

Modelagem e Avaliação de Desempenho

Cadeias de Markov

Prof. Carlos Marcelo Pedroso

2025

Cadeias de Markov: Origem e Princípio Central

- **Pioneirismo de Andrey Markov (1907):** Iniciou o estudo de sequências onde a probabilidade de um evento futuro depende *apenas* do estado **atual**, e não de toda a história passada.
- **Propriedade de Markov (O Princípio "Sem Memória"):**
 - ✓ **Princípio Central :** A probabilidade de transição para o próximo estado ($X_{n+1} = j$) depende **exclusivamente** do estado presente ($X_n = i$), sendo independente de todos os estados anteriores (o *passado*)
 - ✓ Logo: $P(X_{n+1} = j | X_n = i, X_{n-1} = k, \dots) = P(X_{n+1} = j | X_n = i) = p_{ij}$
- **Modelo Discreto:** O processo se move em **passos discretos** (tempo $n = 1, 2, 3, \dots$) entre um conjunto finito ou contável de estados $S = \{S_1, S_2, \dots\}$.

Cadeias de Markov: Hipóteses e Aplicações

➤ Hipóteses de Modelagem:

- ✓ **Propriedade de Markov:** (Conforme slide anterior). Essencial para simplificar o cálculo das probabilidades.
- ✓ **Homogeneidade no Tempo** (*Time-Homogeneous*): As probabilidades de transição p_{ij} são **constantes**, não mudando com o passar do tempo.

➤ Resultados Analíticos Notáveis:

- ✓ **Distribuição Estacionária (W):** Permite calcular a probabilidade de **longo prazo** de o sistema se encontrar em cada estado, útil para análise de equilíbrio.
- ✓ **Matriz Fundamental (N):** Essencial para prever o **número médio de passos** (ou tempo) antes de o sistema atingir um estado final (*absorção*).

➤ Aplicações de Destaque:

- ✓ **Teoria de Filas** (Ex. $M/M/c$): O modelo é o fundamento para a análise de filas com taxas de chegada e serviço **exponenciais** (que possuem a propriedade "sem memória"), como os sistemas $M/M/1$ e $M/M/c$ (múltiplos servidores).
- ✓ **Outras Áreas:** Análise de confiabilidade, Previsão Climática, Finanças e Bioinformática (Modelos Ocultos de Markov - HMM).

Cadeias de Markov: Definição

- Considere um conjunto de estados $S = \{S_1, S_2, \dots, S_n\}$.
- O processo inicia-se em um destes estados e se move sucessivamente de um estado para outro.
- Cada troca é chamada de *passo*.
- Se o estado corrente é S_i , então ela se move para o estado S_j com uma probabilidade denotada por p_{ij} .
- **Importante:** Esta probabilidade p_{ij} não depende dos estados anteriores da cadeia de Markov.
- As probabilidades p_{ij} são chamadas de *probabilidades de transição*.

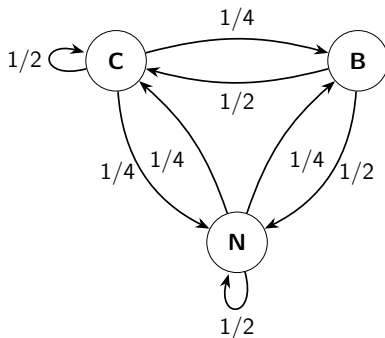
Matriz de Transição P

- P é chamada *matriz de transição* e possui algumas propriedades interessantes.
- As linhas representam a probabilidade de transição de um estado para outro.
- Para calcular a probabilidade da cadeia se encontrar no estado j , mas a n passos adiante, pode-se calcular P^n .

Exemplo: Clima para Astronomia em Curitiba

Cenário

- ▶ Curitiba nunca apresenta dois dias seguidos com bom tempo para observação astronômica.
- ▶ Se há um dia bom (**B**), é mais provável ter chuva (**C**) ou nuvens (**N**) no próximo dia.
- ▶ Se há chuva ou nuvens, existe uma chance de haver tempo bom no próximo dia.



Exemplo: Clima para Astronomia em Curitiba

Matriz de Transição (P):

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} & C & B & N \\ \begin{array}{c} C \\ B \\ N \end{array} & \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

Comandos Máxima

```
P:matrix([1/2,1/4,1/4],[1/2,0,1/2],[1/4,1/4,1/2]);  
Float(P);
```

Exemplo: Potências da Matriz P

Para o caso do exemplo anterior (clima para astronomia em Curitiba), temos:

$$\mathbf{P}^1 = \begin{array}{c} \begin{array}{ccc} & C & B & N \\ \begin{array}{c} C \\ B \\ N \end{array} & \begin{pmatrix} 0.500 & 0.250 & 0.250 \\ 0.500 & 0.000 & 0.500 \\ 0.250 & 0.250 & 0.500 \end{pmatrix} \end{array}$$

$$\mathbf{P}^2 = \begin{array}{c} \begin{array}{ccc} & C & B & N \\ \begin{array}{c} C \\ B \\ N \end{array} & \begin{pmatrix} 0.438 & 0.188 & 0.375 \\ 0.375 & 0.250 & 0.375 \\ 0.375 & 0.250 & 0.438 \end{pmatrix} \end{array}$$

Comandos Máxima

```
float(P^^2);  
float(P^^3);
```

Cálculo da Distribuição de Estados

Para se calcular a probabilidade de se encontrar no estado j dado um vetor de probabilidade inicial $\mathbf{u}$, n passos adiante, pode-se calcular:

$$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n$$

Exemplo (Clima): Suponha que a probabilidade inicial seja $\mathbf{u} = (1/3, 1/3, 1/3)$ e deseja-se a previsão para 3 dias ($n = 3$).

$$\begin{aligned}\mathbf{u}^{(3)} &= \mathbf{u}\mathbf{P}^3 \\ &= (1/3, 1/3, 1/3) \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix} \\ &= (.401, .198, .401)\end{aligned}$$

Comandos Máxima

```
u:matrix([1/3,1/3,1/3]);  
float(u.P^3);
```

Exercício: Universidades Americanas

Cenário

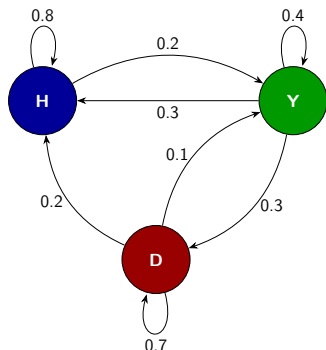
Considere três universidades: **H**arvard, **Y**ale e **D**armouth.

- Filhos de ex-alunos de Harvard: 80% chance de ir para Harvard. Os demais (20%) vão para Yale.
- Filhos de ex-alunos de Yale: 40% chance de ir para Yale. Os demais dividem-se igualmente (30% para Harvard e 30% para Darmouth).
- Filhos de ex-alunos de Darmouth: 70% chance de ir para Darmouth, 20% para Harvard e 10% para Yale.

- 1 Encontre a matriz **P**.
- 2 Encontre a probabilidade de que um *neto* de um ex-aluno de Harvard estude em Darmouth.
- 3 Encontre a probabilidade de que um *bisneto* de um ex-aluno de Darmouth estude em Yale.

