Métodos Probabilísticos para Engenharia Informática

2020-2021

Probabilidade

- Qual a hipótese de chover amanhã?
- Qual a possibilidade de chegar a horas à aula ?
- Qual a probabilidade de eu ganhar o Euromilhões (ou um de vocês) ?

- São questões que colocamos frequentemente…
- e estão relacionadas com o **incerto / não determinístico**

Aleatório

- Em termos qualitativos, “qualquer coisa” que não seja predizível com certeza absoluta

- Acontecimento (evento) cujo resultado não possa ser determinado com certeza absoluta.
  – Caso contrário é determinístico

- adj. Que repousa sobre um acontecimento **incerto**, fortuito: contrato aleatório.
  Diz-se de uma grandeza que pode tomar certo número de valores, a cada um dos quais está ligada uma probabilidade.
    - De: **dicionário online de português**

- http://www.priberam.pt/dlpo/aleat%C3%B3rio

Então qual o interesse ?

- Qual o interesse em estudar algo que não se pode prever ?

- Na maioria das aplicações existe algum tipo de regularidade que se manifesta se o número de observações / experiências for elevado

Problema Exemplo 1

- Qual a probabilidade de acertar num PIN/**password** de 4 dígitos escolhendo um PIN completamente ao acaso?

E de 20 dígitos ?

Problema Exemplo 2

- Qual a probabilidade de 80% ou mais dos alunos de MPEI deste ano letivo, que não reprovem por faltas, ter sucesso na UC?

- **O que precisamos saber**
  - Quantos alunos considerar?
  - Qual a probabilidade de cada um passar ?
    - Talvez simplificar ?
  - O que é isso de probabilidade ?

Probabilidade

"Medida do grau de certeza associado a um resultado proveniente de um fenómeno de acaso"

- Palavra usada pela primeira vez por Bernoulli (1654-1705)

Recordar …

- **Experiência aleatória**
  - Procedimento que deve produzir um resultado
  - Mas mesmo que seja repetido nas mesmas condições não garante que o resultado seja idêntico
  - Exemplo:
    - Escolher aleatoriamente uma letra do alfabeto

- Experiência aleatória é especificada por
  - Espaço de amostragem
  - Conjunto de acontecimentos (ou eventos)
  - Lei de probabilidade

Espaço de amostragem

- Conjunto (S) de todos os resultados possíveis de uma experiência aleatória
  - Em geral representado por S (do inglês "Sample Space")

- Resultados têm de ser mutuamente exclusivos e não divisíveis

- discreto se for contável
  - i.e. se contiver um número finito de elementos ou se contiver um número infinito em que se pode estabelecer uma correspondência biunívoca com o conjunto dos inteiros
- contínuo se não for contável

- Elementos de S são designados por resultados elementares

Acontecimentos / eventos

- Os resultados elementares das experiências não constituem necessariamente os únicos itens de interesse nas experiências
  - Exemplo:
    - No caso da contagem de mensagens de email podemos estar interessados no facto de o número total exceder um determinado limiar (nº > L)

- **Acontecimento (evento)** A é um subconjunto de S
  - S é obviamente um subconjunto de si próprio e constitui o evento certo
  - O conjunto vazio, ϕ, também é subconjunto, o evento impossível

Lei de probabilidade

- **Atribui probabilidade aos vários eventos**

- Probabilidade: número associado a um evento que indica a "verosimilhança" de esse evento ocorrer quando se efetua a experiência
  - valor entre 0 e 1 (às vezes é usada a escala 0 a 100%)
    - 1 para acontecimento certo
    - 0 para acontecimento impossível

Como é que se definem/obtêm as probabilidades associadas a eventos ?

- Através de **medição**

- Através da construção de **modelos** probabilísticos

- Probabilidades **teóricas**

- Probabilidades **empíricas**

- Probabilidades subjectivas
  - Exemplo:
    - Um Médico diz que tem 95 % de certeza de que determinada pessoa tem uma determinada doença
    - Uma casa de apostas estimou em 1/5 a probabilidade de Portugal ser campeão Europeu em 2016
      - E fomos Campeões ☺
  - Não nos interessam nesta UC

Diferentes abordagens

- Teoria **clássica** (de Laplace)
  - Probabilidades teóricas

- **Frequencista**
  - Probabilidades empíricas

- Teoria **matemática**

Teoria Clássica

Noção clássica

Simon de Laplace (1749-1827)

- "*Pour étudiér un phénoméne, il faut réduire tous les evénements du même type à un certain nombre de cas également possibles, et alors la probabilité d'un événement donné est une fraction, dont le numérateur représente le nombre de cas favorables à l'événement e dont le dénominateur représente par contre le nombre des cas possibles*"
  - pg 17 livro "O Acaso"

- Primeiro reduzir o fenómeno a um conjunto de resultados elementares, **"casos", igualmente prováveis**

$$P(evento) = \frac{número\ de\ casos\ favoráveis}{número\ de\ casos\ possíveis}$$

Exemplo

- **Lançamento de 1 DADO**
  – Honesto
    - => qualquer face igualmente provável

- Probabilidade de obter certa face, ex: a 5 ?

- 6 resultados ou <u>eventos elementares</u>
  – Representáveis pelo conjunto {1,2,3,4,5,6}

- Ao evento "saída da face 5" apenas corresponde um caso favorável
  – > P("face 5")=1/6

Variante do problema

- E se 2 faces tivessem o 5 marcado ?

- Espaço de amostragem ?
  - S={1,2,3,4,5} ? => casos possíveis =5 ?
  - S={1,2,3,4,**5**,**5**}

- P("sair 5")=2/6

Regras básicas (OU)

- P("sair face maior que 4") ?

  = P("sair face 5 ou face 6") = P({5,6}) = 2/6

  = P({5})+P({6})

- P("face par")=P({2})+P({4})+P({6})=1/2
- P("qualquer face") = 6 x 1/6 =1

… $P(A \cup B) = P(A) + P(B)$

Sempre ???

Regras básicas

- P("face menor ou igual a 4")

  =1 – P("face maior que 4")

  = 1 – 2/6 = 4/6

Regra do **complemento**

$P(\bar{A}) = 1 - P(A)$

Regras básicas (**E**)

- P("face par E face menor ou igual a 4")=

  = P("face par") x P("face menor ou igual a 4")

  $= \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$

  De facto existem 2 possibilidades em 6 , {2,4}

  … $P(A \cap B) = P(A) \times P(B)$

  Sempre? ….

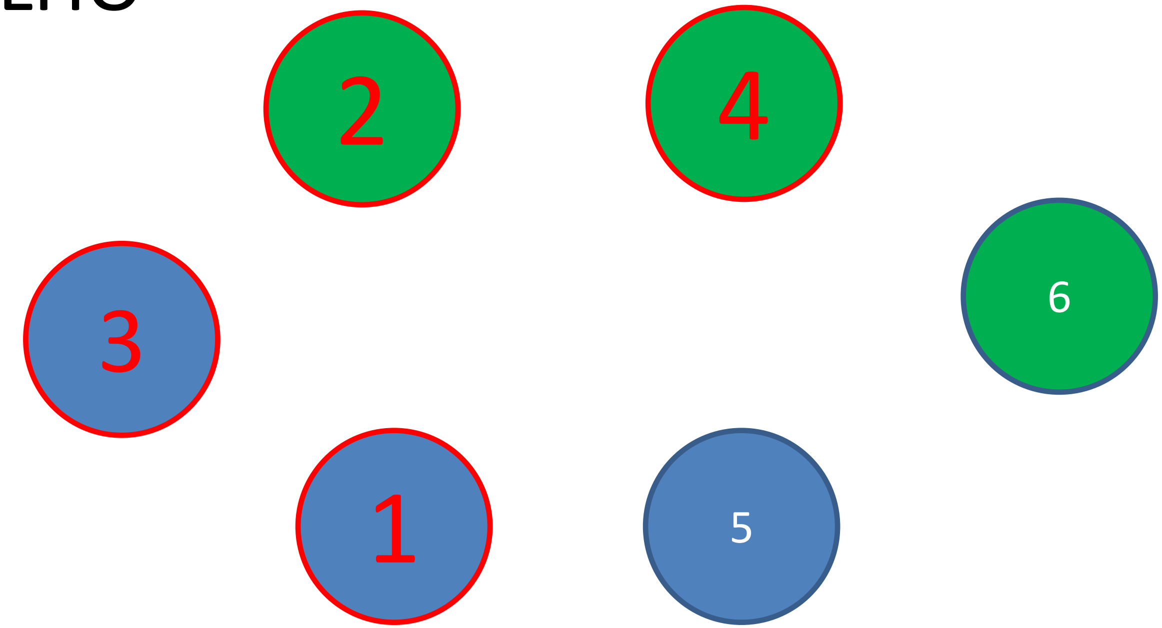
  (só se os acontecimentos forem independentes)

Aplicação das regras (OU novamente)

- P("face par OU face menor ou igual a 4") = ?
- Se fizermos P("face par")+ P("face menos ou igual a 4") dá $7/6 > 1$!!
- Qual o erro ?

Acontecimentos

- A="face par" e fundo VERDE
- B="face menor ou igual a 4" limite e texto a VERMELHO



…

Temos 3 com fundo verde => $P(A) = ½$

Temos 4 com vermelho => $P(B) = 2/3$

… mas temos 2 casos com fundo verde e limite e texto vermelho

- No mínimo perigoso ☺

- Estávamos a contar 2 vezes a intersecção

- $P(A \cup B) = P(A)+P(B) - P(A \cap B)$

  $= ½ + 2/3 - 2/6 = 3/6+4/6-2/6 = 5/6$

Testar as regras num problema

- Considere uma **família com 2 filhos** e que a probabilidade de nascer rapaz é igual à de nascer uma rapariga.
- Designando o nascimento de um filho por M e uma filha por F, qual a probabilidade de MF ?

- Probabilidade de pelo menos 1 rapaz numa família com 2 filhos ?

Resolução

- Pelo menos 1 rapaz => MF ou FM ou MM

- MF é a intersecção (“e”) de M no primeiro e F no segundo = ½ * ½

- Similar para MM e FM

- P(MF) + P(MM) + P(FM) = ¾
  - Devido à união (“ou”)

Problema do Cavaleiro de Méré

- P("sair pelo menos um seis em 4 lançamentos de 1 dado") vs P ("sair pelo menos um DUPLO 6 em 24 lançamentos de 2 dados")

- Melhor usar a regra do complemento...
- P("nenhum 6 em 4 lançamentos ")=
- P("não 6 na primeira E não 6 na segunda E …")

  =P("não 6 na primeira") x P("não 6 na segunda")

…

= 5/6 x 5/6 …. = (5/6) ^ 4

- P(“sair pelo menos um seis em 4 lançamentos de 1 dado”)

  $= 1 - \left(\frac{5}{6}\right)^4$

  $= 0{,}51775$

- P (“sair pelo menos um DUPLO 6 em 24 lançamentos de 2 dados”) =

  $= 1 - \left(\frac{35}{36}\right)^{24}$

  $= 0{,}49141$

Não esquecer

- Estas regras e definição clássica ASSUMEM dados honestos, moedas honestas, igual probabilidade de nascer rapaz e rapariga, equiprobabilidade para os eventos elementares

- Uma questão que surge naturalmente é se na prática tais valores são ou não razoáveis ?

Abordagem Frequencista

Noção **Frequencista**

- Noção introduzida por De Moivre (1718)

- Repete-se a experiência um certo número de vezes (N)
- Seja k o número de vezes que ocorre o acontecimento que nos interessa (ex: "sair face 5 num dado")
- Determina-se f=k/n , ou seja a frequência relativa de ocorrência
- Usa-se esta frequência como uma medida empírica de probabilidade

Frequência relativa

- Definição:
  - Se uma experiência for repetida N vezes nas mesmas condições a frequência relativa do evento A é

$$f(A) = \frac{\#\, ocorrências\ do\ evento\ A}{N}$$

- Se a frequência relativa convergir quando N aumenta, então o limite da frequência relativa é a probabilidade de A

$$p(A) = \lim_{N\to\infty} \frac{\#\, ocorrências\ de\ A}{N}$$

Frequência relativa (cont.)

- $0 \leq f(A) \leq 1$

- Numa experiência com K resultados possíveis em N experiências:
  - S= $\{A_1, A_2, A_3, \dots, A_K\}$
  - o resultado A_i ocorre N_i vezes
  - Cada um dos resultados possíveis terá uma frequência relativa de $f(A_i) = N_i/N$
  - $\sum_{i=1}^{K} f(A_i) = 1$

Exemplo em **Matlab**

- Probabilidade de **sair 2 caras em 3 lançamentos**

- Como se simula a experiência de lançar uma moeda ?
- Como se simula 3 lançamentos ?
- Como se repete “muitas” vezes ?
- Como contar as ocorrências do evento ?

Simular lançamentos …

```
%  simular 1 lançamento (de uma moeda)
lan= rand() >0.5  % assumiremos que 1 = “cara”

% simular os 3 lançamentos
lan_3= rand (3, 1) > 0.5     % ou l3 = rand(1,3) > 0.5

% repetir N vezes
N= 1e6  %  mas comecem com valor pequeno
lancamentos= rand(3,N);  % importante o “;”
```

ocorrências ... freq. relativa

```
% contar num ocorrências de “2 caras”
%          contar num caras (1s) em cada experiência
%          (que se encontra numa coluna da matriz lancamentos)
numCarasNaExperiencia= sum (lancamentos);

%  contar vezes em que esse número de caras é 2
numOcorrencias =    sum (numCarasNaExperiencia ==2)

% calcular freq relativa
fr  = numOcorrencias / N

%  usar como estimativa da probabilidade
pA= fr
```

Variação com N

```
% variação da frequência relativa em função de N

N= 1e5
lancamentos = rand(3,N) > 0.5;
sucessos= sum(lancamentos)==2; % 1 = sucesso
fabsol =  cumsum(sucessos);
frel = fabsol ./ (1:N);

plot(1:N, frel);
```

Problema das aprovações a MPEI

- **Simulação** em Matlab [simulMPEI.m]

```
p=0.85 ;            %  prob aluno passar em 2015 - 2016
n= 164;             % alunos de MPEI
N=1e6;              % experiências
k=fix(0.8 * n);     %  os 80 %

aprovados = rand(n,N) < p;   %%  1 indica aprovado

numOcorrencias =0;
for k1=k:n
   sucessos= sum(aprovados)==k1 ;
   fprintf('%d aprovados -> %8d , p=%.5f\n',k1,sum(sucessos),sum(sucessos)/N);
   numOcorrencias = numOcorrencias +sum(sucessos);
end

probSimulacao= numOcorrencias /N;
fprintf('prob de %d ou mais em %d passsarem é de %.4f\n',k,n,probSimulacao);
```

Problema do Cavaleiro de Méré

- P("sair pelo menos um seis em 4 lançamentos de 1 dado") vs P ("sair DUPLO 6 em 24 lançamentos de 2 dados")

- Simulação **em Matlab** [cavaleiro.m]

- TPC: Simulação em Java

Simples mas não perfeita

- Conceptualmente é extremamente simples e pode ser aplicada praticamente a todas as experiências

- Tem, no entanto, algumas desvantagens:
  - Em muitos casos requer considerável dispêndio de tempo
  - As experiências devem poder ser repetidas em condições idênticas
  - Quando o espaço amostral é infinito surgem questões de fiabilidade uma vez que só podemos efectuar um número finito de repetições da experiência
  - A própria obtenção dos valores coloca algumas questões:
    - Quantos ensaios se tem de efectuar para termos medidas fiáveis ?
    - Como se lida com medições sujeitas a erro ?

Teoria Axiomática de Probabilidade

Definição axiomática de probabilidade

- Em 1933, Kolmogorov estabeleceu a DEFINIÇÃO DE PROBABILIDADE POR AXIOMATIZAÇÃO
  – na sua obra Foundations of the Theory of Probability.

- com base nas propriedades das frequências relativas e das operações sobre conjuntos

O que é uma axiomática?

Em determinado ponto da evolução de uma teoria de pensamento matemático, torna-se imperioso ordenar, sistematizar e relacionar todos os conhecimentos entretanto nela reconhecidos, isto é, proceder à sua **AXIOMATIZAÇÃO**

Axiomática de probabilidades

- Axioma 1- probabilidades são não-negativas
  P(A) > = 0
- Axioma 2 – normalização (S tem probabilidade 1)
  P(S) =1

- Axioma 3a – Se A e B forem mutuamente exclusivos
  P(A U B) = P(A) + P(B)

- Axioma 3b – Se A1, A2, … for uma sequência infinita de acontecimentos mutuamente exclusivos ($\forall_{i \neq j}\, Ai \;\cap Aj = \emptyset$)

$$P(\cup_{k=1}^{\infty} Ak) = \sum_{i=1}^{\infty} P(Ak)$$

Teoremas

- Como sabem, às afirmações que se obtêm dedutivamente a partir dos axiomas, ou de outras já deles obtidas por dedução, chamamos TEOREMAS.

