Cadeias de Markov Ocultas

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Cadeias de Markov Ocultas (*Hidden Markov Models*, HMM)

Seja uma cadeia de Markov $\{X_t, t = 1, 2, \dots\}$ com espaço de estados finito $\{0, 1, \dots, K\}$ e probabilidades de transição,

$$P_{ij} = P(X_t = j | X_{t-1} = i)$$

e distribuição inicial,

$$P(X_1 = i) = p_i, \;\; i = 0, 1, \dots, K.$$

Seja $\{Y_t, t = 1, 2, \dots\}$ um outro processo estocástico definido pelas distribuições condicionais,

$$p(y_t | x_t, x_{t-1}, \dots, x_1, y_{t-1}, \dots, y_1) \;\; = \;\; p(y_t | x_t)$$

Ou seja, a distribuição de Y_t depende apenas de X_t, $\;t = 1, 2, \dots$

Definição

Um HMM é um processo duplamente estocástico composto por,

- Uma sequência $Y_1, Y_2, \ldots,$ *observável* e,
- uma sequência $X_1, X_2, \ldots,$ *não observável*.

Dizemos que $\{X_t, t \geq 1\}$ é uma *cadeia de Markov oculta*.

Sequencialmente temos as seguintes distribuições de interesse,

$$
\begin{array}{ccccccc}
\longrightarrow & X_{t-1} & \longrightarrow & X_t & \longrightarrow & X_{t+1} & \longrightarrow \\
 & \downarrow & & \downarrow & & \downarrow & \\
 & Y_{t-1} & & Y_t & & Y_{t+1} &
\end{array}
$$

Para uma sequência finita $y_1, \ldots, y_n$ a distribuição conjunta é dada por,

$$
\begin{aligned}
p(x_1, \ldots, x_n, y_1, \ldots, y_n) &= p(x_1, \ldots, x_n)\; p(y_1, \ldots, y_n | x_1, \ldots, x_n) \\
&= \left[p(x_1) \prod_{t=2}^{n} p(x_t | x_{t-1}) \right] \prod_{t=1}^{n} p(y_t | x_t).
\end{aligned}
$$

Exemplo. Uma linha de produção pode estar em bom estado (estado 1) ou em mau estado (estado 2). Seja $\{X_t, t \geq 1\}$ uma cadeia de Markov que representa o estado da linha de produção ao longo do tempo com matriz de transição,

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix}.$$

A cada tempo t um item é produzido e pode ter qualidade aceitável quando,

$$\begin{aligned} X_t &= 1 \text{ com probabilidade } 0.99, \text{ ou} \\ X_t &= 2 \text{ com probabilidade } 0.96. \end{aligned}$$

Neste exemplo note que,

- O estado da linha de produção não é observável.
- A qualidade do item produzido é mensurável (observável).

Seja o processo estocástico $\{Y_t, t \geq 1\}$ indicando se o item é observado como aceitável $(Y_t = 1)$ ou inaceitável $(Y_t = 0)$. Então,

$$\begin{aligned} P(Y_t = 1|X_t = 1) &= 0.99 \\ P(Y_t = 0|X_t = 1) &= 0.01 \\ P(Y_t = 1|X_t = 2) &= 0.96 \\ P(Y_t = 0|X_t = 2) &= 0.04 \end{aligned}$$

Na prática queremos saber o estado da linha de produção mas só conseguimos observar o estado de cada item.

Exemplo. No exemplo anterior, suponha que em um tempo n temos a sequência de itens

$$y_1, \ldots, y_n$$

que foram medidos como aceitáveis ou não aceitáveis.

Com base nesta informação podemos aprender alguma coisa sobre o estado da linha de produção.

Possíveis problemas de interesse:

- *Filtragem*. Qual a probabilidade da linha de produção estar em bom estado no tempo t dada a informação até aquele tempo?

$$P(X_t = 1 | y_1, \ldots, y_t).$$

- *Previsão*. Qual a probabilidade de que linha de produção estará em bom estado em um tempo futuro $n + h$?

$$P(X_{n+h} = 1 | y_1, \ldots, y_n), \; h = 1, 2, \ldots$$

- *Suavização*. Qual a probabilidade de que linha de produção esteve em mau estado em um tempo passado $n - h$ dada toda a informação?

$$P(X_{n-h} = 1 | y_1, \ldots, y_n), \; h = 1, 2, \ldots, n.$$

Em um HMM,

- $\{Y_t, t \geq 1\}$ não é uma cadeia de Markov.
- Em um tempo t, dado X_t os valores futuros,

  $$Y_t, X_{t+1}, Y_{t+1}, \ldots$$

  são independentes do passado,

  $$X_1, Y_1, \ldots, X_{t-1}, Y_{t-1}.$$

Suponha que deseja-se obter a distribuição da cadeia de Markov no tempo t dados os sinais observados até o tempo t.