Exercício: Matriz P (Sugestão)

Os estados são H, Y, D (em ordem: Harvard, Yale, Darmouth).



$$\mathbf{P} = \begin{matrix} & \begin{matrix} H & Y & D \end{matrix} \\ \begin{matrix} H \\ Y \\ D \end{matrix} & \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.7 \end{pmatrix} \end{matrix}$$

Comandos Máxima

```
P:matrix([0.8,0.2,0],[0.3,0.4,0.3],[0.2,0.1,0.7]);  
float(P^^2) % Netos  
float(P^^3); % Bisnetos
```

Cadeias de Markov Ergódicas e Regulares

- ▶ Uma cadeia de Markov é chamada de *ergódica* se é possível ir de um estado para qualquer outro da cadeia (não necessariamente em um único passo).
- ▶ Uma matriz de transição $\mathbf{P}$ é dita *regular* se alguma potência de $\mathbf{P}$ ($\mathbf{P}^n$) contém somente entradas positivas (não contém nenhuma entrada igual a zero).
- ▶ Uma cadeia de Markov absorvente *não pode* ser regular.

Exemplo de Cadeia Ergódica não Regular:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{matrix}$$

- ▶ É *ergódica*.
- ▶ Não é regular (se n é ímpar, não é possível atingir um dado estado).

Cadeias de Markov Regulares: Estado Estacionário

- Seja $\mathbf{P}$ uma matriz de transição para uma cadeia de Markov *regular*.
- Conforme n tende a infinito ($n \rightarrow \infty$), as potências $\mathbf{P}^n$ se aproximam da *matriz limite* em que todas as linhas são iguais (vetor $\mathbf{W}$).
- O vetor $\mathbf{W}$ é um vetor de probabilidade onde todos os componentes são positivos e sua soma é igual a 1.

Probabilidade de Longo Prazo

$$\mathbf{W} = \lim_{n \rightarrow \infty} \mathbf{P}^n$$

Para uma cadeia ergódica, existe um único vetor $\mathbf{W}$ tal que $\mathbf{W}\mathbf{P} = \mathbf{W}$, com $\mathbf{W}$ positivo.

Exemplo (Clima):

$$\mathbf{P}^\infty = \begin{pmatrix} .4 & .2 & .4 \\ .4 & .2 & .4 \\ .4 & .2 & .4 \end{pmatrix}$$

- A longo prazo, a permanência em cada estado é dada pelo vetor $\mathbf{W}$, independentemente do estado inicial.

Cadeias Regulares: Cálculo do Vetor W

O valor limite $\mathbf{W} = (w_1, w_2, w_3)$ pode ser determinado resolvendo o sistema:

① $\mathbf{WP} = \mathbf{W}$

② $w_1 + w_2 + w_3 = 1$

Sistema de Equações (Clima):

$$(w_1 \quad w_2 \quad w_3) \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} = (w_1 \quad w_2 \quad w_3)$$

Que leva ao sistema linear:

$$w_1 + w_2 + w_3 = 1$$

$$(1/2)w_1 + (1/2)w_2 + (1/4)w_3 = w_1$$

$$(1/4)w_1 + (1/4)w_3 = w_2$$

$$(1/4)w_1 + (1/2)w_2 + (1/2)w_3 = w_3$$

Solução:

$$\mathbf{W} = (.4, .2, .4)$$

Exemplo: modelo de Gilbert-Elliott

- O Modelo de Gilbert-Elliott é uma Cadeia de Markov de Dois Estados que simula o canal de comunicação.
- O objetivo é gerar padrões realistas de perda de pacotes em rajadas (*bursty loss*) em simulações.
- É amplamente usado para modelar a perda em canais com memória, onde a perda de um pacote aumenta a probabilidade de perda do próximo.
- **Alternativa superior** aos modelos de perda independente (probabilidade fixa), que não capturam o comportamento real das redes.

Exemplo: modelo de Gilbert-Elliot

O modelo opera em dois estados mutuamente exclusivos, que representam as condições do canal:

➤ G: Estado Bom (Good)

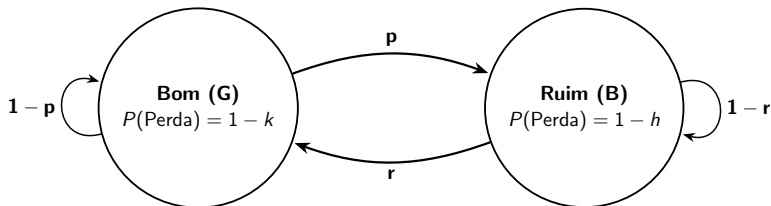
- ✓ Condição de rede **favorável** (baixa interferência, pouco congestionamento).
- ✓ **Perda Baixa:** A probabilidade de um pacote ser perdido é $1 - k$ (onde $k \approx 1$).

➤ B: Estado Ruim (Bad)

- ✓ Condição de rede **desfavorável** (alta interferência, sobrecarga).
- ✓ **Perda Alta:** A probabilidade de um pacote ser perdido é $1 - h$ (onde h é baixo ou $h \approx 0$).

Modelo de Gilbert-Elliot

- O comportamento do modelo é definido por quatro probabilidades:
 - ✓ **p** Transição de **Bom (G)** para **Ruim (B)**.
 - ✓ **r** Transição de **Ruim (B)** para **Bom (G)**.
 - ✓ Probabilidade de Perda no estado G: $1 - k$.
 - ✓ Probabilidade de Perda no estado B: $1 - h$.



- Calcule as probabilidades de estado estacionário.

Exemplo: problema do professor distraído

➤ Problema: Professor Distraído

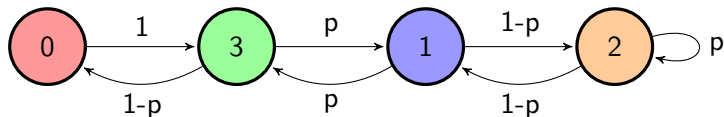
- ✓ Um professor se alterna entre seu gabinete e sua casa.
- ✓ **Regra do Guarda-chuva:** Se estiver chovendo e houver um guarda-chuva disponível (no local de saída), ele o leva para o destino. Se não houver, ele se molha.
- ✓ **p:** probabilidade de chover em Curitiba.
- ✓ **Total de guarda-chuvas:** 3.

➤ Questões:

- 1 Matriz de transição de probabilidade para o sistema e diagrama de transição de estados correspondente.
- 2 Determine a probabilidade do professor se molhar.

Solução para o problema do professor distraído

- A melhor abordagem neste caso é representar o número de guarda chuvas no local onde o professor está.
- A transição de estado representa um dia.



