

3ª Prova de Probabilidade-DEST UFMG

Lucas Rafael Costa Santos
2021017723

Dezembro 2023 Belo Horizonte-MG

1 Introdução

Este trabalho tem como objetivo explorar a convergência da média amostral para a distribuição normal, conforme previsto pelo Teorema do Limite Central (TLC). Para realizar essa análise, serão consideradas três diferentes distribuições de probabilidade: (i) Binomial ou Bernoulli com probabilidades de sucesso $p < 0.1$ e $p = 0.5$; (ii) Exponencial com um parâmetro λ a ser escolhido; e (iii) Uniforme(a, b) com a e b sendo valores apropriados.

O procedimento consiste em gerar m amostras de tamanho n para cada uma das distribuições, permitindo a variação de n e m. Em seguida, serão gerados histogramas para visualizar a distribuição das amostras. Para avaliar a convergência para a distribuição normal, serão calculadas as médias amostrais para cada conjunto de amostras.

A análise prossegue com a geração de histogramas para as médias amostrais e a comparação desses resultados com a distribuição normal teórica, levando em consideração a média e o desvio padrão da distribuição original. Será realizada também a padronização das médias amostrais para compará-las com a distribuição normal padrão (0, 1).

O código desenvolvido permitirá a fácil execução e compreensão, sendo acessível mesmo para pessoas não familiarizadas com programação. A ideia é fornecer uma ferramenta útil para estudantes e profissionais de áreas como Biológicas ou Humanas, que desejam entender a convergência da média amostral para a distribuição normal e suas implicações.

2 Resumo do Código

O código define três funções: `gerar_amostras`, `plotar_histograma` e `analisar_tcl`.

- A função `gerar_amostras` gera amostras de uma distribuição dada, com parâmetros específicos, e retorna uma matriz de amostras.
- A função `plotar_histograma` plota um histograma dos dados fornecidos.
- A função `analisar_tcl` gera amostras de uma distribuição, calcula a média de cada amostra, plota um histograma das médias e compara esse histograma com uma distribuição normal teórica.

É solicitado ao usuário que se insira os valores de n (o tamanho da amostra) e m (o número de amostras). O código, então, gera amostras, calcula médias e plota histogramas para três distribuições diferentes: binomial, exponencial e uniforme.


Os histogramas e as médias amostrais demonstram o Teorema Central do Limite (TCL), que afirma que a distribuição das médias amostrais de uma população com média finita e desvio padrão se aproxima de uma distribuição normal à medida que o tamanho da amostra aumenta, independentemente da forma da distribuição da população. Isso é evidente nos gráficos produzidos pelo código, onde é possível ver que os histogramas das médias amostrais se assemelham a uma curva normal.

3 Gerando os Gráficos

Para gerar os gráficos entre no link abaixo e siga as seguintes instruções:

<https://colab.research.google.com/drive/1F5TXtSnNFtByhUZnrNBHWzKp2YX-q2Z6?usp=sharing>

- Para rodar o programa, clique no ícone de execução presente no canto superior esquerdo do código:



```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import binom, expon, uniform, norm

def gerar_amostras(distribuicao, parametros, m, n):
    return distribuicao.rvs(size=(m, n), *parametros)

def plotar_histograma(data, titulo, xlabel, ylabel):
    plt.hist(data, bins=30, density=True, alpha=0.7)
    plt.title(titulo)
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.show()

def analisar_tcl(dist_name, dist_params, n, m):
    means = []
```

Figure 1: Execução

```
a, b = 0, 1
uniforme = uniform(loc=a, scale=b-a)
amostras_uniforme = gerar_amostras(uniforme, (), m, n)
plotar_histograma(amostras_uniforme.flatten(), 'Amostras - Uniforme', 'Valores', 'Densidade')
medias_uniforme = np.mean(amostras_uniforme, axis=1)
plotar_histograma(medias_uniforme, 'Médias Amostrais - Uniforme', 'Valores', 'Densidade')
analisar_tcl('uniform', (a, b-a), n, m)
```

*** Digite o valor de n (tamanho da amostra):




Figure 2: Valor de 'n'

- Depois de alguns segundos irá surgir, abaixo do código, uma caixinha pedindo o valor de 'n'. Digite o valor e aperte a tecla 'enter'.
- Logo abaixo digite o valor de 'm'.

```
# (iii) Uniforme(a=0, b=1)
a, b = 0, 1
uniforme = uniform(loc=a, scale=b-a)
amostras_uniforme = gerar_amostras(uniforme, (), m, n)
plotar_histograma(amostras_uniforme.flatten(), 'Amostras - Uniforme', 'Valores', 'Densidade')
medias_uniforme = np.mean(amostras_uniforme, axis=1)
plotar_histograma(medias_uniforme, 'Médias Amostrais - Uniforme', 'Valores', 'Densidade')
analisar_tcl('uniform', (a, b-a), n, m)
```