Teoremas / Corolários :

Prob. do acontecimento complementar

- $P(\bar{A}) = 1 - P(A)$

Demonstração

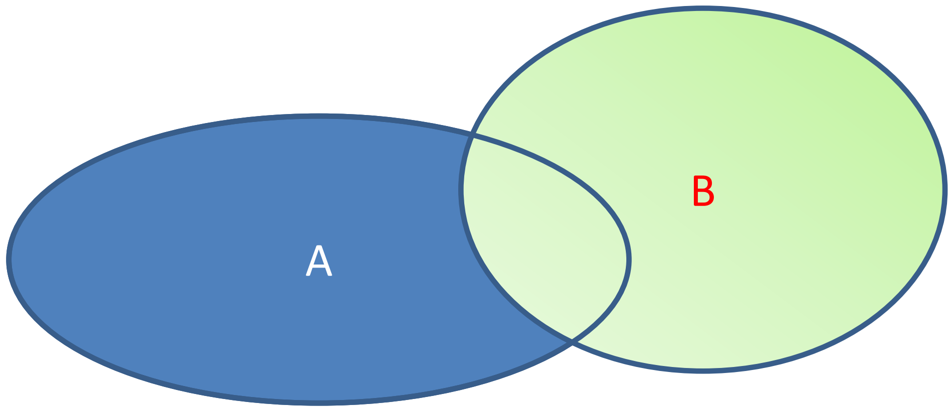
- Como $A \cap \bar{A} = \emptyset$
- E $A \cup \bar{A} = S$

- Pelo axiomas 2 e 3:
- $P\,(A \cup \bar{A}) = P(A) + P(\bar{A}) = P(S)$ ...
- $P(A) + P(\bar{A}) = 1$
- $P(\bar{A}) = 1 - P(A)$

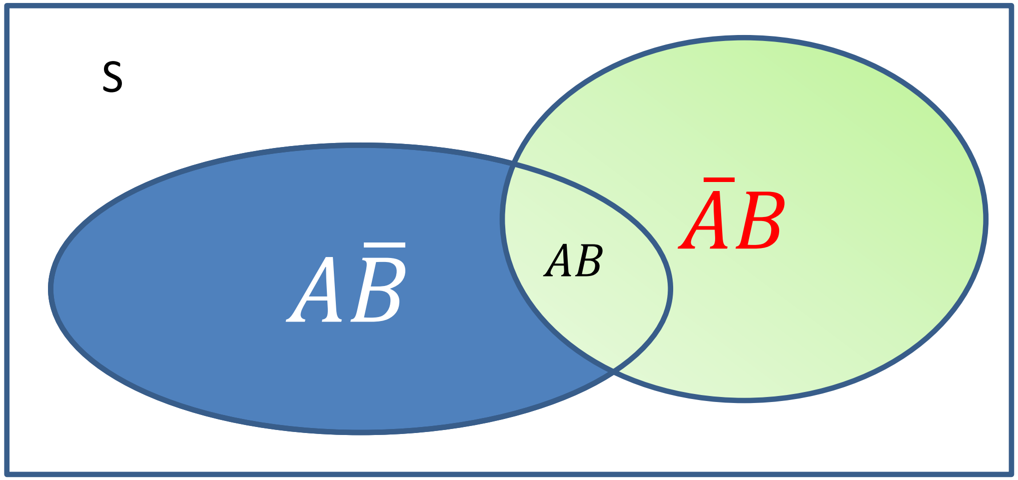
Teorema/Corolário:
Probabilidade da união

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Com AB$\equiv A \cap B$



Demonstração



$A \cup B = A\bar{B} \cup AB \cup \bar{A}B$, disjuntos

Logo (axioma):

$$\mathrm{P}(A \cup B) = P(A\bar{B}) + P(AB) + P(\bar{A}B)$$

Adicionando e subtraindo P(AB)

$$\mathrm{P}(A \cup B)$$
$$= \big(P(A\bar{B}) + P(AB)\big) + P\big((\bar{A}B) + P(AB)\big) - P(AB)$$
$$= P(A) + P(B) - P(AB)$$

Probabilidade em espaços de amostragem não contáveis

- Exemplo:
  - Escolha de um número real no intervalo [a,b]
- Seja o acontecimento A “número pertencer a [c,d]”

- P(A)= (d-c)/ (b-a)

- A probabilidade de qualquer ponto $x \in [a, b]$ é $igual\ \ a\ 0$
  - Ter, por exemplo,]c,d[dará igual

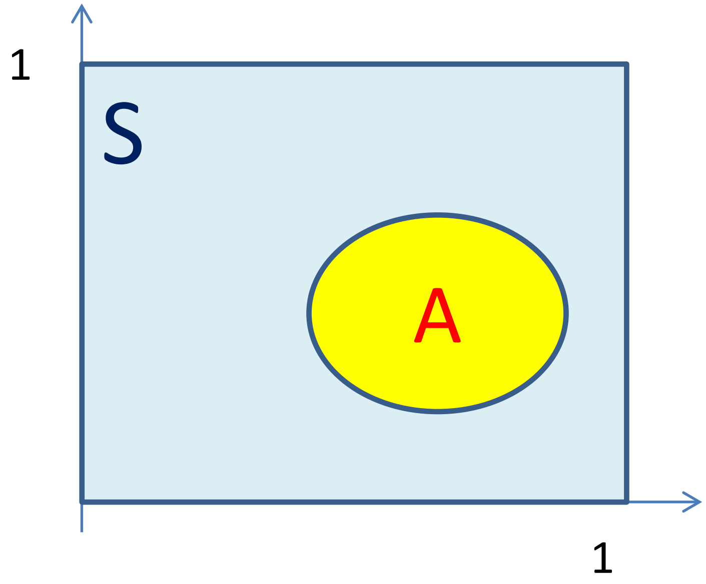
Probabilidade em espaços de amostragem não contáveis

- Exemplo:
  - Escolha de um número real no intervalo [0,90] relativo ao atraso de chegada a uma aula de 1h30m
- Seja o acontecimento A “chegar dentro da tolerância, i.e. [0,15[
- P(A)= (15-0)/ (90-0) = 0.16(6)
  - Assumindo que podem chegar em qualquer altura da aula ☹
    - O que não é válido

Probabilidade em espaços de amostragem não contáveis

- No caso de um par de números reais x,y entre 0 e 1

$$S = \{x, y : x \in [0,1] \; \cap y \;\; \in [0,1]\}$$



$$P(A) = \frac{Área\ (A)}{Área\ (S)}$$

A axiomática é compatível com as teorias anteriores ?

Sim, como era de esperar.

A definição frequencista de probabilidade satisfaz a axiomática

- Satisfaz o axioma 1: «p(A) não negativo»

  – pois se as frequências são números não negativos também convergem para um número não negativo.

A definição frequencista de probabilidade satisfaz a axiomática

- Satisfaz o axioma 2: «p(E)=1»

  - pois as frequências relativas de um acontecimento certo são sempre 1, logo, tendem para 1.

A definição frequencista de probabilidade satisfaz a axiomática

- Satisfaz o axioma 3: «Se (A e B são incompatíveis) então a probabilidade da reunião de A com B é igual à soma das probabilidades de A e de B»
  - pois se os acontecimentos A e B são incompatíveis não têm resultados comuns, a frequência relativa de AUB é a frequência relativa de A mais a frequência relativa de B e o limite da soma das duas sucessões é a soma dos limites.

A Lei de Laplace verifica a axiomática

- Lei de Laplace
  - Num espaço finito de resultados equiprováveis a probabilidade de um acontecimento A é igual ao quociente entre o número de resultados favoráveis a A (#A) e o número de resultados possíveis (#E)

A Lei de Laplace verifica a axiomática

- Satisfaz o axioma 1: «p(A) não negativo»
  - Pois p(A)=#A / #E o que significa que p(A) é o quociente entre um número real não negativo e um número positivo.

A Lei de Laplace verifica a axiomática

- Satisfaz o axioma 2: «p(E)=1»
  - Pois p(E)= #E / #E é o quociente entre dois números iguais.

A Lei de Laplace verifica a axiomática

- Satisfaz o axioma 3: «Se (A e B são incompatíveis) então a probabilidade da reunião de A com B é igual à soma das probabilidades de A e de B»

  – Se A e B são disjuntos

$$\#(A \cup B) = \#A + \#B$$

E então:

$$p(A \cup B) = \frac{\#(A \cup B)}{\#E} = \frac{\#A + \#B}{\#E} = \frac{\#A}{\#E} + \frac{\#B}{\#E} = p(A) + p(B)$$

TPC

- Completar a verificação dos restantes axiomas para as duas teorias (Clássica e Frequencista)

Para aprender mais …

- Capítulos iniciais do Livro "Probabilidades e Processos Estocásticos", F. Vaz, Universidade de Aveiro [ELEARNING]

- Links para material online:
  - http://www.stat.berkeley.edu/~stark/SticiGui/Text/probabilityAxioms.htm
  - https://www.youtube.com/course?list=PL10921DED3A8BFF53

- Capítulos iniciais do Livro "O Acaso" , Joaquim Marques de Sá, Gradiva

Probabilidades Condicionais

Probabilidade condicional

- Por vezes dois acontecimentos estão relacionados
  - A ocorrência de um depende ou faz depender a ocorrência do outro

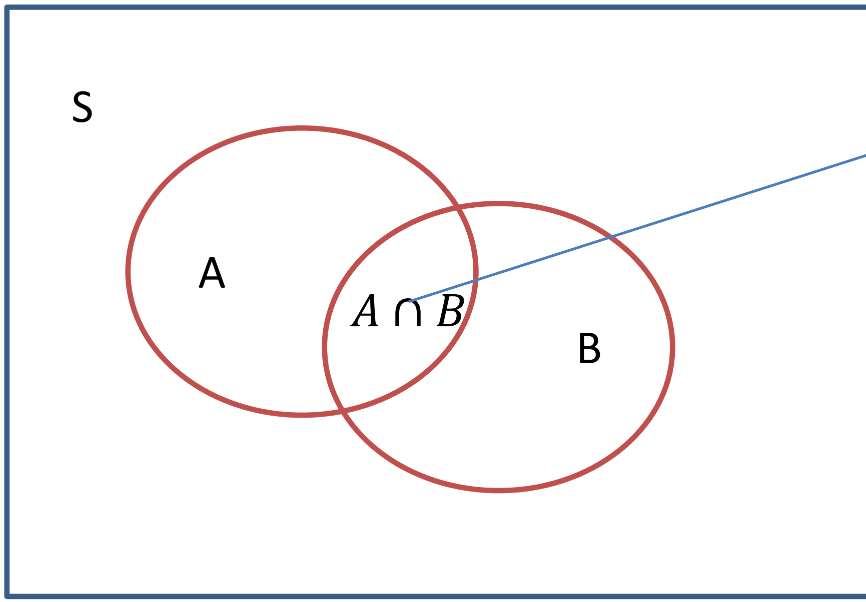
- A Probabilidade de ocorrência de um evento A com a informação de que o evento B ocorreu é a designada **PROBABILIDADE CONDICIONAL** de A dado B
  - Definida por:

$$P(A|B) = \frac{\boldsymbol{P(AB)}}{P(B)} \; se \; P(B) \neq 0$$

Indefinida se P(B)=0

Interpretação da probabilidade condicional

- $P(A|B) = \frac{P(AB)}{P(B)}$ $se\ P(B) \neq 0$



Zona onde A se realiza,
Sabendo que B ocorreu

Exemplo de aplicação

- 2 números de 1 a 4, N1 e N2

- Evento B = "min(N1,N2)=2"
- Evento M = "max(N1,N2)"

- **P(M=1|B) =**
  P("max()=1" & "min()=2") / P("min()=2") =
  …
  =0

- P(M=2|B) = …
  = 1/5

N2→	1	2	3	4
1				
2		B / 2	B /3	B /4
3		B /3		
4		B /4		

Regra da cadeia: P(AB), P(ABC) …

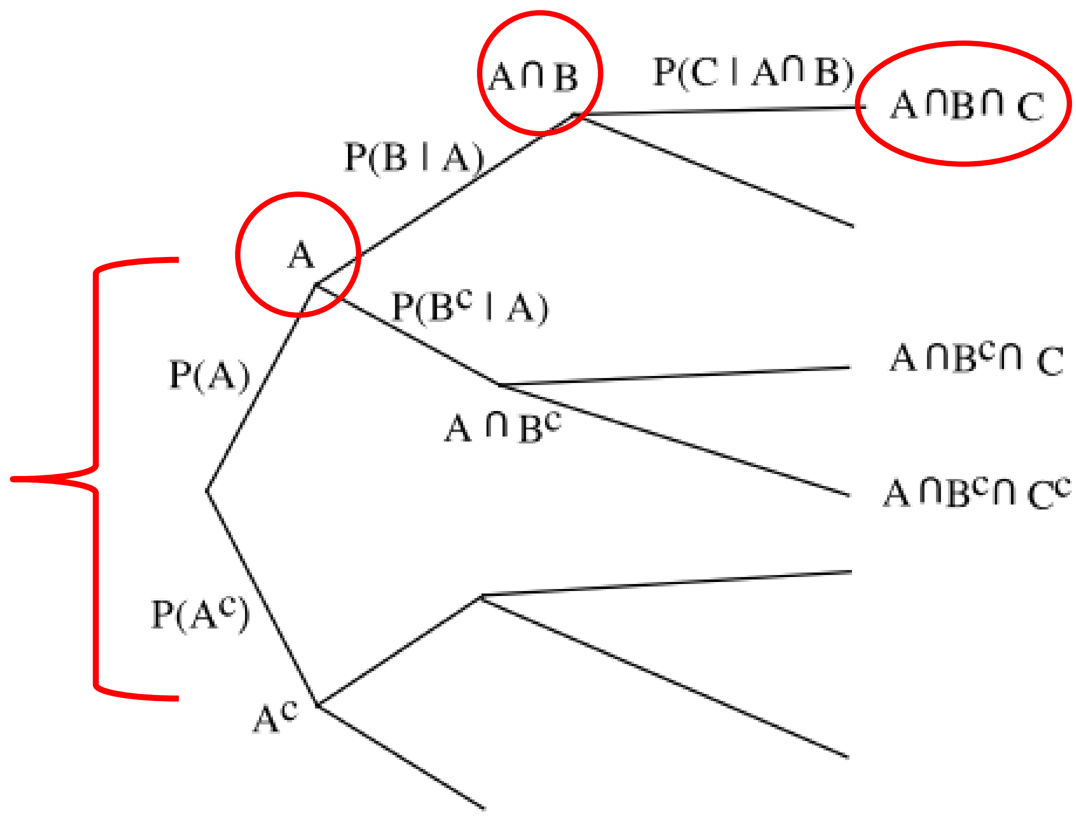
- $P(AB) = P(A|B) \times P(B)$

- Aplicando sucessivamente temos (regra da cadeia)

$$P(A_1 A_2 A_3 \ldots A_n)$$
$$= P(A_1|A_2 \ldots A_n) \times P(A_2 A_3 \ldots A_n)$$
$$= P(A_1|A_2 \ldots A_n)$$
$$\times P(A_2|A_3 \ldots A_n) \ldots P(A_{n-1}|A_n)$$

Regra da cadeia / multiplicação

- $P(ABC) = P(A)P(B|A)\,P(C|AB)$



Problema com 2 urnas...

- Consideremos 2 urnas, designadas por X e Y, contendo bolas brancas e pretas:
  - X contém 4 brancas e 5 pretas e
  - Y contém 3 brancas e 6 pretas.
- Escolhe-se uma urna ao acaso e extrai-se uma bola. Qual a probabilidade de sair bola branca?