Solução para o problema do professor distraído

Comandos Máxima

```
W:matrix([w0,w1,w2,w3]);  
P:matrix([0,0,0,1],[0,0,1-p,p],[0,1-p,p,0],[1-p,p,0,0]);  
W.P=W;  
# [(1 - p) w3  p w3 + (1 - p) w2  p w2 + (1 - p) w1  p w1 + w0 ] = [ w0  w1  
    w2  w3 ]
```

Sistema de equações

$$(1 - p) w_3 = w_0 \quad (1)$$

$$w_3 + 1 - p w_2 = w_1 \quad (2)$$

$$p w_2 + (1 - p) w_1 = w_2 \quad (3)$$

$$p w_1 + w_0 = w_3 \quad (4)$$

Solução para o problema do professor distraído

Eliminando (4) e em seu lugar usando $w_0 + w_1 + w_2 + w_3 = 1$ podemos resolver:

Comandos Máxima

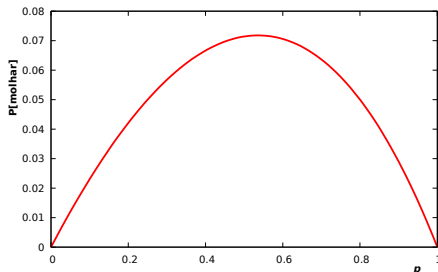
```
solve( [(1-p)*w3-w1, p*w3 + (1 - p)*w2 - w1, p*w2 + (1 - p)*w1 - w2, w0+w1+  
        w2+w3-1 ],  
        [w1,w2,w3,w4]);  
#  
#          p - 1          1          1          1  
#          [w0 = -----, w1 = - ----, w2 = - ----, w3 = - ----]]  
#          p - 4          p - 4          p - 4          p - 4
```

Resposta: $w_0 * p = 0.07059$

Resposta: $P_{\text{molhar}} = p w_0 = p \left[\frac{p-1}{p-4} \right]$

Solução para o problema do professor distraído

- Substituindo $p = 0.6$ obtemos $P_{\text{molhar}} = 0.07059$
- Uma breve análise. Plotando o gráfico de P_{molhar} obtemos



- Paradoxalmente, quanto maior a probabilidade de chover, menor a probabilidade do professor se molhar.
- Por curiosidade, derivando a expressão $p \cdot \frac{p-1}{p-4}$ e igualando a zero, obtemos um ponto de máximo em 0.68

Problema de Estado Estacionário

Suponha que um experimento possui a seguinte matriz de transição $\mathbf{P}$:

$$\mathbf{P} = \begin{pmatrix} .5 & .5 \\ p & 1 - p \end{pmatrix}$$

O valor de p é desconhecido.

- Suponha que o número de repetições realizado foi grande o suficiente para atingir o **estado estacionário** $\pi = (\pi_1, \pi_2)$.
- Foi observado que em 20% das vezes o sistema encontram-se no estado 1 e 80% no estado 2.
- **Tarefa:** Encontre o valor de p .

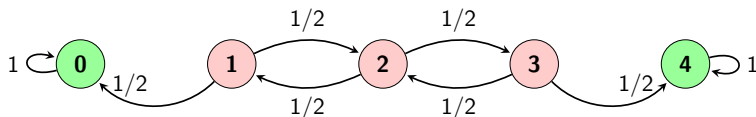
Cadeias de Markov Absorventes

- São cadeias onde existem estados para os quais não é possível realizar a transição para nenhum outro estado.
- Este estado é denominado *estado absorvente*.
- Um estado absorvente apresenta $p_{ii} = 1$.
- Em uma cadeia de Markov absorvente, o número de passos até atingir o estado absorvente é chamado *transiente*.

Exemplo: O Bêbado na Rua

Cenário

- Números 1 a 3: Quarteirões (estados transientes).
- Número 0: Casa (*estado absorvente*, $p_{00} = 1$).
- Número 4: Bar (*estado absorvente*, $p_{44} = 1$).



Matriz P correspondente:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Questões e Matriz Canônica

Questões sobre Cadeias Absorventes:

- Qual a probabilidade de que o processo seja eventualmente absorvido?
- Na média, quantos passos serão dados até que o processo seja absorvido?
- Na média, quantas vezes um dado estado transiente será visitado até que o processo seja absorvido?
- **Matriz P Canônica:** Considere r estados absorventes (ABS) e t estados transientes (TR).

$$\mathbf{P} = \begin{array}{c} \begin{array}{cc} & \begin{array}{cc} TR & ABS \end{array} \\ \begin{array}{c} TR \\ ABS \end{array} & \left(\begin{array}{cc} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{array} \right) \end{array}$$

- **I:** matriz identidade $r \times r$.
- **0:** matriz zero $r \times t$.
- **R:** matriz $t \times r$.
- **Q:** matriz $t \times t$.

Matriz Fundamental N

Definição

Para uma cadeia de Markov absorvente, a *matriz fundamental* **N** é dada por:

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$$

- Um elemento n_{ij} de **N** fornece o *número esperado de vezes* que o processo estará no estado transiente S_j caso o estado inicial seja o estado S_i .

Exemplo (Bêbado): Estados transientes são $\{1, 2, 3\}$.

$$\mathbf{Q} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix} \implies \mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}$$

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{matrix}$$

► Interpretação:

- ✓ Iniciando-se no estado 1, o número médio de vezes que o sistema permanece nos estados 1, 2 e 3 é, respectivamente, 3/2, 1 e 1/2.
- ✓ Iniciando-se no estado 2, o número médio de vezes que o sistema permanece nos estados 1, 2 e 3 é, respectivamente, 1, 2 e 1.
- ✓ Iniciando-se no estado 3, o número médio de vezes que o sistema permanece nos estados 1, 2 e 3 é, respectivamente, 1/2, 1 e 3/2.

Número Médio de Passos para Absorção

- Seja **t** o vetor coluna que armazena o número médio de passos para absorção a partir dos estados transientes.
- Seja **c** um vetor coluna com todos os elementos iguais a 1.

Fórmula

$$\mathbf{t} = \mathbf{N}\mathbf{c}$$

Cálculo para o Exemplo (Bêbado):

$$\mathbf{t} = \mathbf{N}\mathbf{c} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix}$$

Interpretação:

- Do estado 1, a média é de 3 passos para a absorção.
- Do estado 2, a média é de 4 passos para a absorção.
- Do estado 3, a média é de 3 passos para a absorção.

Probabilidade de Absorção

- O analista pode estar interessado em calcular a probabilidade do sistema encerrar em um dos estados absorventes.
- Seja b_{ij} a probabilidade da cadeia ser absorvida por um estado absorvente S_j , caso o estado inicial seja um estado transiente S_i .

Matriz **B** (Probabilidade de Absorção)

A matriz **B** ($t \times r$) será dada por:

$$\mathbf{B} = \mathbf{NR}$$

Exemplo (Bêbado): Estados Absorventes $\{0, 4\}$.

$$\mathbf{R} = \begin{matrix} & \begin{matrix} 0 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix} \end{matrix}$$

Probabilidade de Absorção

$$\mathbf{B} = \mathbf{NR} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix} = \begin{matrix} & \begin{matrix} 0 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{matrix}$$

Comandos Máxima

```
Q:matrix([0,1/2,0],[1/2,0,1/2],[0,1/2,0]);  
I:matrix([1,0,0],[0,1,0],[0,0,1]);  
N:invert(I-Q);  
c:matrix([1],[1],[1]);  
t:N.c;  
R:matrix([1/2,0],[0,0],[0,1/2]);  
B:N.R;
```

- **Interpretação:** Do estado 2, a probabilidade do estado absorvente ser a casa (0) é 1/2 e para o bar (4) é 1/2.

Número Médio de Passos para Primeira Passagem

- É o número médio de passos para ir de um estado para outro.
- Para determinar o número médio de passos para atingir um determinado estado i (o estado alvo), a cadeia original, que é ergódica, deve ser transformada em uma *cadeia absorvente*.
- Isso é feito transformando o estado i em um estado absorvente.
- Em seguida, a teoria de cadeias de Markov absorventes pode ser aplicada (cálculo de $\mathbf{t} = \mathbf{Nc}$).

Número Médio de Passos para Recorrência

Definição

É o número médio de passos para retornar a um determinado estado S_i , dado que o sistema está em S_i .

