MPEI 2018-2019

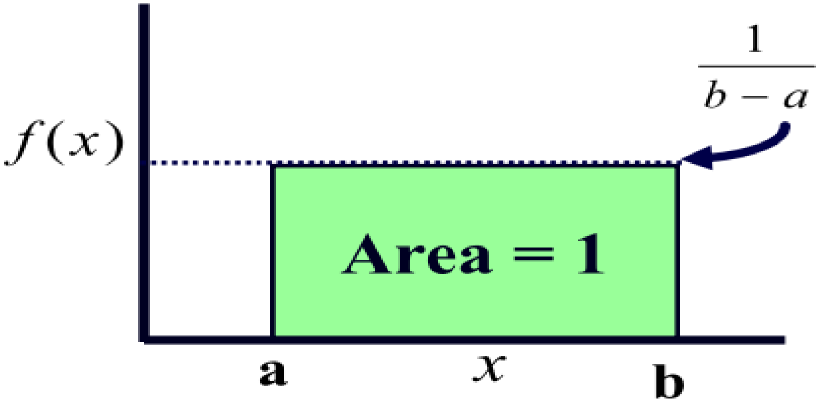
Aula 7

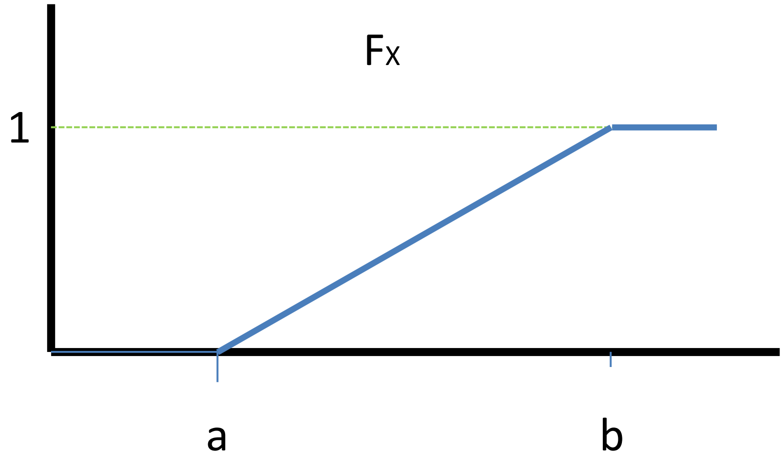
Distribuições (continuação)

Distribuições contínuas

Variável aleatória uniforme

- U(a,b) é definida por:
- $f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & caso\ contrário \end{cases}$
- $F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$
- $E[X] = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$



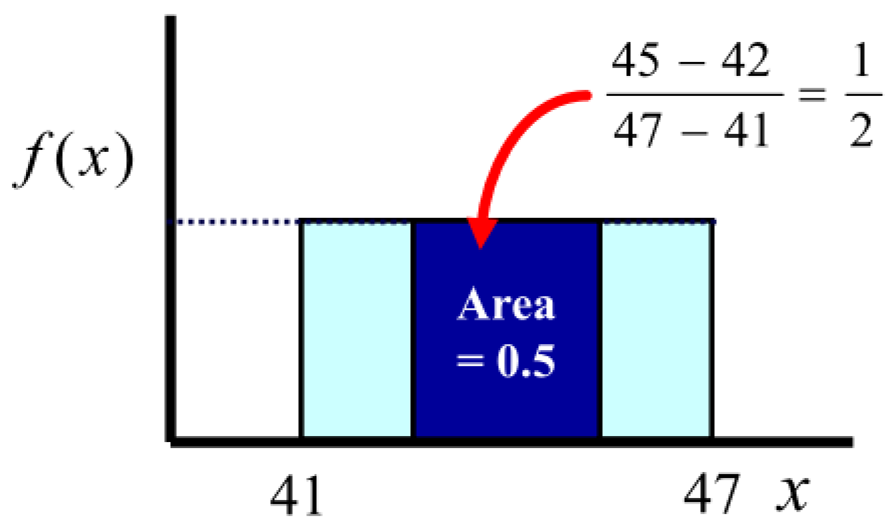


Exemplo

- P(42<=X<=45)   com U(41,47)

$$P(x_1 \le X \le x_2) = \frac{x_2 - x_1}{b - a}$$

$$P(42 \le X \le 45) = \frac{45-42}{47-41} = \frac{1}{2}$$



Função rand() do Matlab

- A função rand() do Matlab gera números obedecendo a uma distribuição uniforme

  – Com $a = 0$ e $b = 1$

- Para ter U(a,b) basta usar:

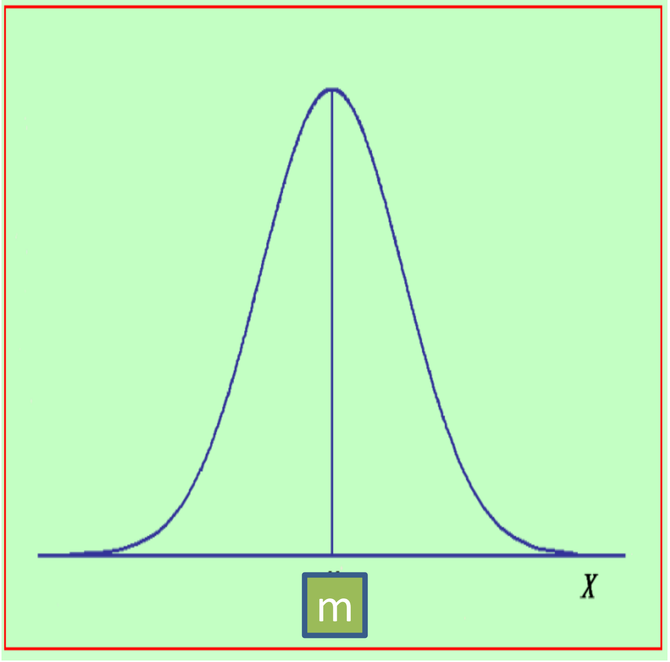
a+rand()*(b-a)

Variável aleatória Normal (ou Gausseana)

- Uma V.A. diz-se normal ou Gausseana se

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

- Frequentemente usa-se a notação $N(m, \sigma^2)$
- Curva em forma de sino, simétrica em torno da média (m) e com alargamento σ