Ou seja deseja-se calcular,

$$P(X_t = j | y_1, \ldots, y_t), \ t = 1, 2, \ldots$$

ou equivalentemente,

$$\begin{aligned}
&P(X_1 = j | y_1), \\
&P(X_2 = j | y_1, y_2), \\
&P(X_3 = j | y_1, y_2, y_3), \\
&\vdots
\end{aligned}$$

Portanto é um problema de Filtragem.

Note que,

$$\begin{aligned} P(y_1,\ldots,y_t,x_t) &= \sum_i P(y_1,\ldots,y_t,X_{t-1}=i,X_t=j) \\ &= \sum_i P(y_1,\ldots,y_{t-1},X_{t-1}=i,y_t,X_t=j) \end{aligned}$$

$$= \sum_i P(y_t,X_t=j|y_1,\ldots,y_{t-1},X_{t-1}=i)\, P(y_1,\ldots,y_{t-1},X_{t-1}=i).$$

Denotemos a distribuição conjunta de $y_1,\ldots,y_t,x_t$ como,

$$F_t(j) = p(y_1,\ldots,y_t,x_t),$$

Segue então que,

$$
\begin{aligned}
F_t(j) &= \sum_i P(y_t, x_t | y_1, \ldots, y_{t-1}, x_{t-1})\ F_{t-1}(i) \\
&= \sum_i P(y_t | y_1, \ldots, y_{t-1}, X_t = j) P(X_t = j | X_{t-1} = i)\ F_{t-1}(i) \\
&= \sum_i p(y_t | j)\ P_{ij}\ F_{t-1}(i).
\end{aligned}
$$

Portanto, iniciando o processo com,

$$F_1(i) = P(X_1 = i, Y_1 = y_1) = P(X_1 = i)p(y_1|i) = p_i p(y_1|i)$$

a equação anterior pode ser usada recursivamente para calcular $F_2(i), F_3(i), \ldots, F_n(i)$.

A probabilidade desejada então fica,

$$P(X_t = j|y_1, \ldots, y_t) = \frac{F_t(j)}{\sum_i F_t(i)}, \;\; t = 1, 2, \ldots$$

Exemplo. No exemplo anterior suponha que $P(X_1 = 1) = 0.8$ e observou-se as condições dos 3 primeiros itens como,

$$y_1 = 1, y_2 = 0, y_3 = 1.$$

- Calcule $P(X_3 = 1|y_1, y_2, y_3)$.
- Calcule $P(X_4 = 1|y_1, y_2, y_3)$.
- Calcule $P(Y_4 = a|y_1, y_2, y_3)$.

Temos que,

$$
\begin{aligned}
F_1(1) &= P(X_1 = 1)P(Y_1 = 1|X_1 = 1) = 0.8 \times 0.99 = 0.792 \\
F_1(2) &= P(X_1 = 2)P(Y_1 = 1|X_1 = 2) = 0.2 \times 0.96 = 0.192
\end{aligned}
$$

$$
\begin{aligned}
F_2(1) &= P(y_2|1)\,[P_{11}F_1(1) + P_{21}F_1(2)] \\
&= P(Y_2 = 0|X_2 = 1)P_{11}F_1(1) \\
&= 0.01 \times 0.9 \times 0.792 = 0.007128
\end{aligned}
$$

$$
\begin{aligned}
F_2(2) &= P(y_2|2)\,[P_{12}F_1(1) + P_{22}F_1(2)] \\
&= P(Y_2 = 0|X_2 = 2)[P_{12}F_1(1) + P_{22}F_1(2)] \\
&= 0.04[\times 0.1 \times 0.792 + 1 \times 0.192] = 0.010848
\end{aligned}
$$

$\ldots$

A probabilidade desejada fica,

$$P(X_3 = 1|y_1, y_2, y_3) = \frac{F_3(1)}{F_3(1) + F_3(2)}.$$

Ver Exemplo 4.43.