Digite o valor de n (tamanho da amostra): 30
Digite o valor de m (número de amostras):




Figure 3: Valor de 'm'

- Após alguns segundos os gráficos serão gerados.
- A seguir temos as imagens geradas tomando $n = 30$ e $m = 1000$:

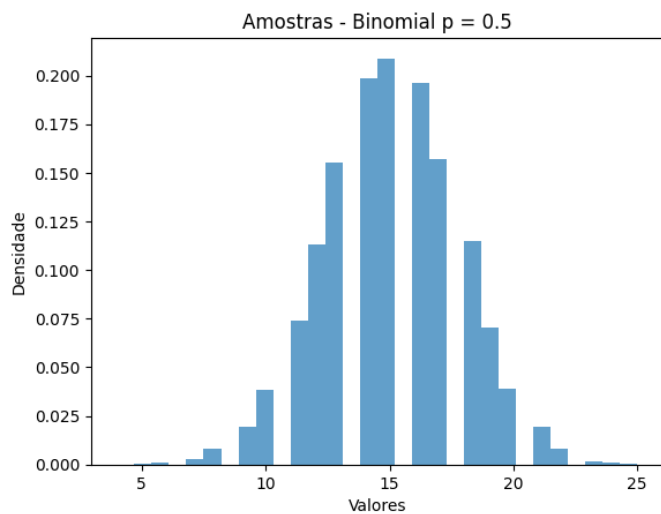


Figure 4: *Amostras - Binomial*

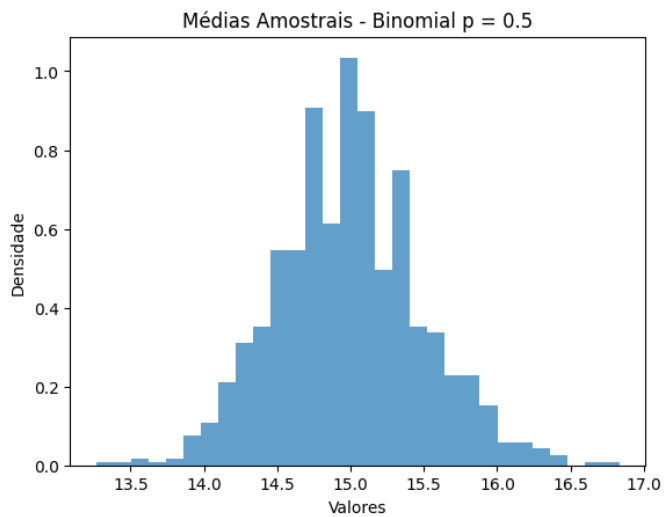


Figure 5: *Médias Amostras - Binomial*

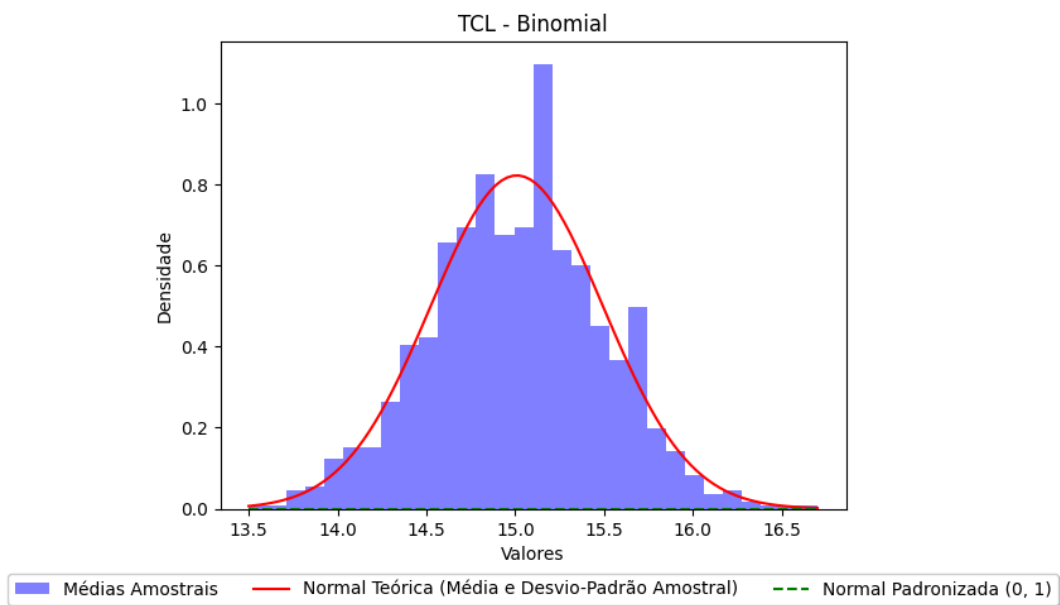


Figure 6: *TCL - Binomial*

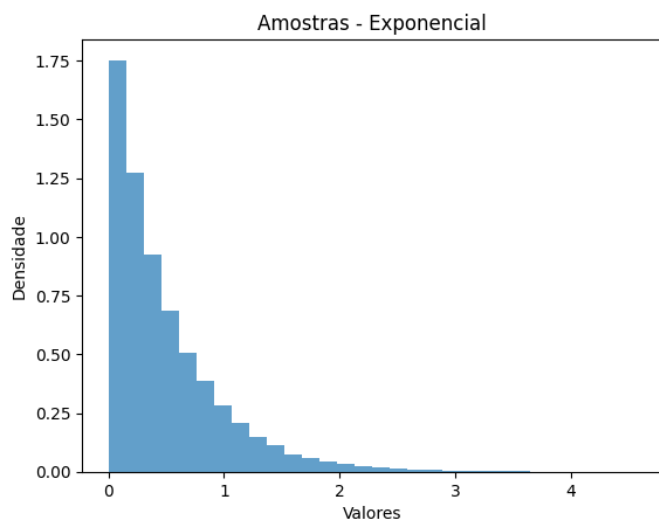


Figure 7: *Amostras - Exponencial*

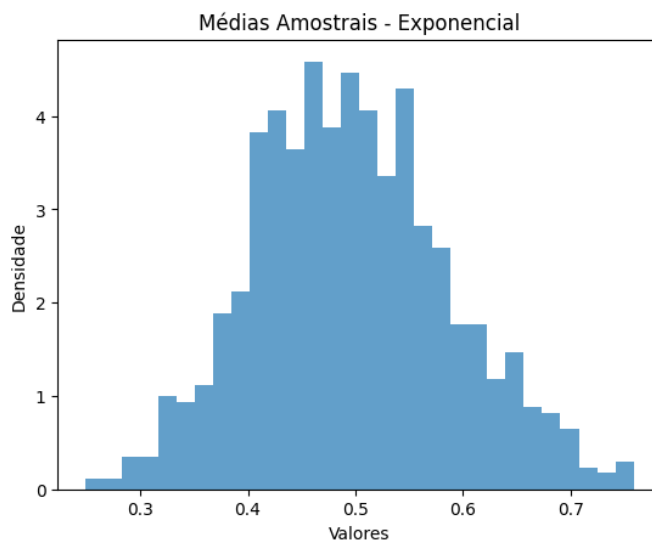


Figure 8: *Médias Amostrais - Exponencial*

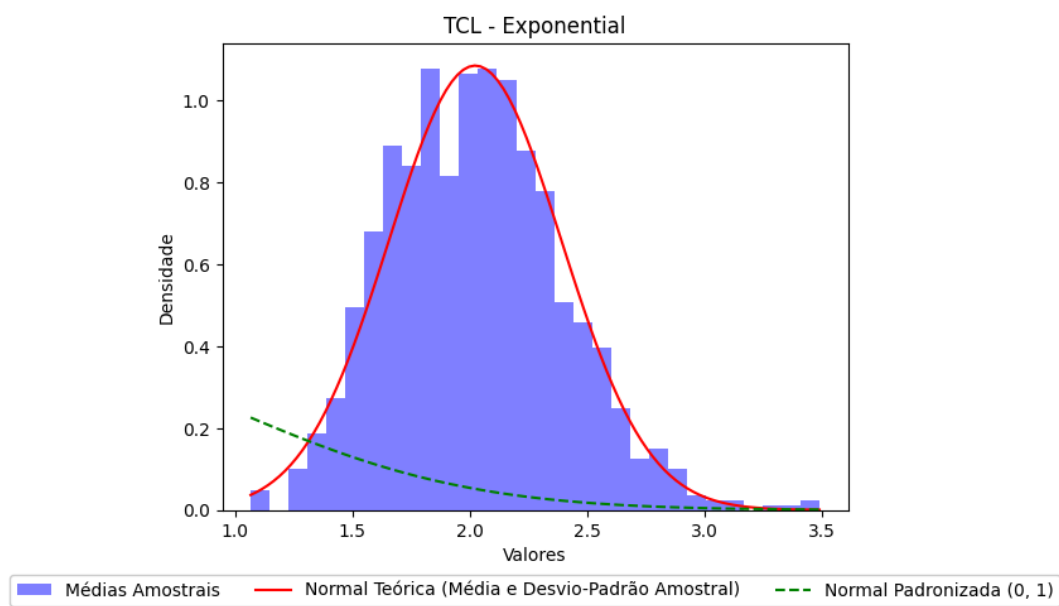


Figure 9: *TCL - Exponencial*

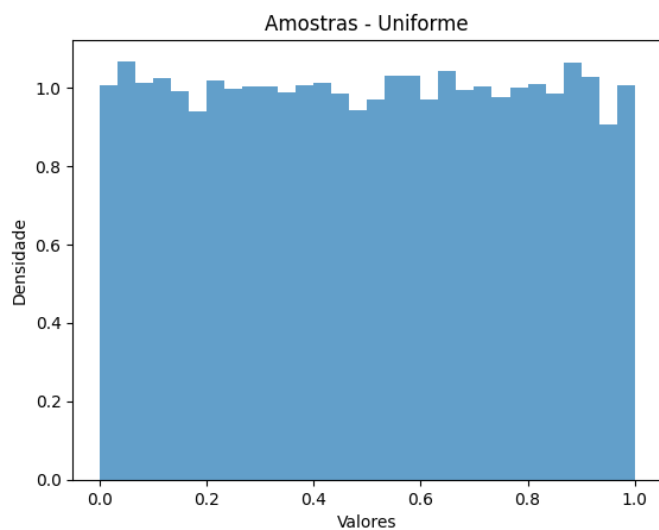


Figure 10: *Amostras - Uniforme*