- **P("bola branca")**

= P("branca da urna X OU branca da urna Y")

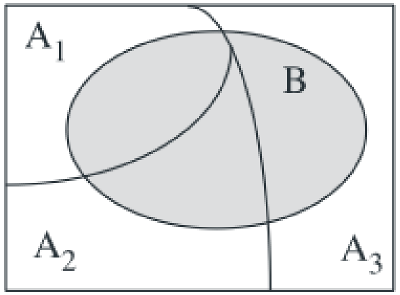
= P("branca E urna X")+P("branca E urna Y")

= P("branca"|"urna X") x P("urna X") + P("branca"|"urna Y") x P("urna Y")

= $(4/9)x(1/2)+(3/9)x(1/2) \approx 0{,}39$

Lei da Probabilidade total

- Dividir para conquistar
- Partição do espaço de amostragem A_1, A_2, A_3
- Ter $P(B|A_i)$, para todos os i



- $P(B) = P(B|A_1)P(A_1)$
  $+P(B|A_2)P(A_2)$
  $+ P(B|A_3)P(A_3)$

Em geral: $\mathrm{P(B)} = \sum_j P(B|A_j)P(A_j)$

Condicionamento inverso

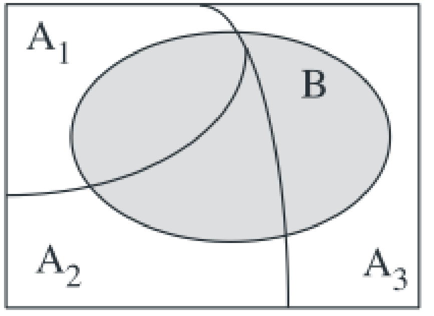
- Continuando com as urnas …
- **Problema Inverso** (condicionamento inverso)

P("urna X" | "bola branca")

Resolvido pela primeira vez pelo Reverendo Thomas Bayes (1702-1761)

Regra de Bayes

- Probabilidades *a priori* $P(A_i)$
- Sabemos $P(B|A_i) \;\; \forall i$



- Pretendemos calcular $\boldsymbol{P(A_i|B)}$
  - i.e. $P(A_i)$ dado que B ocorreu

- $P(A_i|B) = \dfrac{P(A_i \cap B)}{P(B)}$

$= \dfrac{P(B|A_i)P(A_i)}{P(B)}$

$= \dfrac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$

Aplicando ao problema das urnas

- **P(“urna X” | “bola branca”)**

$$= \frac{P(\text{"bola branca"}|\text{"}urna\ X\text{"}) \times P(\text{"}urna\ X\text{"})}{P(\text{"bola branca"})}$$

$$= \frac{\left(\frac{4}{9}\right)\times\left(\frac{1}{2}\right)}{7/18} = \frac{4}{7}$$

- P(“urna Y” | “bola branca”)=

$$\ldots = \frac{\left(\frac{3}{9}\right)\times\left(\frac{1}{2}\right)}{7/18} = \frac{3}{7}$$

$$= 1 - P(\text{"urna X"}|\text{"bola branca"})$$

Causa e efeito

- No evento “urna X se bola branca” podemos considerar que a saída de bola branca é o EFEITO da causa “urna X”
- A Regra de Bayes, em consequência, pode ser escrita da seguinte forma:

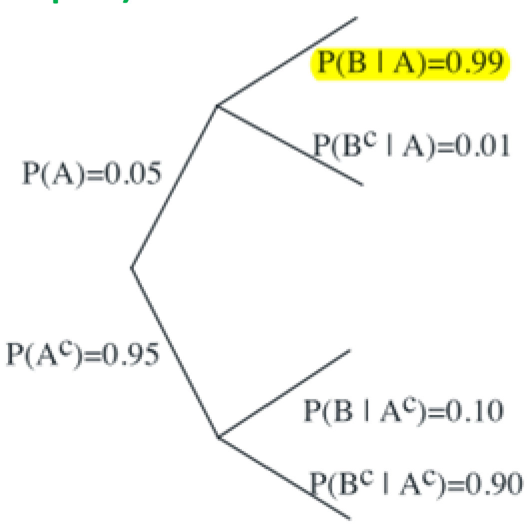
$$P(\text{"causa"}|\text{"efeito"}) = \frac{P(\text{"efeito"}|\text{"causa"}) \times P(\text{"causa"})}{P(\text{"efeito"})}$$

Exemplo de aplicação Regra de Bayes - radar

Evento A: avião voando na zona do radar, P(A)= 0.05

Evento B: Aparece algo no ecrã do radar, P(B|A)= 0.99

P(A | B)= ?



- $P(A \cap B) =$
$=P(B|A)\,P(A)$
$=0{,}99 \times 0{,}05$

- $P(B) =$
$= P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})$

- $P(A \mid B) =$
$= \frac{P(B|A) \times P(A)}{P(B)}$
$= \frac{0{,}99 \times 0{,}05}{0{,}1445} = 0.3426$
(valor baixo)

Outro exemplo – sistema de comunicação/transmissão

- 0 ou 1 na entrada, transmissão sujeita a erros, entradas equiprováveis.
- No receptor uma decisão é tomada (0 ou 1):
  - Se ε for a probabilidade de erro, qual a entrada mais provável se na saída obtemos 1 ?

- Seja A_k o acontecimento "entrada é k", k=0,1

  A_0 $e A_1$ constituem uma partição de S
- Seja B_1 o acontecimento "saída = 1"

…

- $P(B_1) = P(B_1|A_0) \times P(A_0) + P(B_1|A_1) \times P(A_1)$

$$= \varepsilon\frac{1}{2} + (1-\varepsilon)\frac{1}{2} = \frac{1}{2}$$

- Probabilidade de entrada ter sido zero dado que saída igual a 1 ?
  $P(A_0|B_1) = P(B_1|A_0) \times P(A_0)/P(B_1)$
  $= \frac{\varepsilon}{2} \;/\; \frac{1}{2} = \varepsilon$

- De forma similar $P(A_1|B_1) = \ldots = 1 - \varepsilon$

- Se ε <1/2 a entrada mais provável é 1 (se saída for 1)
  - Que é o que se pretende em geral.

Independência

Independência

- 2 acontecimentos são independentes sse P(AB) = P(A)P(B)
  - Simétrico relativamente a A e B
  - Aplica-se mesmo que P(A)=0
  - Implica P(A|B)=P(A) [mas não é a definição]
    - Ocorrência de B não fornece informação sobre ocorrência de A

- Generalização…

  os acontecimentos $A_1, A_2, A_3 \ldots A_n$ são independentes sse

  $$P(A_1 \cap A_2 \cap A_3 \ldots \cap A_n) = \mathrm{P}(A_1)P(A_2)P(A_3)\ldots P(A_n)$$

Independência vs independência 2 a 2

- Experiência:
  2 lançamentos de moeda
- Acontecimentos
  - A: primeira é caras
  - B: segunda é caras
  - C: mesmo resultado em ambas

HH	HT
TH	TT

- $P(C)$? P(A) ? P(B) ?
  2/4=1/2
- $P(C \cap A) =$
  1/4 = 1/2 x 1/2
  C e A indep.
- $P(C \cap B) =$
  1/4 = 1/2 x 1/2
  C e B indep.
- $P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$ …. $A\ e\ B\ ind.$
- $P(C \cap B \ \cap A) =$
  $$\frac{1}{4} \neq \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$$

Independência 2 a 2 não implica independência

Independência e acontecimentos mutuamente exclusivos

- Em geral dois acontecimentos que sejam mutuamente exclusivos e tenham probabilidade não nula não podem ser independentes

  $$0 = P(A \cap B) = P(A)\,P(B)$$

  implicaria que um deles tenha probabilidade nula

- Em geral considera-se que acontecimentos em experiências distintas são independentes (experiências independentes)

Sequências de experiências independentes

- Se uma experiência aleatória for composta por *n* experiências independentes e se A_k for um acontecimento que diga respeito à experiência *k*, é razoável admitir que os *n* acontecimentos são independentes

- Então:

$$P(A_1 \cap A_2 \cap A_3 ... \cap A_n) = \mathrm{P}(A_1)P(A_2)P(A_3)...P(A_n)$$

Experiências de Bernoulli

- Uma **experiência de Bernoulli** consiste em realizar uma experiência e registar se um dado acontecimento se verifica (sucesso) ou não (falha)

Qual a **probabilidade de k sucessos em n ensaios independentes** ?

- Seja *p* a probabilidade de sucesso
  - E (1-*p*) a de falha
- A probabilidade de *k* sucessos e (*n*-*k*) falhas é:

$$p^k\ (1-p)^{n-k}$$

- *k* sucessos em *n* experiências podem ocorrer de C_k^n maneiras

- Então a probabilidade pedida é:

$$\boldsymbol{P_n(k) = C_k^n\ \ p^k\ \ (1-p)^{n-k}}$$

Lei Binomial

Visão frequencista e probabilidade condicional

- Do ponto de vista frequencista, considerando os eventos A e B e N experiências, podemos escrever:

- $P(A|B) \approx \frac{k_{A\ e\ B}/_{N}}{k_{B}/_{N}} = \frac{k_{A\ e\ B}}{k_B}$

- Onde $k_{A\ e\ B}$ é o número de ocorrência de "A e B"
  - Que também se pode denominar por frequência absoluta e representar por f_{AB}

Simulação

- Como fazer para ter P(A|B) ?

- Realizar N experiências

- Contar o número de ocorrências de AB

  Será fAB (frequência absoluta)
- Contar número de ocorrências de B

  fB

- $P \cong \dfrac{fAB/_N}{fB/_N} = \dfrac{fAB}{fB}$

Exemplo de simulação (Independência vs independência 2 a 2)

- Relembremos:

  2 lançamentos de moeda

  A: primeira é caras

  B: segunda é caras

  C: mesmo resultado em ambas

- $P(C\ |A \cap B)$

Principais assuntos

- Probabilidade
- Teorias de Probabilidade (Clássica, Frequencista, Axiomática)
- Probabilidade Condicional
- 3 Ferramentas muito importantes
  - Regra da multiplicação
  - Teorema da Probabilidade total
  - Regra Bayes
- Aplicação da teoria frequencista a probabilidades condicionais

Não esquecer

- Independência de 2 eventos
- Independência de uma coleção de eventos
- Probabilidade condicional

$$P(A|B) = {P(A \cap B)}/{P(B)}$$

- Regra da multiplicação:

$$P(A \cap B) = P(B) \times P(A|B) = P(A) \times P(B|A)$$

- Probabilidade total:

$$\mathrm{P(B)} = \sum_j P(B|A_j)P(A_j)$$

Regra de Bayes

$$P(A_i|B) = \frac{P(A_i) \;\; P(B|A_i)}{P(B)}$$

Para aprender mais …

- Capítulos iniciais do Livro “Probabilidades e Processos Estocásticos”, F. Vaz, Universidade de Aveiro
  - Disponíveis no Elearning da UC

MPEI 2020-2021

Variáveis Aleatórias

Motivação

- A probabilidade é uma função sobre eventos (conjuntos)
- Utilização das ferramentas da análise matemática (ex: derivação) não é imediata
  - Especialmente se os resultados da experiência não forem números

- Se conseguirmos **mapear o espaço de amostragem (S) para a recta real** facilita o uso das ferramentas de análise e aritmética
- Na maioria dos casos o mapeamento não é artificial
  - Muitas vezes não nos interessa os eventos mas uma grandeza numérica relacionada
    - Exemplo: número de caras em N lançamentos de uma moeda

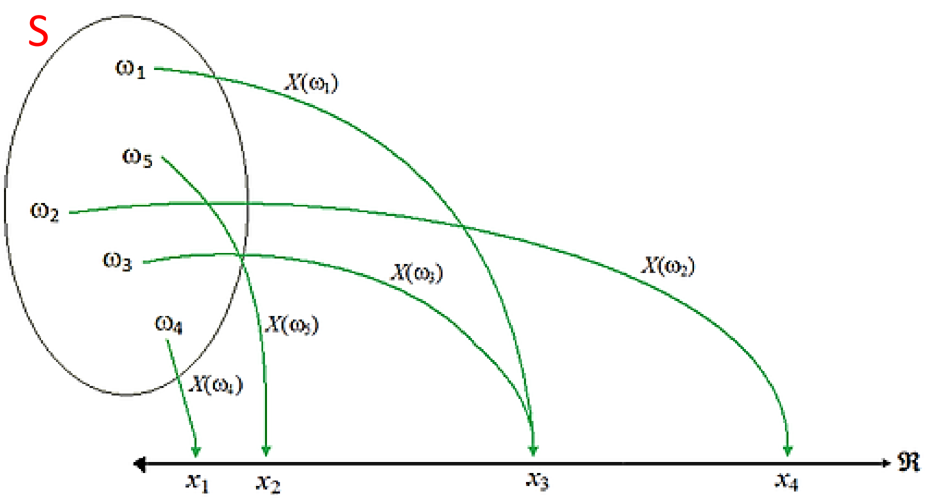
Conceito de variável aleatória

- Uma função que mapeia o espaço de amostragemna recta real é designada de **VARIÁVEL ALEATÓRIA**
  - *Random Variable* em Inglês

- Numa definição "informal":
- uma Variável Aleatória é o resultado numérico das nossas experiências (aleatórias)

Variável Aleatória - Definição

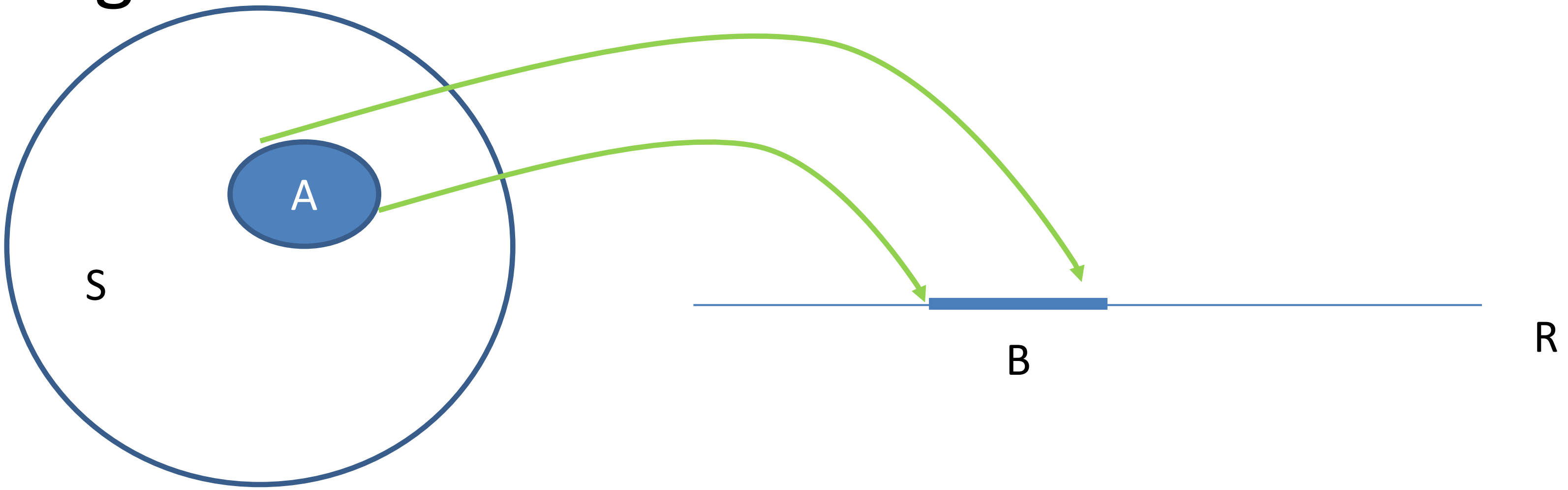
- Uma variável aleatória escalar X é formalmente definida como sendo um mapeamento de um espaço amostral S para a recta real



  – A qualquer elemento ω de S associa-se uma imagem $X(\omega)$ na recta real

Caso contínuo

- Se os conjuntos que representam os eventos forem contínuos, o mapeamento é para um segmento da recta real



- A e B são acontecimentos equivalentes

Tipos de Variáveis aleatórias

- **Discreta** : se os valores que a variável aleatória pode assumir forem finitos
  - ou infinitos mas contáveis
  - Exemplo: número de acessos por minuto a uma página web
- **Contínua** : se os valores que pode assumir formarem um ou vários intervalos disjuntos
  - Exemplo: Duração de uma conferência no Skype
- **Mista**: onde se verificam os atributos que definem os 2 tipos anteriores

Tipos

- Discreta/contínua ou mista ?