Fórmula: Dado o vetor $\mathbf{W} = (w_1, w_2, \dots, w_n)$ (probabilidade limite), o número médio de passos para recorrência ($\mathbf{r}_i$) é dado por:

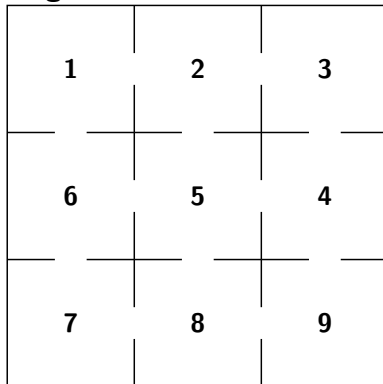
$$\mathbf{r}_i = \frac{1}{w_i}$$

Labirinto: Modelo Inicial - Definição e Diagrama

Cenário 1: Cadeia Regular (Sem Memória)

- O rato se move entre as 9 salas.
- A matriz **P** descreve esta cadeia regular onde todos os estados são transientes e recorrentes.

Diagrama do Labirinto



Labirinto: Matriz P Original (Cadeia Regular)

Matriz de Transição P (Regular)

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \left(\begin{array}{ccccccccc} 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 1/4 & 0 & 1/4 & \mathbf{0} & 1/4 & 0 & 1/4 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \end{array} \right) \end{matrix}$$

Nota: A entrada p_{55} é $\mathbf{0}$, garantindo que o rato continue em movimento.

Labirinto: Definição do Estado Absorvente

- Considere que um Queijo na Sala 5 é agora um **Estado Absorvente**.
- **Estados Transientes (T):** $\{1, 2, 3, 4, 6, 7, 8, 9\}$
- **Estado Absorvente (A):** $\{5\}$
- **Matriz P na Forma Canônica (Reorganizada):**

$$P = \begin{matrix} & \begin{matrix} T & 5 \end{matrix} \\ \begin{matrix} T \\ 5 \end{matrix} & \begin{pmatrix} Q & R \\ 0 & 1 \end{pmatrix} \end{matrix}$$
$$P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 6 & 7 & 8 & 9 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Labirinto: Matriz Fundamental e Visitas Médias

Matriz Fundamental (N)

- Fornece o número médio de vezes que o rato visita o estado j , dado que começou em i , antes de ser absorvido.

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$$

Matriz $\mathbf{I} - \mathbf{Q}$ (8 x 8):

$$\mathbf{I} - \mathbf{Q} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{pmatrix} 1 & -1/2 & 0 & 0 & -1/2 & 0 & 0 & 0 \\ -1/3 & 1 & -1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & 1 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1/3 & 1 & 0 & 0 & 0 & -1/3 \\ -1/3 & 0 & 0 & 0 & 1 & -1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1/3 & 1 & -1/3 \\ 0 & 0 & 0 & -1/2 & 0 & 0 & -1/2 & 1 \end{pmatrix} \end{matrix}$$

Labirinto: Matriz Fundamental e Visitas Médias

Resultado da Matriz N (Inversa Numérica):

$$N = \left(\frac{1}{8} \right) \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{pmatrix} 14 & 9 & 4 & 3 & 9 & 4 & 3 & 2 \\ 6 & 14 & 6 & 4 & 4 & 2 & 2 & 2 \\ 4 & 9 & 14 & 9 & 3 & 2 & 3 & 4 \\ 2 & 4 & 6 & 14 & 2 & 2 & 4 & 6 \\ 6 & 6 & 4 & 2 & 2 & 14 & 6 & 4 & 2 \\ 4 & 3 & 2 & 3 & 9 & 14 & 9 & 4 \\ 2 & 2 & 2 & 4 & 4 & 6 & 14 & 6 \\ 2 & 3 & 4 & 9 & 3 & 4 & 9 & 14 \end{pmatrix} \end{matrix}$$

Iniciando-se no estado 1, o número médio de vezes em que o sistema permanece nos estados 1, 2, 3 , ... etc., será, respectivamente, 14/8, 9/8, 4/8, etc

Labirinto: Número Médio de Passos para Absorção

Vetor de Tempo Médio de Absorção (t)

- Fornece o número médio total de passos até o rato atingir a sala 5.
- $t = N \cdot c$

$$t = \begin{matrix} & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix} & \begin{pmatrix} 6 \\ 5 \\ 6 \\ 5 \\ 5 \\ 6 \\ 5 \\ 6 \end{pmatrix} \end{matrix}$$

- Iniciando-se no estado 1, o sistema leva em média 6 passos para atingir o estado 5 (absorvente). Iniciando-se no estado 2, o sistema leva 5 passos para atingir o estado absorvente e assim por diante.
- Iniciando-se no estado 2, o sistema leva 5 passos para atingir o estado absorvente e assim por diante.

Labirinto: Número médio de passos para recorrência

- No exemplo do labirinto, pode ser calculado $w.P=w$ (acrescentado somatório de $w_i=1$), obtendo-se:

$$W = \left(\frac{1}{12} \quad \frac{1}{8} \quad \frac{1}{12} \quad \frac{1}{8} \quad \frac{1}{6} \quad \frac{1}{8} \quad \frac{1}{12} \quad \frac{1}{8} \quad \frac{1}{12} \right)$$

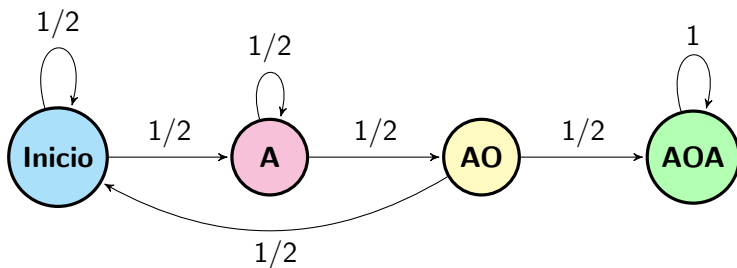
- De onde pode ser deduzido o vetor r (número médio de passos para recorrência):

$$\mathbf{r} = (12 \quad 8 \quad 12 \quad 8 \quad 6 \quad 8 \quad 12 \quad 8 \quad 12)$$

Exemplo: Lançamento de moedas

- Considere o lançamento de uma moeda equilibrada.
- O resultado pode ser cara (A) ou coroa (O), com uma probabilidade de $1/2$.
- Utilizando cadeias de Markov, determine o número médio de lançamentos para obter a sequência AOA.

Solução: Lançamento de moedas



Solução: Lançamento de moedas

Comandos Máxima

```
Q:matrix([0.5,0.5,0],[0,0.5,0.5],[0.5,0,0]);  
I:matrix([1,0,0],[0,1,0],[0,0,1]);  
X:I-Q;  
N:invert(X);  
c:matrix([1],[1],[1]);  
t:N.c;
```

$$\mathbf{t} = \begin{matrix} & \text{Início} & \text{AOA} \\ \begin{matrix} A \\ AO \end{matrix} & \begin{pmatrix} 10.0 \\ 8.0 \\ 6.0 \end{pmatrix} \end{matrix}$$

Serão necessárias 10 jogadas em média.

Cadeias de Markov em Tempo Contínuo

- A variável **tempo** é considerada.
- O tempo de permanência em cada transição é considerado como *exponencialmente distribuído* (hipótese básica para validade).
- O parâmetro que determina a taxa de transição do estado i para o próximo estado j é dado por q_{ij} .

