

Processos Estocásticos

Adriana Eva Fernandes da Silva

Sumário

1	Parte I	2
1.1	Introdução.	2
1.2	Objetivo	3
1.3	Descrição da cadeia de markov em relação ao avanço dos estágios da doença Crônica de Alzheimer, obtenção da matriz de transição da Cadeia e diagrama de transição.	3
1.4	Obtenção da classe de estados, classificação dos estados, distribuição limite distribuição estacionária da cadeia de Markov	5
1.5	Obtenção da distribuição estacionária e interpretação dos resultados . . .	5
1.6	Algoritmo para simulação da cadeia de markov em estudo com o objetivo de simular uma amostra de tamanho n da cadeia.	6
1.7	Simulação de uma amostra de tamanho n da cadeia de markov em estudo para diferentes valores iniciais da cadeia (α) utilizando o R-Studio.	6
1.8	Estimação da distribuição estacionária da cadeia e comparação com o resultado teórico obtido.	8
1.9	Estimação e análise dos tempos de retorno a cada estado da cadeia.	9
1.10	Estimação da probabilidade de transição (\hat{P}_{ab}) para diferentes tamanhos amostrais.	10
2	Parte II	11
2.1	Dados	11
2.2	Análise Descritiva dos dados	12
3	Conclusão:	21

Capítulo 1

Parte I

1.1 Introdução.

A doença de Alzheimer é a forma de demência mais comum no mundo, e se caracteriza pelo seu aparecimento em etapas de acordo com o estágio da doença. Representa a causa de 60-70% dos casos de demência, cuja causa não é totalmente compreendida.

Esse distúrbio é definido como uma doença neurodegenerativa crônica, que se manifesta lentamente e piora com o tempo. O sintoma inicial mais comum é a perda de memória de curto prazo, tornando difícil lembrar de eventos recentes. Conforme a doença progride, os sintomas pioram e podem levar à morte.

Porém, muitas vezes é impossível identificar corretamente os sintomas iniciais da doença, o que pode levar a atrasos no diagnóstico e na orientação adequada ao paciente e seus familiares.

Portanto, compreender a condição da doença de Alzheimer pode ajudar as famílias a detectar os sinais da doença em um estágio inicial e garantir o acompanhamento adequado das pessoas afetadas. A doença se desenvolverá em um tempo incerto e pode permanecer assintomática por muitos anos.

Assim, teoricamente, a evolução da doença pode ser dividida em quatro etapas:

1. **Pré-demência:** Nessa fase, os sintomas são muito sutis e muitas vezes atribuídos, de forma equivocada, ao envelhecimento natural ou ao estresse. Nesse momento, quando se suspeita de doença de alzheimer, pode-se aplicar testes neuropsicológicos que permitem revelar a doença até oito anos antes da pessoa cumprir os critérios para diagnóstico de Alzheimer. Além da perda recente de memória, esse estágio também pode incluir sintomas como mudanças sutis na atenção, apatia, irritabilidade e depressão.
2. **Estágio inicial:** No estágio leve, para muitas pessoas consideradas o primeiro estágio, os sintomas antes da demência pioram e os pacientes podem ter dificuldade em encontrar palavras, perder o caminho no tempo e no espaço e ter dificuldade para tomar decisões.

3. **Estágio intermediário:** No estágio intermediário, são comuns dificuldades mais evidentes no dia-a-dia, com esquecimento de fatos mais marcantes, como o nome de pessoas próximas; dificuldade com higiene pessoal e autocuidados; incapacidade de cozinhar e cuidar da casa; maior dificuldade para falar e se expressar com clareza; alucinações e alterações de comportamento.
4. **Estágio final:** O estágio avançado, ou estágio terminal, cursa com prejuízo gravíssimo da memória, com incapacidade de registrar dados e muita dificuldade em recuperar informações antigas, como o reconhecimento de pessoas e locais familiares. Nessa fase, o paciente pode já não conseguir alimentar-se normalmente, apresentando dificuldade em deglutir, além de poder haver incontinência urinária e fecal e comportamento inadequado intensificado. A doença pode progredir ainda com dificuldades motoras, sendo necessário auxílio para caminhar, como o uso de cadeira de rodas ou até mesmo ficar acamado. Nesse estágio, o paciente apresenta alto risco de desenvolver infecções, escaras e necessidade de hospitalização.

1.2 Objetivo

Apresentar como o avanço dos estágios da doença crônica de Alzheimer pode ser modelado por uma cadeia de Markov de acordo com o estágio em que a doença se encontra, utilizando os conhecimentos adquiridos no decorrer da disciplina de Processos Estocásticos até a presente data neste relatório.