VA	Tipo ? (D,/C,/M)
Número de palavras com erro numa página	
Atraso com que chega às aulas TP	
Número de caixas abertas no supermercado	
Tempo de espera numa caixa de supermercado	
Número de páginas relevantes para uma procura num motor de pesquisa (ex: Google)	
Número de “bugs” num módulo de código	
…	

Caracterização das variáveis aleatórias

Parte 1

Distribuição de probabilidades

- As variáveis aleatórias são caracterizáveis pelo conjunto de valores que podem assumir e as probabilidades associadas

  – Ou seja pela “distribuição de probabilidades”

Função (massa) de probabilidade

- Uma variável aleatória discreta escalar $\boldsymbol{X}$ é especificada por:

1. Conjunto de valores que pode assumir: $x_i, i = 1,2, \ldots$

2. Probabilidade associada a cada um desses valores: $\boldsymbol{p_X(x_i)}$
   - **Denominada de função massa de probabilidade**
     - **Probability Mass Function em Inglês**
   - **ou mais simplesmente função de probabilidade**

$$\boldsymbol{p_X(x_i) = P(X = x_i)}$$

Onde $\boldsymbol{P(X = x_i) = P(w\!: X(w) = x_i)}$
 i.e. A probabilidade do evento cujos resultados $\boldsymbol{w}$ satisfazem $\boldsymbol{X(w) = x_i}$

Função de probabilidade

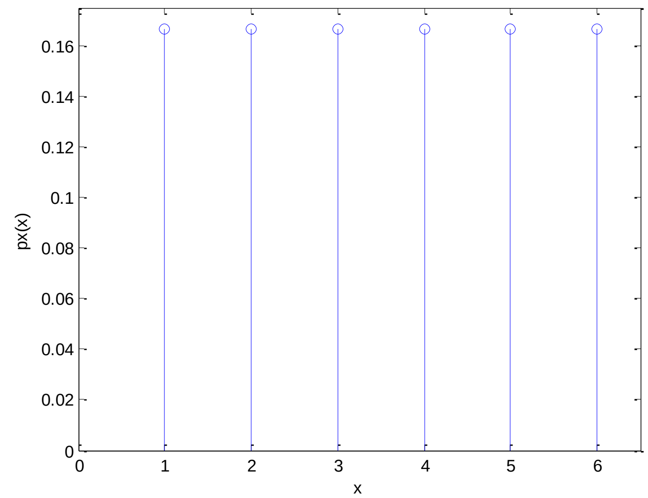
- Os axiomas da probabilidade implicam:

- $\boldsymbol{p_X(x_i) \geq 0}$

- $\sum_i \boldsymbol{p_X(x_i)} = 1$

Exemplo de função de probabilidade

- Lançamento de dado equilibrado e $\boldsymbol{X}$ igual ao número que sai
- $\boldsymbol{X}$:Variável aleatória discreta



- *Função de probabilidade*

  x_i={1,2,3,4,5,6}

  $\boldsymbol{p_X(x_i)}$=1/6

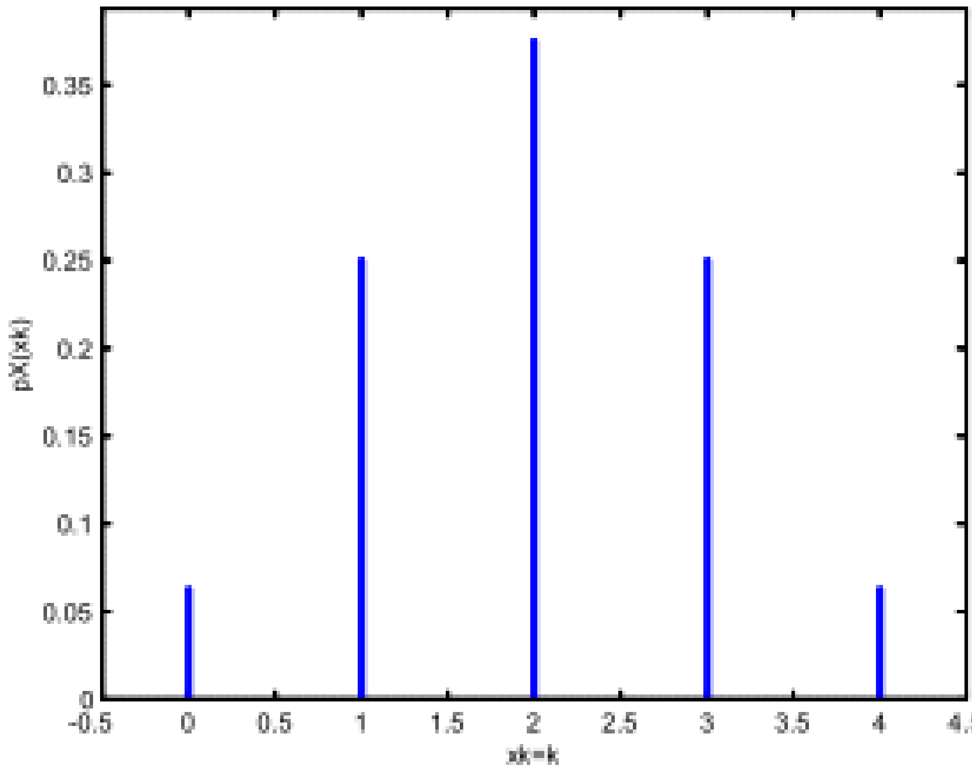
- %% Matlab

```
xi = 1:6; p=ones(1,6)/6;
stem(xi,p), xlabel('x'), ylabel('px(x)');
```

Outro exemplo

Função massa de probabilidade

para a variável aleatória representando o número de “caras” em 4 lançamentos de uma moeda



Função distribuição acumulada (discreta)

- Uma variável aleatória (discreta) pode ser também especificada pela sua função distribuição acumulada (fda), definida como
- $F_X(x) = p_X(X \leq x) = \sum_{i:x_i \leq x} p_X(x_i)$

- Dos axiomas e corolários:

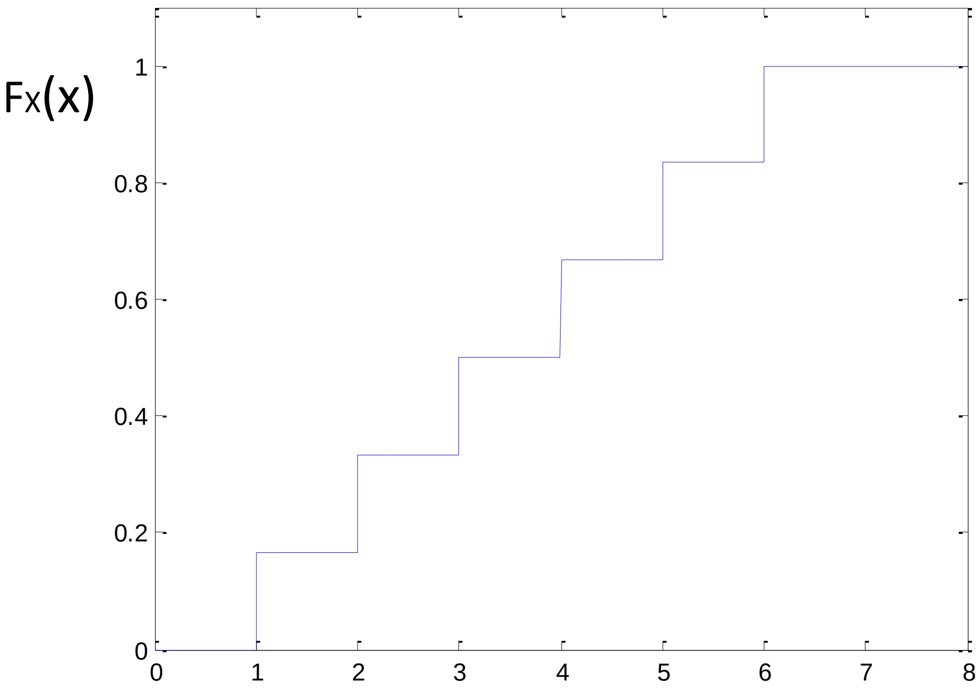
  É uma função não decrescente

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

Exemplo de função de distribuição

- Para uma variável aleatória discreta a função distribuição acumulada é uma função em escada

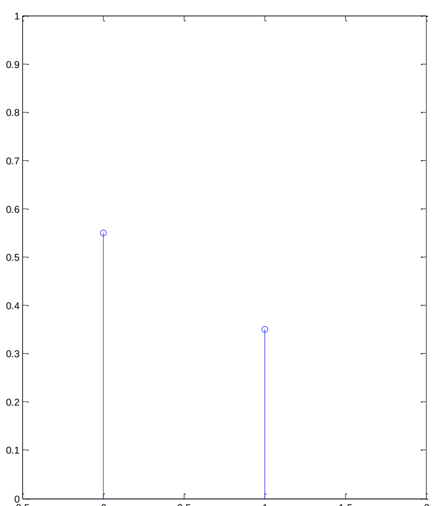


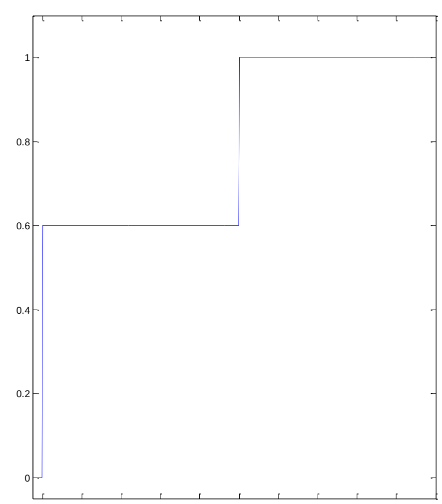
Outro exemplo

- Cada símbolo transmitido num sistema de transmissão pode ser interpretado como uma variável aleatória que toma os valores $x_1 = 0$ com probabilidade $1 - p$ e $x_2 = 1$ com probabilidade p

$\boldsymbol{p_X(x)}$

função probabilidade





$F_X(x)$

Função Distribuição (acumulada)

Variáveis aleatórias contínuas

- Também pode ser especificada pela sua função distribuição acumulada
- A definição é idêntica para o caso contínuo e discreto

$$F_X(x) = Prob(X \leq x)$$

- $F_X(x)$ é agora contínua
- Propriedades:

$$0 \leq F_X(x) \leq 1$$

$$\lim_{x \to \infty} F_X(x) = 1 \qquad \lim_{x \to -\infty} F_X(x) = 0$$

$$a < b \Rightarrow F_X(a) \leq F_X(b)$$

$$\boldsymbol{P[a < X \leq b] = F_X(b) - F_X(a)}$$

Variáveis aleatórias contínuas

- Podem ser especificada pela sua função de densidade de probabilidade $f_X(x)$

Probability density function (pdf) em Inglês

- Obtém-se derivando a função de distribuição

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Função de DENSIDADE de probabilidade

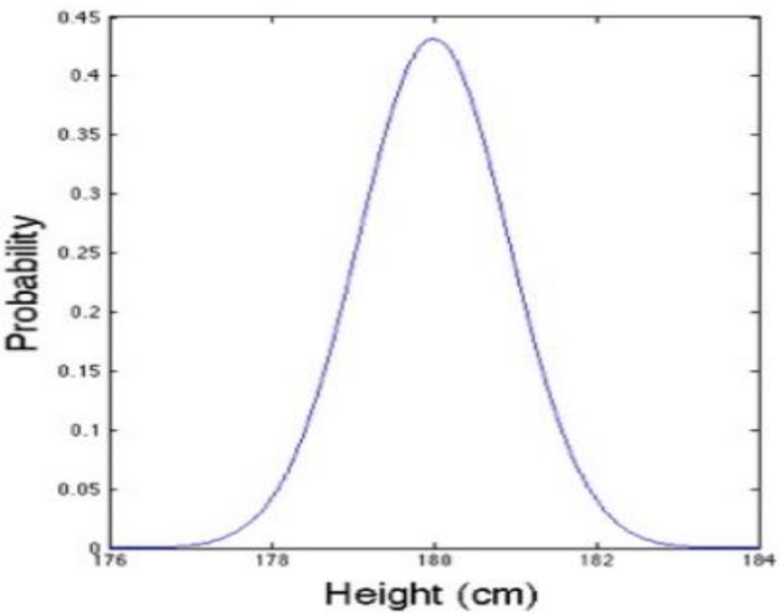
- $f_X(x)$ não é uma probabilidade ...
  - Apenas define os valores de probabilidade quando integrada num intervalo
    - $p(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$

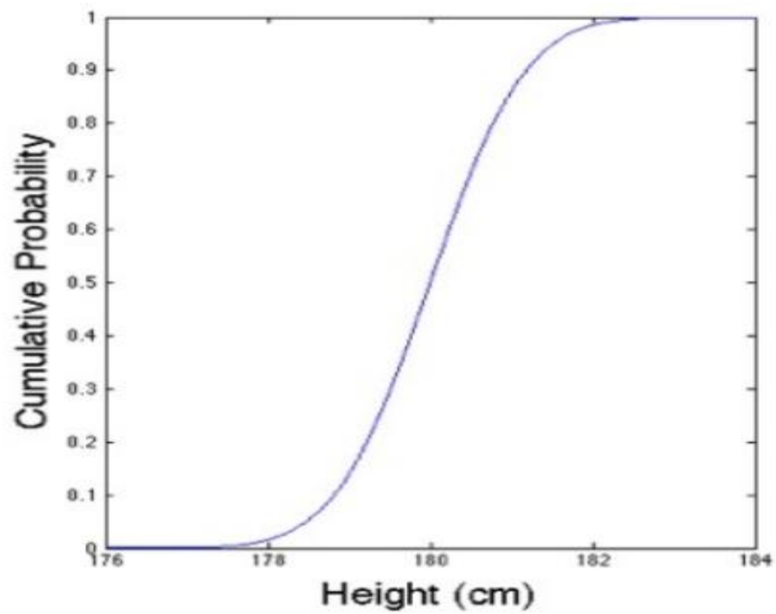
- $f_X(x)dx$ é a probabilidade da variável $\boldsymbol{X}$ pertencer ao intervalo $(x, x + dx)$, sendo dx um acréscimo infinitesimal

- $f_X(x) \equiv \frac{prob}{dx}$ → daí o nome "densidade"

Relações entre funções de densidade e de distribuição (caso contínuo)

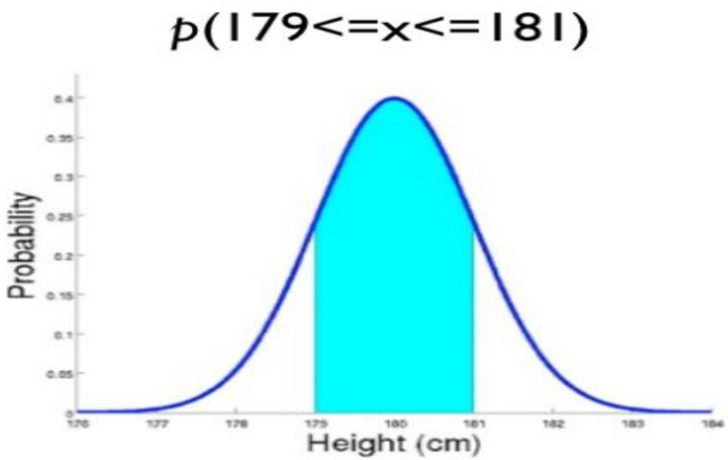
- $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$
- Exemplo de par de funções de densidade e de distribuição





Probabilidades e função de densidade

- $\mathrm{P}(a < X \leq b) = \int_a^b f_X(x)dx$



- A probabilidade é a área debaixo da curva
- Área total da curva =1

Caracterização das variáveis aleatórias

Parte 2

Motivação

- As funções apresentadas anteriormente fornecem uma descrição completa de uma variável aleatória

- Mas em muitos casos não necessitamos de toda a informação
  - Exemplo:
    - no caso dos "bugs" em módulos de código saber o valor médio pode ser suficiente

Média ou Valor esperado

- Consideremos N lançamentos de um dado

Ex: 4 1 6 6 5 5 5 3 4 2 …

Média = (1+1+..1 + 2+2+…+2 + …) / N
= (numDe1s x 1 + numDe2s x 2 …) /N
= numDe1s/N x 1 + NumDe2s/N x 2 + …

Assumindo que N tende para infinito
= p(1) x 1 + p(2) x 2 + p(3) x 3 … +p(6) x 6
= $\sum_i p(x_i) x_i \qquad com\ x_i = 1,2, \ldots 6$

Valor esperado

- Consideremos as experiências na origem da variável aleatória $\boldsymbol{X}$:
- Dizemos que o valor esperado de $\boldsymbol{X}$ é o valor médio de $\boldsymbol{X}$ ao repetirmos as experiências indefinidamente
- Representando por X_i o valor de $\boldsymbol{X}$ na experiência i, este valor é:

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Valor esperado (continuação)

- Só existe valor esperado se existir o limite
  - O limite existe se X_i tiver limite inferior e superior finitos, o que é verdade no mundo real
    - Ex: o peso de uma pessoa nunca é negativo

- Representando por x_i os m diferentes valores que X_i pode assumir e por $K_{i,n}$ o número de vezes que ocorre cada x_i , o nosso limite passa a:

$$\lim_{n \to \infty} \frac{x_1\, K_{1,n} + x_2\, K_{2,n} + \cdots + x_m\, K_{m,n}}{n}$$

$$\sum_{i=1}^{m} x_i \lim_{n \to \infty} \frac{K_{i,n}}{n} = \sum_{i=1}^{m} x_i\, P(X = x_i\,)$$

