

Rio de Janeiro, 22 de Setembro de 2017.

Trabalho 3 de Fundamentos de Data Science

PROFESSOR: EDUARDO LABER

Considere uma fonte de geração de números aleatórios que gera números

$$\min\{\lfloor 1/z^2 \rfloor, 100.000.000.000\},$$

onde  $z$  é um número aleatório gerado uniformemente entre  $[0, 1]$ .

Para as questões abaixo considere  $N = 10^{10}$  e sempre utilize a mesma seed para geração das sequências de números aleatórios.

Passo 1 Gere uma sequência de tamanho  $N$  e calcule o total de números distintos criados.

Passo 2 Gere uma sequência de tamanho  $N$  e estime o total de números distintos criados através da abordagem vista na aula. Estude o impacto de variar o número de funções de hash.

Passo 3 Gere uma sequência de tamanho  $N$  e calcule o surprise number exato da sequência.

Passo 4 Gere uma sequência de tamanho  $N$  e estime o surprise number via o AMS. Estude o impacto de variar o número de funções de hash.

Passo 5 Repita o passo 5 assumindo que o valor de  $N$  não é conhecido a priori.

Escreva um breve relatório explicando as estruturas de dados utilizadas, como a experimentação foi realizada, os resultados encontrados e as principais observações.