1.3 Descrição da cadeia de markov em relação ao avanço dos estágios da doença Crônica de Alzheimer, obtenção da matriz de transição da Cadeia e diagrama de transição.

Seja X_n o estágio em que a doença se encontra. Para predizer em qual estágio a doença vai estar em $n+1$, ou seja, em X_{n+1} , basta apenas saber em qual estágio ela se encontra no instante n , isto é, X_n . Desse modo, qualquer informação sobre o passado é irrelevante se sabemos o estágio da doença no instante n para indicar o estágio da doença no instante $n+1$.

Nesse caso, podemos descrever o processo através de uma cadeia de Markov com espaço

de estados :

$$S = \{1, 2, 3, 4\}$$

Para caracterizar a cadeia, foi necessário definir a matriz de transição P, ou seja, foi calculada cada uma das probabilidades de transição com base no avanço dos estágios da doença, da seguinte maneira:

1. Probabilidade de estar no Estágio 1 e continuar no Estágio 1:

- $\mathbb{P}(X_{n+1} = 1 | X_n = 1) = 0.8$

2. Probabilidade de estar no Estágio 1 e ir para o Estágio 2:

- $\mathbb{P}(X_{n+1} = 2 | X_n = 1) = 0.2$

3. Probabilidade de estar no Estágio 2 e continuar no Estágio 2:

- $\mathbb{P}(X_{n+1} = 2 | X_n = 2) = 0.4$

4. Probabilidade de estar no Estágio 2 e ir para o Estágio 3:

- $\mathbb{P}(X_{n+1} = 3 | X_n = 2) = 0.6$

5. Probabilidade de estar no Estágio 3 e continuar no Estágio 3:

- $\mathbb{P}(X_{n+1} = 3 | X_n = 3) = 0.5$

6. Probabilidade de estar no Estágio 3 e ir para o Estágio 4:

- $\mathbb{P}(X_{n+1} = 4 | X_n = 3) = 0.5$

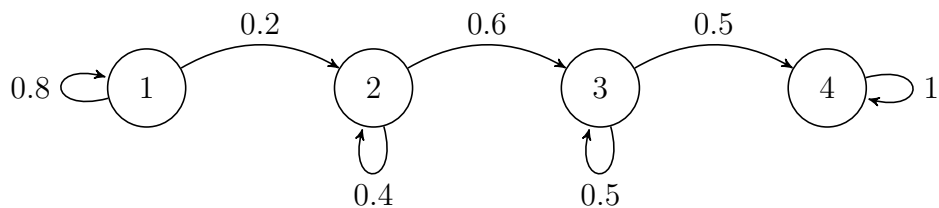
7. Probabilidade de estar no Estágio 4 e continuar no Estágio 4:

- $\mathbb{P}(X_{n+1} = 4 | X_n = 4) = 1$

Assim, a matriz de transição da cadeia de Markov é dada por:

$$P = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Diagrama de transição de estados referente a matriz P:



1.4 Obtenção da classe de estados, classificação dos estados, distribuição limite distribuição estacionária da cadeia de Markov

A cadeia possui quatro classes de estados:

$$S_1 = \{1\} \quad S_2 = \{2\} \quad S_3 = \{3\} \quad S_4 = \{4\}$$

onde S_1, S_2, S_3 , são classificadas como transientes e S_4 absorvente.

Não há cura para o Alzheimer, mas, atualmente, existem medicamentos que estabilizam a doença, retardando o seu desenvolvimento. Não havendo a cura, podemos perceber que nunca o estado atual irá regredir, ou seja, o processo não volta ao estado anterior. Assim, os estados nunca poderão se comunicar e quando a doença está no seu estágio final, ela permanecerá no mesmo, vemos que este último é um estado absorvente. Como esta cadeia não possui uma única classe de estados, concluímos que ela não é irredutível

Verificando a periodicidade dos estados:

$$\begin{aligned}d(1) &= \text{mdc}(1) = 1 \\d(2) &= \text{mdc}(1) = 1 \\d(3) &= \text{mdc}(1) = 1 \\d(4) &= \text{mdc}(1) = 1\end{aligned}$$

Note que todos os estados são aperiódicos, portanto a cadeia é aperiódica.

Dado que a cadeia não possui apenas uma classe de estados, constatamos que não existe distribuição limite, pois ela não é irredutível e também não é recorrente positiva.

