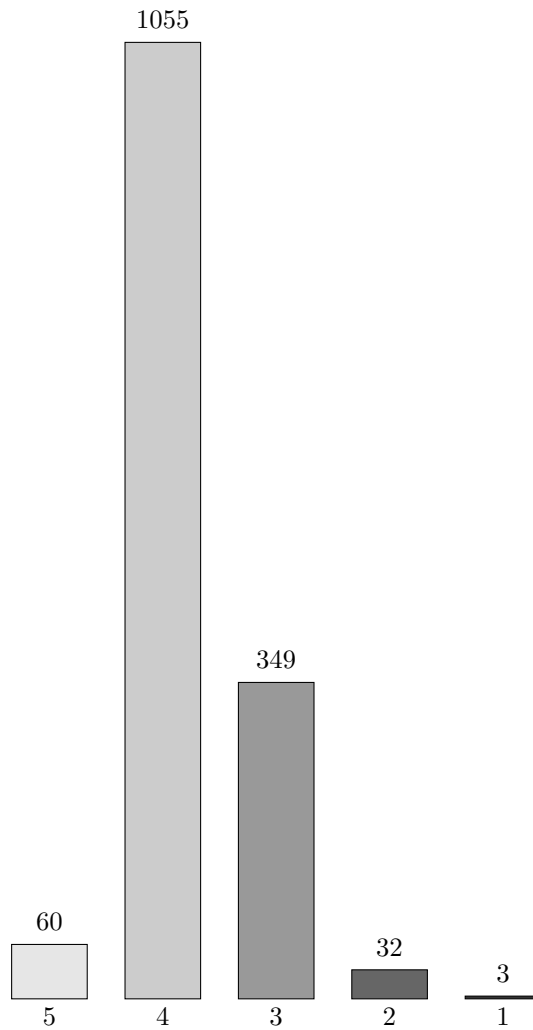


Entropia de dados

Kauan Toledo Camargo

1 Database utilizada:

Foi-se utilizado como base de dados qualitativos uma lista de jogos eletrônicos datados desde 1980 até 2023, contendo datas de lançamento, nota média e críticas. Neste trabalho iremos analisar a entropia das notas médias de usuários entre esses 1499 jogos.



Os dados foram arredondados para facilitar o cálculo.

<https://www.kaggle.com/datasets/arnabchaki/popular-video-games-1980-2023>

2 Calculando a entropia de dados

$$P(nota5) = \frac{60}{1499} = 0.04$$

$$P(nota4) = \frac{1055}{1499} = 0.704$$

$$P(nota3) = \frac{349}{1499} = 0.233$$

$$P(nota2) = \frac{32}{1499} = 0.021$$

$$P(nota1) = \frac{3}{1499} = 0.002$$

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i)$$

$$H(X) = - (0.04 \log_2 0.04 + 0.704 \log_2 0.704 + 0.233 \log_2 0.233 \\ + 0.021 \log_2 0.021 + 0.002 \log_2 0.002)$$

$$H(X) = - (0.04(-4.644) + 0.704(-0.506) + 0.233(-2.102) \\ + 0.021(-5.573) + 0.002(-8.966))$$

$$H(X) = 1.168$$

$$H_{max} = \log_2 5$$

$$H_{max} = 2.322$$

Conclusão: Os dados possuem um grau médio de aleatoridade, se distanciando de forma moderada da distribuição uniforme comparados com a entropia máxima possível.

3 Python - entropy.py

Código:

```
import math

entropy = 0.0
total = 0
classes = []
n = int(input('Quantidade de classes: '))

for i in range(n):
    classes.append(int(input(f"Valor da classe {i+1}: ")))
    total += classes[i]

for i in range(n):
    probability = classes[i] / total
    entropy -= probability * math.log2(probability)

maxEntropy = math.log2(n)

print(f'Valor Total: {total}')
print(f'Entropia dos dados: {entropy:.3f}')
print(f'Entropia máxima: {maxEntropy:.3f}')
```

Console:

```
Quantidade de classes: 5
Valor da classe 1: 3
Valor da classe 2: 32
Valor da classe 3: 349
Valor da classe 4: 1055
Valor da classe 5: 60
Valor Total: 1499
Entropia dos dados: 1.168
Entropia máxima: 2.322
```