Equações de Estado Estacionário:

$$\mathbf{\Pi} \mathbf{Q} = \mathbf{0}, \quad \sum_{i=1}^n \pi_i = 1$$

- $\mathbf{Q}$ é a *matriz de transição de taxas*.
- $\mathbf{\Pi}$ é o vetor de estado estacionário ($\mathbf{\Pi} = (\pi_1, \dots, \pi_n)$).
- O elemento q_{ii} (diagonal principal) é obtido fazendo-se o complemento do somatório dos demais elementos da linha.

Exemplo: Cluster de Servidores

O Problema do Cluster

Um sistema consiste em **dois servidores idênticos** operando em paralelo (cluster).

- Deseja-se determinar a **probabilidade de falha** (ou indisponibilidade) do sistema.
- A falha ocorre quando **ambos os servidores estão inoperantes**.

Modelo de Confiabilidade

Assumimos que as falhas e os reparos são modelados por taxas de transição:

- **Taxa de Falha** (de um servidor): λ (Exponencial)
- **Taxa de Reparo** (de um servidor): μ (Exponencial)

Exemplo: Cluster de Servidores

Definição dos Estados

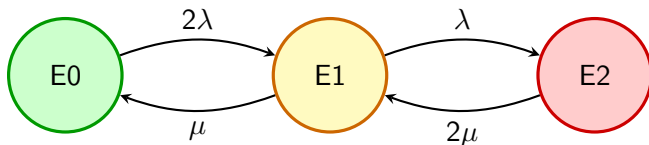
O sistema é modelado por uma Cadeia de Markov discreta, onde o estado n representa o **número de servidores FUNCIONAIS** no sistema.

- **Estado 0** (E_0): Ambos os servidores estão funcionando (Sistema Plenamente Operacional).
- **Estado 1** (E_1): Um servidor está funcionando e o outro está sendo reparado.
- **Estado 2** (E_2): Nenhum servidor está funcionando (Sistema Falho).

Objetivo Final

Nosso objetivo é encontrar a probabilidade estacionária do estado E_0 , ou seja, $P(E_0)$.

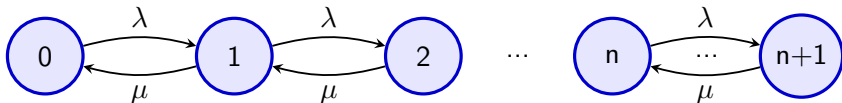
3. Diagrama de Transição de Estados



- O primeiro passo é construir a Matriz de Taxas de Transição ($\mathbf{Q}$).
- Para resolver as probabilidades estacionárias $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$, resolvemos o sistema de equações: $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$
- Resolver...

Exemplo: Fila M/M/1

- **Modelo M/M/1:** Fila com chegadas de Poisson (taxa λ) e tempos de serviço exponenciais (taxa μ).
- **Estados:** O estado n representa o número de clientes no sistema ($n = 0, 1, 2, \dots$).
- **Fator de Utilização:** $\rho = \lambda/\mu$. O sistema é estável se $\rho < 1$.



Matriz de Taxas de Transição (Q)

- A matriz **Q** (Matriz de Taxas de Transição) é infinita, mas o padrão é claro.
- Os elementos da diagonal são as taxas de saída (soma das taxas da linha).

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{matrix} & \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & 0 & \dots \\ 0 & \mu & -(\lambda + \mu) & \lambda & \dots \\ 0 & 0 & \mu & -(\lambda + \mu) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{matrix}$$

Equações de Balanço de Fluxo

A distribuição de probabilidades de estado $\mathbf{w} = [w_0, w_1, w_2, \dots]$ satisfaz a equação de equilíbrio:

$$\mathbf{w}\mathbf{Q} = \mathbf{0} \quad \text{e} \quad \sum_{n=0}^{\infty} w_n = 1$$

Dedução das Probabilidades de Estado Estacionário (w_n)

► A equação $\mathbf{wQ} = \mathbf{0}$ nos fornece as seguintes equações:

① **Estado 0:** $-\lambda w_0 + \mu w_1 = 0 \Rightarrow w_1 = \frac{\lambda}{\mu} w_0 = \rho w_0$

② **Estado 1:** $\lambda w_0 - (\lambda + \mu) w_1 + \mu w_2 = 0$

③ **Estado 2:** $\lambda w_1 - (\lambda + \mu) w_2 + \mu w_3 = 0$

④ **Estado 3:** $\lambda w_2 - (\lambda + \mu) w_3 + \mu w_4 = 0$

⑤ ...

⑥ **Estado n ($n \geq 1$):** $\lambda w_{n-1} - (\lambda + \mu) w_n + \mu w_{n+1} = 0$

Relação Recursiva (por indução)

Usando $w_1 = \rho w_0$ na equação do Estado 1 e simplificando:

$$\rho w_0 - (\rho + 1) w_1 + w_2 = 0 \Rightarrow w_2 = (\rho + 1) \rho w_0 - \rho w_0 = \rho^2 w_0$$

Generalizando:

$$\mathbf{w_n} = \rho^n \mathbf{w_0}$$

Resolvendo

- A soma das probabilidades de estado deve ser igual a 1.

$$\sum_{n=0}^{\infty} w_n = 1$$

Substituindo $w_n = \rho^n w_0$:

$$w_0 + \rho w_0 + \rho^2 w_0 + \cdots = w_0 \sum_{n=0}^{\infty} \rho^n = 1$$

Série Geométrica

Como $\rho < 1$, a soma da série geométrica é:

$$\sum_{n=0}^{\infty} \rho^n = \frac{1}{1 - \rho}$$

Portanto, $w_0 \cdot \frac{1}{1 - \rho} = 1$, o que implica:

$$w_0 = 1 - \rho$$

Esperança do Número de Clientes no Sistema ($E[n]$)

- $E[n]$ é o número médio de clientes no sistema (fila + serviço).

$$E[n] = \sum_{n=0}^{\infty} n \cdot w_n = \sum_{n=0}^{\infty} n \cdot \rho^n (1 - \rho)$$

$$E[n] = (1 - \rho) \cdot \sum_{n=1}^{\infty} n \rho^n \quad (\text{Existe se } \rho < 1)$$

- Resolvendo o somatório (Série Derivada). Sabemos que $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$. Derivando em relação a x e multiplicando por x :

$$\sum_{k=0}^{\infty} k x^k = \frac{x}{(1-x)^2} \quad \text{Fazendo } x = \rho \text{ Obtemos } \sum_{n=1}^{\infty} n \rho^n = \frac{\rho}{(1-\rho)^2}$$

- **Resultado Final:**

$$E[n] = (1 - \rho) \cdot \frac{\rho}{(1 - \rho)^2} \quad \Rightarrow \quad \mathbf{E[n]} = \frac{\rho}{1 - \rho}$$

Número Médio na Fila ($E[q]$) e Lei de Little

- O número de clientes no sistema (N) é a soma do número na fila (N_q) e o número em serviço (N_s).

$$E[n] = E[q] + E[s]$$

- $E[s]$: Esperança do número de clientes em serviço. Em um M/M/1, isso é a probabilidade de o servidor estar ocupado, que é ρ (utilização do sistema). Substituindo:

$$E[q] = E[n] - \rho = \frac{\rho}{1 - \rho} - \rho = \frac{\rho - \rho(1 - \rho)}{1 - \rho} = \frac{\rho - \rho + \rho^2}{1 - \rho}$$

- **Resultado Final:**

$$E[q] = \frac{\rho^2}{1 - \rho}$$

Modelos Ocultos de Markov (HMM)

- É um sistema modelado por uma cadeia de Markov onde os *estados não podem ser observados diretamente*.
- Uma sequência de *eventos observáveis* pode ser utilizada para inferir o estado atual da cadeia de Markov.
- Uma HMM é definida por:
 - 1 Conjunto de estados e probabilidades de transição.
 - 2 Eventos observáveis.
 - 3 Probabilidades de emissão.