1.5 Obtenção da distribuição estacionária e interpretação dos resultados

Para obter a distribuição estacionária fazemos:

$$\pi = \pi \cdot P \Rightarrow (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4) \cdot \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} = (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4)$$

Assim, temos o seguinte sistema de equações lineares:

$$\begin{cases} \pi_1 = 0.8\pi_1 \\ \pi_2 = 0.2\pi_1 + 0.4\pi_2 \\ \pi_3 = 0.6\pi_2 + 0.5\pi_3 \\ \pi_4 = 0.5\pi_3 + \pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases}$$

Resolvendo o sistema acima, apresentamos os seguintes resultados:

$$\pi_1 = 0 \quad \pi_2 = 0 \quad \pi_3 = 0 \quad \pi_4 = 1$$

Logo, a distribuição estacionária é dada por:

$$\pi = (0 \ 0 \ 0 \ 1)$$

É possível observar que a distribuição estacionária existe, e é única. Por esse resultado, temos que a probabilidade da doença estar em um determinado estágio depois de um longo período de tempo não vai depender de seu estágio inicial.

1.6 Algoritmo para simulação da cadeia de markov em estudo com o objetivo de simular uma amostra de tamanho n da cadeia.

1. Entrada: estado atual (X_n)
2. Saída: estado futuro (X_{n+1})
3. Se $X_n=1$: Gere X_{n+1} da distribuição multinomial($1, (0.8, 0.2, 0, 0)$)
4. Se $X_n=2$: Gere X_{n+1} da distribuição multinomial($1, (0, 0.4, 0.6, 0)$)
5. Se $X_n=3$: Gere X_{n+1} da distribuição multinomial($1, (0, 0, 0.5, 0.5)$)
6. Se $X_n=4$: Gere X_{n+1} da distribuição multinomial($1, (0, 0, 0, 1)$)
7. return $X_n + 1$

1.7 Simulação de uma amostra de tamanho n da cadeia de markov em estudo para diferentes valores iniciais da cadeia (α) utilizando o R-Studio.

Utilizando o software R-Studio e um algoritmo para a simulação de uma amostra de tamanho n da cadeia em estudo para diferentes valores iniciais da cadeia, foram obtidos os seguintes resultados:

[illegible]

Figura 1.1: Cadeia simulada para uma amostra de tamanho $n=500$ e distribuição inicial $x_0 = 1$

[illegible]

Figura 1.2: Cadeia simulada para uma amostra de tamanho $n=500$ e distribuição inicial $x_0 = 2$

[illegible]

Figura 1.3: Cadeia simulada para uma amostra de tamanho $n=500$ e distribuição inicial $x_0 = 3$

5 e Figura 6, podemos observar que a medida que n aumenta, as probabilidades de transição da cadeia se aproximam da distribuição estacionária obtida teoricamente.

1.9 Estimação e análise dos tempos de retorno a cada estado da cadeia.

Utilizando o R-studio podemos implementar um código para que possamos estimar o tempo esperado de retornos:

[illegible]

Figura 1.8: Algoritmo para estimativa do números de retornos para os Estágios 1,2,3 e 4 da doença.

```
> estado0 <- which(cadeia.sim==0)
> tempo.retorno0 <- diff(estado0)
> mean(tempo.retorno0)
[1] 1
>
> estado1 <- which(cadeia.sim==1)
> tempo.retorno1 <- diff(estado1)
> mean(tempo.retorno1)
[1] 1
>
> estado2 <- which(cadeia.sim==2)
> tempo.retorno2 <- diff(estado2)
> mean(tempo.retorno2)
[1] 1
>
> estado3 <- which(cadeia.sim==3)
> tempo.retorno3 <- diff(estado3)
> mean(tempo.retorno3)
[1] 1
>
```

Figura 1.9: Estimativa do números de retornos para os Estágios 1,2,3 e 4 da doença.

Observando os resultados obtidos na **Figura 9**, temos as seguintes estimativas:

- Para o Estágio 1 da doença, uma estimativa de retorno em 1 passo.
- Para o Estágio 2 da doença, uma estimativa de retorno em 1 passo.
- Para o Estágio 3 da doença, uma estimativa de retorno em 1 passo.
- Para o Estágio 4 da doença, uma estimativa de retorno em 1 passo.

Note que, para qualquer estado que a cadeia esteja, o tempo estimado de retorno é igual a 1. De fato, podemos observar que o Alzheimer é uma doença crônica e

progressiva, de modo que, seja qual for o estágio em que a doença se encontre, ela só poderá se manter em seu estágio atual ou ir para o próximo estágio, fazendo com que não seja possível o retorno ao avançar para um outro estado.