Valor esperado

- O termo “valor esperado” é algo enganador…
- Não é na realidade algo que devemos esperar que ocorra
  - Pelo contrário, muitas vezes é muito pouco provável ou mesmo impossível de ocorrer
    - Exemplo: valor médio do lançamento do dado (=3,5)

- Apesar desta dificuldade com o seu nome, o valor esperado desempenha um papel central em Probabilidades e Estatística

Valor esperado : E[X]

- O valor esperado de uma variável designa-se por E[X]

- No caso discreto: $\mathrm{E}[X] = \sum_i x_i p(x_i)$

- No caso contínuo: $E[X] = \int_{-\infty}^{+\infty} x\, f_X(x) dx$

Propriedades do valor esperado

- E[X] é um operador linear

Sendo a e c constantes ($\in R$) e X e Y variáveis aleatórias:

$E[aX] = a\,E[X]$

$E[X+Y] = E[X] + E[Y]$

$E[X+c] = E[X] + c$

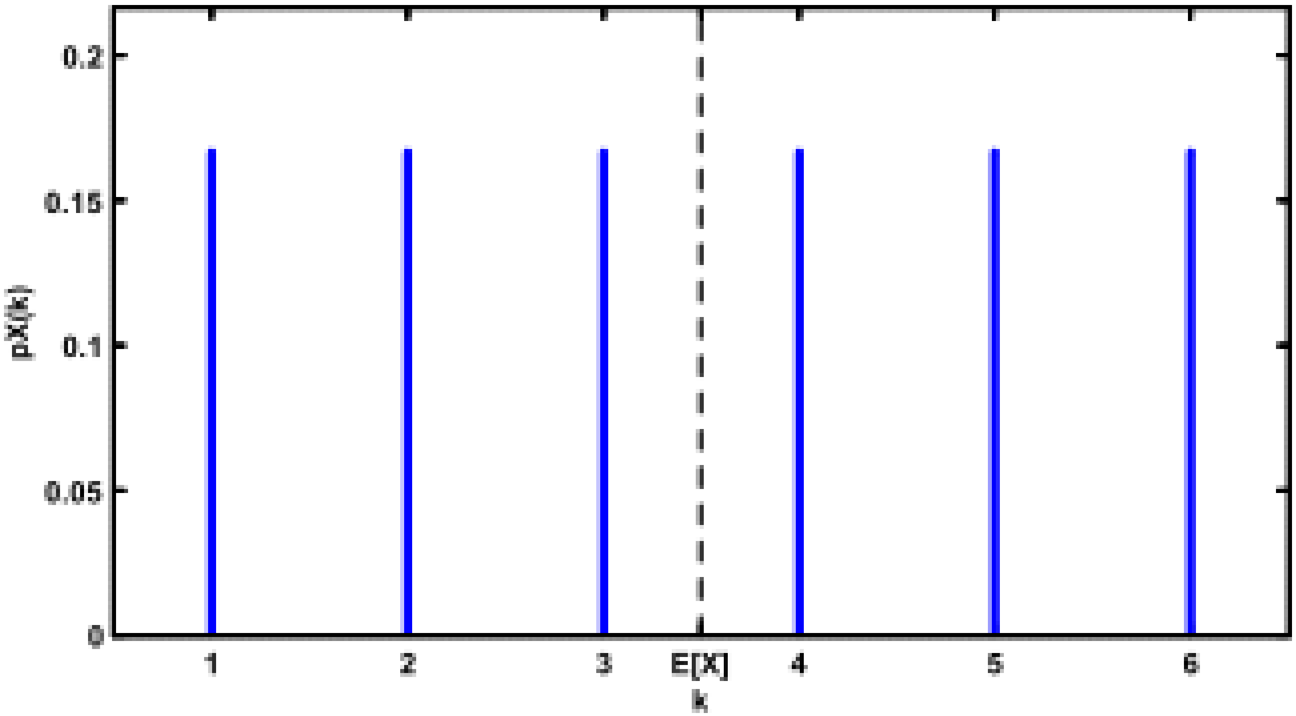
Exemplo de cálculo de E[X]

x_i	$p_X(x_i)$	$x_i\, p_X(x_i)$
-1	.1	-.1
0	.2	.0
1	.4	.4
2	.2	.4
3	.1	.3
		1.0

$\boxed{E[X] = 1.0}$

Exemplo: lançamento de 1 dado

- Função de probabilidade para o resultado do lançamento de uma dado e respetivo valor esperado



A Média pode não ser suficiente

- Se pretendermos comparar as classificações de duas turmas práticas de MPEI é suficiente sabermos a média ?

- Posso ter a mesma média e turmas muito diferentes:
  - Uma turma com a generalidade dos alunos próximos dessa média
  - Outra turma com classificações muito mais dispersas entre 0 e 20

- Uma medida dessa "dispersão" é dada pela variância

Variância

- Ideia base:

  Usar a diferença dos valores da variável para a média (valor esperado) e fazer a sua média

- Para evitar o cancelamento de diferenças negativas e positivas, em vez de usar diretamente o valor da diferença utilizar o seu valor quadrático

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

Variância

- Aplicando a definição de valor esperado temos:
- var(X)=σ^2 = $\sum_i [\boldsymbol{x_i} - \boldsymbol{E}(\boldsymbol{X})]^{\mathbf{2}}\ p(x_i)$

- Propriedade importante:

  var(X)= $E[X^2] - E^2[X]$

  – Demonstra-se facilmente de $E[(X - E(X))^2]$ usando as propriedades de E[X]

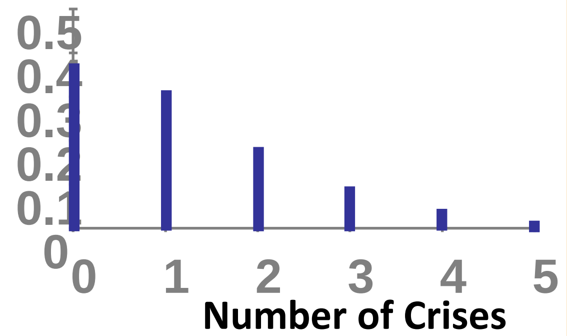
  – Facilita muitos cálculos, evitando uso direto da definição

Desvio padrão

- A raiz quadrada da variância é o desvio padrão

- Muitas vezes representado por σ

Exemplo (discreto)



xi	p(xi)	(xi-μ)	$(xi-E(X))^2$	$(xi-E(X))^2$ p(xi)
0	.37	-1.15	1.32	.49
1	.31	-0.15	0.02	.01
2	.18	0.85	0.72	.13
3	.09	1.85	3.42	.31
4	.04	2.85	8.12	.32
5	.01	3.85	14.82	.15
				1.41

Variância - propriedades

- Sendo X uma variável aleatória e c uma constante :

- Soma de uma constante:

  $\text{var}(X+c)=\text{var}(X)$

- Multiplicação por um factor de escala

  $\text{var}(c\,X) = \boldsymbol{c^2}\ \text{var}(X)$

Média e variância - interpretação

- E[X] pode ser interpretado como:
  - Valor médio de X
  - Centro de gravidade da função massa de probabilidade (caso discreto) ou função de densidade de probabilidade

- Desvio padrão / Variância dá uma medida da dispersão da variável aleatória
  - Pequenos valores indicam var. aleatória muito concentrada em torno da média
    - Se for zero não temos var. aleatória (todos valores iguais à média

Momentos de ordem *n*

- Os conceitos de média e variância podem ser generalizados …

- Momento de ordem n (caso discreto):

$$m_n = E[X^n] = \sum_i \boldsymbol{x_i}^{\boldsymbol{n}} \, p_X(x_i)$$

- Exemplo (dados)

$$E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \frac{1}{6} + 3^2 \frac{1}{6} + \ \ldots$$

$$= \frac{1+2+4+9+16+25+36}{6} = 15{,}1667$$

Momentos centrados de ordem *n*

- A generalização da variância resulta nos momentos centrados de ordem *n*

- $E[(X - E[X])^n] = \sum_i (x_i - E[X])^{\boldsymbol{n}} \; p_X(x_i)$

- A variância é o momento centrado de 2ª ordem

Exemplo de aplicação

- Qual o valor da variância dos valores obtidos no lançamento de um dado honesto ?

- var(X) ?
- var(X) $= E[X^2] - E^2[X]$
- $E[X^2] = ?$
- $E^2[X]$= ?

Tópicos da aula (resumo)

- Variável aleatória (conceito e definição)

- Função massa de probabilidade e função densidade de probabilidade
- Função de distribuição acumulada

- Valor esperado
- Média e Variância
- Momentos

Para saber mais…

- Link(s)

  http://www.stat.berkeley.edu/~stark/SticiGui/Text/randomVariables.htm

- Capítulo 3 do livro “Probabilidades e Processos Estocásticos”, F. Vaz

MPEI 2020-2021

Variáveis aleatórias (continuação): Distribuições

Distribuições - Motivação

- As funções de massa de probabilidade e de densidade de probabilidade (para o caso contínuo) podem assumir as mais variadas formas
- Mas existe um conjunto de "formas" (distribuições) que aparecem repetidamente em muitos e variados problemas
  - Formam um conjunto de ferramentas base que é muito útil conhecer …

Existem muitas distribuições

- Discretas
  - **Bernoulli**
  - **Binomial**
  - **Poisson**
  - **Geométrica**
  - …
- Contínuas
  - **Uniforme**
  - **Normal**
  - **Exponencial**
  - Qui-quadrado
  - T de Student …

- Ver Wikipedia
  - https://en.wikipedia.org/wiki/List_of_probability_distributions

Distribuições Discretas

Distribuição de Bernoulli

- Distribuição directamente relacionada com as experiências de Bernoulli
- Seja A um acontecimento relacionado com o resultado de uma experiência aleatória
- A variável de Bernoulli define-se como
- $I_A(\omega) = \begin{cases} 1 & se\ \omega \in A \\ 0 & caso\ contrário \end{cases}$

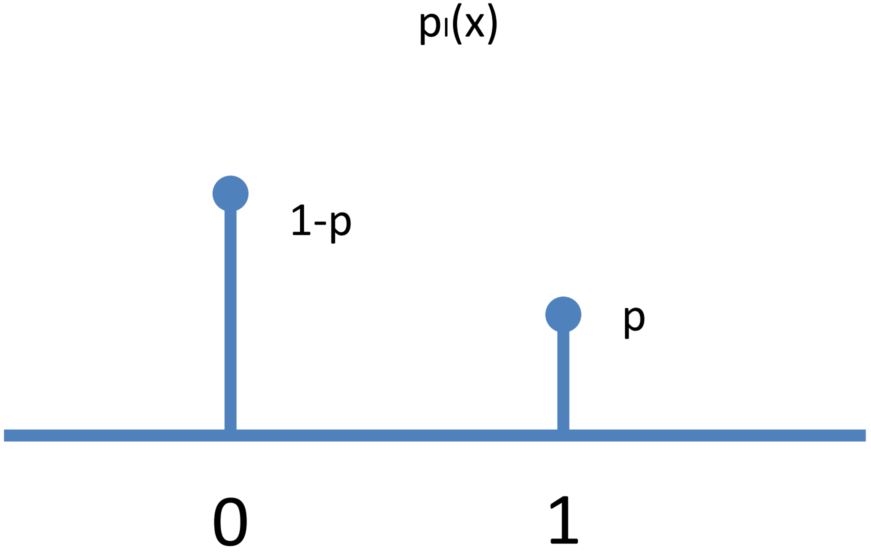
Distribuição de Bernoulli

- O I que usamos para a designar resulta de ser usada muitas vezes como indicadora da ocorrência/não ocorrência de um evento

- Quando o evento ocorre a variável aleatória I assume o valor 1
  - caso contrário o valor 0

Distribuição de Bernoulli

- $S_I = \{0,1\}$
- p= Pr(A)



- $p_I(1)$ =p
- $p_I(0)$ =1-p

- Valor esperado E[I] ?
- Var(I) = ?

Distribuição de Bernoulli

- $\mathrm{E}[I] = \sum_i x_i p(x_i)$
  $= 0 \times (1 - p) + 1 \times p$
  $= \boldsymbol{p}$
- $Var(I) = E[I^2] - (E[I])^2$

- $E[I^2]=0^2 \times (1 - p) + 1^2 \times p = \mathrm{p}$

- $Var(I) = p - p^2 = \boldsymbol{p\,(1 - p)}$

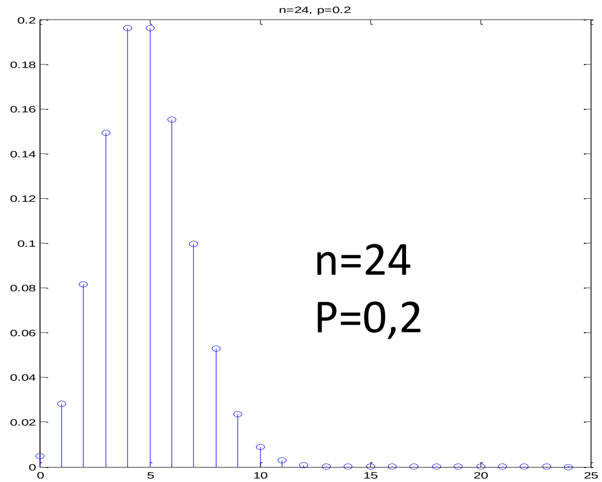
Distribuição Binomial

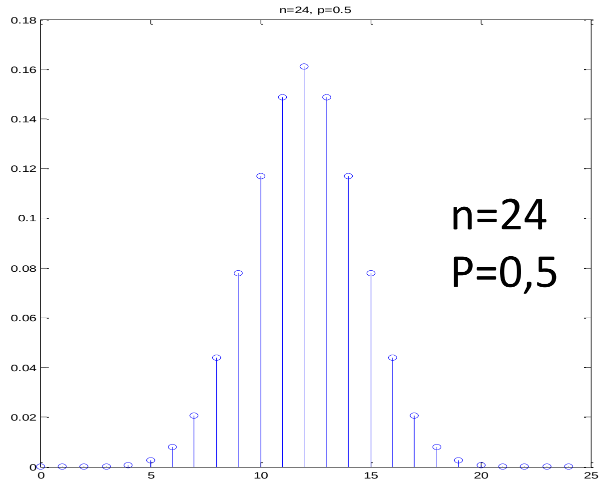
- Directamente relacionada com a Lei Binomial
- Seja X o número de vezes que um acontecimento A ocorre em n experiências de Bernoulli
  - isto é, X representa o número de sucessos em n experiências (observações)

- $X = \sum_{j=1}^{n} I_j \quad \rightarrow S_X = \{0,1,2,..,n\}$

Distribuição Binomial

- $p_X(k) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$





- $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$

Distribuição Binomial – Média e Variância

- Fácil derivar usando o facto de termos **n variáveis de Bernoulli** independentes, que designamos por I_i

- $E[X] = E[\sum I_i] = \sum E[I_i] \quad pq\,?$
$$= p + p + \cdots + p = \boldsymbol{n}\, p$$
- De forma similar

$\text{Var}(X) =$
$$Var(\sum I_i) = \sum Var(I_i) = \cdots = \boldsymbol{n}\, \underbrace{p\,(1-p)}$$

(as variáveis aleatórias I_i são independentes)

Distribuição Binomial - Exemplos

- Têm distribuição Binomial, por exemplo:
  - Número de peças defeituosas num lote de um determinado tamanho (ex: 50 peças)
  - Número de respostas certas num exame de verdadeiro falso
  - Número de clientes que efectuaram compras em 100 que entraram numa loja

Distrib. Binomial - Áreas de aplicação

- A distribuição surge em muitas áreas cientifico-tecnológicas:
  - Engenharia de produção: Muitas vezes as medidas de controlo de qualidade são baseadas na distribuição binomial
    - O caso Binomial aplica-se a qualquer situação industrial em que o resultado é binário e os resultados de ensaio são independentes e com probabilidades constantes
  - Medicina: Por exemplo os resultados "cura" ou "não cura" são importantes na indústria farmacêutica
  - Indústria Militar: "acerta" "falha" é muitas vezes a interpretação do lançamento de um míssil ou de uma missão
  - **Informática**: "acerto" e "falha" é uma interpretação possível para detectores de SPAM, testes a métodos/funções de um programa, procura de informação na web ...

Exemplo de aplicação 1: Transmissão digital

- Um sistema de transmissão digital envia um pacote de 1 kByte através de canal com ruído sendo a probabilidade de erro de cada bit 10^{-3} (ou seja 1 bit em cada mil).
- Considerando que os erros são independentes, determine:
  - Probabilidade de haver 1 erro ?
  - Probabilidade de haver erro ?

