

Projeto 1 - Resultados

1 Parte A

Iniciamos a primeira parte do projeto com a leitura e visualização dos 20 arquivos disponibilizados. A Figura 1 apresenta cada um dos 20 conjuntos de dados dispostos em gráficos de barras. Visualmente, podemos separar os dados em dois conjuntos. Dos arquivos 1 a 10, encontra-se valores compreendidos no intervalo de -0.5 a 0.5. Já nos arquivos de 11 a 20, tem-se valores maiores no intervalo -4.0 a 3.0. As Figuras 2 e 3 exemplificam os tipos de dados que temos em cada um dos subconjuntos dos 20 arquivos fornecidos. É importante notar que a partir dessa primeira análise, temos a hipótese de que tais dados podem apresentar alguma característica que possibilita o agrupamento em subconjuntos com similaridade. Testando tal hipótese, observamos o histograma de cada um dos dados (apresentado na Figura 4). Vemos então que novamente os dados aparentam ser divididos em duas categorias (justamente como havíamos dito anteriormente, de 1 a 10 e de 11 a 20). As Figuras 5 e 6 apresentam em detalhes o histograma de um arquivo de cada subgrupo proposto.

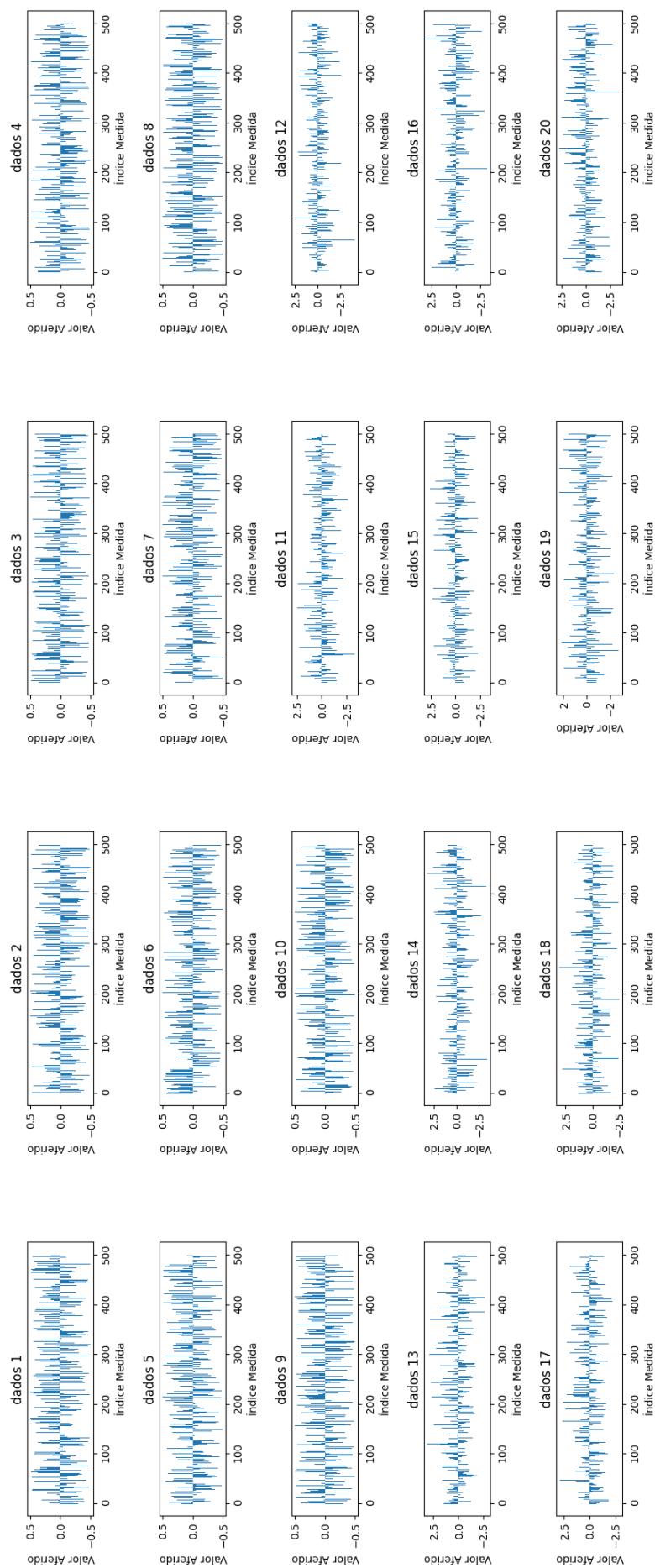


Figura 1: Gráfico de barras dos dados disponibilizados

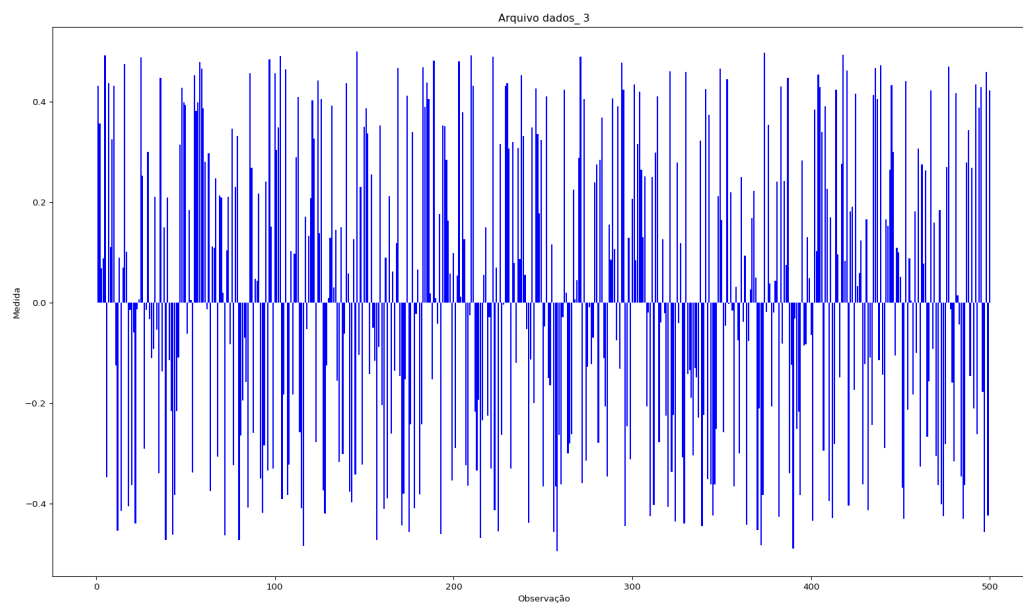


Figura 2: Gráfico de barras do arquivo "dados_3"