1.10 Estimação da probabilidade de transição (\hat{P}_{ab}) para diferentes tamanhos amostrais.

```
> P <- rbind( c(0.8,0.2,0,0),c(0,0.4,0.6,0),c(0,0,0.5,0.5),c(0,0,0,1))
>
> matrix.power(P,100)
      [,1]      [,2]      [,3] [,4]
[1,] 2.037036e-10 1.018518e-10 2.037036e-10 1
[2,] 0.000000e+00 1.606938e-40 4.733165e-30 1
[3,] 0.000000e+00 0.000000e+00 7.888609e-31 1
[4,] 0.000000e+00 0.000000e+00 0.000000e+00 1
> #carregar o pacote para calcular potência de matrizes
> require(matrixcalc)
```

Figura 1.10: Probabilidade de transição com $n = 100$.

```
> P <- rbind( c(0.8,0.2,0,0),c(0,0.4,0.6,0),c(0,0,0.5,0.5),c(0,0,0,1))
>
> matrix.power(P,1000)
      [,1]      [,2]      [,3] [,4]
[1,] 1.230232e-97 6.15116e-98 1.230232e-97 1
[2,] 0.000000e+00 0.00000e+00 5.599582e-301 1
[3,] 0.000000e+00 0.00000e+00 9.332636e-302 1
[4,] 0.000000e+00 0.00000e+00 0.000000e+00 1
```

Figura 1.11: Probabilidade de transição com $n=1000$.

```
> P <- rbind( c(0.8,0.2,0,0),c(0,0.4,0.6,0),c(0,0,0.5,0.5),c(0,0,0,1))
>
> matrix.power(P,10000)
      [,1] [,2] [,3] [,4]
[1,] 9.881313e-324 0 0 1
[2,] 0.000000e+00 0 0 1
[3,] 0.000000e+00 0 0 1
[4,] 0.000000e+00 0 0 1
```

Figura 1.12: Probabilidade de transição com $n=10000$.

Com as probabilidades de transição obtidas nas figuras **10**, **11** e **12**, podemos observar que conforme aumentamos o número de passos da nossa cadeia, as probabilidades de transição tendem a se estabilizar, se aproximando da distribuição estacionária, mas nunca de fato se estabilizarão, pois nossa cadeia não tem distribuição limite.

Capítulo 2

Parte II

Os coronavírus são vírus de RNA de fita simples, com cerca de 120 nanômetros de diâmetro. Eles são suscetíveis a mutações e recombinações e são, portanto, muito diversos. Há cerca de 40 variedades diferentes e eles infectam principalmente humanos, mamíferos não-humanos e aves. Nesta pesquisa, usamos banco de dados onde, foram comparados variantes nos registros de RNA com a variante do SARS-CoV-2.

2.1 Dados

Para este trabalho selecionamos dois bancos de dados da mesma região mas com um certo espaço de tempo de coleta dos dados. Os primeiros dados que iremos utilizar tem como código /USA/FL-CDC-STM-000030529/2021, datado de 07 de março de 2021. Sua estimativa para as probabilidades de transição é apresentada abaixo pela matriz P_1 :

$$P_1 = \begin{bmatrix} 0.27 & 0.25 & 0.20 & 0.28 \\ 0.38 & 0.15 & 0.08 & 0.39 \\ 0.28 & 0.20 & 0.17 & 0.35 \\ 0.27 & 0.16 & 0.29 & 0.28 \end{bmatrix}$$

O espaço de estados é dado por $S = \{A, C, G, T\}$

Agora vamos utilizar os dados de /USA/MD-MDH-2365/2021, datado de 10 de maio de 2021. Sua estimativa para as probabilidades de transição é apresentada abaixo pela matriz P_2 :

$$P_2 = \begin{bmatrix} 0.27 & 0.24 & 0.21 & 0.28 \\ 0.39 & 0.15 & 0.08 & 0.38 \\ 0.28 & 0.20 & 0.17 & 0.35 \\ 0.27 & 0.16 & 0.29 & 0.28 \end{bmatrix}$$

Com o espaços de estados dado por $S = \{A, C, G, T\}$

O código utilizado para obter as probabilidades de transição foi:

```

1 palavra = 'tt'
2 with open('teste.txt') as f:
3     ocorrencias = f.read().count(palavra)
4     print(ocorrencias)
5
6

```

Figura 2.1: Código para obter as estimativas das probabilidades de transição

2.2 Análise Descritiva dos dados

Apresentando se houve diferença entre as transições, vamos exemplificar usando o gráfico de barras. Para o banco de dados da matriz de transição P_1 temos:

Do estado A indo para os estados A, C, G e T:

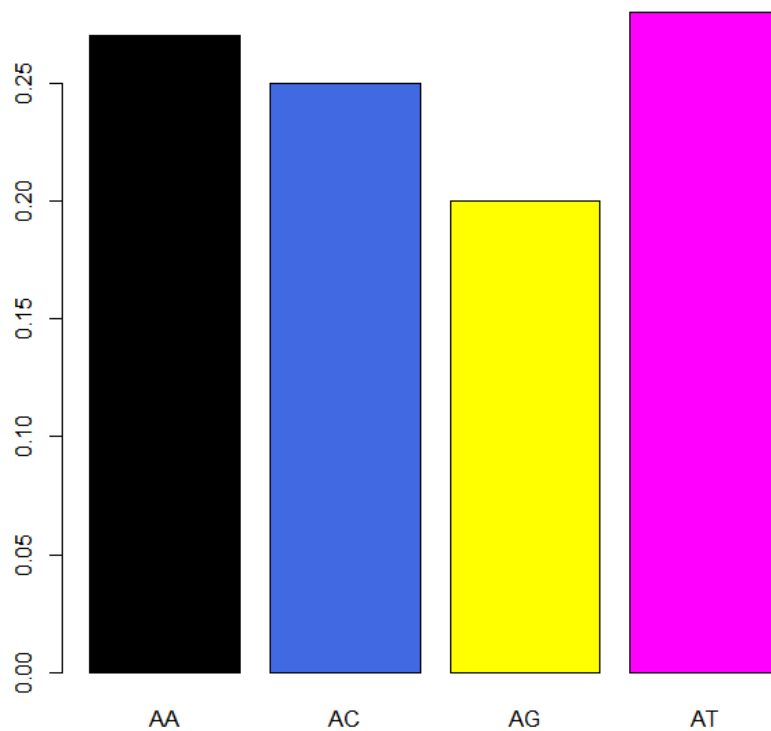


Figura 2.2: probabilidades de transição A para C, G e T

Podemos observar pela figura 2.2, que as probabilidades de transição estão bem distribuídas, no entanto, existe uma maior probabilidade de transição do estado A para o T e do estado A para o A, se comparado com as probabilidades de ir do estado A para os estados C e G.

Do estado C indo para os estados A, C, G e T:

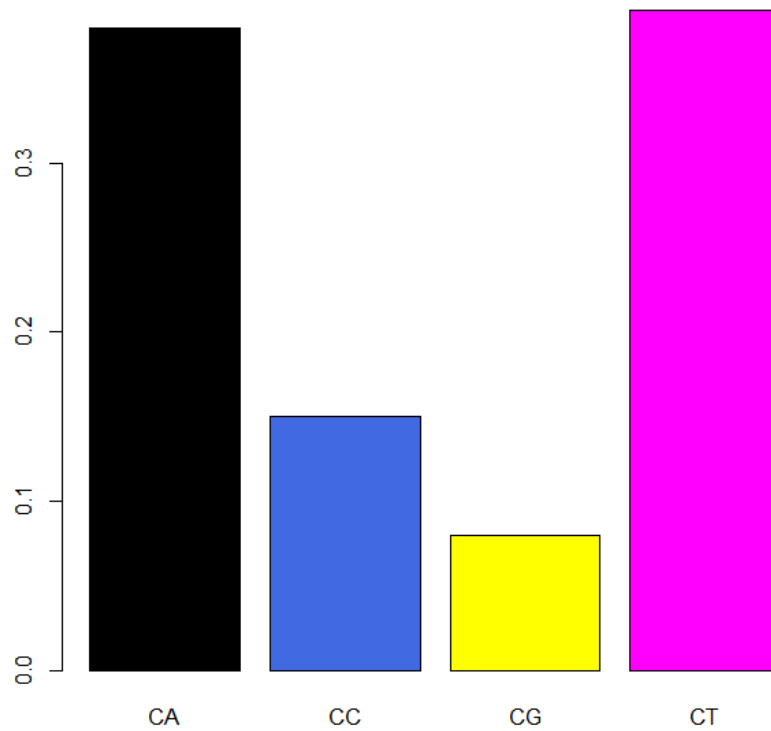


Figura 2.3: probabilidades de transição de C para A, C, G e T

Podemos observar pela figura 2.3, que as probabilidades de transição não estão bem distribuídas. Existe uma maior probabilidade de transição do estado C para o T e do estado C para o A, se comparado com as probabilidades de ir do estado C para os estados C e G.