Exemplo 2 – segurança de aviões

- Considere que um motor de avião pode falhar com probabilidade p e que as falhas em motores distintos são independentes.
- Se um avião se despenha quando mais do que 50% dos motores falham, é mais seguro voar num avião de 4 motores ou de 2 motores ?
- Faz parte de um dos guiões Práticos
- Como resolver ?

- Sugestão: calcular a probabilidade de cair um avião com 2 motores, repetir para o de 4 motores e comparar os resultados (será função da probabilidade de falha de um avião)

Possível resolução

- O de 2 motores despenha-se se os 2 motores falharem. Qual a probabilidade de 2 falhas em 2 motores ?
- $p2 = p_X(2, n = 2) = \binom{2}{2} p^2 (1-p)^{2-2} = p^2$

- O de 4 despenha-se se 3 ou 4 falharem. Qual a probabilidade ?
- $p4 = p_X(3, n = 4) + p_X(4, n = 4)$
- $= \binom{4}{3} p^3 (1-p)^{4-3} + \binom{4}{4} p^4 (1-p)^{4-4}$
- = $4\,p^3(1-p) \ + p^4 \ = 4\,p^3 - 3\,p^4$

- Relação entre $p2$ e $p4$
- $\frac{p4}{p2} = 4p - 3p^2 = p\,(4 - 3p)$
  - NOTE que depende de p

…

p	p2	p4	p4/p2
0,01	0,0001	0,00000397	0,0397
0,02	0,0004	0,00003152	0,0788
0,03	0,0009	0,00010557	0,1173
0,04	0,0016	0,00024832	0,1552
0,05	0,0025	0,00048125	0,1925
0,06	0,0036	0,00082512	0,2292
0,07	0,0049	0,00129997	0,2653
0,08	0,0064	0,00192512	0,3008
0,09	0,0081	0,00271917	0,3357
0,1	0,01	0,0037	0,37
0,2	0,04	0,0272	0,68
0,3	0,09	0,0837	0,93
0,4	**0,16**	**0,1792**	**1,12**
0,5	0,25	0,3125	1,25
0,6	0,36	0,4752	1,32
0,7	0,49	0,6517	1,33
0,8	0,64	0,8192	1,28
0,9	0,81	0,9477	1,17

p	p2	p4	p4/p2
0,3	0,09	0,0837	0,93
0,31	0,0961	0,091458	0,9517
0,32	0,1024	0,099615	0,9728
0,33	0,1089	0,10817	0,9933
0,34	**0,1156**	**0,117126**	**1,0132**
0,35	0,1225	0,126481	1,0325
0,36	0,1296	0,136236	1,0512
0,37	0,1369	0,146387	1,0693
0,38	0,1444	0,156934	1,0868
0,39	0,1521	0,167873	1,1037

O que significam p4/p2 < 1 ?

é mais seguro voar num avião de 4 motores ou de 2 motores ?

Exemplo de aplicação III

- According to the U.S. Census Bureau, approximately 6% of all workers in Jackson, Mississippi, are unemployed.
- In conducting a random telephone survey in Jackson, what is the probability of getting two or fewer unemployed workers in a sample of 20?

- De: Business Statistics, Ken Black, 6th ed, John Willey & Sons (cap 5)

Resolução

- 6% desempregado => $p = 0{,}06$
- Tamanho da amostra é 20 => $n = 20$
- 94% têm emprego => $1 - p = 0{,}94$
- x é o número de sucessos que se pretende

- Qual é a probabilidade de termos 2 ou menos desempregados na amostra de 20 ?

- Neste tipo de problemas o importante e muitas vezes o mais difícil é identificar o p, n e x

Resolução

$$n = 20$$
$$p = 0{,}06$$
$$q = 1 - p = 0{,}94$$
$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0{,}2901 + 0{,}3703 + 0.2246 = 0{,}8850$$

$$P(X = 0) = \frac{20!}{0!(20-0)!}\left(0{,}06\right)^{0}\left(0{,}94\right)^{20-0} = (1)(1)(0{,}2901) = 0{,}2901$$

$$P(X = 1) = \ldots$$

$$P(X = 2) = \ldots.$$

Distribuição Geométrica

- Seja X o número de vezes que é necessário repetir uma experiência de Bernoulli até obter um sucesso
  – Prob. Sucesso: p   prob. Falha = 1-p

- $p_X(k) = \boldsymbol{p}(1-p)^{k-1}\,, k = 1,2,3, \ldots$

  Porque teremos k-1 insucessos e depois sucesso

- $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} p(1-p)^{k-1}$

Exemplo de aplicação – Helpdesk UA

- Problema:
- Considere o serviço de atendimento via telefone do Helpdesk da UA.
- Supondo que a probabilidade de se conseguir contactar o suporte é p=0,1 (só ao fim de 10 tentativas ☹).
- Determine a probabilidade de necessitar de menos de 3 chamadas até conseguir expor o seu problema ?

- Solução:
- $\Pr(n^{\text{o}}\ chamadas < 3) =$ $\Pr(1\ chamada\ OU\ 2\ chamadas)$
- $= p(1-p)^{1-1} + p(1-p)^{2-1} = p(2-p) = 0{,}19$

Distribuição Geométrica – Média e Variância

- Demontra-se que:

- $E[X] = \frac{1}{p}$
  - Resultado de $\sum_{i=1}^{\infty} i \;\; p(1-p)^{i-1}$
  - Intuitivo: no exemplo do Helpdesk, por exemplo, quanto mais provável atenderem menos chamadas teremos que fazer (em média)

- $Var(X) = (1-p)/p^2$

Dist. Binomial para valores de n elevados

- Consideremos o seguinte cenário:
- Num conjunto de programas a probabilidade de haver pelo menos um erro ao analisar um conjunto de **1000 linhas** de código é **p** (p<1)
  - Não nos interessa o número de erros, apenas se existe algum ou não

- Se o número total de linhas dos programas for **N x 1000** e os dividirmos em blocos de 1000 linhas a probabilidade de **k** blocos terem erros segue a distribuição **Binomial com parâmetros N e p**

(continuação)

- Se quisermos analisar células de 100 linhas, e considerarmos que a distribuição dos erros é uniforme, a probabilidade desce para p/10
  - Teríamos então uma Binomial com parâmetros 10 N e p/10
  - Teoricamente temos a forma de cálculo mas basta N ser um número moderado e 10N começa a ser elevado e os cálculos complicados [mesmo em computador]

- Exemplo: Blocos de 100 linhas; 1000 blocos ; p=0,98/10
- Qual a probabilidade do número de blocos com erro ser inferior ou igual a 100 ?

$$P = F_x(100) = \sum_{k=0}^{100} \binom{1000}{k} 0{,}098^k 0{,}902^{1000-k}$$

☹

...

- As coisas ainda se complicam mais de reduzirmos mais o tamanho dos blocos (100 linhas, 10 linhas ..)

- Será que conseguimos arranjar maneira(s) eficiente(s) de calcular quando o tamanho é muito pequeno ?

- No limite teremos apenas um bloco de uma linha que vai ter, ou não, um erro
  - Número de blocos "infinitesimais" com erro = número de erros

Distribuição de Poisson

- Considere-se que temos uma variável Binomial,n cresce e p decresce por forma a $np \to \lambda > 0$

- Para n grande pode fazer-se as seguintes aproximações: $p \cong \frac{\lambda}{n}$ e $1 - p \cong 1 - \frac{\lambda}{n}$

- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ → Binomial

- $\lim_{n\to\infty} P(X = k) = \lim_{l\to\infty} \binom{n}{k} p^k (1-p)^{n-k}$=
- = ... =$\frac{\lambda^k e^{-\lambda}}{k!}$
- $p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

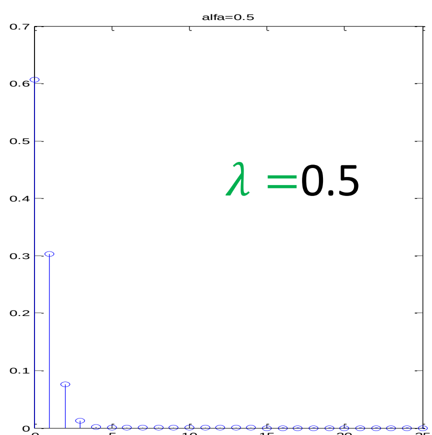
é a função de massa de probabilidade da distribuição de Poisson , com k=0, 1, 2, ...

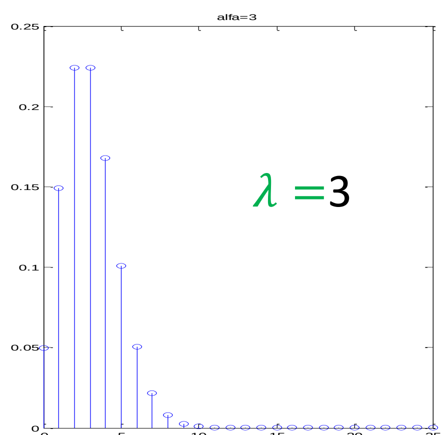
Distribuição para vários valores do parâmetro λ

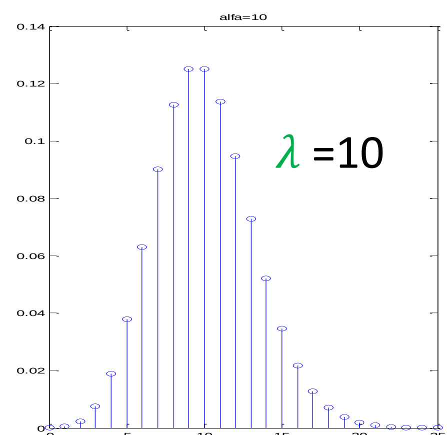
- Função de probabilidade:

$$p_X(k) = \Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0,1,2,3 \; \ldots$$

- Tem apenas um parâmetro, o lambda







Distribuição de Poisson: Média e Variância

- $E[X] = \lambda$
  - Relembre que λ é aproximado por np e o valor esperado da Binomial é np

- $Var(X) = \lambda$

Distribuição de Poisson

- A distribuição de Poisson foca-se apenas no número de ocorrências (discreto) num intervalo de tempo contínuo (ou região do espaço).

- Esta distribuição não tem um número de experiências (n) como na Binomial
  - As ocorrências são independentes das outras ocorrências

Aproximação de Poisson à distribuição Binomial

- Problemas envolvendo a distribuição Binomial em que **n é grande e o valor de p é pequeno**, gerando desta forma **eventos raros**, são os candidatos à utilização da distribuição de Poisson

- Regra prática (“rule of thumb”) :
  - Se n>20 e np <=7   a aproximação de Poisson é suficientemente próxima para ser usada em vez da Binomial

Aproximação de Poisson à distribuição Binomial

- Procedimento para aproximar a Binomial por Poisson:
  1. Calcular a média da Binomial **μ = np**

  2. Como μ é o valor esperado da Binomial, passa a ser o **λ** (=E[X]) de Poisson

  3. Usar a fórmula de Poisson (ou uma tabela)

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Aplicações da Distribuição de Poisson

- As distribuições de Poisson surgem em experiências onde se verificam as seguintes propriedades:
  - O número de resultados que ocorrem num determinado intervalo de tempo ou região é independente do número que ocorre em qualquer outro intervalo temporal ou região espacial disjunta
  - A probabilidade que um resultado ocorra durante um intervalo ou região infinitesimal é proporcional ao comprimento do intervalo ou dimensão da região e não depende das ocorrências fora desse intervalo ou região
  - A probabilidade de haver mais que um resultado numa região infinitesimal é desprezável

Exemplo de aplicação

- Bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes.
- What is the probability of having more than 7 customers in a 4-minute interval on a weekday afternoon?

- De: Business Statistics, Ken Black, 6th ed, John Willey & Sons (cap 5)

Resolução

- Consideremos que o número de clientes (em intervalos de 4 minutos) é representado pela variável aleatória X

- Pretendemos

  P("X > 7 clientes /4 minutos")

- λ= ?

- λ = 3,2 [nº médio de clientes em 4 minutos]

Resolução (continuação)

- A solução requer que calculemos para   k = 8, 9, 10, 11, 12, 13, 14, ... até o valor ser aproximadamente zero
  - Ou usemos o complemento e calculemos k =0,1,2,3,4,5,6,7

- Depois é só somar as probabilidades

- O resultado (0,0168) mostra que é pouco provável que um banco que tem em média 3,2 clientes a cada 4 minutos receba mais de 7 clientes num período de 4 minutos
  - TPC: confirmar este valor

Resolução (continuação)

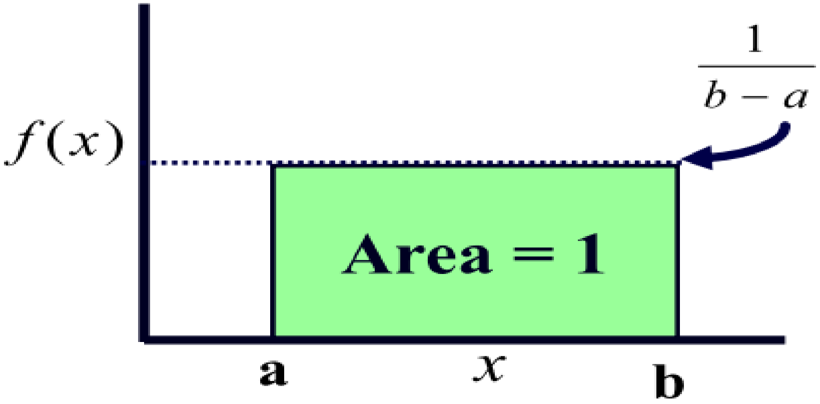
- Este tipo de probabilidades são muito úteis para os gestores de Bancos (e outras instituições com atendimento ao público) dimensionarem o número de pessoas e postos de atendimento

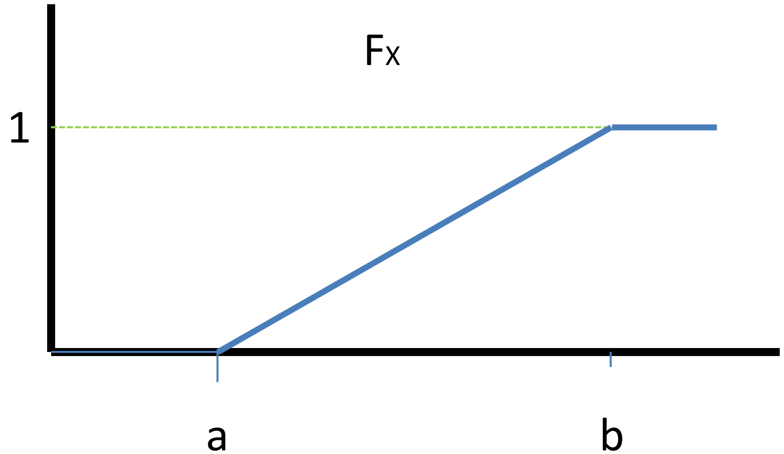
- A distribuição de Poisson é também muito útil na modelação da chegada de mensagens (ou outros tipos de eventos) em redes de computadores

Distribuições contínuas

Distribuição uniforme

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



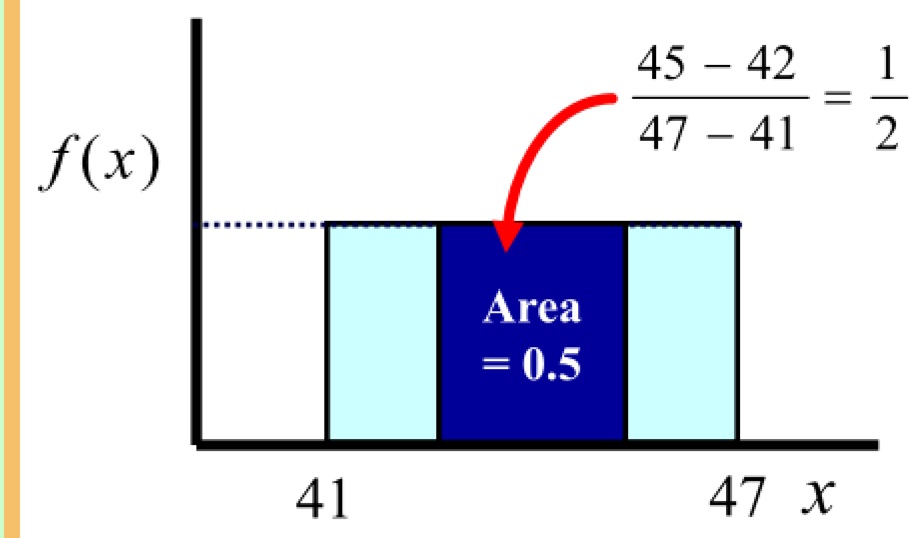


Exemplo

- 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:

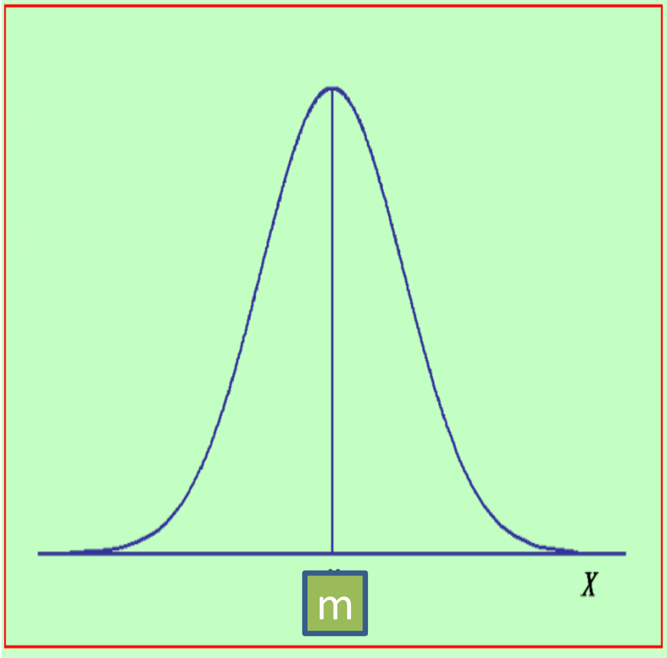
a+rand()*(b-a)

Distribuição Normal (ou Gaussiana)

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



Distribuição Normal

- Função de distribuição acumulada

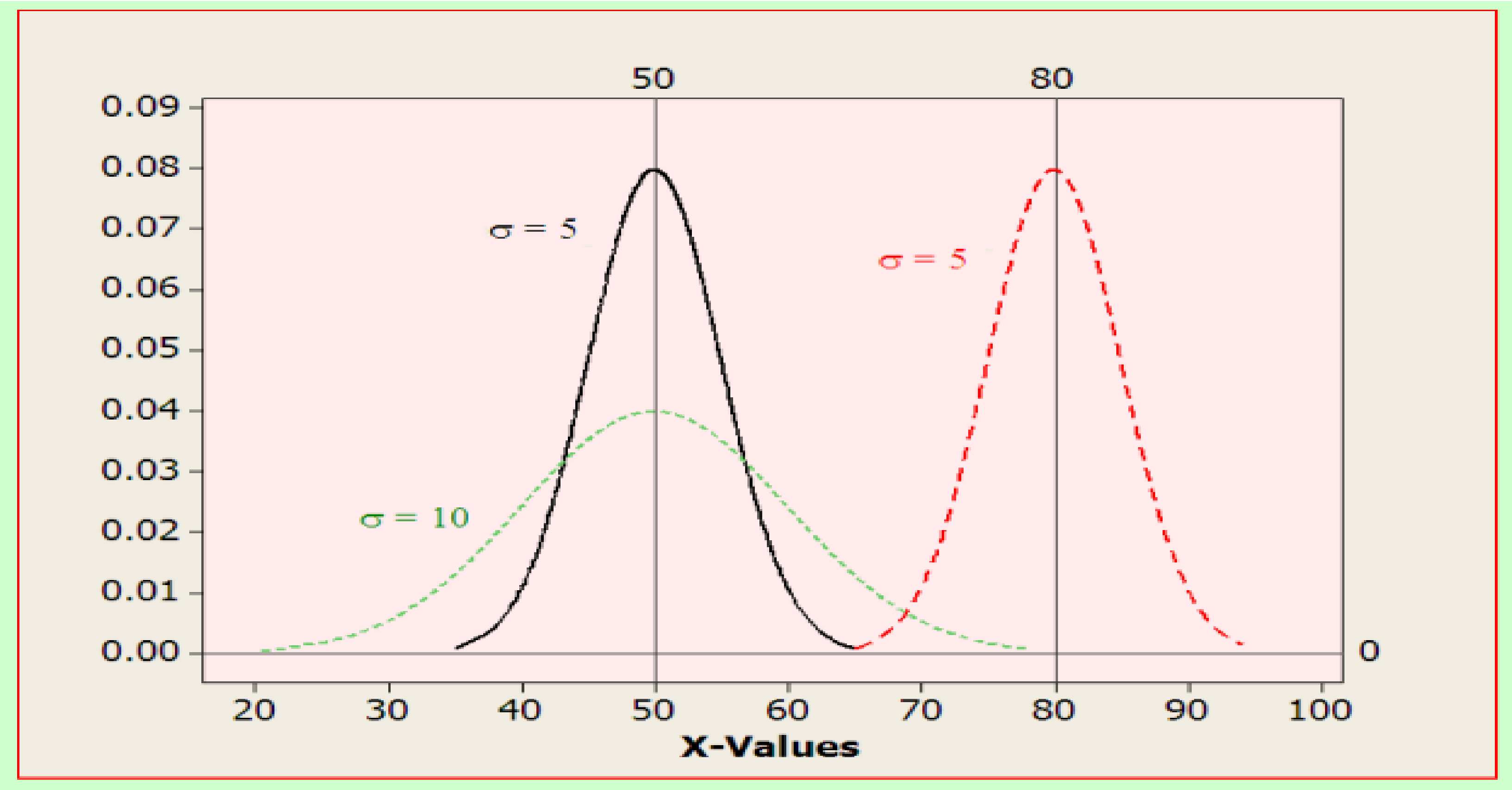
$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}dt$$

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

- 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:

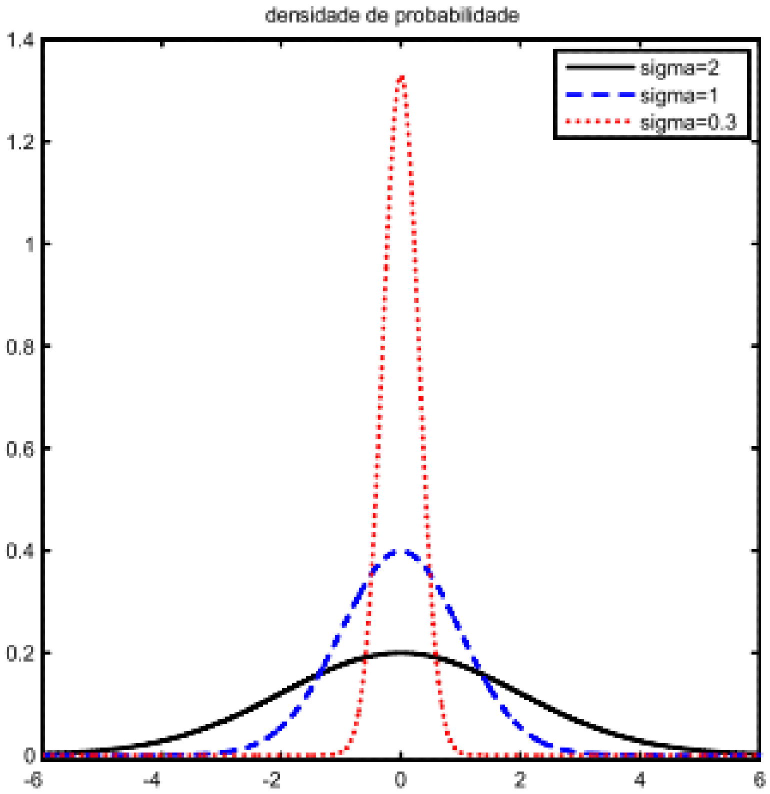
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x)^2}{2}}$$

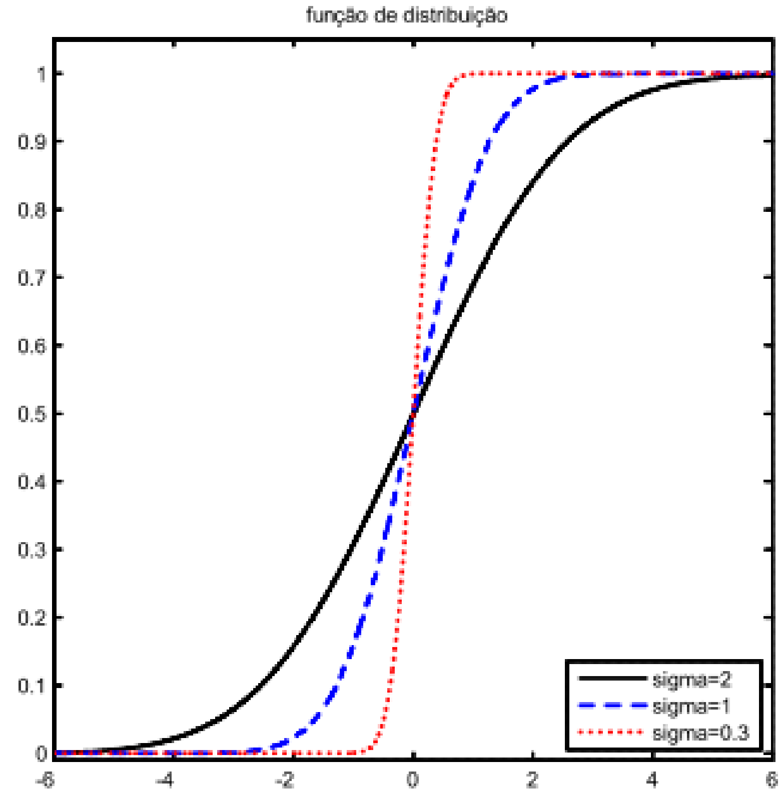
- A função de distribuição acumulada $\Phi(x)$:

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

- $\Phi(x)$ 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]

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

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 função massa de probabilidade da 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

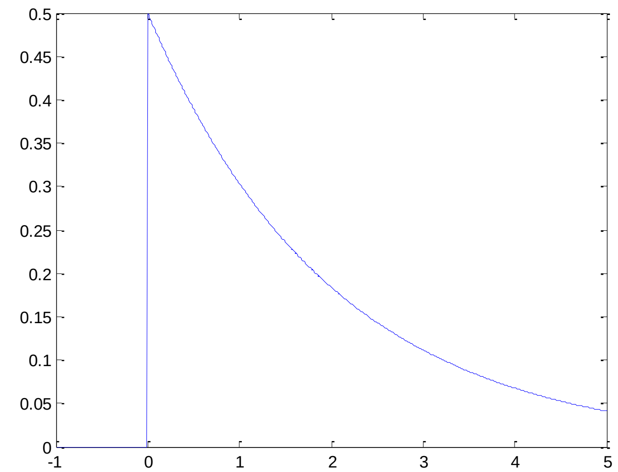
- É não negativa (prob. 0 para $x<0$)
- Está relacionada com a distribuição (discreta) de Poisson
  - Se o número de acontecimentos que ocorrem num intervalo seguem distribuição de Poisson, o tempo entre eles segue distribuição exponencial

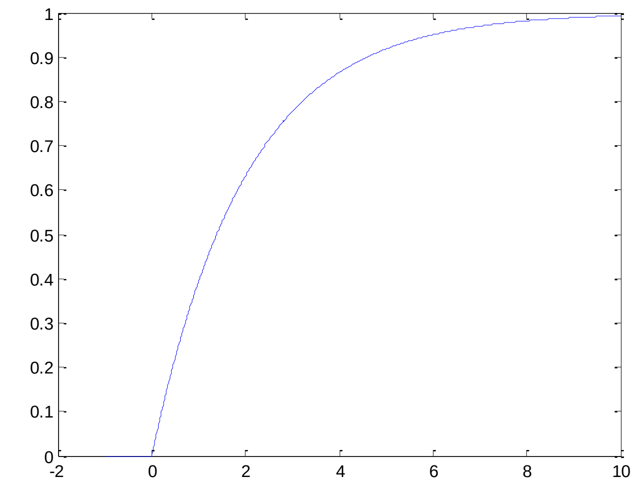
Distribuição exponencial

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

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

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





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?
- $\lambda = \frac{1}{10} = 0{,}1$

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

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

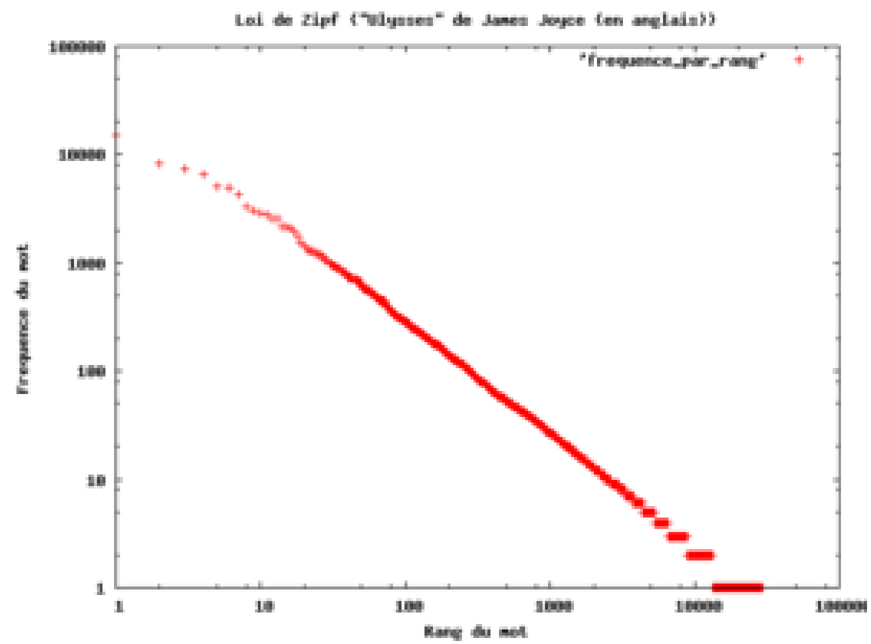
- A **Lei de Zipf** é uma lei empírica que rege a dimensão, importância ou frequência dos elementos de uma lista ordenada
  - formulada na década de 1940 por Zipf, na sua obra *Human Behaviour and the Principle of Least-Effort* ("Comportamento Humano e o Principio do Menor Esforço"),
- Trata-se de uma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

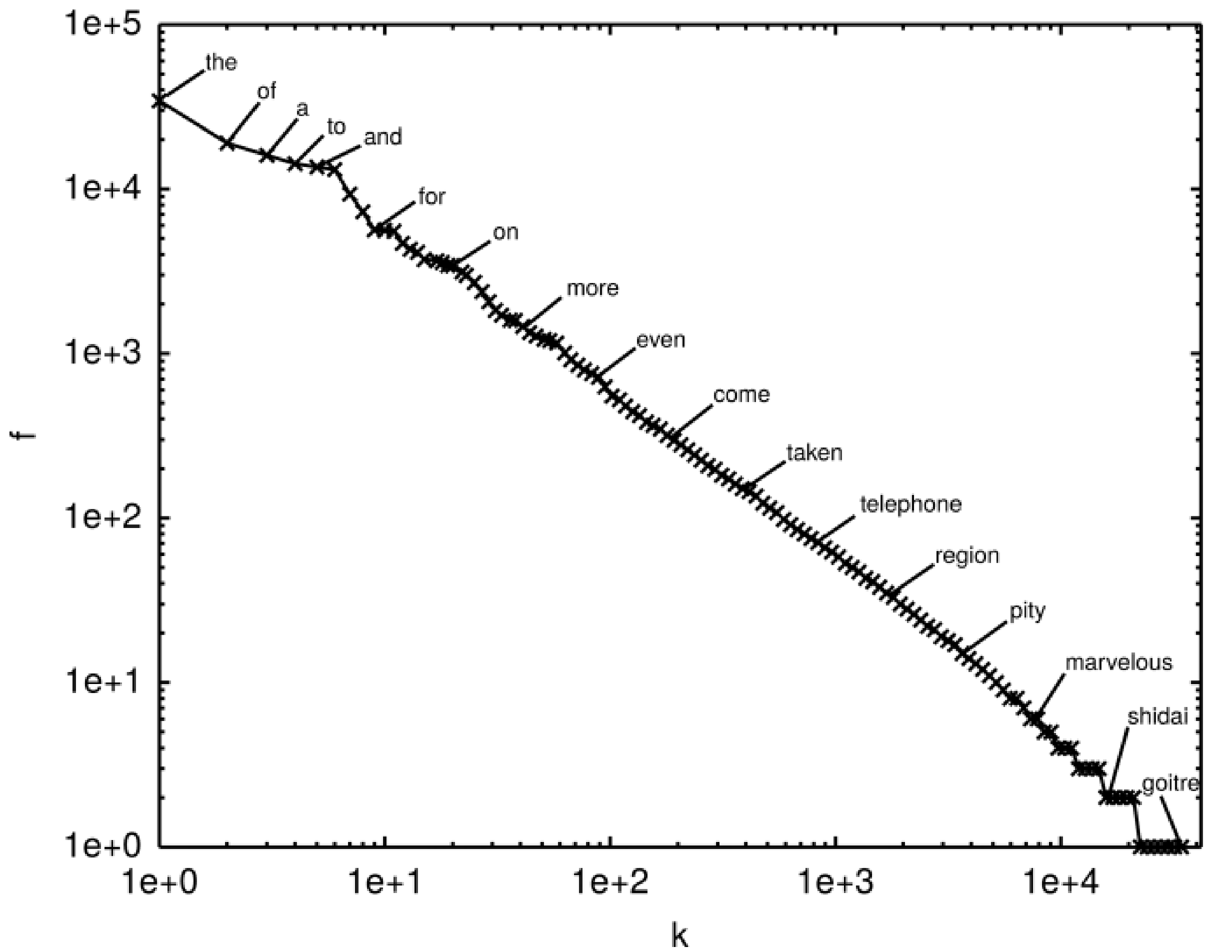


Figura 1: Lei de Zipf no Inglês escrito (dados do OANC). *Rank* (k) versus frequência de ocorrência (f).