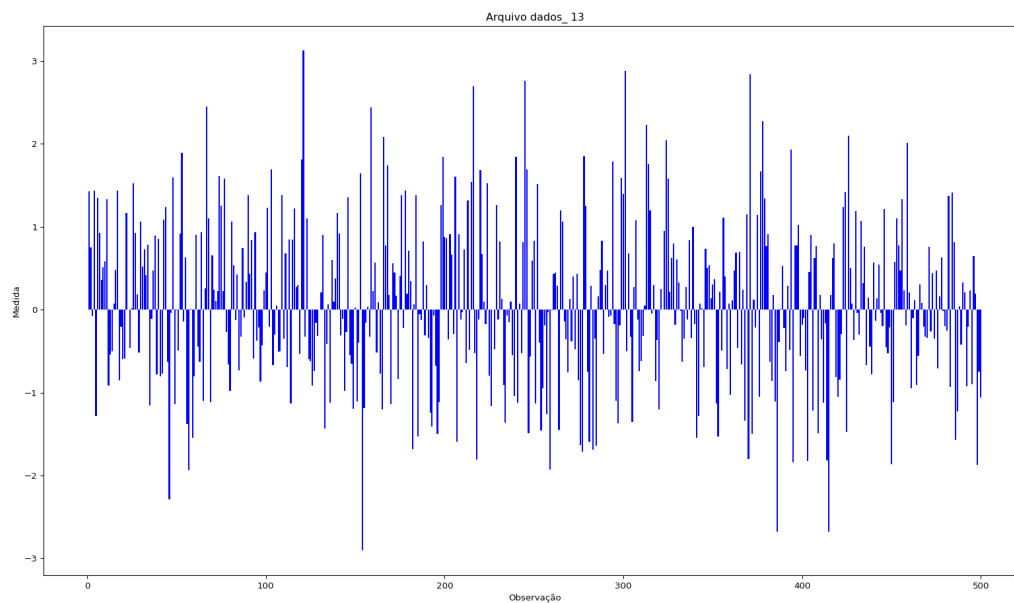


Figura 3: Gráfico de barras do arquivo "dados_13"