Variável Normal

- Função de distribuição acumulada

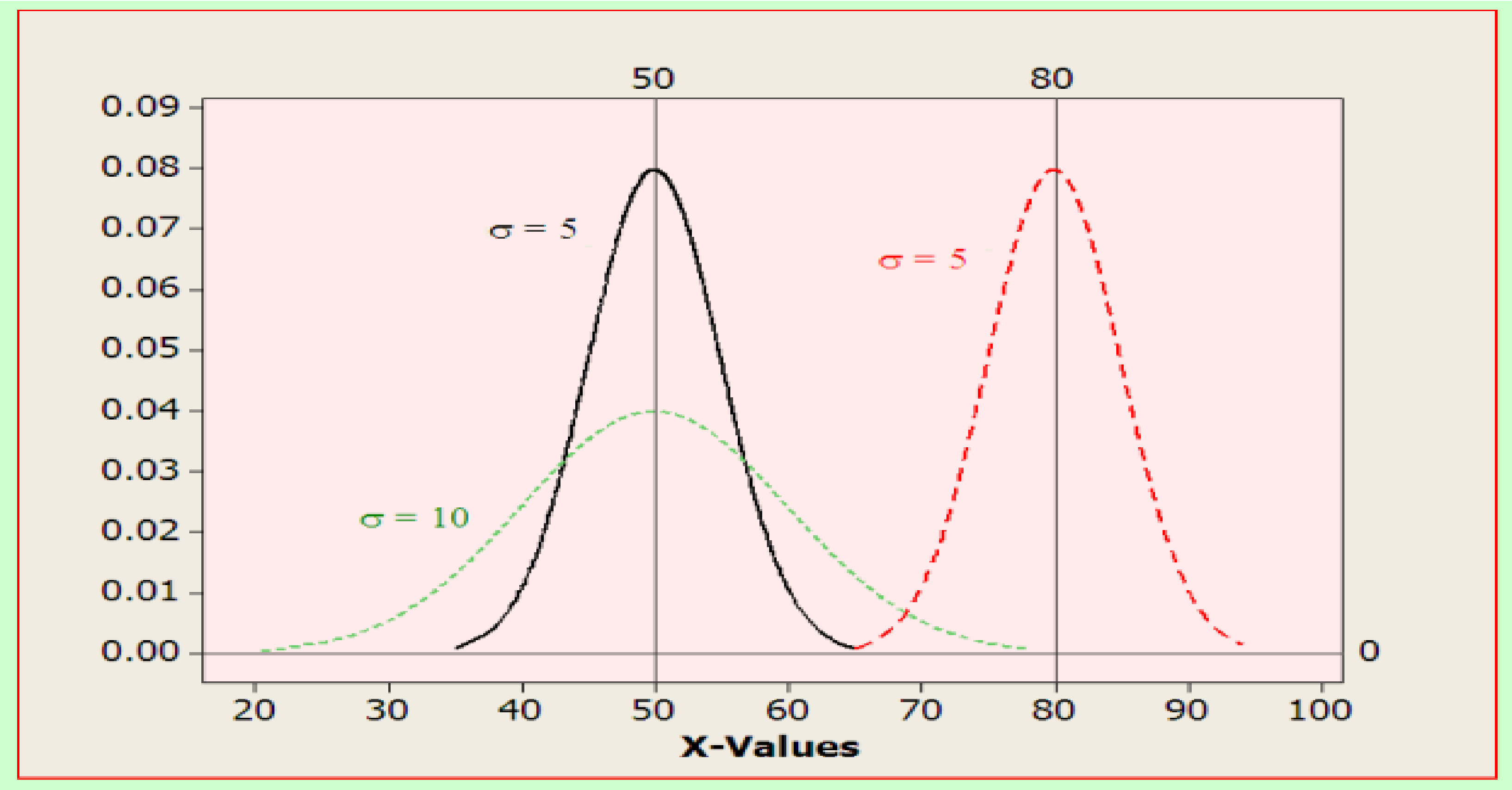
$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt$$

- $E[X] = m$
- $Var(X) = \sigma^2$

- Nota: é muito comum utilizar-se µ em vez de m para representar a média

Família de curvas

- Variando os 2 parâmetros …



Gaussiana normalizada

- Como existe um número infinito de combinações para m e σ pode gerar-se uma família infinita de curvas
  - Sendo pouco prático lidar com esta situação, em especial antes da existência de computadores

- Foi desenvolvido um mecanismo pelo qual qualquer distribuição normal pode ser convertida numa distribuição única, a Gaussiana normalizada $N(0,1)$

- A fórmula de conversão é:

$$\mathrm{z} = \frac{x-m}{\sigma}$$

Ou seja, subtrair a média e dividir pelo desvio padrão

Gaussiana normalizada

- Função densidade de probabilidade:

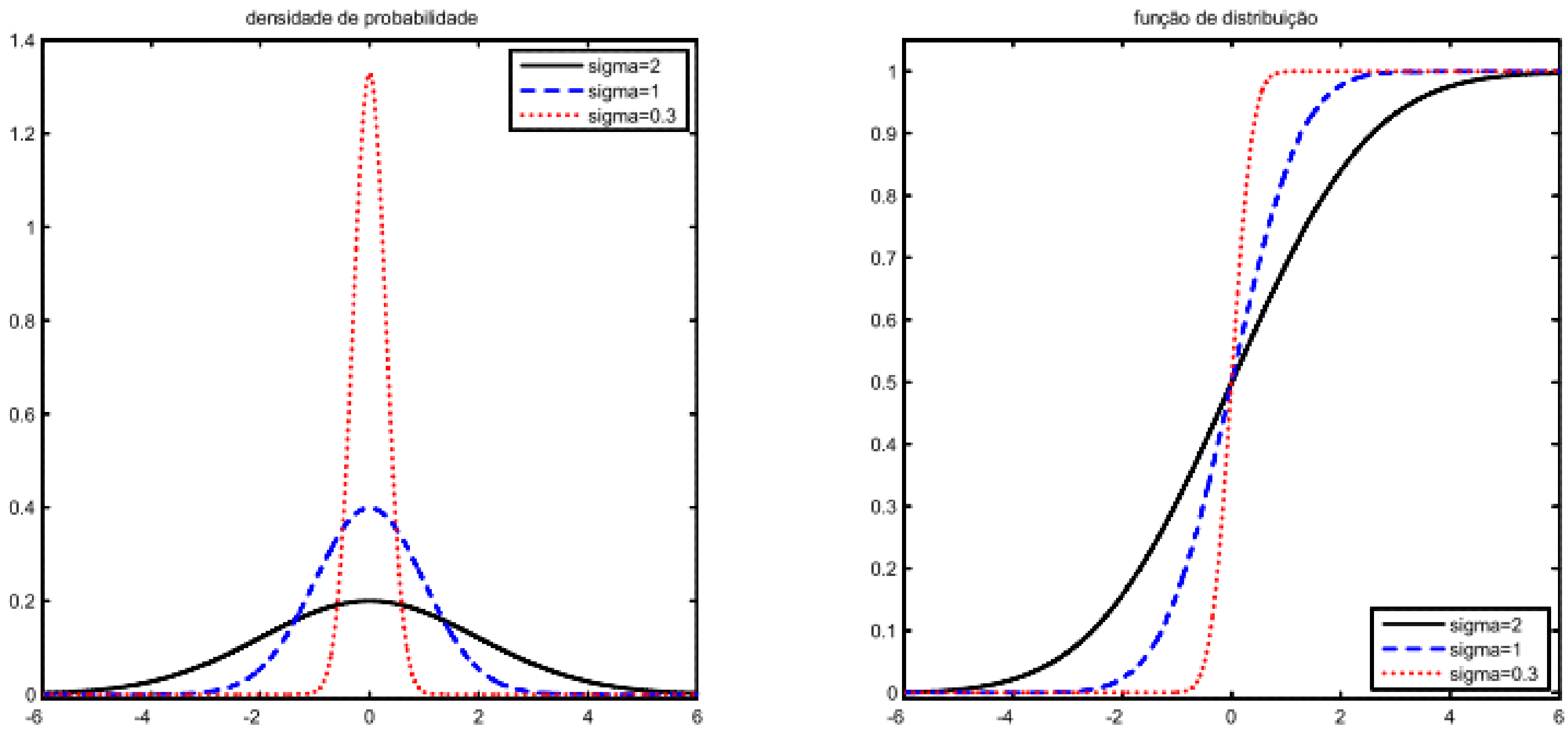
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x)^2}{2}}$$

- A função de distribuição acumulada $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t)^2}{2}} dt$$

- $\Phi(x)$ encontra-se frequentemente tabelada
  - Outras tabelas comuns são as de $Q(x) = 1 - \Phi(x)$

Gaussiana normalizada



Função distribuição acumulada

- A função de distribuição (acumulada) de N(m,σ^2) pode ser expressa em termos de $\Phi(x)$

$$F_X(x) = \Phi\left(\frac{x - m}{\sigma}\right)$$

Exemplo de uso de $Q(x)$

- Uma empresa, monopolista do mercado de um determinado produto, tem uma procura mensal *X* que segue uma distribuição normal N(75,100). Determine P[78<*X*<80]

$$
\begin{aligned}
P\left[78 \leq X \leq 80\right] &= P\left[\frac{78-75}{10} \leq U_X \leq \frac{80-75}{10}\right] = \\
&= P\left[0.3 \leq U_X \leq 0.5\right] = Q(0.3) - Q(0.5) = 0.074
\end{aligned}
$$

Distribuição Normal

- É muito provavelmente a mais conhecida e utilizada de todas as distribuições (contínuas)
- Adequa-se/ajusta-se a muitas características humanas
  - Altura, peso, velocidade, resultados de testes de inteligência, esperança de vida…
- Também se adequa a muitas outras coisas da natureza
  - Árvores, animais etc têm muitas características que seguem a distribuição normal
- Surge quando vários efeitos acumulados e independentes se sobrepõem

Distribuição Normal e a Binomial

- Demonstra-se que a Binomial de média $\mathrm{m} = np$ e $\sigma^2 = np(1-p)$ com m não muito pequeno e n elevado pode ser aproximada por:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-m)^2}{2\sigma^2}}$$