Do estado G indo para os estados A, C, G e T:

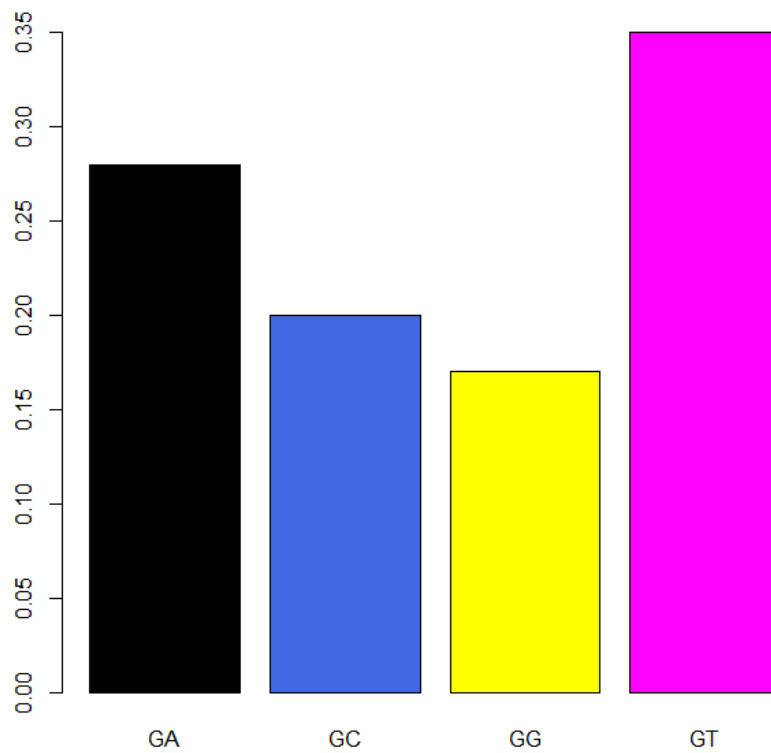


Figura 2.4: probabilidades de transição de G para A, C, G e T

Podemos observar pela figura 2.4, que as probabilidades de transição do estado G para os estados T, A, C e G estão diminuindo, respectivamente.

Do estado T indo para os estados A, C, G e T:

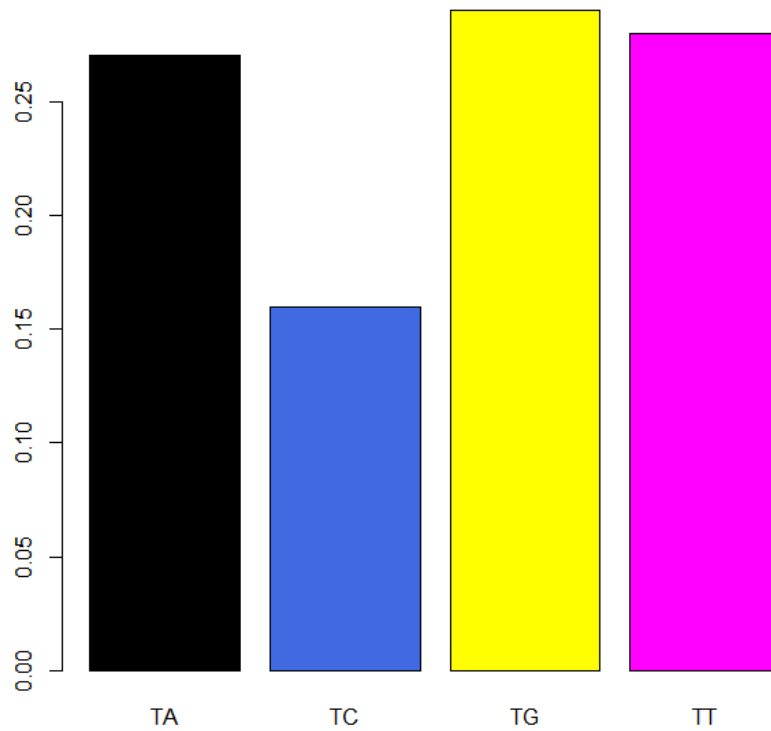


Figura 2.5: probabilidades de transição de T para A, C, G e T

Podemos observar pela figura 2.5, que as probabilidades de transição estão bem distribuídas entre os estados T para A, T para G e T para T, no entanto, existe uma probabilidade menor de transição do estado T para o C.