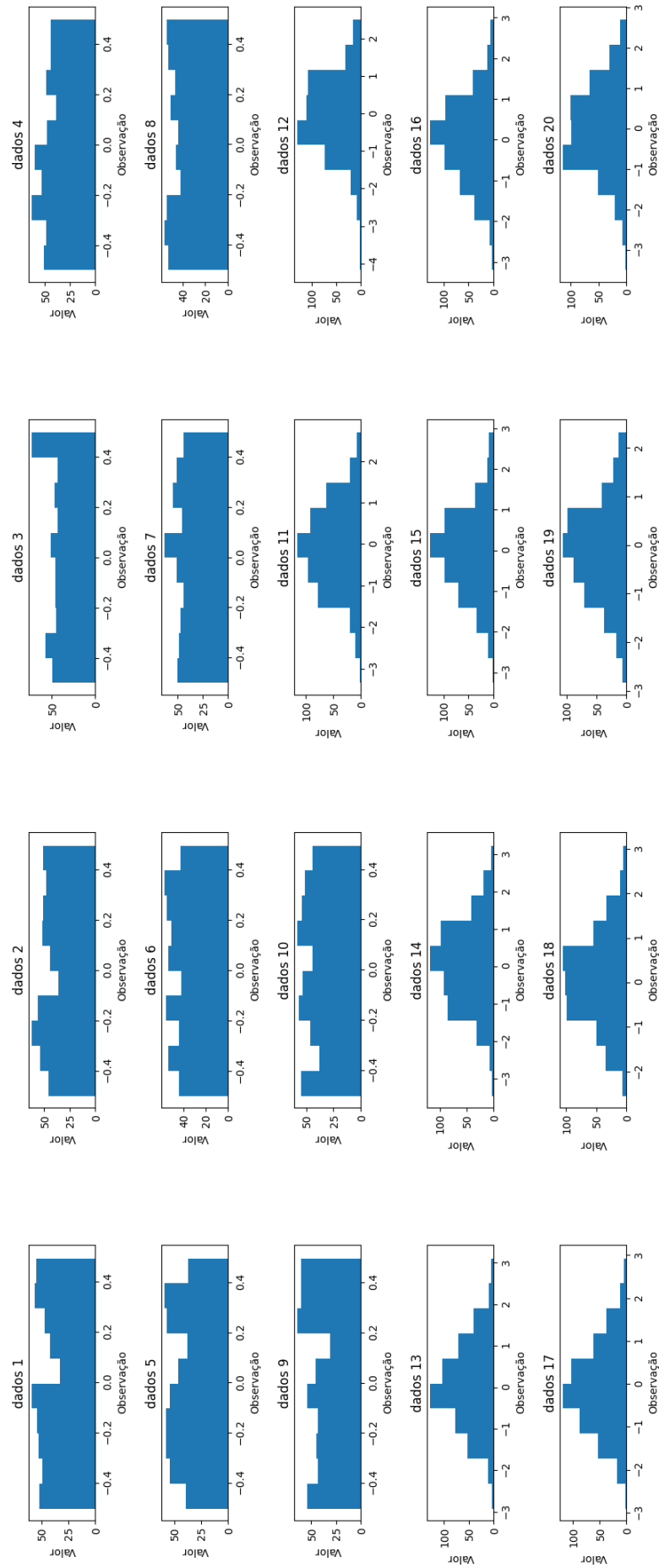


Figura 4: Histogramas dos dados disponibilizados

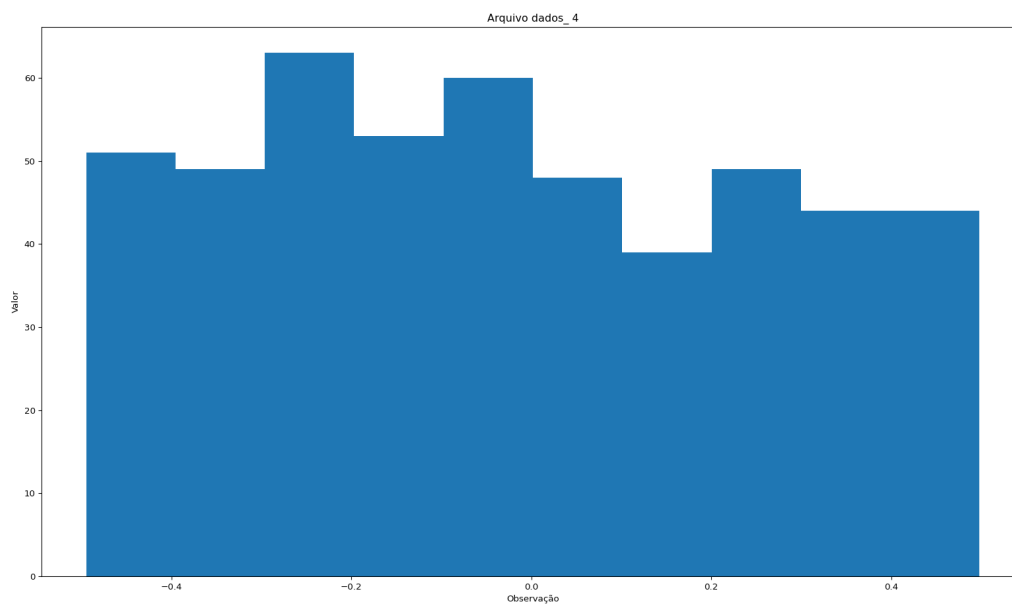


Figura 5: Histograma do arquivo "dados_4"

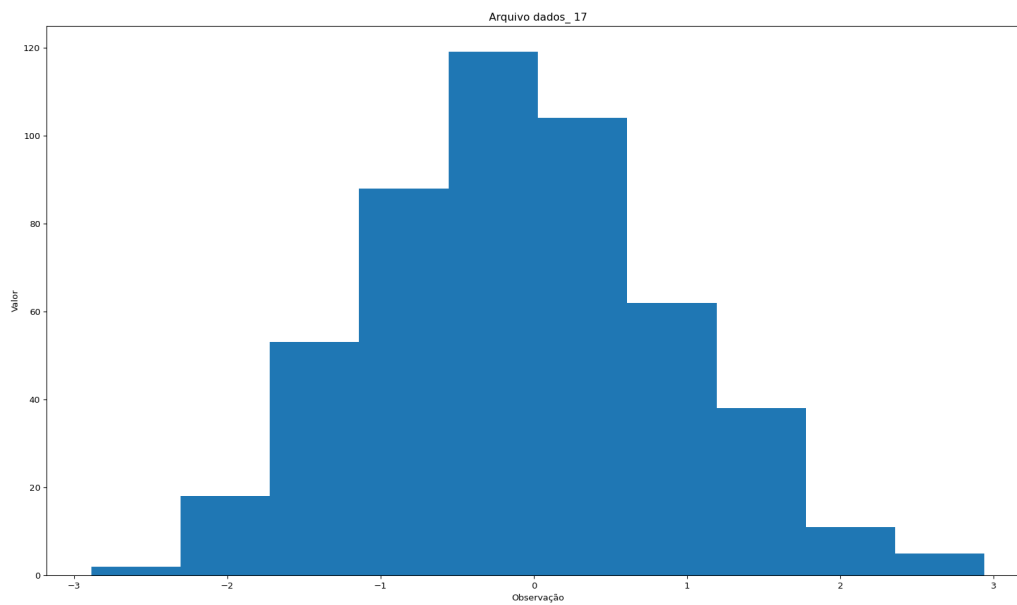


Figura 6: Histograma do arquivo "dados_17"

Nesse ponto, temos a tendência de inferir que os dados contém valores aleatórios seguindo duas diferentes distribuições: Uniforme e Gaussiana. Para testar tal hipótese, podemos calcular a média e desvio padrão de cada conjunto de dados e comprovar tal hipótese gerando novos dados a partir dos valores estatísticos extraídos dos dados. Seguindo nesse processo, calculamos a média e desvio de cada um dos arquivos obtendo a Figura 7. Nela podemos ver um comportamento similar justamente entre os conjuntos de arquivos discernidos anteriormente, principalmente pelo desvio padrão dos dados. A linha laranja separa visualmente os arquivos a partir do comportamento estatístico observado.

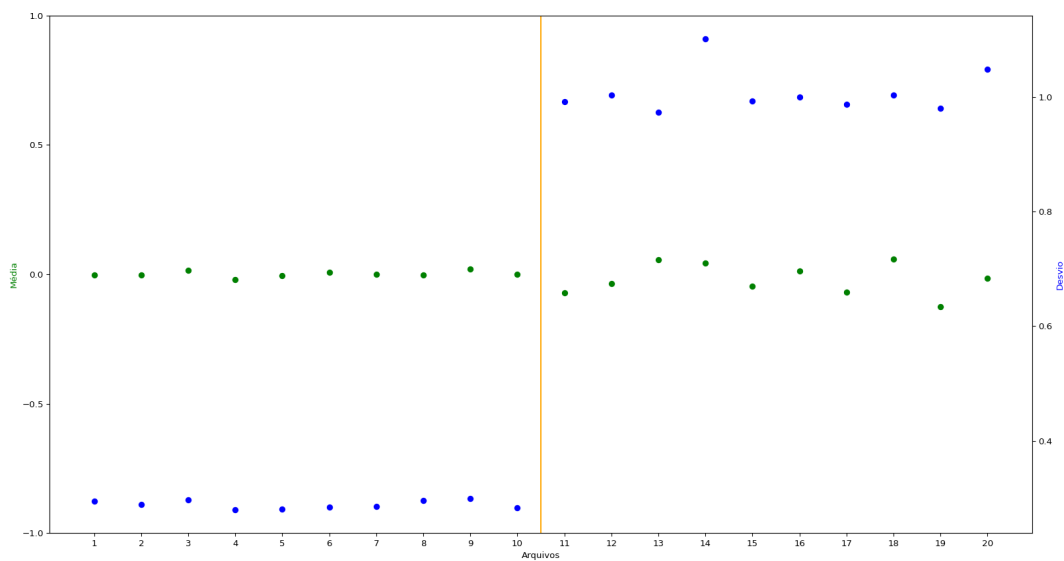


Figura 7: Média e desvio padrão dos dados de cada um dos arquivos. Linha laranja representa a separação visual entre os arquivos

Com tais dados em mãos, propusemos dois geradores de dados seguindo a media e o desvio encontrados para cada um dos conjuntos de dados. O primeiro gerador nos retorna um conjunto de 500 dados seguindo uma distribuição uniforme com valores aleatórios dentro do conjunto $[-0.5, 0.5]$. Já o segundo gerador nos retorna 500 dados gerados a partir de uma distribuição gaussiana aleatória com parâmetros média zero e desvio padrão um. Com esses geradores fomos capazes de obter 20 novos arquivos seguindo a hipótese de distribuições uniformes e gaussianas e construir os gráficos apresentados nas figuras a seguir.