- Ou seja a distribuição normal
  - Desde que $m = np$ e variância igual $np(1-p)$

Distribuição exponencial

- Surge frequentemente em problemas envolvendo filas de espera e fiabilidade
  - Exemplos:
    - Tempo até um computador avariar
    - Tempo entre chegada de utentes à urgência de um Hospital

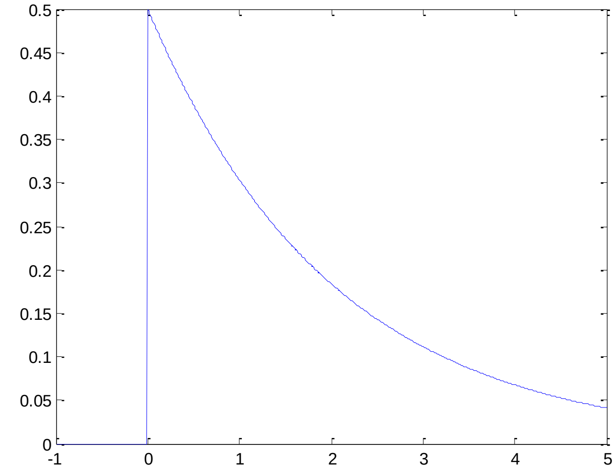
- É não negativa (prob. 0 para x<0)
- Está relacionada com a distribuição (discreta) de Poisson
  - Se o número de acontecimentos que ocorrem num intervalo seguem distribuição de Poisson, o tempo entre eles segue distribuição exponencial

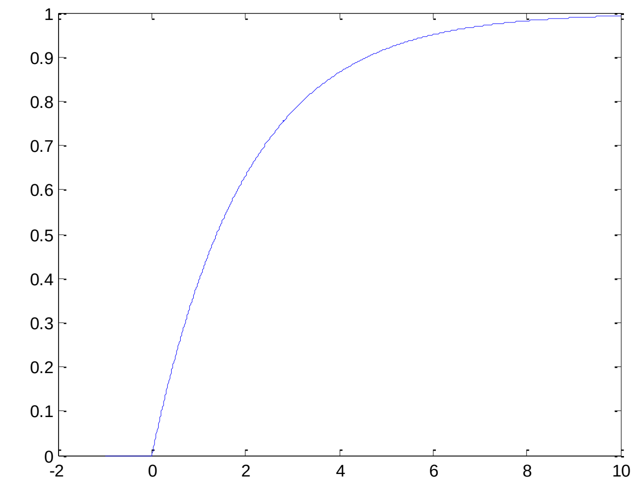
Distribuição exponencial

- $f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$

- $F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$

- $E[x] = \frac{1}{\lambda}$
- $\mathrm{Var}(x) = \frac{1}{\lambda^2}$





Exemplo de aplicação

- A vida útil, em milhares de horas, de um componente de um robô é uma variável aleatória com distribuição exponencial de valor médio 10 (milhares de horas)
- Qual a probabilidade de um desses componentes selecionado ao acaso durar menos de 4000 horas?
- $\lambda = \frac{1}{10} = 0{,}1$

$$P\left[X < 4\right] = \int_0^4 0.1 e^{-0.1x} dx = F_X(4) = 0.33$$

Outras distribuições

Distribuição dos primeiros dígitos

- Em 1881, um matemático e astrónomo americano, Simon Newcomb, percebeu que as primeiras páginas dos livros de logaritmos das bibliotecas estavam mais gastas que o resto, intrigado, investigou o assunto e...
- percebeu que em amostras aleatórias de dados reais o dígito 1 aparece quase 1/3 das vezes
  - Em lugar dos 1/9 se seguissem uma distribuição uniforme (discreta)

- Mais tarde, em 1938, o físico **Frank Benford** após uma investigação mais profunda chegou à mesma conclusão que Newcomb, indo mais além aplicando a fórmula numa variedade de números

Lei/Distribuição de Benford

- Função probabilidade →

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(\frac{d+1}{d}\right) = \log_{10}\left(1+\frac{1}{d}\right).$$

d	*P*(*d*)	Valor Relativo de P(d)
1	30.1%	
2	17.6%	
3	12.5%	
4	9.7%	
5	7.9%	
6	6.7%	
7	5.8%	
8	5.1%	
9	4.6%	