Agora para o banco de dados da matriz de transição P_2 temos:
Do estado A indo para A, C, G e T:

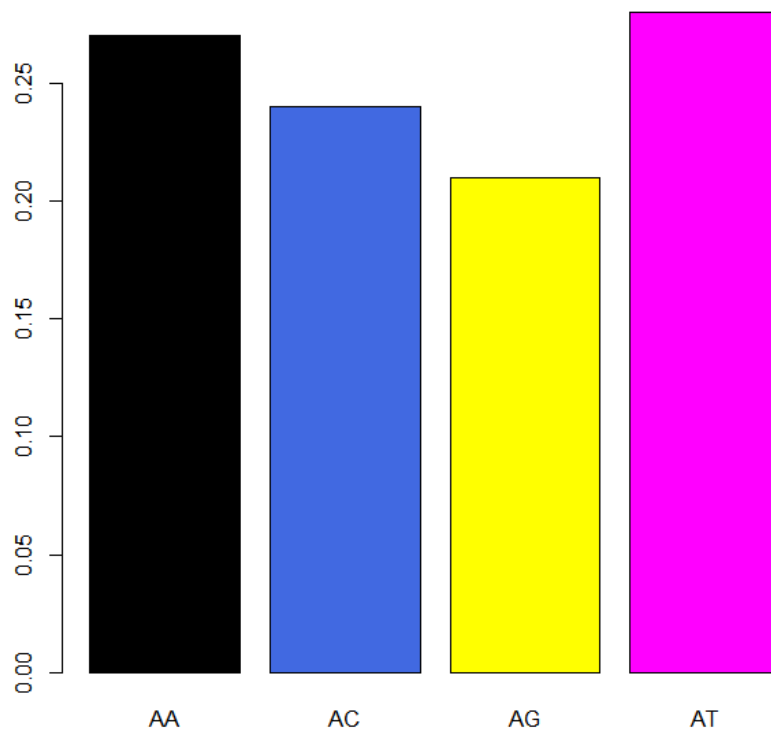


Figura 2.6: probabilidades de transição A para C, G e T

Podemos observar pela figura 2.6, que as probabilidades de transição do estado A para os estados T, A, C e G estão diminuindo, respectivamente.

Do estado C indo para A, C, G e T:

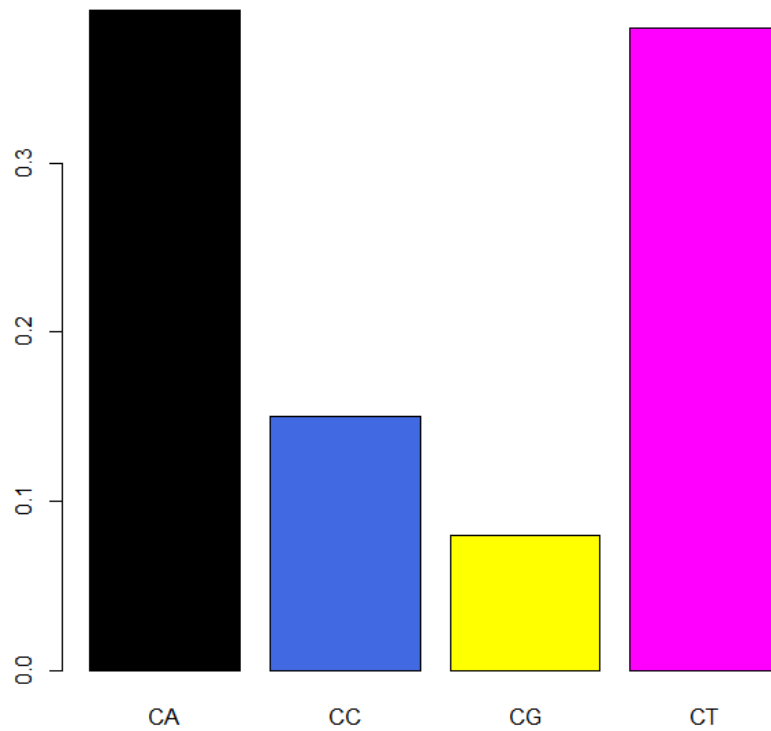


Figura 2.7: probabilidades de transição de C para A, C, G e T

Podemos observar pela figura 2.7, que as probabilidades de transição estão bem distribuídas dos estados C para A e C para T, note que existe uma menor probabilidade do estado C continuar no C e menor ainda do estado C ir para o estado G.