Problemas Principais:

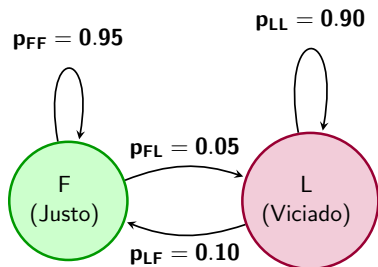
- 1 Determinar o modelo (número de estados, transições, eventos observáveis).
- 2 Realizar o treinamento (encontrar as probabilidades de transição e de emissão). (Algoritmo de *Baum-Welch*)
- 3 Decodificar o estado mais provável a partir de uma sequência de observação. (Algoritmo de *Viterbi*)

HMM: Exemplo do Cassino Desonesto

- **Problema:** Um jogador observa uma sequência de resultados de dados em um cassino e suspeita que o crupiê está, em certos momentos, trocando um *dado justo* por um *dado viciado* (que favorece o número 6).
- **Objetivo do Jogador:** Dada a sequência de lançamentos (os números observados), determinar qual foi a sequência mais provável de dados utilizados (Justo ou Viciado) para gerar esses resultados.
- **Componentes do HMM:**
 - ① **Estados Ocultos (S):** Aquilo que não é diretamente observável.
 - ✓ F : Dado Justo (*Fair*)
 - ✓ L : Dado Viciado (*Loaded*)
 - ② **Observações (O):** Os resultados visíveis do processo.
 - ✓ $O = \{1, 2, 3, 4, 5, 6\}$: Os números na face do dado.

HMM: Exemplo do Cassino Desonesto

1. Diagrama de Transição (Estados Ocultos)



2. Matrizes de Probabilidade

A. Transição (A)

$$P = \begin{matrix} & \begin{matrix} F & L \end{matrix} \\ \begin{matrix} F \\ L \end{matrix} & \begin{pmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{pmatrix} \end{matrix}$$

B. Emissão (B)

Observação	$P(O F)$	$P(O L)$
1 a 5	$1/6 \approx 0.167$	0.10
6	$1/6 \approx 0.167$	0.50

Diagrama do HMM: Cassino Trapaceiro

2. Probabilidades de Emissão (**B**)

➤ Dado Justo (*F*):

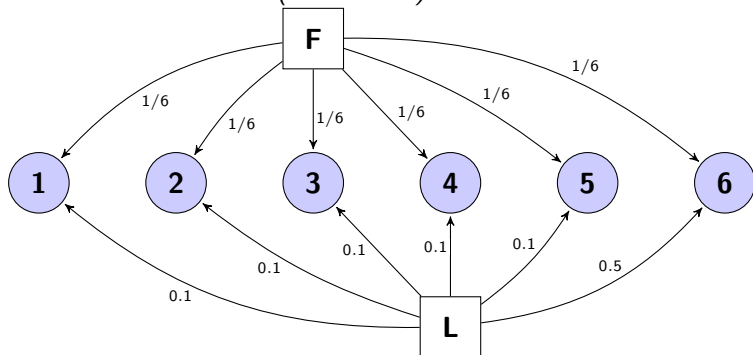
✓ $P(O_i|F) = 1/6 \approx 0.167$ para $i \in \{1, 2, \dots, 6\}$

➤ Dado Viciado (*L*): (Bias para 6)

✓ $P(O_6|L) = 0.50$

✓ $P(O_{1-5}|L) = 0.10$

Estrutura de Emissão (Conceitual)



Matrizes de Transição (A) e Emissão (B)

Definição do HMM no R (Exemplo: Pacote HMM)

```
# Estados: 1=Justo, 2=Viciado. Observacoes: 1 a 6.

# 1. Matriz de Transicao A
A <- matrix(c(
  0.95, 0.05, # Linha 1: Transicao de Justo (F)
  0.10, 0.90 # Linha 2: Transicao de Viciado (L)
), nrow=2, byrow=TRUE)

# 2. Matriz de Emissao B (Colunas: 1, 2, 3, 4, 5, 6)
B <- matrix(c(
  1/6, 1/6, 1/6, 1/6, 1/6, 1/6, # Emissao do Justo (F)
  0.1, 0.1, 0.1, 0.1, 0.1, 0.5 # Emissao do Viciado (L: bias para 6)
), nrow=2, byrow=TRUE)

# Probabilidade Inicial Pi (Assume 50% de chance de comecar em qualquer estado)
Pi <- c(0.5, 0.5)

# Criar o objeto HMM (Funcao Conceitual)
library(HMM)
hmmDados <- initHMM(
  States=c("F","L"), Symbols=1:6,
  transProbs=A, emissionProbs=B, startProbs=Pi
)
```

O Algoritmo de Viterbi

- Problema de Decodificação: Dado um modelo HMM $\lambda = (A, B, \pi)$ e uma sequência de observações O , encontrar a sequência de estados ocultos $S = (s_1, s_2, \dots, s_T)$ que **maximiza a probabilidade** $P(S|O, \lambda)$.
- O Viterbi resolve este problema de forma eficiente usando Programação Dinâmica.
- Ele calcula a probabilidade máxima de um caminho terminar em cada estado para cada passo no tempo.

Aplicação

- Dada uma longa sequência de lançamentos de dados, o Viterbi pode determinar, para cada lançamento, se o dado mais provável de ter sido usado era o justo ou o viciado.

Viterbi no R

1 Sequência de Observações (Exemplo)

- ✓ Sequência de 20 lançamentos:
(6, 6, 6, 5, 6, 1, 2, 3, 4, 5, 6, 6, 6, 6, 6, 6, 1, 2, 3, 4)

2 Comandos no R

```
# Sequencia de Observacoes (O)
observacoes <- c(6, 6, 6, 5, 6, 1, 2, 3, 4, 5, 6, 6, 6, 6, 6, 6, 1, 2, 3, 4)

# Rodar o algoritmo de Viterbi
# Funcao viterbi esta disponivel no pacote HMM
viterbiResultado <- viterbi(hmmDados, observacoes)

# Imprimir a sequencia de estados mais provavel
print(viterbiResultado)
```

3 Saída Esperada:

- ✓ O resultado seria uma sequência de estados ('L' ou 'F') que corresponde à sequência de observações.
- ✓ Picos de '6' consecutivos tendem a ser decodificados como estado S_L (Viciado).

Algoritmo de Baum-Welch: O Treinamento do HMM

- **Propósito:** O Baum-Welch é usado para **treinar** ou *aprender* os parâmetros desconhecidos de um HMM a partir de uma longa sequência de observações.
- **O Que Aprender?** O algoritmo ajusta as duas matrizes do modelo:
 - ✓ Matriz de **Transição (A)**: A probabilidade de mudar de um estado oculto para outro ($P(S_{t+1} | S_t)$).
 - ✓ Matriz de **Emissão (B)**: A probabilidade de gerar uma observação em um dado estado oculto ($P(O_t | S_t)$).
- **O Desafio Central:**

O Oculto no HMM

Não sabemos qual foi a sequência real de estados ocultos (**S**) que gerou as observações. O Baum-Welch resolve isso estimando a **probabilidade de todos os caminhos possíveis** e ajustando os parâmetros para o que for mais provável.