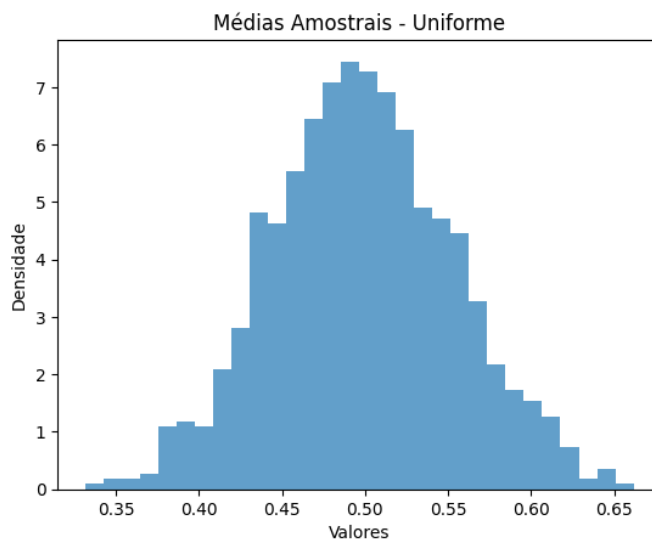


Figure 11: *Médias Amostrais - Uniforme*

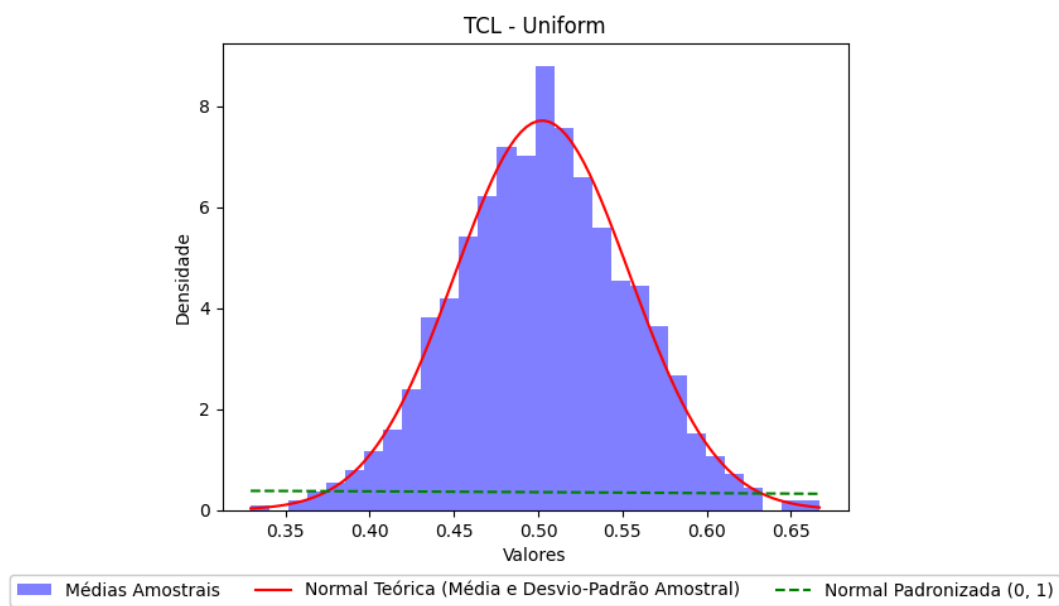


Figure 12: *TCL - Uniforme*

4 Análise

Para analisar os gráficos gerados, vamos nos concentrar em alguns pontos principais:

- Histogramas das amostras de cada distribuição:

Aqui, estamos olhando para a distribuição dos dados. Na distribuição binomial, é possível ver um pico no meio (perto de $n/2$ para $p=0.5$) e a distribuição deve ser simétrica. Na distribuição exponencial, a maioria dos dados devem estar perto de zero, com uma cauda longa à direita. Por fim, na distribuição uniforme, os dados devem ser aproximadamente igualmente distribuídos em todo o intervalo $[a, b]$.

- Histogramas das médias amostrais:

Como vimos anteriormente, de acordo com o Teorema Central do Limite (TCL), a distribuição das médias amostrais deve se aproximar de uma distribuição normal à medida que o tamanho da amostra aumenta, independentemente da forma da distribuição original. Portanto, podemos ver um pico no meio com caudas simétricas em ambos os lados.

- Comparação com a distribuição normal teórica:

Aqui, verificamos se a distribuição das médias amostrais segue o TCL. A linha vermelha no gráfico representa a distribuição normal teórica com a mesma média e desvio padrão que as médias amostrais. Se o TCL se mantiver, a distribuição das médias amostrais (barras azuis) deve se aproximar desta linha vermelha à medida que o tamanho da amostra aumenta.

- Comparação com a distribuição normal padronizada (0,1):

A linha verde representa a distribuição normal padrão (média 0, desvio padrão 1). As médias amostrais foram padronizadas (subtraídas da média e divididas pelo desvio padrão) para comparar com esta distribuição. Novamente, se o TCL se mantiver, a distribuição das médias amostrais padronizadas deve se aproximar desta linha verde à medida que o tamanho da amostra aumenta.

Em geral, estamos procurando ver se os histogramas das médias amostrais se assemelham a uma distribuição normal e se aproximam cada vez mais de uma distribuição normal à medida que o tamanho da amostra aumenta. Isso seria uma confirmação visual do Teorema Central do Limite.

Vamos rodar novamente o nosso código, dessa vez aumentando os valores de $n(5000)$ e $m(50000)$. Observe que com o aumento do tamanho das amostras a distribuição das médias amostrais se aproxima de uma distribuição normal.