Do estado G indo para A, C, G e T:

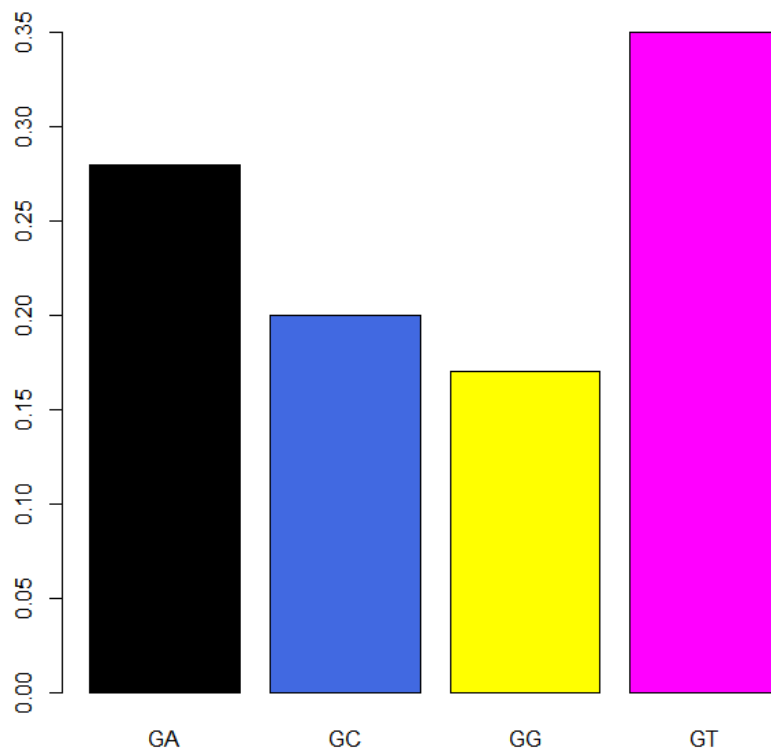


Figura 2.8: probabilidades de transição de G para A, C, G e T

Podemos observar pela figura 2.8, que as probabilidades de transição do estado G para os estados T, A, C e G estão diminuindo, respectivamente.

Do estado T indo para A, C, G e T:

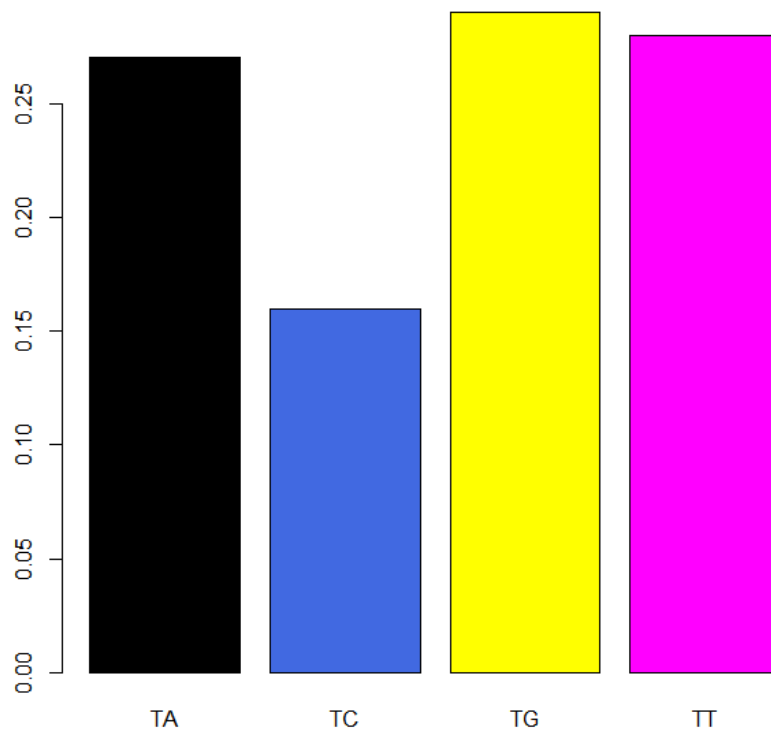


Figura 2.9: probabilidades de transição de T para A, C, G e T

Podemos observar pela figura 2.9, que as probabilidades de transição estão bem distribuídas entre os estados T para A, T para G e T para T, no entanto, a probabilidade do estado T ir para o estado C é menor

Capítulo 3

Conclusão:

Ao analisar os resultados obtidos neste trabalho, foi notado que entre as datas de coleta do banco de dados, datados de 7 de março e 10 de maio, ambos de 2021, em um período de quase 2 meses, não houve uma diferença significativa entre as variantes nos registros de RNA com a variante do SARS-CoV-2.