O Algoritmo de Baum-Welch (Treinamento Não Supervisionado)

- Problema de Treinamento: Dado um conjunto de observações O , ajustar os parâmetros do modelo $\lambda = (A, B, \pi)$ para **maximizar a probabilidade** $P(O|\lambda)$.
- É um algoritmo iterativo baseado no método EM (Expectation-Maximization).
- Objetivo: Estimar as matrizes de transição (A) e emissão (B) sem saber qual foi o dado usado em cada lançamento (dados não rotulados/não supervisionados).

Cenário de Teste: Novo Vício

- Geramos uma nova sequência de lançamentos (observações O^{novo}) com um dado viciado diferente: vício para o número 1.
- Vício Novo (S_L^{novo}): $P(O_1|S_L^{novo}) = 0.5$ e $P(O_i|S_L^{novo}) = 0.1$ para $i = 2..6$.
- O Baum-Welch deve aprender este novo padrão a partir dos dados.

Treinamento Não Supervisionado

- Novo Modelo Viciado para O_1
- Simulação de Novo Modelo Viciado (Bias para 1):

```
Anovo <- matrix(c(  
  0.90, 0.1, # Linha 1: Transicao de Justo (F)  
  0.30, 0.70 # Linha 2: Transicao de Viciado (L)  
) , nrow=2, byrow=TRUE)  
  
# Definir a nova matriz B (bias para 1, linha 2)  
Bnovo <- matrix(c(  
  1/6, 1/6, 1/6, 1/6, 1/6, 1/6, # Dado Justo (F)  
  0.5, 0.1, 0.1, 0.1, 0.1, 0.1 # Dado Viciado (L)  
) , nrow=2, byrow=TRUE)  
  
hmmNovo <- initHMM(  
  States=c("F","L"), Symbols=1:6,  
  transProbs=Anovo, emissionProbs=Bnovo, startProbs=Pi  
)  
  
# 2. Gerar uma longa sequencia de observacoes O_novo  
# (Assumimos que o modelo original, desconhecido, eh o hmm_model_novo)  
simulacao <- simHMM(hmmNovo, length=1000)  
observacoesNovo <- simulacao$observation  
estadosOcultosNovo <- simulacao$states
```

- Estamos aqui produzindo um conjunto de dados simulados...

Baum-Welch no R (Treinamento)

- **Treinamento Não Supervisionado:** Partimos de um modelo inicial com parâmetros aleatórios. O algoritmo iterativamente ajustará A e B até a convergência.

```
# Partir de um modelo inicial (parametros uniformes)
A <- matrix(c(
  0.3, 0.7, # Linha 1: Transicao de Justo (F)
  0.8, 0.2 # Linha 2: Transicao de Viciado (L)
), nrow=2, byrow=TRUE)

B <- matrix(c(
  1/10, 1/10, 1/10, 1/10, 1/10, 1/2, # Dado Justo (F)
  1/6, 1/6, 1/6, 1/6, 1/6, 1/6 # Dado Viciado (L)
), nrow=2, byrow=TRUE)

hmmInicial <- initHMM(
  States=c("F", "L"), Symbols=1:6,
  transProbs=A, emissionProbs=B, startProbs=Pi
)

# Rodar o algoritmo Baum-Welch
hmmTreinado <- baumWelch(
  hmmInicial, observacoesNovo,
  maxIterations=100)

# 4. Analise dos Resultados
print("Matriz de Emissao B Estimada:")
print(hmmTreinado$emissionProbs)
```

- **Resultado Esperado:** O Baum-Welch descobriu o novo vício e as probabilidades de transição sem nunca ter sido informado sobre ele.

Parâmetros de Controle do Baum-Welch

Melhorando a Convergência e a Generalização

O processo de treinamento é ajustado por três parâmetros chave, que controlam a otimização e a estabilidade:

❶ Número Máximo de Iterações (`max_iter`)

- ✓ Limita o número de passos E (Expectation) e M (Maximization).
- ✓ Necessário para evitar que o algoritmo rode indefinidamente, já que a convergência é assintótica.

❷ Critério de Parada (`delta`)

- ✓ Define a tolerância para a convergência.
- ✓ O treinamento para quando a mudança na Log-Likelihood for menor que δ : $|\mathcal{L}_{t+1} - \mathcal{L}_t| < \delta$.

❸ Pseudocontagem (`pseudocount`)

- ✓ Aplica **suavização** (smoothing) nos cálculos de probabilidade.
- ✓ Evita o problema de **probabilidades zero** (overfitting), garantindo que todos os estados possam ser visitados e todas as observações emitidas.
- ✓ Essencial para a estabilidade e generalização do modelo.

Critério de Avaliação: Log-Likelihood

A Métrica de Treinamento

- **Definição:** A Log-Likelihood é o logaritmo da probabilidade da sequência de observações (O) ter sido gerada pelo modelo (λ).

$$\mathcal{L} = \log P(O|\lambda)$$

- **Interpretação:** É uma medida de **ajuste** (fit). Quanto maior o $\mathcal{L}$ (ou seja, menos negativo, mais próximo de zero), melhor o modelo explica os dados de treinamento.
- **Uso no Treinamento:**
 - ✓ Serve como o objetivo do algoritmo Baum-Welch (EM).
 - ✓ É usada para selecionar a melhor parametrização para o modelo.

O Treinamento de HMMs: O Máximo Local

Justificativa para Múltiplas Soluções Iniciais

- O algoritmo Baum-Welch (EM) é iterativo e baseado em gradientes implícitos.
- Ele pode convergir para um **máximo local** da função de Log-Verossimilhança.
- Não há garantia de que este seja o **máximo global** (o melhor modelo possível).

Estratégia Robusta

Uso de Múltiplas Soluções Iniciais:

- Iniciar o treinamento com λ_0 gerados aleatoriamente (Transição A e Emissão B).
- Repetir o treinamento N vezes e selecionar o HMM que resultar na Log-Likelihood mais alta ($\max(\text{LogLikelihood})$).

Ver:

<https://www.eletrica.ufpr.br/pedroso/2025/EELT7055/hmm/hmm/>

Critério de Avaliação: AIC

Selecionando o Número Ideal de Estados

A Log-Likelihood pura favorece modelos complexos (com mais estados), pois eles sempre se ajustam melhor aos dados de treinamento (overfitting). O Critério de Informação de Akaike (AIC) resolve este problema, penalizando a complexidade.

Critério de Informação de Akaike (AIC)

$$AIC = 2k - 2 \cdot \mathcal{L}$$

Onde:

- k é o **número de parâmetros** livres no HMM (penalidade de complexidade).
- $\mathcal{L}$ é o Log-Likelihood (medida de ajuste).
- **O melhor modelo é aquele que minimiza o AIC.**
 - ✓ O HMM que atinge o melhor equilíbrio entre **ajuste aos dados** ($\mathcal{L}$ alto) e **simplicidade** (k baixo).
 - ✓ Usado para comparar modelos HMM com diferentes **números de estados** (N).

HMM: Inferindo o Clima de dentro de um Laboratório sem Janelas

- **Cenário:** Um Engenheiro Eletricista está há dois dias em um laboratório sem janelas (no estilo de Curitiba, onde o clima muda rapidamente) montando suas atividades.
- **O Problema:** O estudante não consegue observar o **estado verdadeiro** do ambiente externo (Frio, Chuva ou Calor).
- **A Solução HMM:** Utilizar as observações dos demais estudantes que entram no laboratório para inferir a sequência de estados climáticos mais provável ao longo do tempo.
- **Objetivo:** Determinar a sequência de clima mais provável (S^*) que gerou a sequência de vestimentas observada (O), utilizando o **Algoritmo de Viterbi**.

