbb{N}\cdot$$

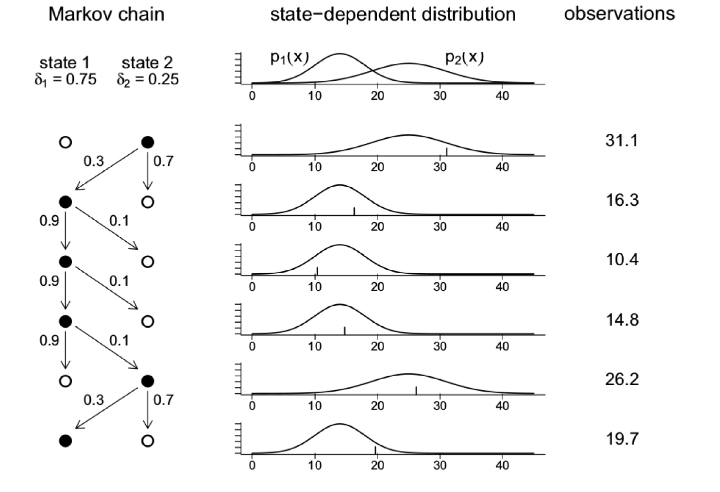
Assim, um Modelo Oculto de Markov ou HMM é uma combinação de dois processos, uma Cadeia de Markov que determina o estado no tempo t, $C_t = c_t$ e um processo dependente do estado que gera a observação $S_t = s_t$ em dependência do estado atual $C_t = c_t$. De fato, para cada estado possível do espaço de estados $\{1, 2, \cdots, m\}$ temos uma distribuição diferente para S_t.

Em geral, assumimos que a Cadeia de Markov é homogênea e irredutível com a matriz de probabilidades de transição Γ. Pela irredutibilidade do $\{C_t\}$, existe uma distribuição estacionária única da Cadeia de Markov δ. Vamos supor que $\{C_t\}$ é estacionária, de modo que δ é para todo t a distribuição de C_t.

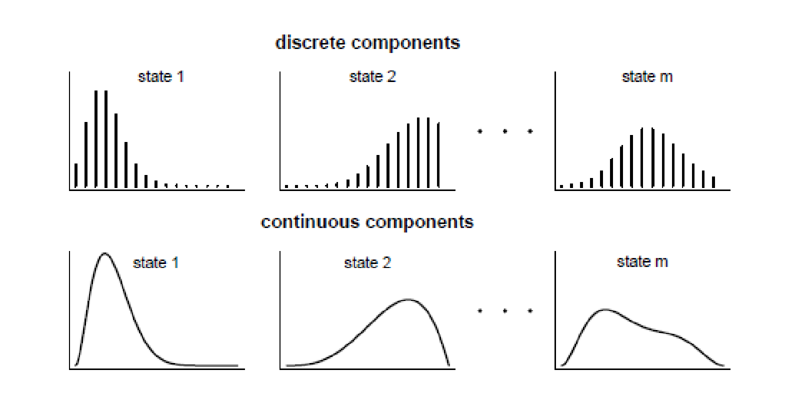
Observe que um HMM é uma construção bastante teórica. Na realidade, apenas o processo dependente de estado $\{S_t\}$ é observado enquanto a Cadeia de Markov subjacente $\{C_t\}$ permanece desconhecida ou oculta. No entanto, em muitas aplicações, há uma interpretação razoável para os estados subjacentes.

Considere, por exemplo, que uma série de retorno diários são modeladas com um HMM de dois estados. Então, os estados da Cadeia de Markov subjacente podem ser interpretados como estados gerais do mercado financeiro, ou seja, um estado com baixa atividade comercial e um estado com alta atividade comercial.

Em contraste com a figura a seguir que mostra a estrutura do processo de um modelo de mistura independente de dois componentes aqui, as probabilidades para o estado C_{t+1} dependem do estado C_t, pois o processo paramétrico é modelado através de uma Cadeia de Markov. No entanto, como no caso de misturas independentes para cada estado, existe uma distribuição diferente, dependente do estado, para a variável aleatória S_t no tempo t, discreta ou contínua, como mostra a figura posterior.



Estrutura do processo de uma distribuição de mistura de dois componentes.
Figura retirada do livro Zucchini MacDonald (2009).