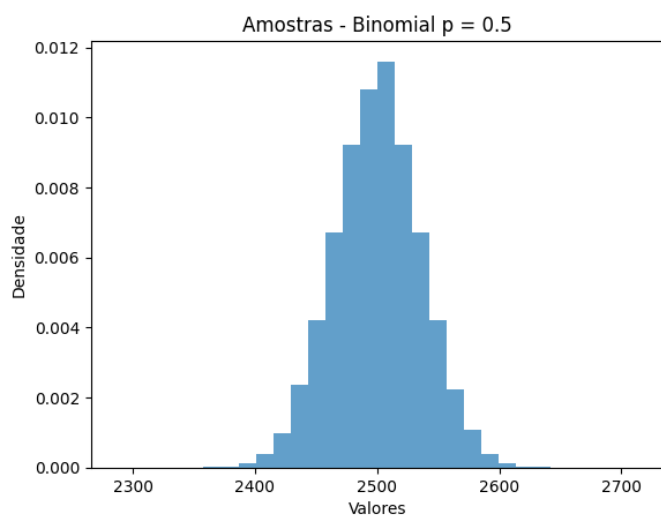


Figure 13: *Amostras - Binomial*

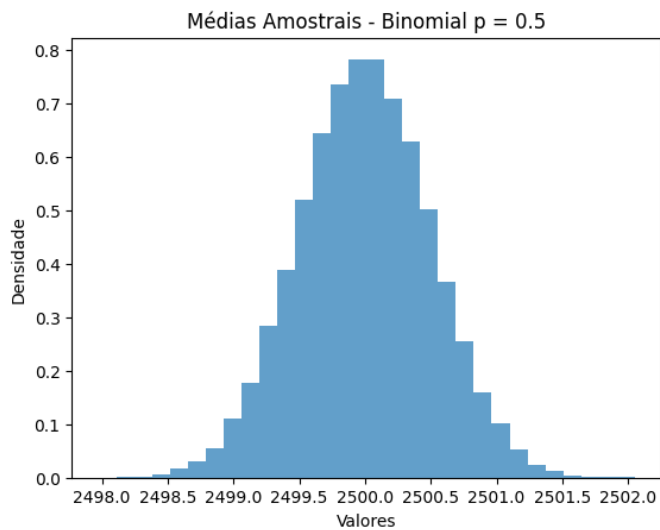


Figure 14: *Médias Amostrais - Binomial*

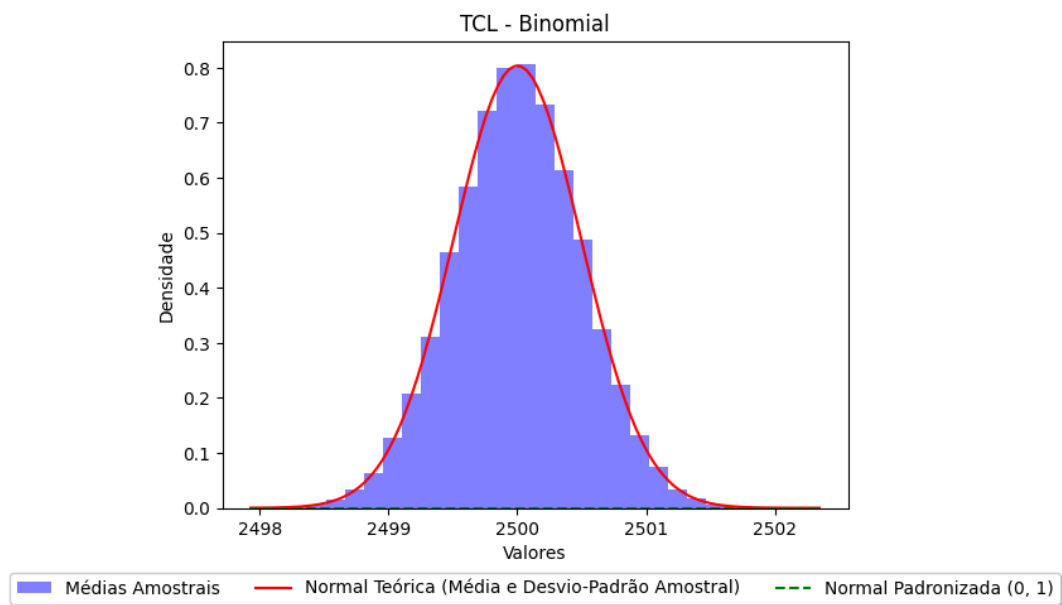


Figure 15: *TCL - Binomial*

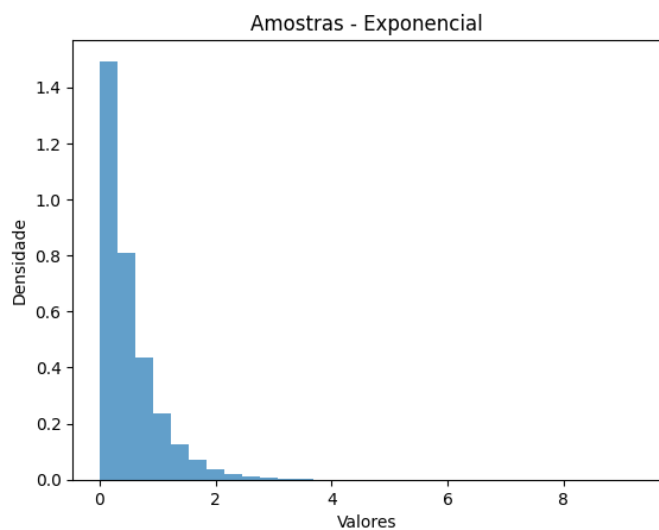


Figure 16: *Amostras - Exponencial*

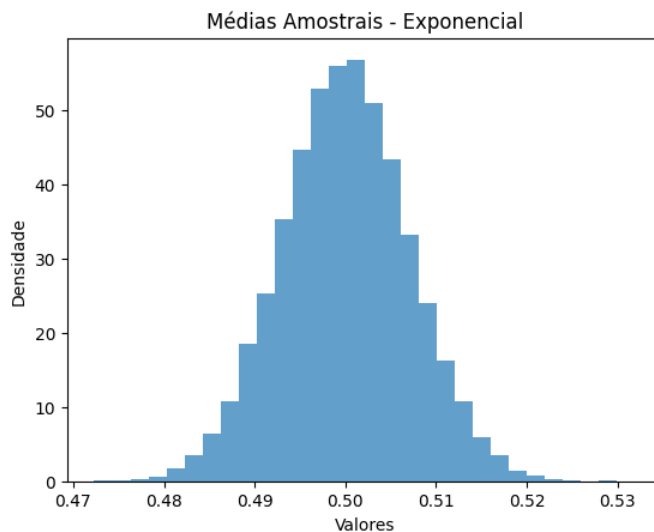


Figure 17: *Médias Amostrais - Exponencial*

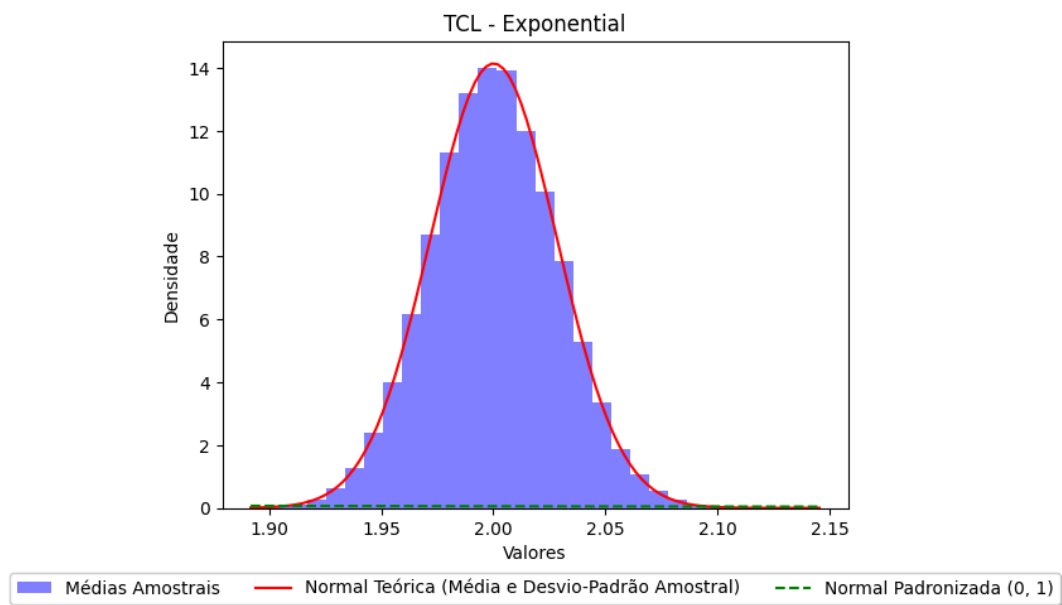


Figure 18: *TCL - Exponencial*

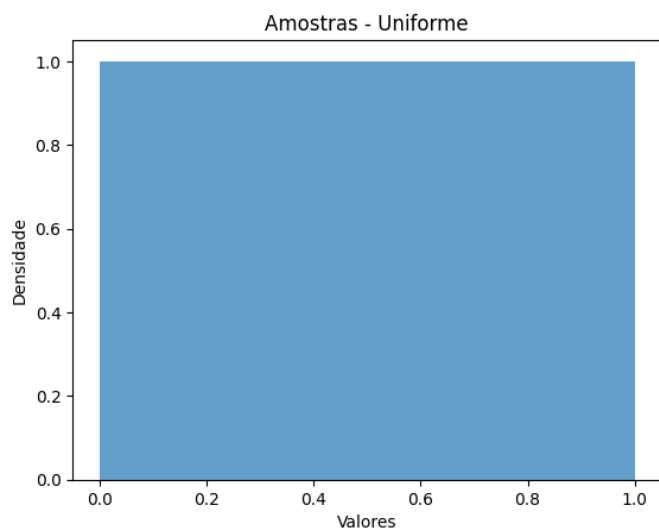


Figure 19: *Amostragem Uniforme*

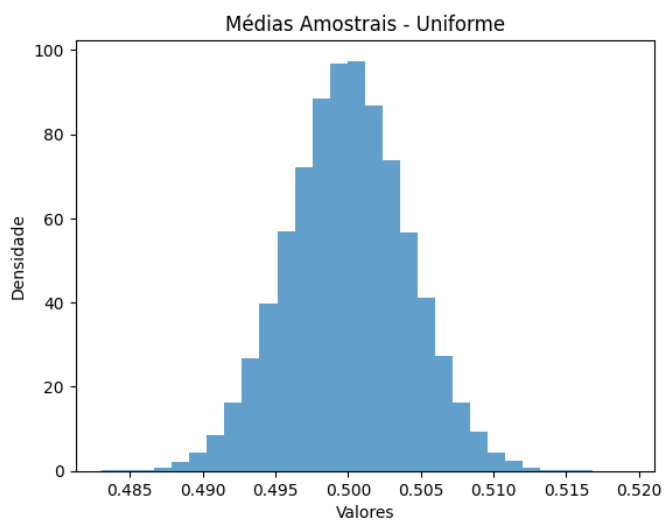


Figure 20: *Médias Amostrais - Uniforme*

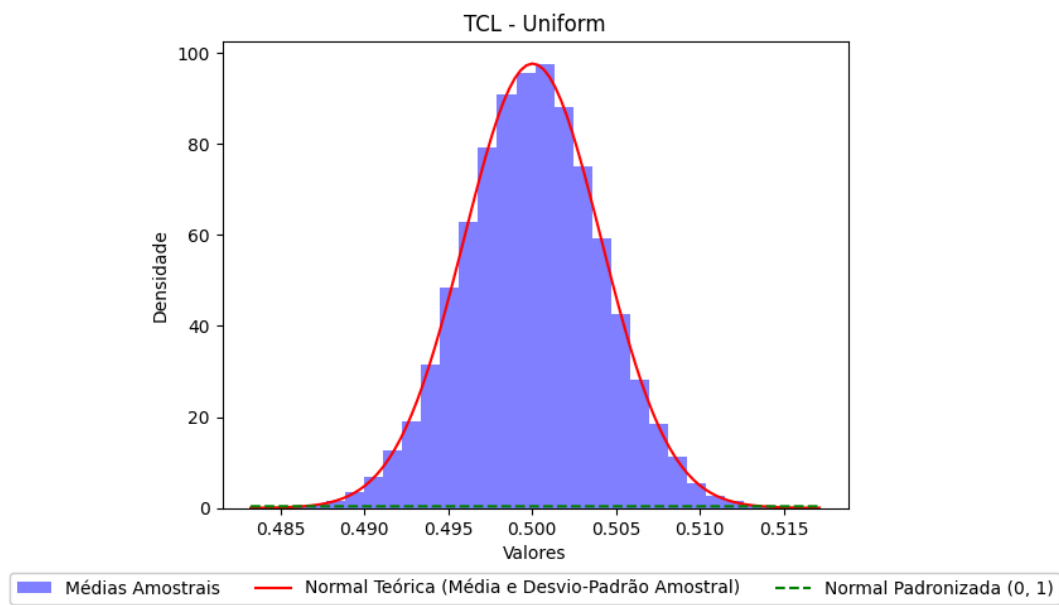


Figure 21: *TCL - Uniforme*

4.1 Conclusão

A análise realizada por meio dos histogramas das amostras, das médias amostrais e das comparações com distribuições teóricas no código fornece evidências consistentes com o Teorema Central do Limite (TCL). Como podemos ver, o aumento dos tamanhos de amostra (n) e do número de amostras (m) resulta em histogramas de médias amostrais que se aproximam cada vez mais de uma distribuição normal.

A observação das características esperadas, como o pico central e as caudas simétricas nos histogramas das médias amostrais, confirma a validade do TCL. Além disso, a sobreposição entre o histograma das médias amostrais e a distribuição normal teórica, bem como a consistência ao comparar com a distribuição normal padronizada, reforça a robustez da aplicação do TCL para diferentes tipos de distribuições (binomial, exponencial e uniforme).

Portanto, os resultados obtidos demonstram de maneira visual e quantitativa que, conforme previsto pelo TCL, a distribuição das médias amostrais se aproxima de uma distribuição normal à medida que o tamanho da amostra aumenta. Essa conclusão tem implicações significativas para a inferência estatística, proporcionando uma base sólida para generalizações sobre médias populacionais independentemente da forma da distribuição original da população.