

## Lista 5: Fundamentos Estatísticos para Ciência dos Dados

Ricardo Pagoto Marinho

3 de abril de 2018

1. O código e os dados estão em <https://github.com/ricardopmarinho/UFGM/tree/master/M>.

O problema proposto consiste em olhar o perfil de tráfego em uma rede Wi-Fi durante a utilização do site *Facebook*. Desta forma, olharei para o protocolo utilizado em cada pacote, *i.e.*, TCP, UDP, DNS, ARP, etc. Além disto verificarei o tamanho de cada pacote. Para isso, utilizarei o programa *Wireshark*, que captura todos os pacotes que estão passando (tanto de entrada como de saída) em uma interface do computador analisado. Dentre os vários protocolos que possivelmente são utilizados em uma rede, os da camada de transporte tentem a se repetir mais vezes, já que usualmente são utilizados apenas dois: TCP e UDP, enquanto que nas outras, a quantidade de protocolos pode variar mais. Além disso, olharei o tamanho dos pacotes em bytes. O site é acessado antes de começar o monitoramento, logo, os pacotes maiores, que contem as imagens e dados do site, não serão capturados.

- Protocolos

Foram identificados 13 protocolos no teste com as seguintes frequências:

cias:	Protocolo	Frequência	Considerando a distribuição
	ARP	2	
	DB-LSP-DISC	8	
	DHCPv6	1	
	DNS	4	
	ICMP	4	
	LLMNR	4	
	MDNS	2	
	NBNS	3	
	QUIC	20	
	SSL	5	
	TCP	28	
	TLSv1.2	17	
	UDP	2	

dos dados como uma Poisson com  $\lambda = 5$  pois é sabido que a chegada de pacotes em uma rede segue essa distribuição e dividindo os protocolos em grupos de 2 (com exceção do UDP, que ficou sozinho), os conjuntos de protocolos e suas frequências fo-

ram:	Protocolos	Frequência	Utilizando a distri-
	ARP& DB-LSP-DISC	10	
	DHCPv6& DNS	5	
	ICMP& LLMNR	8	
	MDNS& NBNS	5	
	QUIC& SSL	25	
	TCP& TLSv1.2	45	
	UDP	2	

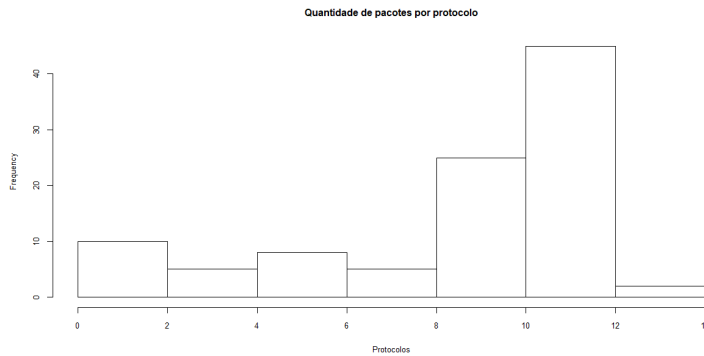


Figura 1: Protocolos

buição de Poisson com a configuração dita, os valores esperados foram:

Protocolos	Frequência
ARP& DB-LSP-DISC	12.4
DHCPv6& DNS	40.6
ICMP& LLMNR	49.7
MDNS& NBNS	31.5
QUIC& SSL	11.9
TCP& TLSv1.2	2.9
UDP	0.5

A Figura 1, mostra a distribuição em relação aos protocolos.

- Tamanho dos pacotes Para o tamanho dos pacotes, foram divididos em 14 grupos, variando de 0 a 1400 bytes, ou seja, cada grupo

Janela	Frequência
[0,100]	70
(100,200]	7
(200,300]	9
(300,400]	5
(400,500]	0
(500,600]	2
(600,700]	0
(700,800]	0
(800,900]	0
(900,1000]	0
(1000,1100]	0
(1100,1200]	0
(1200,1300]	0
(1300,1400]	7

é uma janela de 100 bytes. Suas frequências foram:

Utilizando a distribuição de Poisson com  $\lambda = 50$ , os valores espe-

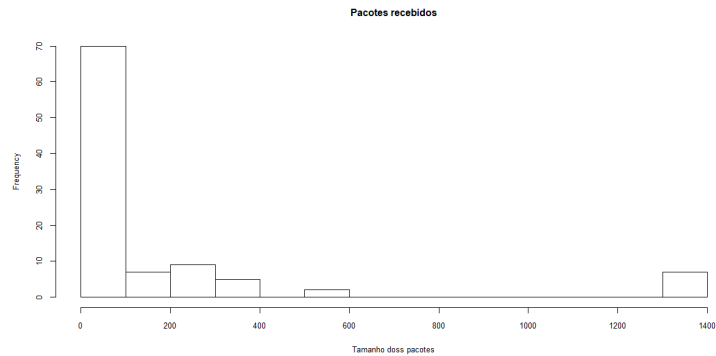


Figura 2: Tamanho dos pacotes

	Janela	Frequência
	[0,100]	100
	(100,200]	$3 \times 10^{-8}$
	(200,300]	0
	(300,400]	0
	(400,500]	0
	(500,600]	0
rados foram:	(600,700]	0
	(700,800]	0
	(800,900]	0
	(900,1000]	0
	(1000,1100]	0
	(1100,1200]	0
	(1200,1300]	0
	(1300,1400]	0

A Figura 1 mostra a distribuição das frequências dos tamanhos dos protocolos.