Distribuições componentes de um HMM (Zucchini MacDonald, 2009).

Pode-se pensar em qualquer distribuição possível para modelar as distribuições dependentes do estado. Por exemplo, assumindo distribuições de Poisson como distribuições dependentes do estado, produz um HMM Poisson. Se as distribuições dependentes do estado seguem uma distribuição normal, obtém-se um HMM normal.

Seja C_t o estado da Cadeia de Markov no tempo t, S_t denota a distribuição dependente do estado no tempo t e s é qualquer realização possível de S_t. Então, no caso de componentes discretos, usamos a seguinte notação para a distribuição dependente do estado:

$$\begin{array}{rclcrcl} C_t & = & 1 & \qquad & p_1(s) & = & P(S_t = s \,|\, C_t = 1) \\ C_t & = & 2 & & p_2(s) & = & P(S_t = s \,|\, C_t = 2) \\ & \vdots & & & & \vdots & \\ C_t & = & m & & p_m(s) & = & P(S_t = s \,|\, C_t = m) \cdot \end{array}$$

Portanto, $p_i(s), i \in \{1, 2, , m\}$ representa a distribuição de S_t dado que $C_t = i$. Observe que, em geral, é necessário adicionar outro índice para o tempo t, ou seja, escrever $p_{t,i}(s)$, no entanto, como assumimos que as distribuições dependentes de estado não mudam ao longo do tempo, omitimos o índice de tempo. A notação para o caso contínuo é; semelhante:

$$\begin{array}{rcl@{\qquad}rcl} C_t & = & 1 & f_1(s) & = & f(S_t = s \,|\, C_t = 1) \\ C_t & = & 2 & f_2(s) & = & f(S_t = s \,|\, C_t = 2) \\ & \vdots & & & \vdots & \\ C_t & = & m & f_m(s) & = & f(S_t = s \,|\, C_t = m)\cdot \end{array}$$

Exemplo

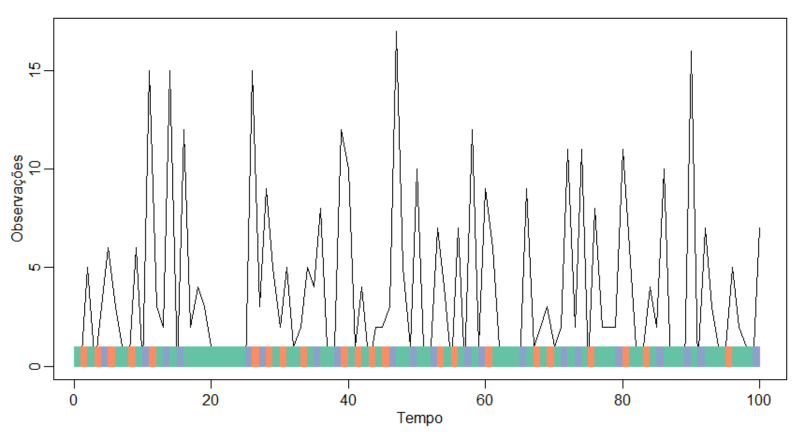
Considere um Modelo Oculto de Markov $\{S_t, C_t \,:\, t \in \mathbb{N}\}$ com distribuições dependentes de estado Poisson de parâmetros 1, 5 e 10, isto é, no estado 1 a observação S_t têm distribuição Poisson(1). Considere também a matriz de probabilidades de transição

$$\Gamma = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 0 & 1/3 \\ 1/2 & 1/2 & 0 \end{pmatrix}\cdot$$

Exemplo

Vamos utilizar o pacote **R mhsmm** escrito e mantido por David Harte. Neste pacote de funções vamos utilizar duas delas: **hmmspec**, a qual cria um objeto tipo **mhsmm** a tempo discreto no qual o processo observado pode ser multivariado e a função **simulate**, a qual permite obter o processo simulado.

```
> library(mhsmm)
> modelo = hmmspec(init = c(1/3, 1/3, 1/3), trans = Gama,
      parms.emis = list(lambda=c(1, 5, 10)), dens.emis = dpois.hsmm)
> modelo
Hidden Markov Model specification:
init:
0.3333333 0.3333333 0.3333333
transition:
          [,1]      [,2]      [,3]
0.3333333 0.3333333 0.3333333
0.6666667 0.0000000 0.3333333
0.5000000 0.5000000 0.0000000
> par(mfrow = c(1,1), mar=c(3,3,1,1), mgp=c(1.6,0.6,0), pch=19)
> plot(treino, xlab="Tempo", ylab="Observaccedil;otilde;es", col = 1:J)
```



A barra horizontal mostra os estados diferentes enquanto a curva mostra os valores simulados da distribuição de emissões. As cores correspondem aos estados da seguinte forma: 1=verde, 2=azul e 3=laranja.