- A Lei/Distribuição de Benford, também conhecida como a "Lei dos Primeiros Dígitos", é uma ferramenta muito poderosa e muito simples que aponta suspeitas de fraudes, erros de digitação etc

- Mais info:
  – https://pt.wikipedia.org/wiki/Lei_de_Benford
  – http://gigamatematica.blogspot.pt/2011/07/lei-de-benford.html

Lei/Distribuição de Zipf

- George Kingsley **Zipf**, linguista da Universidade de Harvard, analisou a obra monumental de James Joyce, *Ulisses*, e contou as palavras distintas, ordenando-as por frequência

- Verificou que:
  - a palavra mais comum surgia 8000 vezes;
  - a décima, 800 vezes;
  - a centésima, 80 vezes;
  - a milésima, 8 vezes.

Lei de Zipf

- A **Lei de Zipf** é uma lei empírica que rege a dimensão, importância ou frequência dos elementos de uma lista ordenada
  - formulada na década de 1940 por Zipf, na sua obra *Human Behaviour and the Principle of Least-Effort* ("Comportamento Humano e o Principio do Menor Esforço"),

- Trata-se de uma lei de potências sobre a distribuição de valores de acordo com o nº de ordem numa lista.
  - Numa lista, o membro n teria uma relação de valor com o 1º da lista segundo 1/n.

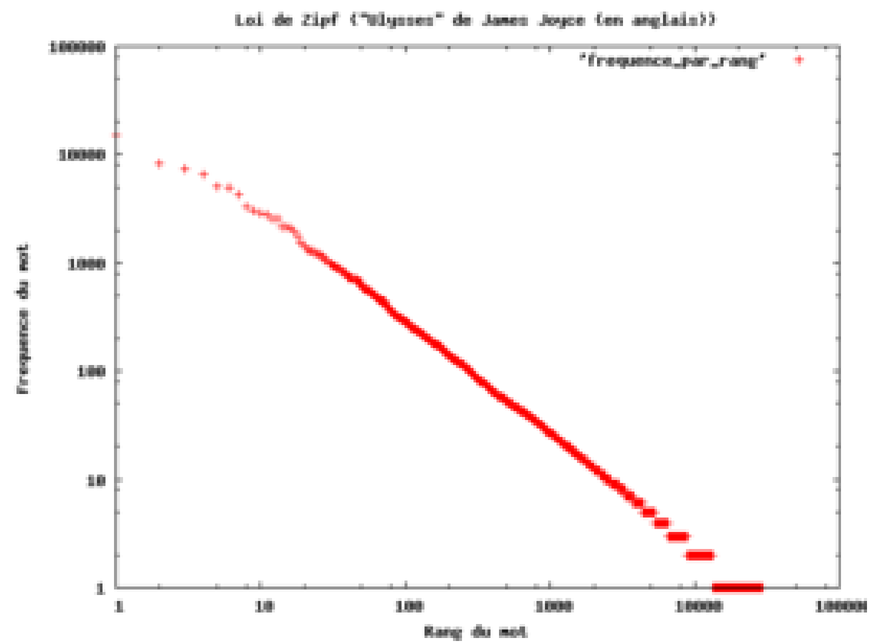
- Mais info: https://pt.wikipedia.org/wiki/Lei_de_Zipf

Lei de Zipf

- A lei de Zipf prevê que num dado texto, a probabilidade de ocorrência *p*(*n*) de uma palavra esteja ligada à sua ordem *n* na ordem das frequências por uma lei da forma:

$$p(n) = \frac{K}{n}$$

- Sendo K uma constante dependente da língua
- $p(n)$ é estimada com base na contagem de ocorrências de palavras num texto ou conjunto de textos



Frequência das palavras em função da ordem na versão original de Ulisses *de James Joyce*.
De: Wikipedia