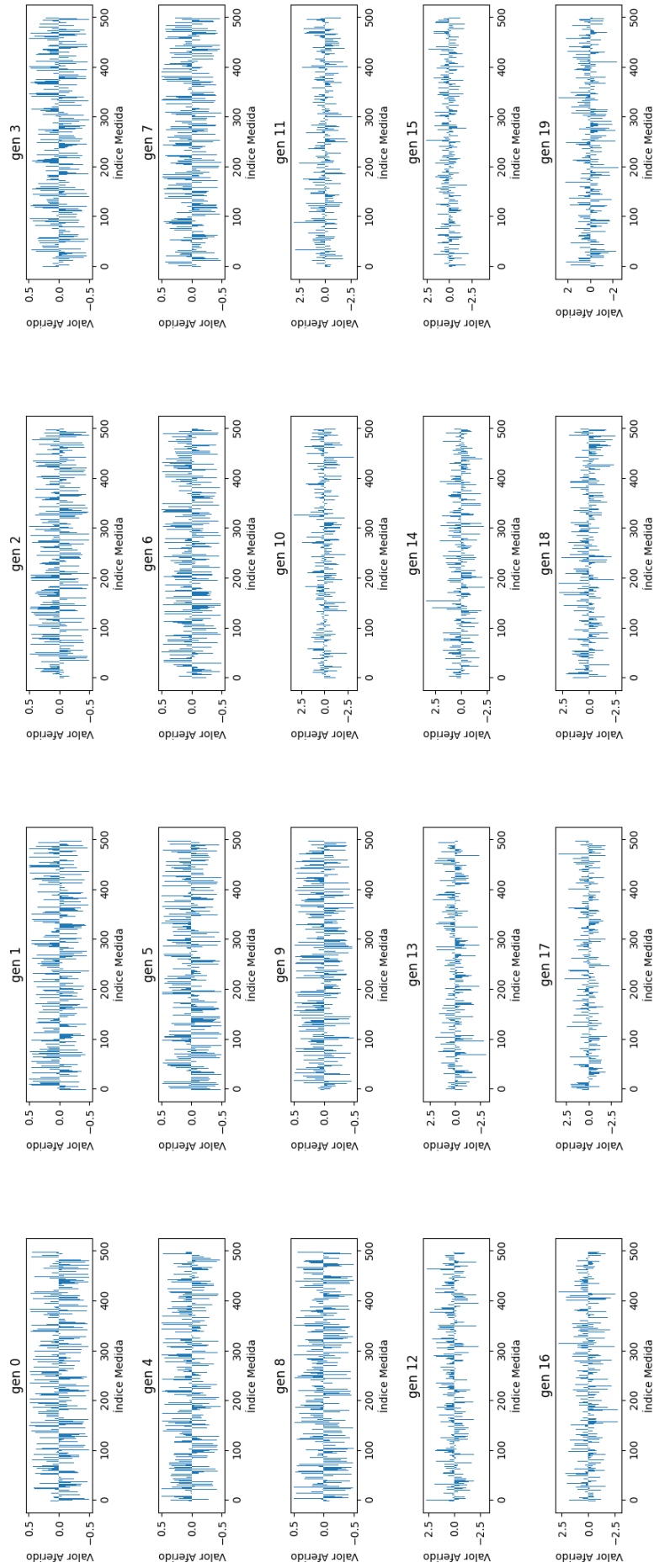


Figura 8: Gráficos de barras dos arquivos gerados

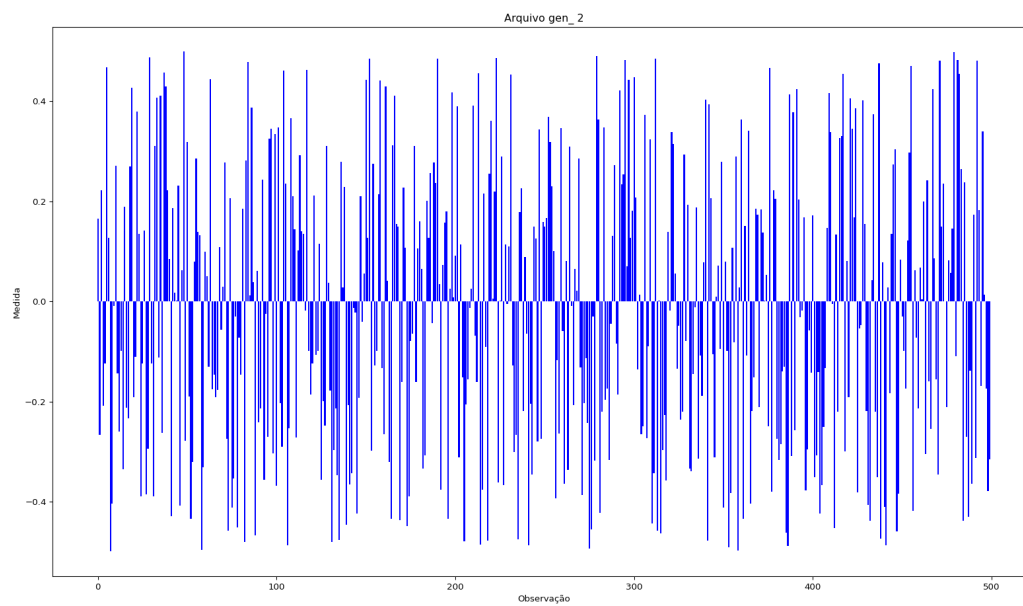


Figura 9: Gráfico distribuição uniforme com dados gerados

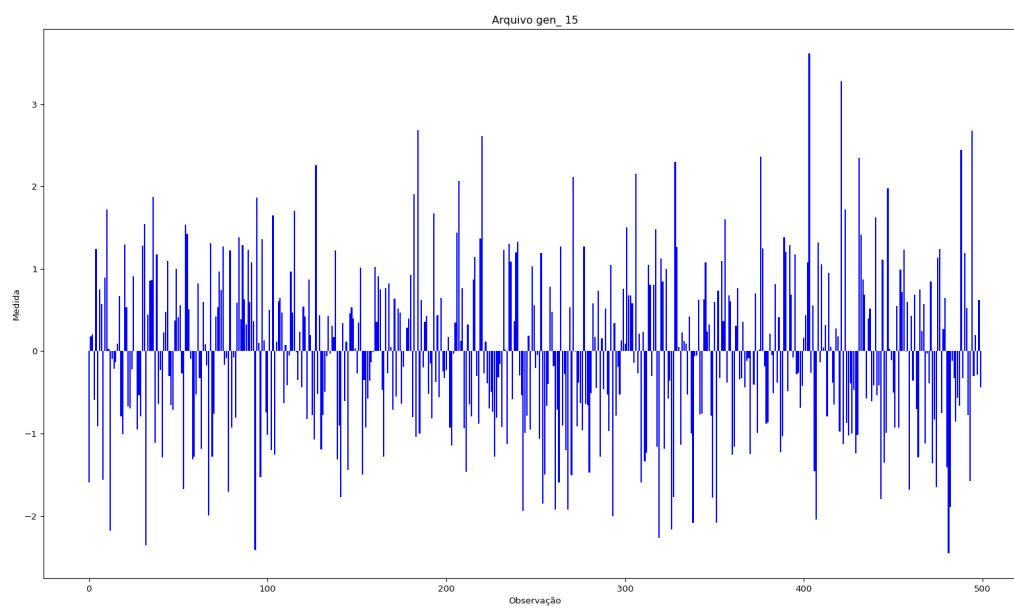


Figura 10: Gráfico distribuição gaussiana com dados gerados

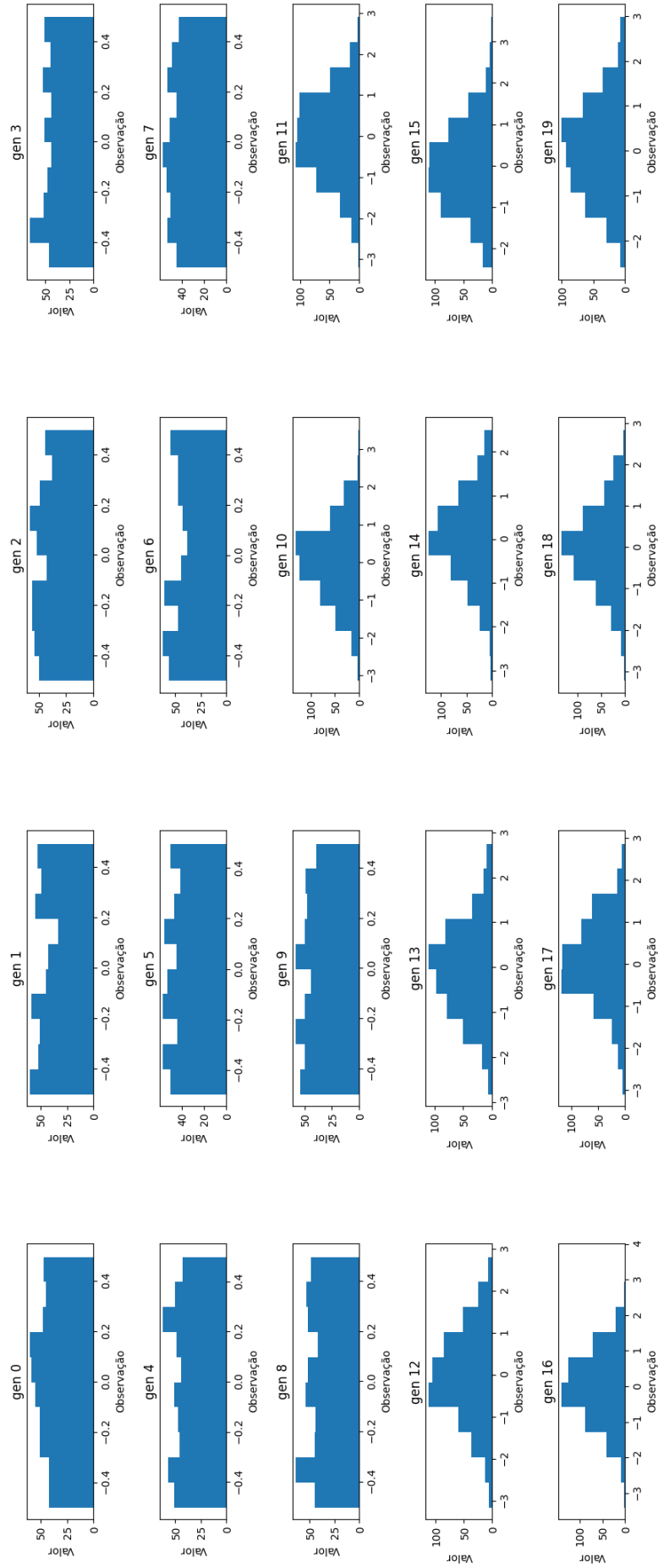


Figura 11: Histogramas dos dados gerados

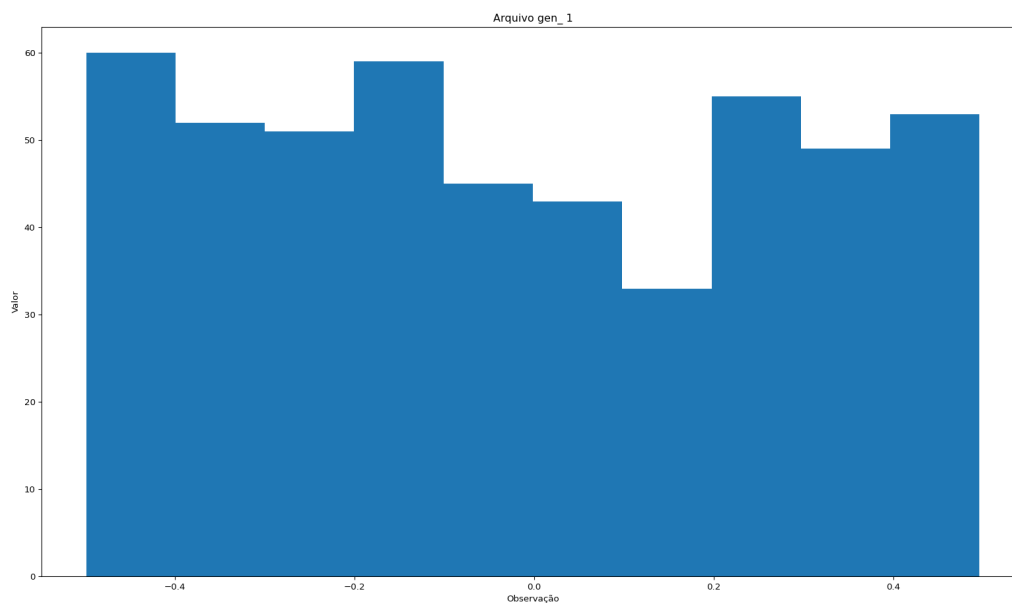


Figura 12: Histograma da distribuição uniforme gerada

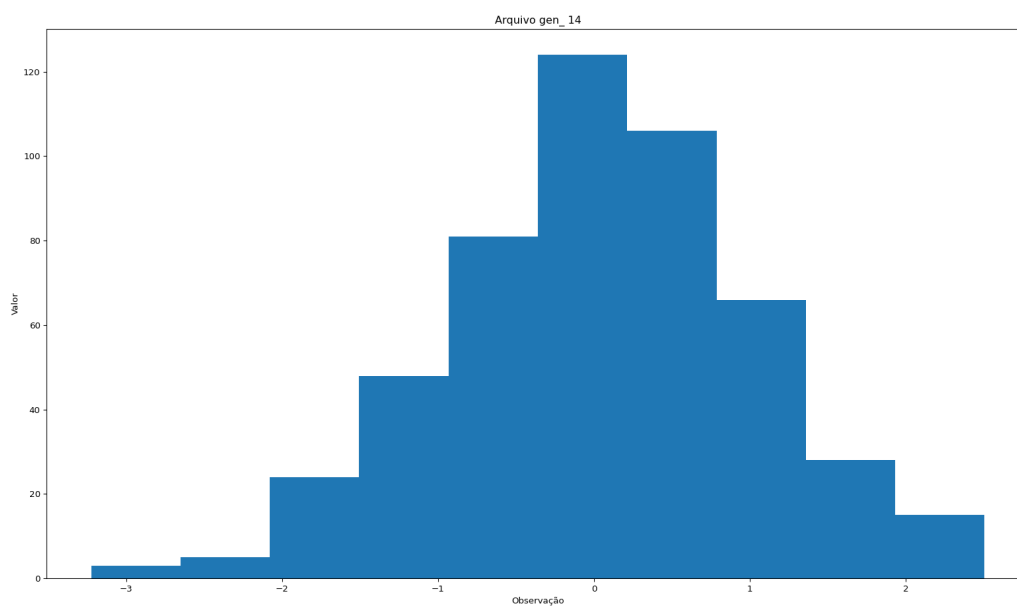


Figura 13: Histograma da distribuição gaussiana gerada