Exemplo (continuação)

```
> treino = simulate(modelo, nsim = 100, seed = 1234, rand.emission = rpois.hsmm)
> str(treino)
List of 3
 $ s: int [1:100] 1 2 1 2 3 2 1 1 2 1 ...
 $ x: int [1:100] 0 5 0 3 6 3 1 0 6 0 ...
 $ N: num 100
- attr(*, "class")= chr "hsmm.data"
> for (i in 1:modelo$J) print(mean(treino$x[treino$s==i]))
[1]  1.157895
[1]  4.545455
[1] 10.90476
```

Utilizando os seguintes comandos do pacote **markovchain** verificamos detalhes da cadeia simulada

```
> library(markovchain) > createSequenceMatrix(treino$s, sanitize=FALSE)
   1  2  3
1 41 27 33
2 35  0 15
3 25 23  0
```

Exemplo (continuação)

```
> markovchainFit(treino$s)
$estimate
MLE Fit
A 3 - dimensional discrete Markov Chain defined by the following states:
  1,  2,  3
The transition matrix (by rows) is defined as follows:
          1         2         3
1 0.4059406 0.2673267 0.3267327
2 0.7000000 0.0000000 0.3000000
3 0.5208333 0.4791667 0.0000000

$standardError
           1          2          3
1 0.06339727 0.05144705 0.05687686
2 0.11832160 0.00000000 0.07745967
3 0.10416667 0.09991316 0.00000000

$confidenceLevel
[1] 0.95
```

Exemplo (continuação)

$lowerEndpointMatrix

	1	2	3
1	0.2816842	0.1664923	0.2152561
2	0.4680939	0.0000000	0.1481818
3	0.3166704	0.2833404	0.0000000

$upperEndpointMatrix

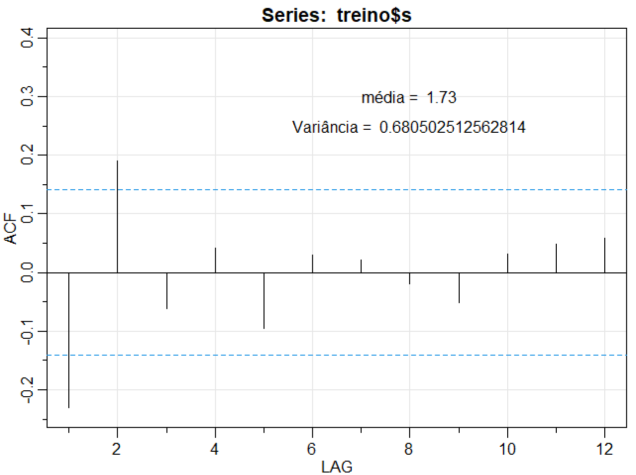
	1	2	3
1	0.5301970	0.3681611	0.4382093
2	0.9319061	0.0000000	0.4518182
3	0.7249963	0.6749929	0.0000000

$logLikelihood
-173.271

Exemplo (continuação)

```
> hittingProbabilities(markovchainFit(treino$s)$estimate)
  1 2 3
1 1 1 1
2 1 1 1
3 1 1 1
> library(astsa)
> acf1(treino$s, max.lag = 12)
 [1] -0.23 0.19 -0.06 0.04 -0.10 0.03 0.02 -0.02 -0.05 0.03 0.05 0.06
> text(8,0.3,paste("média = ",mean(treino$s)))
> text(8,0.25,paste("Variância = ",var(treino$s)))
```

Na figura a continuação mostramos a função de autocorrelação amostral da Cadeia de Markov simulada pelo HMM Poisson.



Função de autocorrelação amostral da cadeia subjacente.