Lei de Zipf – Inglês escrito

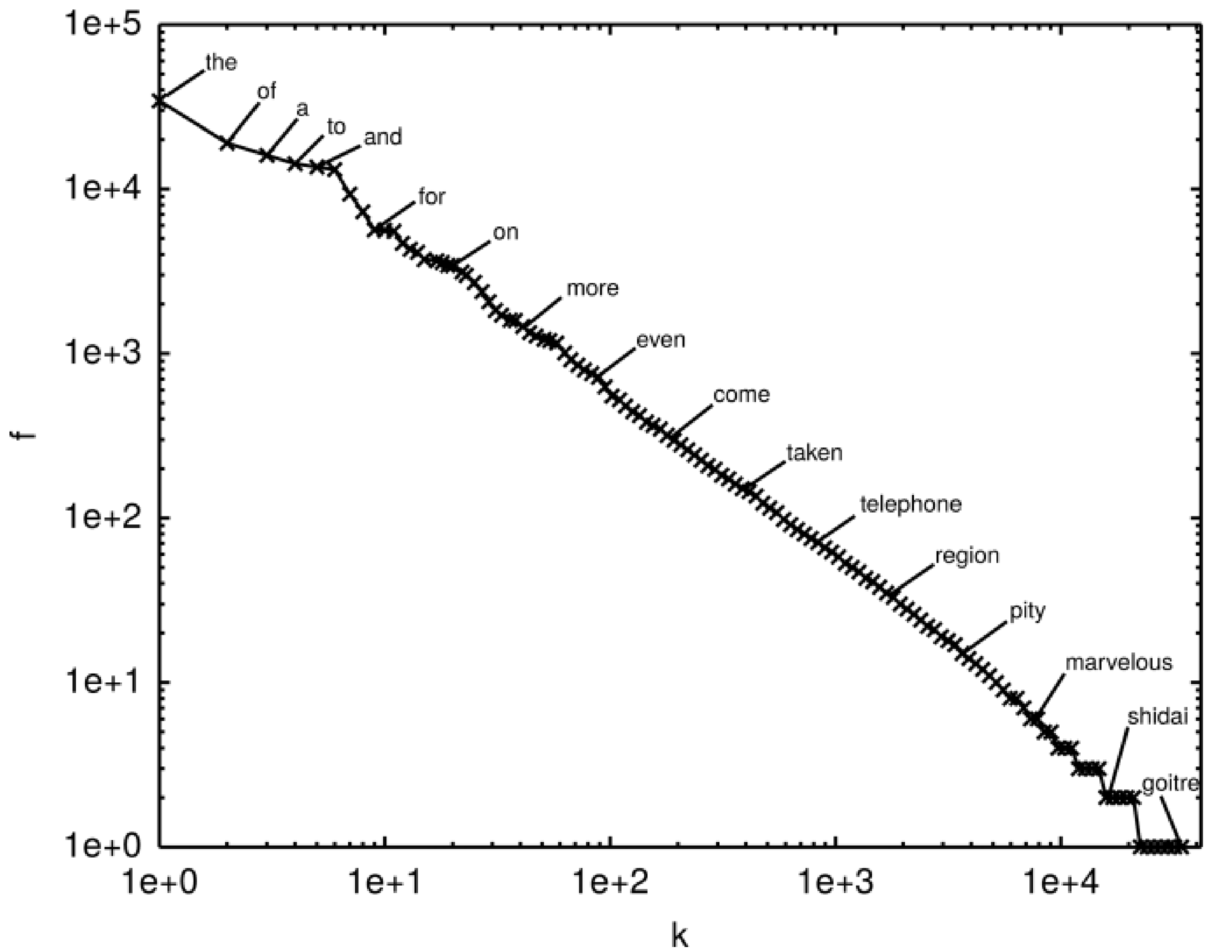


Figura 1: Lei de Zipf no Inglês escrito (dados do OANC). *Rank* (k) versus frequência de ocorrência (f).