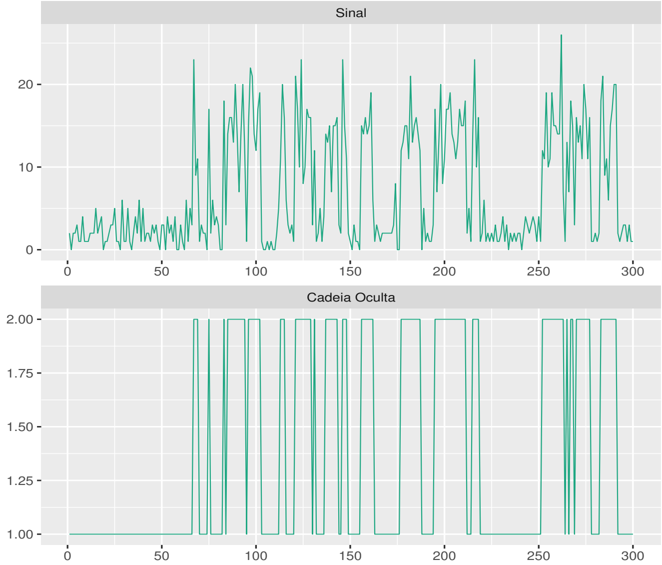
Exemplo. No exemplo anterior suponha que o estado da linha de produção $\{X_t, t \geq 1\}$ tem matriz de transição,

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}.$$

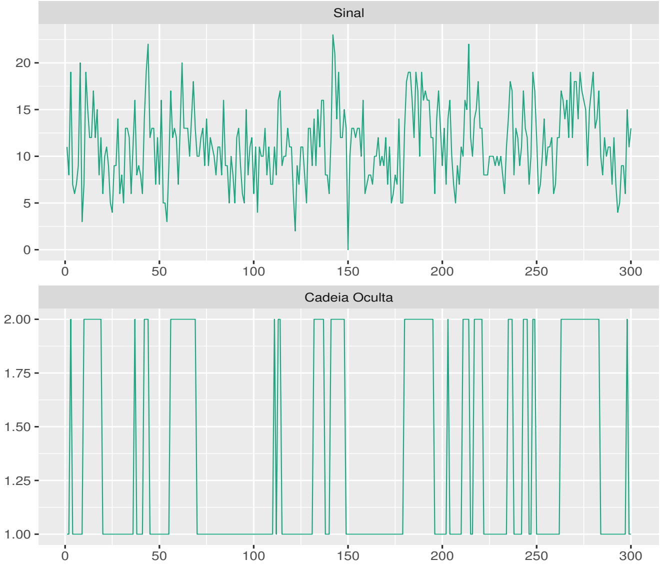
A cada tempo t um item é produzido e pode apresentar um certo número de defeitos segundo uma distribuição de Poisson. Seja $\{Y_t, t \geq 1\}$ tal que,

$$\begin{aligned} Y_t|X_t = 1 &\sim \text{Poisson}(\lambda_1) \\ Y_t|X_t = 2 &\sim \text{Poisson}(\lambda_2). \end{aligned}$$

Realizações simuladas da cadeia oculta e sinais com $X_0 = 1$, $\lambda_1 = 2$ e $\lambda_2 = 15$.



Realizações simuladas da cadeia oculta e sinais com $X_0 = 1$, $\lambda_1 = 10$ e $\lambda_2 = 15$.



- Na primeira figura o processo observado parece estar mesmo seguindo um padrão com 2 regimes: contagens altas e contagens baixas.
- Na segunda figura é mais difícil identificar que pode haver regimes diferentes e quando eles ocorrem.

Comandos do R para simular a cadeia observada e a cadeia oculta.

```
> N= 300
> lambda= c(2,15)
> x= array(1,N+1)
> y= array(0,N+1)
> for (n in 2:(N+1)) {
+     p= ifelse(x[n-1]==1,0.9,0.2)
+     x[n]= sample(c(1,2),size=1,replace=T,prob=c(p,1-p))
+     y[n]= rpois(1,lambda[x[n]])
+ }

> z= cbind(s[2:(N+1)],x[2:(N+1)])
> colnames(z)=c("Sinal","Cadeia Oculta")
> color_scheme_set("viridis")
> mcmc_trace(z,facet_args = list(nrow = 2))
```

Exemplo. No exemplo anterior suponha agora que a linha de produção pode estar em 3 estados diferentes e $\{X_t, t \geq 1\}$ tem matriz de transição,

$$P = \begin{bmatrix} 0.1 & 0.7 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

A cada tempo t um item é produzido e pode apresentar um certo número de defeitos segundo uma distribuição de Poisson. Seja $\{Y_t, t \geq 1\}$ tal que,

$$\begin{aligned} Y_t | X_t = 1 &\sim \text{Poisson}(\lambda_1) \\ Y_t | X_t = 2 &\sim \text{Poisson}(\lambda_2) \\ Y_t | X_t = 3 &\sim \text{Poisson}(\lambda_3). \end{aligned}$$

Vamos usar o pacote `HiddenMarkov` para simular as cadeias sendo $(\lambda_1, \lambda_2, \lambda_3) = (2, 10, 15)$.

```
> library(HiddenMarkov)
> P = matrix(c(0.1,0.7,0.2,
+              0.1,0.7,0.2,
+              0.1,0.1,0.8),nrow=3,ncol=3,byrow=TRUE)
> lambda = c(2,10,15)
> mod= dthmm(NULL, Pi=P, delta = c(1/3,1/3,1/3),
+           distn="pois",pm=list(lambda=lambda))
> data = simulate(mod, nsim=200)
```