Derivamos as distribuições marginais de um HMM e os respectivos momentos. Falando em distribuições marginais, entende-se o seguinte: Dado um modelo para $C_1, \cdots, C_n, S_1, \cdots, S_n$, qual é a distribuição de S_t ou mesmo a distribuição conjunta de S_t e S_{t+k}? Isto é, estamos interessados na distribuição incondicional de S_t, $P(S_t = s)$, em vez da distribuição condicional $p_i(s) = P(S_t = s \,|\, C_t = i)$.

Teorema II.1

Seja $\{S_t, C_t \,:\, t \in \mathbb{N}\}$ um Modelo Oculto de Markov. Então a distribuição incondicional de S_t é

$$P(S_t = s) = \sum_{i=1}^{m} \delta_i p_i(s),$$

onde δ representa a distribuição estacionária da Cadeia de Markov C_t.

Assim, $P(S_t = s)$ é simplesmente uma combinação linear das respectivas probabilidades das distribuições de componentes, com a distribuição estacionária δ como vetor de pesos.

Usando a notação de matricial, o resultado pode ser reescrito como

$$P(S_t = s) = \begin{pmatrix} \delta_1 & \cdots & \delta_m \end{pmatrix} \begin{pmatrix} p_1(s) & & 0 \\ & \ddots & \\ 0 & & p_m(s) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \delta P(s) \mathbf{1}^\top,$$

O cálculo da distribuição marginal bivariada de S_t e S_{t+k}, $P(S_t = u, S_{t+k} = v)$ é um pouco mais desafiador, mas segue o mesmo procedimento na demonstração do teorema anterior. Novamente, o resultado pode ser reescrito usando a notação matricial:

$$P(S_t = u, S_{t+k} = v) = \delta P(u) \Gamma^k P(v) \mathbf{1}^\top.$$

Teorema II.2

Seja $\{S_t, C_t \ : \ t \in \mathbb{N}\}$ um Modelo Oculto de Markov. Então a distribuição marginal multivariada de S_t é

$$P(S_t = s, S_{t+k} = u, S_{t+k+l} = v) = \delta P(s) \Gamma^k P(u) \Gamma^l P(v) \mathbf{1}^\top.$$

Teorema II.3

Seja $\{S_t, C_t \,:\, t \in \mathbb{N}\}$ um Modelo Oculto de Markov. Então a esperança de uma função g da distribuição conjunta $\mathrm{E}\big(g(S_t, S_{t+k})\big)$ é

$$\mathrm{E}\big(g(S_t, S_{t+k})\big) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{m} \delta_i \gamma_{ij}(k)\mathrm{E}\big(g(S_t, S_{t+k}) \,|\, C_t = i, C_{t+k} = j\big),$$

onde $\gamma_{ij}(k) \;=\; (\Gamma^k)_{ij}$, é a probabilidade de transição k-passos do estado i para o estado j da Cadeia de Markov subjacente, ou seja, o elemento na i-ésima linha e j-ésima coluna da matriz de probabilidade de transição k-passos $\Gamma(k) = \Gamma^k$.

Do ponto de vista teórico, também é possível derivar fórmulas para os momentos centralizados, por exemplo, a variância $\mathrm{Var}(S_t)$, mas caso as distribuições reais dos componentes sejam desconhecidas, as fórmulas podem se tornar desagradáveis. Portanto, gostaríamos apenas de mencionar que, em geral, $\mathrm{Var}(S_t)$ pode ser calculado usando

$$\mathrm{Var}(S_t) \;=\; \mathrm{E}(S_t^2) - \mathrm{E}^2(S_t)\cdot$$

Teorema II.4 (Verossimilhança)

Seja $\{S_t, C_t \,:\, t \in \mathbb{N}\}$ um Modelo Oculto de Markov. A função de verossimilhança assume a forma

$$L_T \;=\; P(\{S_1 = s_1, S_2 = s_2, \cdots, S_T = s_T) = \delta P(s_1)\Gamma P(s_2)\Gamma \cdots \Gamma P(s_T)\mathbf{1}^\top \cdot$$

Esta expressão é composta pela matriz de probabilidades de transição Γ, a probabilidade estacionária δ, estas da Cadeia de Markov e a matriz

$$P(s) \;=\; \begin{pmatrix} p_1(s) & & & 0 \\ 0 & p_2(s) & & 0 \\ & & \ddots & \\ 0 & & & p_m(s) \end{pmatrix},$$

composta das distribuições dependentes dos estados.