Observamos então uma grande similaridade entre os dados gerados e os dados fornecidos. Além das análises visuais dos gráficos de barras e histogramas calculamos ainda a média e desvio de cada um dos conjuntos de dados como anteriormente, obtendo a Figura 14. Nela, podemos notar a média ao redor de zero e os desvios padrões similares aos obtidos com os dados fornecidos.

Tais dados podem ser encontrados na tabela 1. As médias e desvios apresentados são referente às médias de todos os conjuntos de dados de cada distribuição (média das médias e média dos desvios). Podemos notar grande similaridade entre as informações estatísticas obtidas. Dessa forma, podemos dizer que realmente os dados fornecidos concordam com as hipóteses propostas.

Dados	μ uniforme	σ uniforme	μ gaussiana	σ gaussiana
Fornecido	-0.0060	0.3882	-0.0120	0.9086
Gerado	-0.0066	0.3877	-0.0020	0.8986

Tabela 1: Comparação de características estatísticas dos dados fornecidos e gerados.

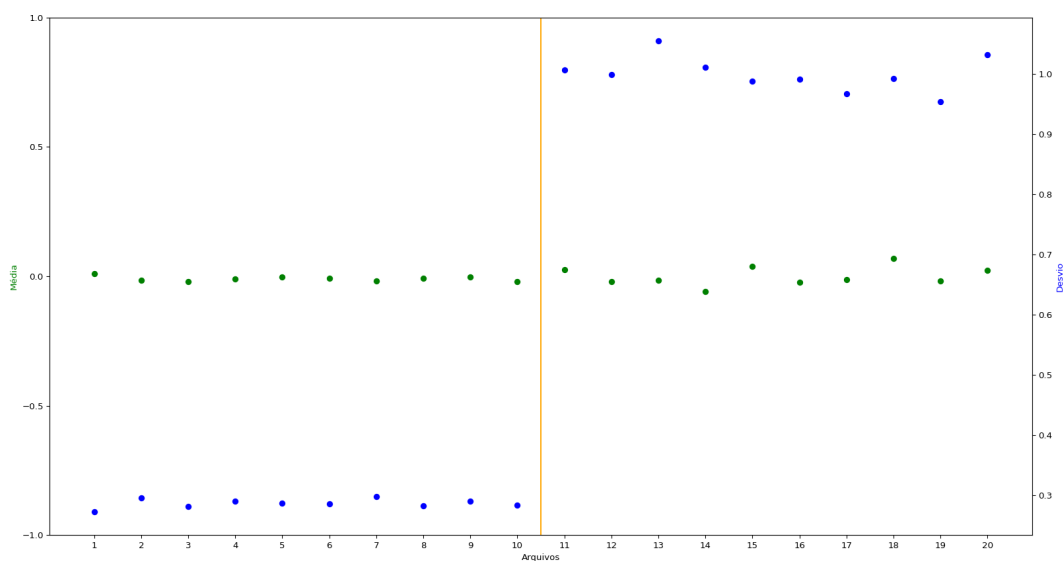


Figura 14: Média e desvio padrão dos arquivos gerados

Para fins de visualização, geramos as imagens a seguir para cada uma das distribuições abordadas.