Olhando para as figuras e para os resultados, podemos concluir que a distribuição utilizada não foi a mais adequada, fazendo que um olhar melhor sobre o tema se faça necessário.

2.
  - 4.3-a) 0.6666667
  - 4.3-b) 0.0625
  - 4.3-c) 0.1428
  - 4.14-a)  $\text{cdf} = \sum_{x=0}^1 20x^3 \times (1-x)$
  - 4.14-b) Olhando a Figura 2,  $P(X \leq \frac{2}{3}) = 0.5$
3.
  -

- o ponto em que  $f(x)$  assume o valor máximo varia, logo depende de  $\delta$ . A altura também depende.
  - $\text{pnorm}(10+2*\text{sqrt}(5),10,5)-\text{pnorm}(10-2*\text{sqrt}(5)-0.01,10,5)$   
0.629441
  - Utilizei a configuração  $N(2,3)$ . As distribuições se parecem sim, a altura que difere.
4. Todos os valores são iguais a 0.87297962. Nenhum valor ficou menor do que 0.05
5.  $E(X) = \frac{1}{\lambda} = \frac{1}{1/3} = 3$

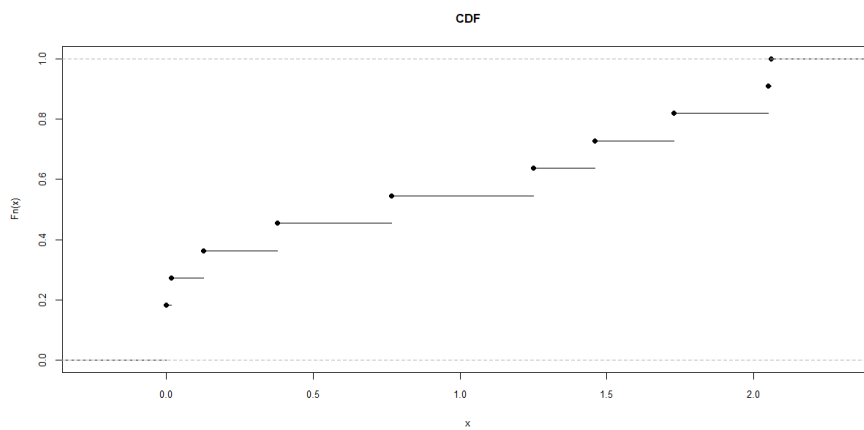


Figura 3: CDF

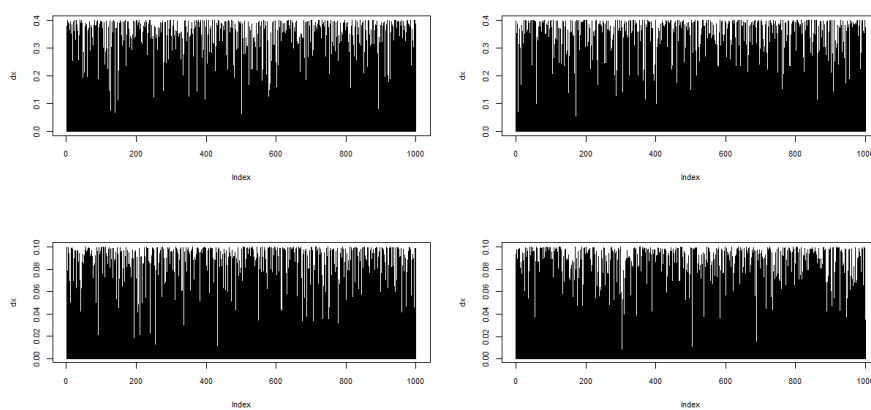


Figura 4: Distribuição normal

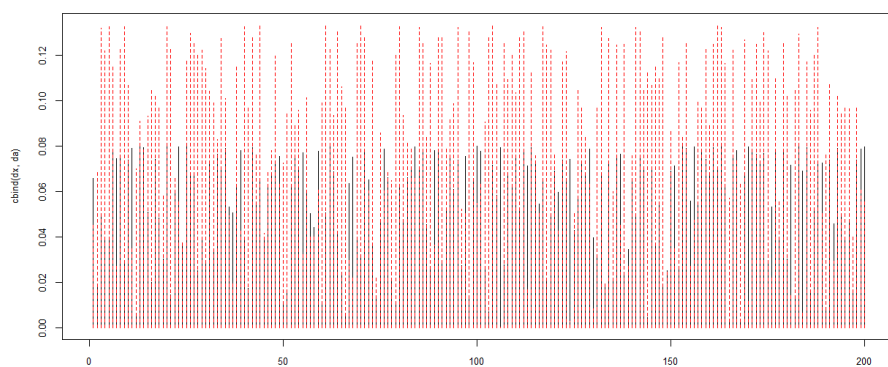


Figura 5: Sobreposição