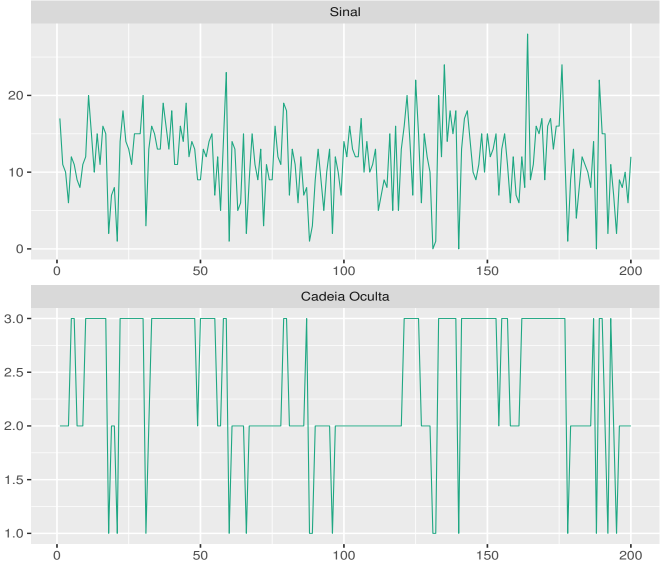
O argumento `delta` contem a distribuição inicial da cadeia oculta, i.e.

$$P(X_1 = 1) = P(X_1 = 2) = P(X_1 = 3) = 1/3.$$

```
> names(data)
```

```
[1] "x"         "Pi"        "delta"     "distn"     "pm"        "pn"
[8] "nonstat"   "y"
```

Realizações simuladas da cadeia oculta e sinais com $X_0 = 1$, $\lambda_1 = 2$, $\lambda_1 = 10$ e $\lambda_2 = 15$.



Resumindo.

- Na prática observa-se somente a sequência de sinais $y_1, y_2, \ldots$
- Supõe-se que esta sequência foi gerada segundo uma cadeia de Markov oculta $\{X_t, t = 1, 2, \ldots\}$ e estamos interessados no processo oculto.
- A distribuição de $Y_t | X_t$ pode ser discreta ou contínua.
- A cadeia oculta pode ter $K \geq 2$ estados.
- Os parâmetros da distribuição de $Y_t | X_t$ podem ser desconhecidos.
- A matriz de transição da cadeia oculta pode ser desconhecida.
- O número de estados K da cadeia oculta pode ser desconhecido.

Probabilidades de transição desconhecidas

Na prática a cadeia de Markov oculta $\{X_t, t \geq 1\}$ pode ter probabilidades de transição P_{ij} desconhecidas.

Uma possivel abordagem consiste em considerar P uma matriz aleatória com as seguintes suposições,

- As linhas da matriz P são vetores aleatórios independentes e,
- cada linha da matriz P é um vetor aleatório com distribuição Dirichlet,

$$(P_{i1}, P_{i2}, \ldots, P_{iK}) \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_K), \; i = 1, 2, \ldots, K.$$

Definição

Um vetor aleatório $\mathbf{X} = (X_1, \ldots, X_K)$ tem distribuição Dirichlet com parâmetros $\alpha_1, \ldots, \alpha_K$, se sua função de densidade de probabilidade conjunta é,

$$f(\mathbf{x}|\alpha_1, \ldots, \alpha_K) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1), \ldots, \Gamma(\alpha_K)} x_1^{\alpha_1 - 1} \ldots x_K^{\alpha_K - 1},$$

sendo,

$$\alpha_1 > 0, \ldots, \alpha_K > 0,$$
$$\alpha_0 = \sum_{j=1}^{K} \alpha_j,$$
$$0 < x_j < 1, \; j = 1, 2, \ldots, K,$$
$$\sum_{j=1}^{K} x_j = 1.$$

No caso geral, a distribuição de $Y_t|X_t$ depende de parâmetros associados ao estado da cadeia oculta, ou seja

$$Y_t|X_t = k \sim F(\boldsymbol{\theta}_k),$$

sendo $F(\cdot)$ uma distribuição discreta ou continua e $\boldsymbol{\theta}_k$ parâmetros desconhecidos.

No exemplo anterior poderiamos escrever,

$$Y_t|X_t = k \sim \text{Poisson}(\lambda_k),$$

sendo $\lambda_1, \lambda_2, \lambda_3$ parâmetros desconhecidos.