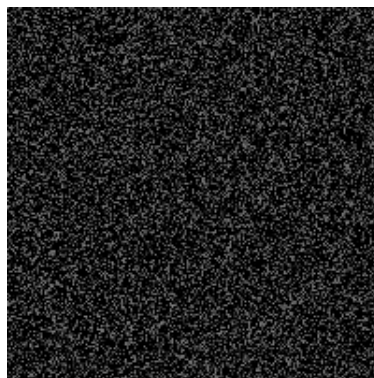


Figura 15: Dados gerados em duas dimensões (200x200) com distribuição uniforme

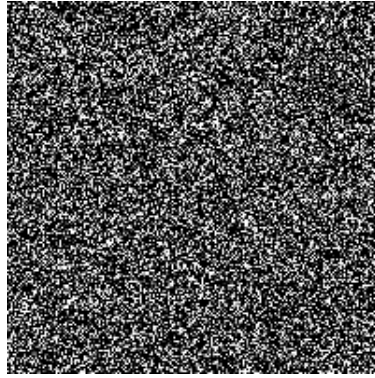


Figura 16: Dados gerados em duas dimensões (200x200) com distribuição gaussiana

2 Parte B

Na segunda parte desse projeto, iniciamos o estudo de grafos com a construção de uma função (método da classe que iniciamos na parte A desse projeto) que aceita como parâmetro uma matriz estocástica representando todas as probabilidade de, em um passo, o grafo estar em um novo nó. A partir dessa função, podemos simular diversos tipos de grafos em diferentes números de passos. Seguindo o pedido no enunciado do projeto, seguimos para a implementação de alguns autômatos baseados na figura 6 do *CDT-22*. As matrizes estocásticas utilizadas para tal implementação, podem ser conferidas nas tabelas a seguir.

A	0	1
0	0.9	0.9
1	0.1	0.1

Tabela 2: Matriz estocástica A.

B	0	1
0	0.2	0.2
1	0.8	0.8

Tabela 3: Matriz estocástica B.

C	0	1
0	0.5	0.5
1	0.5	0.5

Tabela 4: Matriz estocástica C.

Com tais matrizes estocásticas, conseguimos gerar os três diferentes padrões mostrados na Figura 18 com 500 passos por autômato.

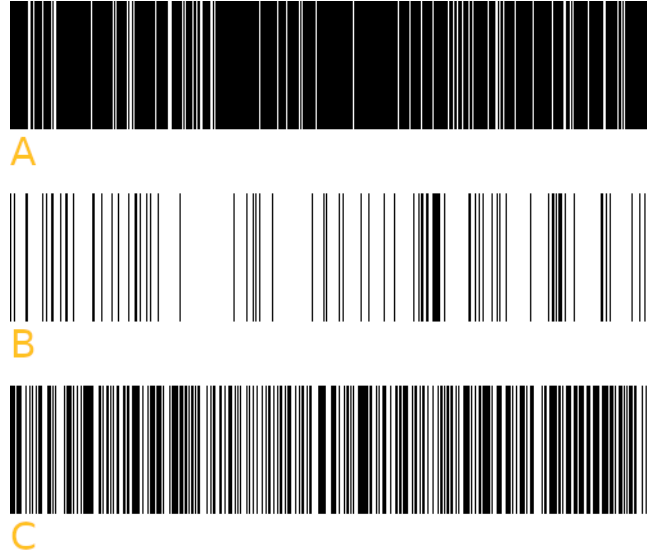


Figura 17: Autômatos criados a partir das matrizes estocásticas

Seguindo com o a geração de autômatos, construímos a matriz estocástica apresentada na tabela 5 e a partir dela foi possível a construção da Figura 18 que apresenta os diferentes níveis gerados pelo autômato. Pode-se notar que há muitos zeros e uns no decorrer dos passos de tal autômato. Isso ocorre porque a única "passagem" para os valores maiores que 1, vem justamente do nó 1 e como a probabilidade de se sair do nó zero e ir para o nó 1 é baixa, todos os outros valores têm uma pequena possibilidade de aparecer. No entanto, uma vez que essa "barreira" é transpassada, os outros valores aparecem em sequência até que se volta ao nó zero e a "barreira" age novamente.

D	0	1	2	3	4	5
0	0.9	0.882	0	0	0	0.01
1	0.1	0.098	0	0	0	0
2	0	0.2	0.2	0.194	0	0
3	0	0	0.8	0.776	0	0
4	0	0	0	0.03	0.5	0.495
5	0	0	0	0	0.5	0.495

Tabela 5: Matriz estocástica D.

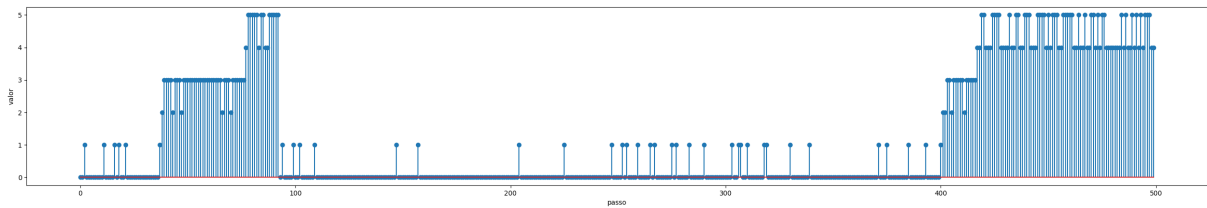


Figura 18: Stem plot a partir do automato gerado pela tabela 5

Fomos capazes de alterar a implementação do gerador de autônomos genéricos afim de construir um encadeamento de autônomos binários. Esse encadeamento pode ser conferido na Tabela 6. A alteração necessária tem a ver com paridade do nó atual. O valor de cada nó é zero se o nó for ímpar e 1 se o nó for par. O passeio de 500 passos por tal autônomo pode ser conferido na Figura 19

E	0	1	0	1	0	1
0	0.9	0.882	0	0	0	0.01
1	0.1	0.098	0	0	0	0
0	0	0.2	0.2	0.194	0	0
1	0	0	0.8	0.776	0	0
0	0	0	0	0.03	0.5	0.495
1	0	0	0	0	0.5	0.495

Tabela 6: Matriz estocástica E.

