# Estimando o número de subdomínios

Stanley Cortes de Sousa

June 25, 2020

## Queremos estimar o número de subdomínios no formato formato $[a-z](k)$

Seja $X$ o conjunto de todas palavras no formato $[a-z](k)$, onde $k$ é a quantidade máxima de caracteres. Considere $Y$ como o subconjunto de $X$ formado pelas palavras que são domínios válidos da UFRJ, dados por http://www.$X_i$.ufrj.br

$W$ é outro subconjunto de $X$, onde $W_i = f(j)$. Seja $j$ um número com até $k$ dígitos em base 26, $f(j)$ mapeia $j$ à uma palavra de $X$ com até $k$ letras. Dada a tabela 0=a ... $P(25)$=z, cada dígito de $j$ é mapeado por $f$ em uma letra, assim:

$j =$ 0 $\longmapsto i = 0,\quad W_0 = f(0) =$ a

$j =$ B $\longmapsto i = 11,\quad W_{11} = f(B) =$ k

$j =$ 20 $\longmapsto i = 52,\quad W_{52} = f(20) =$ ba

$j =$ 2C $\longmapsto i = 64,\quad W_{64} = f(2C) =$ bl

$j =$ A45D $\longmapsto i = 178607,\quad W_{178607} = f(A45D) =$ jefm

$j =$ PPPP $\longmapsto i = 456975,\quad W_{456975} = f(PPPP) =$ zzzz

Note a representação numérica de $i$ e $j$ pode descartar '0's à esquerda, neste caso $W$ conteria a palavra 'a' mas não as sequências com 'a' à esquerda, portanto:

$|X| = 26^k$

$|W| = 26^k - 26^{k-1} ... - 26^0$

Entretanto, considere uma quantidade $n$ de palavras de $X$, se $n < |W|$ podemos utilizar $n$ amostras de $W$ para representar $X$, sem prejuízo, já que $W \in X$

Por fim, seja H a fração de palavras de $X$ que são domínios da UFRJ, H é dado por:

$H = \dfrac{|Y|}{|X|}$ (I)

Seja $Z$ uma variável aleatória uniforme em $[0, 1]$, e seja $g(z) = i$:

$g(z) = int(max_i * z) + 1$

$W_i = f(i_{26})$

Seja $h(W_i)$ uma função indicadora se a palavra $W_i$ pertence a $Y$, ou seja, $h(W_i)$ verifica se http://www.$W_i$.ufrj.br é um endereço válido

De (I) sabemos que:

$E[h(X_i)] = \dfrac{|Y|}{|X|}$, estamos assumindo que para $n < |W| \longmapsto E[h(X_i)] = E[h(W_i)]$

Precisamos estimar $E[h(X_i)]$, pela lei dos grandes números sabemos que:

$M_n \to E[h(X_i)] = \dfrac{|Y|}{|X|}$, e

$M_n = \dfrac{1}{n}\sum_{i=1}^{n} h(X_i)$, por definição

Logo:

$\dfrac{|Y|}{|X|} \to \dfrac{1}{n}\sum_{i=1}^{n} h(X_i)$

$|Y| \to \dfrac{|X|}{n}\sum_{i=1}^{n} h(X_i)$