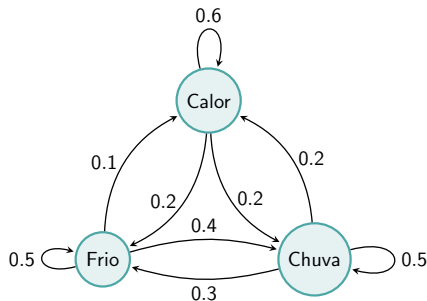
HMM: Estados e Observações do Clima Oculto

- ❶ **Estados Ocultos (S):** Representam a variável que *não pode ser observada* diretamente pelo estudante. São as únicas condições que afetam as observações.
 - ✓ F: **Frio** (Tempo frio, mas seco)
 - ✓ C: **Chuva** (Tempo chuvoso, com ou sem frio)
 - ✓ K: **Calor** (Tempo quente e seco)

- ❷ **Observações (O):** Representam o que o estudante *pode ver* nos colegas que entram no laboratório.
 - ✓ GC: **Guarda-chuva/Capa** (Indicativo forte de chuva)
 - ✓ CSG: **Casaco/Moletom** (Indicativo de frio, mas sem chuva)
 - ✓ CB: **Camiseta/Bermuda** (Indicativo de calor)

HMM: Transições e Probabilidades (Modelo Ilustrativo)

A. Diagrama de Transição (A)



B. Probabilidade de Emissão (B)

$P(O S)$	Frio (F)	Chuva (C)	Calor (K)
GC (Guarda-chuva)	0.05	0.70	0.05
CSG (Casaco/s. Chuva)	0.70	0.20	0.15
CB (Camiseta/Bermuda)	0.25	0.10	0.80

Interpretação: Se o estado oculto for CHUVA (C), a chance de ver alguém com Guarda-chuva (GC) é de 70%.

Decodificação com o Algoritmo de Viterbi

- **Sequência de Observação (O):** O estudante registra a vestimenta de 5 colegas que entraram.

$$O = \{\text{CSG}, \text{GC}, \text{GC}, \text{CB}, \text{CSG}\}$$

- **Caminho de Estados (S):** O Algoritmo de Viterbi calcula o caminho ótimo (a sequência de clima mais provável).

Resultado do Viterbi (Exemplificado)

Passo (t)	1	2	3	4	5
Observação (o_t)	CSG	GC	GC	CB	CSG
Estado Mais Provável (s_t^*)	Frio	Chuva	Chuva	Calor	Frio

- **Inferência:** O estudante conclui que o clima começou frio (CSG alta chance em Frio), mudou para chuva por dois períodos (GC alta chance em Chuva), aqueceu rapidamente (CB alta chance em Calor) e voltou a esfriar (CSG).

Problemas para resolver na aula

Marcas de chope no bar 1

- Um bar oferece quatro tipos de chope (draft beer) em seu menu, que variam em qualidade e preço:
 - ① **Premium** (S_P): O chope mais caro e de melhor qualidade.
 - ② **Artesanal** (S_A): Chope de alta qualidade, preço intermediário.
 - ③ **Lager** (S_L): O chope padrão, preço acessível.
 - ④ **Econômico** (S_E): O chope mais barato e de menor qualidade (o *pior*).
- O padrão de escolha de um cliente em uma visita é modelado como uma Cadeia de Markov, onde o estado atual é o tipo de chope consumido na visita anterior.
- Os estados da cadeia de Markov são os quatro tipos de chope:

$$\text{Estados } S = \{S_P, S_A, S_L, S_E\}$$

Problemas para resolver na aula

Marcas de chope no bar 1

- O gerente (que tem uma boa formação de fundamentos de cálculo) observou o seguinte comportamento dos clientes para o próximo chope solicitado:
 - ✓ **Premium (S_P):** 95% sempre escolhe esta categoria e 5% escolhe Artesanal no próximo.
 - ✓ **Artesanal (S_A):** 10% escolhe o prêmio, 70% Artesanal, 20% Lager no próximo.
 - ✓ **Lager (S_L):** 20% escolhe o Artesanal, 50% escolhe o Lager e 30% escolhe o Econômico.
 - ✓ **Econômico (S_E):** 15% escolhe o Artesanal, 60% escolhe o Lager e 25% escolhe o Econômico.
- Monte o diagrama de transição de estados e a matriz P

Problemas para resolver na aula

Marcas de chope no bar 1

- Responda as seguintes questões:
- 1 Determine as probabilidades de estado estacionário.
 - 2 Interprete o significado dessas probabilidades para o bar, indicando qual chope dominará as vendas. Se o bar tem 1000 clientes regulares, quantos podem ser esperados, em média, para comprar cada tipo a longo prazo?
 - 3 Considerando um cliente que inicia tomando o chope **Premium** (S_P). Qual o número de médio de chopes que ele precisa tomar para acabar experimentando o chope **Econômico** (S_E)?
 - 4 Considerando um cliente que inicia tomando o chope **Econômico** (S_E). Qual o número de médio de chopes que ele precisa tomar para acabar experimentando o chope **Premium** (S_P)?

Problemas para resolver na aula

Marcas de chope no bar 2

- O gerente de outro bar adotou uma abordagem diversa. Ao invés de registrar a preferência de um pedido para outro, ele agora monitora a **taxa** com que os clientes trocam de um tipo de chope para outro (tempo contínuo).
 - ✓ O gerente fez isso anotando o tempo que o cliente permanece pedindo cada tipo de chope (ele sabe que a taxa é o inverso deste tempo):
 - ✓ **De Premium:** A taxa de saída para Artesanal é 0.05 e para Lager é 0.05. A transição para o Econômico é nula.
 - ✓ **De Artesanal :** A taxa de saída para Premium é de 0.05, em direção à Lager é de 0.2.
 - ✓ **De Lager:** A taxa de saída para Premium é de 0.05, em direção à Artesanal é de 0.2 e em direção à Econômico é 0.05.
 - ✓ **De Econômico:** A taxa de saída para Artesanal é de 0.05, para Lager é de 0.3.
 - ✓ Responda:
 - 1 Determine as probabilidades de estado estacionário.
 - 2 Interprete o significado dessas probabilidades para o bar, indicando qual chope dominará as vendas. Se o bar tem 1000 clientes regulares, quantos podem ser esperados, em média, para comprar cada tipo a longo prazo?

Problemas para resolver na aula

Marcas de chope no bar 3

- Em um terceiro bar, o gerente deseja estabelecer um modelo de forma mais automática, sem que seja necessário fazer probabilidades ou anotar tempos.
- Responda:
 - ① O modelo oculto de Markov pode ser utilizado neste caso?
 - ② Existem alguma particularidades neste caso, os dados estarão em sequências - descreva o problema, onde este problema difere de uma HMM *tradicional*.
 - ✓ Helske S. and Helske J. (2019). Mixture Hidden Markov Models for Sequence Data: The seqHMM Package in R, Journal of Statistical Software, 88(3), 1-32. doi:10.18637/jss.v088.i03
 - ③ Esboce o modelo, com os estados ocultos e eventos observáveis.
 - ④ Estabeleça o que os funcionários devem anotar para que seja possível treinar o modelo.
 - ⑤ Uma vez treinado o modelo, o que será possível deduzir?