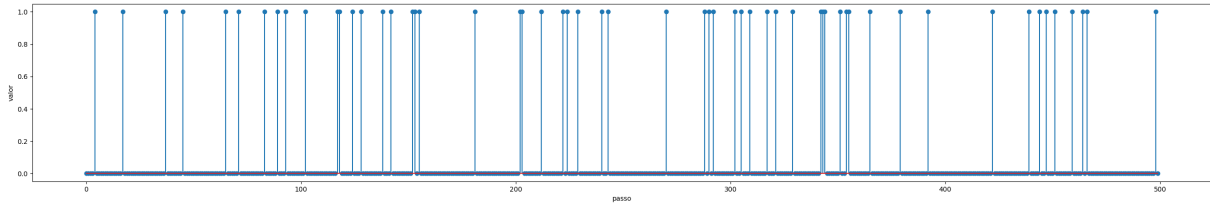


Figura 19: Stem plot a partir do automato gerado pela Tabela 6

Nesse autônomo, podemos ver que os valores se resumem a zeros e uns, porém seguindo ainda as probabilidades da Tabela 5. Novamente os valores de zero se sobressaem sobre os valores de um devido ao mesmo mecanismo de "barreiras" já discutido.

Outras formas de observarmos os resultados dos autômatos das Tabelas 2, 3 e 4, podem ser conferidos nas figuras abaixo:

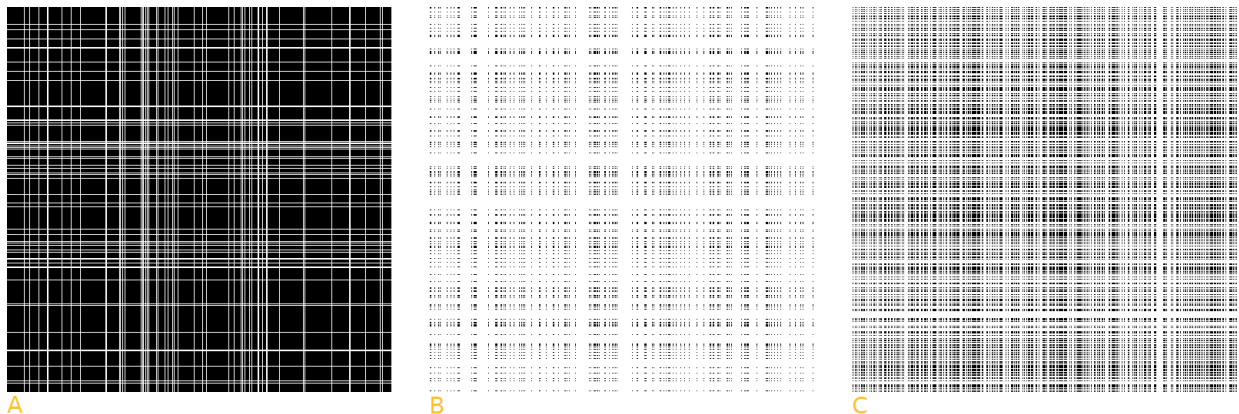


Figura 20: Figura similar à Figura 18 com sobreposição em 90 graus das linhas de cada passo do automato

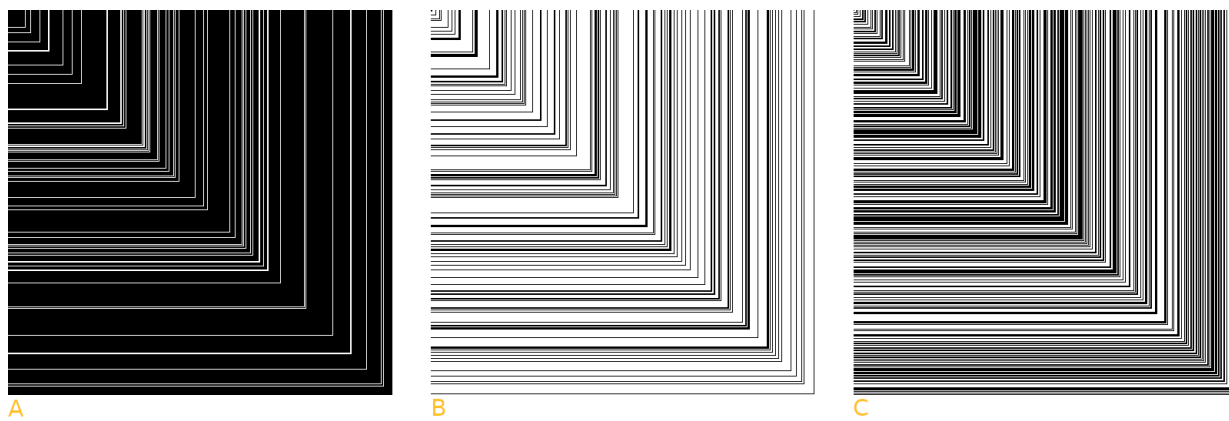


Figura 21: Variação da sobreposição de cruces da Figura 21

Por fim, utilizando ainda os autômatos das Tabelas 2, 3 e 4, calculamos a densidade relativa de 1's para 200, 500, 750, 1000 e 2000 passos. Essas densidades podem ser observadas na Figura 22

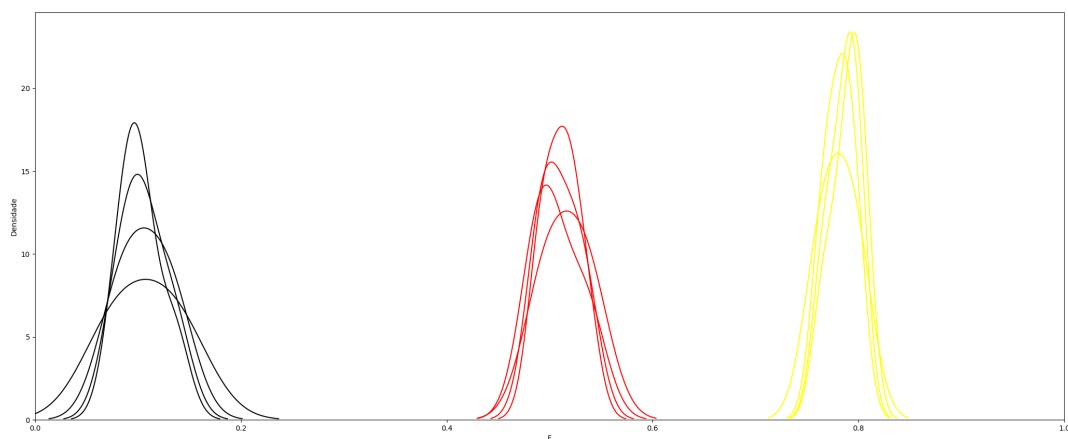


Figura 22: Densidade relativa de 1's para diferentes quantidades de passos. Os autômatos preto, amarelo e vermelho são respectivos às Tabelas 2, 3 e 4