Exemplo de aplicação da Lei de Zipf

- Aplicação na área da segurança:

Revista Brasileira de Ensino de Física, vol. 38, nº 1, 1313 (2016)
www.scielo.br/rbef
DOI: http://dx.doi.org/10.1590/S1806-11173812125

Artigos Gerais

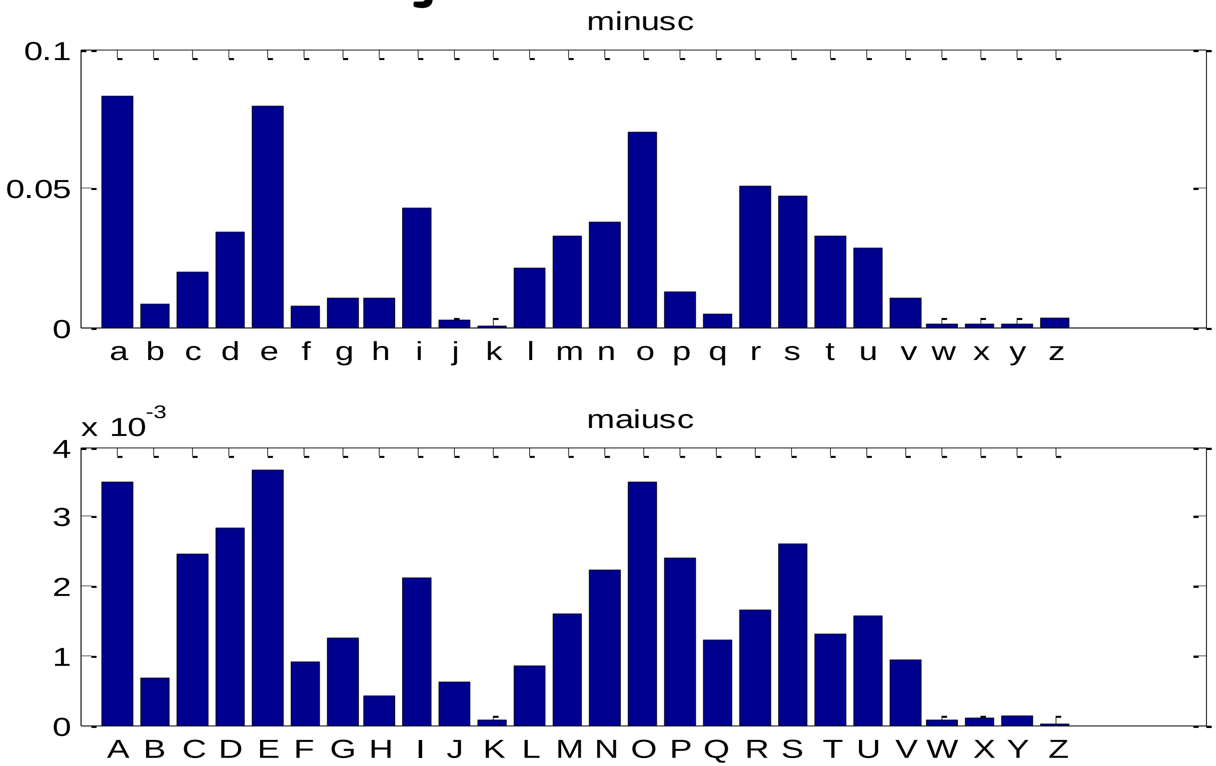
Influência da lei de Zipf na escolha de senhas

Influence of Zipf's law on the password choices

Leonardo Carneiro de Araújo*[1], João Pedro Hallack Sansão[1], Hani Camille Yehia[2]

- Artigo em PDF disponível em http://www.scielo.br/pdf/rbef/v38n1/1806-9126-rbef-38-01-S1806-11173812125.pdf

Outra distribuição: Distribuição das letras em Português …



char	prob.
a	0.083
b	0.008
c	0.020
d	0.034
e	0.079
f	0.007
g	0.010
h	0.011
i	0.043

…..

- Probabilidades estimadas usando o texto pg21209.txt do projecto Gutemberg

Demo Matlab

Mais informação

- Material online
  - Slides relativos aos cap. 5 e 6 do livro “Business Statistics”, Ken Black, 4ed
    - http://business.uni.edu/slides/ECON-1011_Luk/ch05.pdf
    - http://business.uni.edu/slides/ECON-1011_Luk/ch06.pdf
  - Lectures:
    - http://www.stat.berkeley.edu/~stark/SticiGui/Text/randomVariables.htm
  - Wikipedia
- Capítulo 3 do livro “Probabilidades e Processos Estocásticos”, F. Vaz