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

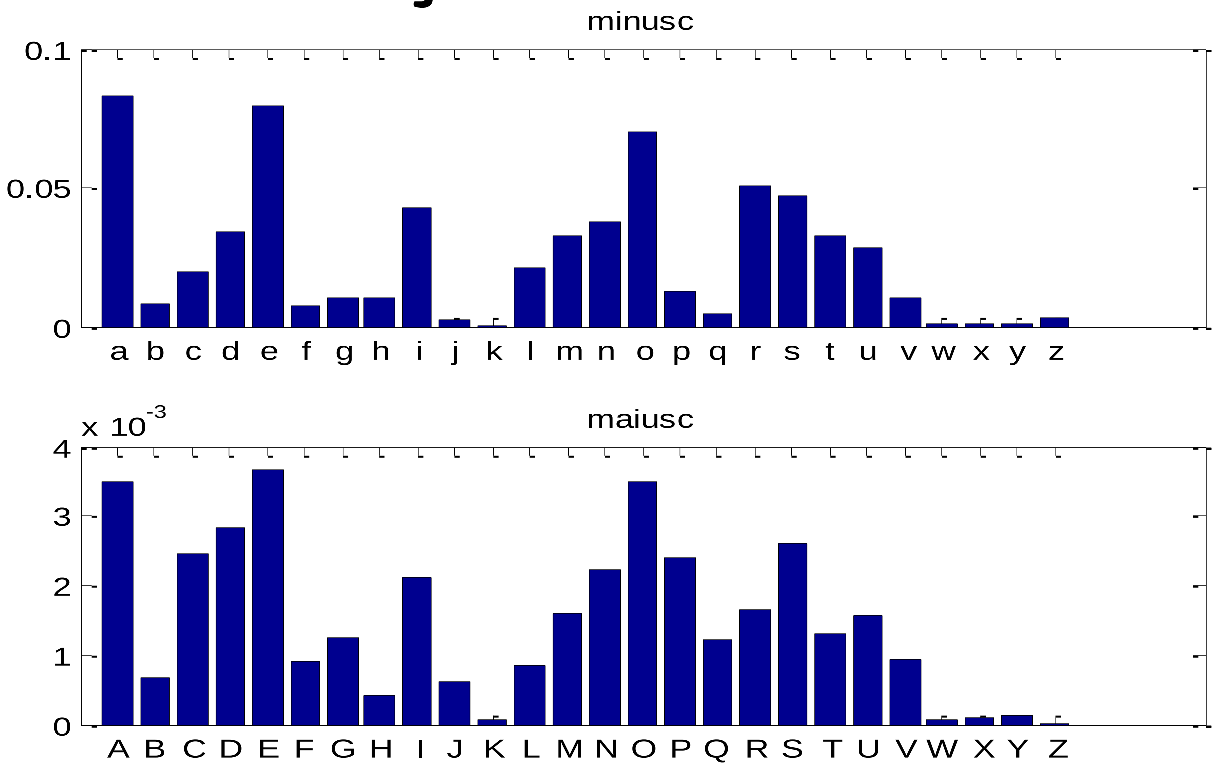
Influência da lei de Zipf na escolha de senhas

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

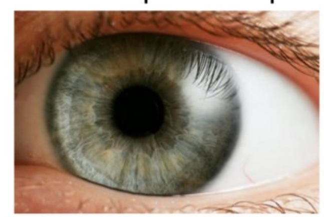
MPEI

Variáveis aleatórias multidimensionais

Motivação

- Trabalhamos frequentemente com grupos de variáveis relacionadas

120 million photoreceptors



256 EEG sensors

- Exemplos:
  - Peso e altura das pessoas
  - Número de temporais em vários meses

$X1$ = número de temporais em Junho (0, 1, ou 2)

$X2$ = número de temporais em Julho (0, 1, ou 2)

Variáveis aleatórias multidimensionais

- Frequentemente temos situações em que os resultados possíveis são conjuntos de várias variáveis aleatórias, *X1*, *X2*,..

- Dois tipos de casos:
  - Experiência aleatória produz várias saídas
  - Repetições da experiência aleatória (com uma única saída)

- A um vector n-dimensional em que as componentes são as variáveis aleatórias $X_1, X_2, \dots, X_n$ chama-se **vector aleatório** ou v.a. Vectorial

$$\mathbf{X} = (X_1 \quad X_2 \quad \dots \quad X_n)$$

Vector aleatório

- **Um vector aleatório** X é uma função que atribui um vector de números reais a todos os resultados ζ em S, o espaço de amostragem da experiência aleatória.
- Exemplo: $\mathbf{X} = (\,H \quad W \quad A)$ com

$$H(\zeta) = \text{altura do estudante } \zeta \text{ em metros,}$$
$$W(\zeta) = \text{peso do estudante } \zeta \text{ em Kg , e}$$
$$A(\zeta) = \text{idade do estudante } \zeta \text{ em anos.}$$

Como caracterizar estas variáveis aleatórias com n-dimensões ?

Funções de distribuição conjuntas

- Para lidar com estas situações envolvendo 2 ou mais variáveis, definem-se, estendendo as definições para uma variável:
  - Função massa de probabilidade conjunta
  - Função de distribuição cumulativa conjunta
  - Função de densidade de probabilidade conjunta

Função probabilidade de massa conjunta

- Para duas variáveis discretas, *X* e *Y*:

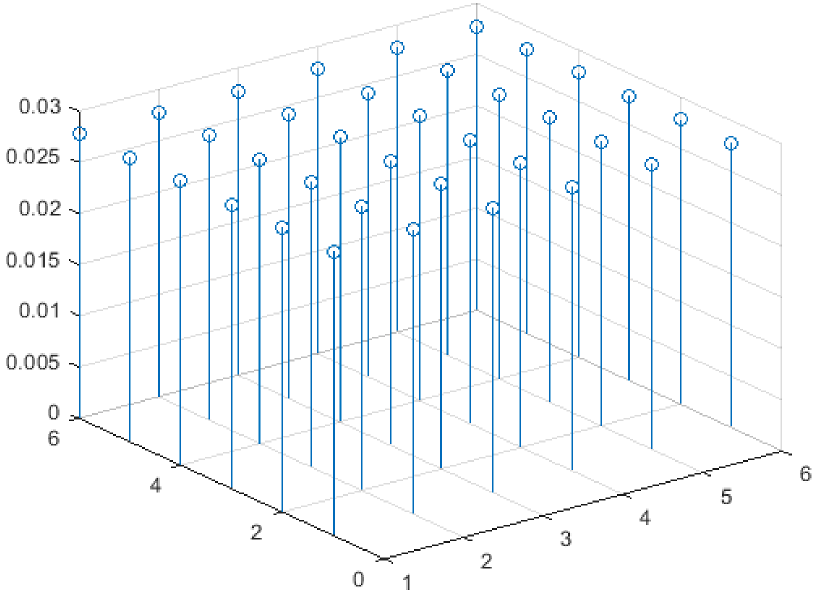
- $p_{X,Y}(i,j) = P(X = i \wedge Y = j)$

- Exemplo: *X*= dado 1; *Y*= dado 2

$$p_{X,Y}(1,1) = p_{X,Y}(1,2) = \ldots = p_{X,Y}(6,6) = 1/36$$

Exemplo (continuação)

- Representação 3D



Função massa de probabilidade conjunta

- A expressão generaliza para mais de 2 variáveis:

- $p_{X_1,X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
- Uma função em $\mathbb{R}^n$, não-negativa

- $\sum_{x_1,x_2,\ldots,x_n} p_{X_1,X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n) = 1$

função de distribuição acumulada conjunta

- Tal como no caso escalar, pode definir-se uma função de distribuição acumulada conjunta
  - Simples extensão

- Para duas variáveis, *X* e *Y*:

$$F_{X,Y}(x,y) = P(X \le x \wedge Y \le y)$$

- Para n variáveis:

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n).$$

No caso discreto é uma função em terraços …

Exemplo 1

- Caso discreto

  Y_1 = número de temporais em Junho (0, 1, ou 2)

  Y_2 = número de temporais in Julho (0, 1, ou 2)

- Tabela com probabilidades

$p_{y1\ y2}(0, 2)$

		Julho (y_2)		
		0	1	2
Junho (y_1)	0	0.05	0.1	0.15
	1	0.1	0.15	0.20
	2	0.15	0.05	0.05

Distribuição de cada uma das variáveis

- A distribuição de cada uma das variáveis pode ser obtida da distribuição conjunta

- Por exemplo, no caso com duas variáveis *X* e *Y*:
- $F_X(a) = P(X \leq a)$
- = $P(X \leq a, Y < \infty)$
- = $F_{X,Y}(a, \infty)$

- De forma similar:
- $F_Y(b) = P(Y \leq b) = F_{X,Y}(\infty, b)$

Funções de probabilidade marginais

- Também se pode obter facilmente a função de massa de probabilidade de cada uma das variáveis

- As fórmulas para o caso discreto são:

- $p_X(x) = \sum_y p_{X,Y}(x, y)$
- $p_Y(y) = \sum_x p_{X,Y}(x, y)$

Funções de probabilidade marginais

- No caso de duas variáveis (X e Y):

- Para obter a função massa de probabilidade de X somamos as linhas apropriadas da tabela representando a função de probabilidade conjunta

- De forma similar obtém-se Y somando as colunas

Exemplo 1

- Para o exemplo introduzido antes..

	Julho (y_2)				
		0	1	2	$p\ (y_1)$
Junho (y_1)	0	0.05	0.1	0.15	0.30
	1	0.1	0.15	0.20	0.45
	2	0.15	0.05	0.05	0.25
	$p\ (y_2)$	0.30	0.30	0.40	1.00

$p_{Y1}(y1) =$

y_1	$p_{Y1}(y_1)$
0	0.30
1	0.45
2	0.25
TOTAL	1.00

y_2	$p_{Y2}(y_2)$
0	0.30
1	0.30
2	0.40
TOTAL	1.00

Generalização

- O caso de *n* variáveis discretas é uma generalização simples

- Se $X_1, X_1, \ldots, X_n$ são variáveis aleatórias discretas no mesmo espaço de amostragem com função massa de probabilidade conjunta:

$$p_{X_1,\ldots X_n}(x_1,\ldots x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

- A função de probabilidade marginal para X_1 é:

$$p_{X_1}(x_1) = \sum_{x_2,\ldots,x_n} p_{X_1,\ldots X_n}(x_1,\ldots x_n)$$

- A função (bidimensional) para a função de probabilidade marginal de X_1 e X_2 :

$$p_{X_1 X_2}(x_1, x_2) = \sum_{x_3,\ldots,x_n} p_{X_1,\ldots X_n}(x_1, x_2, x_3, \ldots, x_n)$$

Independência

- Duas variáveis aleatórias X e Y são independentes se, para qualquer a, b se verificar

- $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$

- Ou seja, são independentes se os eventos $E_a = \{X \leq a\}$ e $E_b = \{Y \leq b\}$ são independentes

Independência

- Em termos de função de distribuição acumulada conjunta:

  X e *Y* são independentes **se e só se**

$$F_{X,Y}(a,b) = F_X(a)F_Y(b)$$

  qualquer que sejam *a* e *b*

- Também, no caso discreto, *X* e *Y* são independentes **se e só se**

$$p(x,y) = p_X(x)\, p_Y(y)$$

- E no caso contínuo $f_{XY}(x,y) = f_X(x)\ \ f_Y(y)$

Generalização – independência de n variáveis aleatórias

- *n* variáveis $X_1, X_2, \dots, X_n$ são independentes se

$$f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall x_1, \cdots, x_n \in \mathbb{R}$$

Exemplo 1

- Y_1 e Y_2 são independentes ?

Julho (y_2)					
Junho (y_1)		0	1	2	$P(y_1)$
	0	0.05	0.1	0.15	0.30
	1	0.1	0.15	0.20	0.45
	2	0.15	0.05	0.05	0.25
	$f(y_2)$	0.30	0.30	0.40	1.00

Julho (y_2)					
Junho (y_1)		0	1	2	$p(y_1)$
	0	0.09			0.30
	1				0.45
	2				0.25
	$p(y_2)$	0.30	0.30	0.40	1.00

Esperança matemática

Extensão das definições

- Os momentos de ordem *j k* das variáveis X, Y definem-se como sendo,
- Caso discreto:

$$E\left[X^j Y^k\right] = \sum_m \sum_n {x_m}^j {y_n}^k \, p_{XY}(x_m, y_n)$$

- Caso contínuo:

$$E\left[X^j Y^k\right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^j y^k f_{XY}(x, y) dx \, dy$$

- Se *j*=1 e *k*=0 ou *j*=0 e *k*=1 temos os valores médios de *X* e *Y*
- Se *j*=2 e *k*=0 ou *j*=0 e *k*=2 temos os valores quadráticos médios

...

- Os momentos centrais conjuntos de ordem *j k* das variáveis X, Y definem-se como:

$$E\left[(X - E[X])^j (Y - E[Y])^k\right]$$

- Para *j*=2 e *k*=0 ou *j*=0 e *k*=2 obtemos as variâncias de X e Y

Correlação

- O momento de ordem *j*=*k*= 1, $E[XY]$, é designado de correlação das variáveis X e Y

- Quando $E[XY] = 0$ as variáveis são ortogonais

$E[XY]$ e Independência

- Sendo *X* e *Y* independentes

$$E[XY] = E[X]E[Y]$$

- Demonstração (caso discreto):

$$E[XY] = \sum_{x,y} xy \, p_{X,Y}(x,y)$$

$=\sum_{x,y} xy \, p_{X,Y}(x,y)$

$= \sum_{x,y} xy \, p(x) p_Y(y)$

$= [\sum_x x \, p_X(x)] \; [\sum_y y \, p_Y(y)]$

$= E[X]E[Y]$

Covariância

- A covariância de duas variáveis X e Y é o seu momento central de ordem *j*= *k*= 1
  - Ou seja $E\left[(X - E[X])\quad(Y - E[Y])\quad\right]$
  - Designa-se por Cov(*X*,*Y*)

- $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$
  $= E\big[XY - XE[Y] - YE[X] + E[X]E[Y]\big]$
  $= E[XY] - 2E[X]E[Y] + E[X]E[Y]$
  $= E[XY] - E[X]E[Y]$

- $E[X] = 0$ ou $E[Y] = 0 \quad \Rightarrow Cov(X, Y) = E[XY]$

Covariância

- É uma generalização da Variância
  $$Cov(X,X) = E[(X - E[X])(X - E[X])]$$
  $$= Var(X)$$

- A covariância é uma medida de relação linear entre as variáveis aleatórias

- Se a relação for não linear, a covariância pode não ser sensível à relação.

Covariância e independência

- Se X e Y são independentes então $\mathrm{Cov}(X,Y) = 0$

- "Demonstração":

- Como vimos $Cov(X,Y) = E[XY] - E[X]E[Y]$
- *X* e *Y* são independentes implica

$$E[XY] = E[X]E[Y]$$

Nota: o contrário não é verdadeiro

pode ter-se Cov(*X*,*Y*)=0 e as variáveis não serem independentes

Propriedades da Covariância

- $Cov(X,X) = Var(X)$
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(cX,Y) = c\, Cov(X,Y)$
- $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$

Demonstração:

$$= E[X(Y+Z)] - E[X]E[Y+Z] =$$
$$= E[XY] + \mathrm{E[XZ]} - \mathrm{E[X]E[Y]} - \mathrm{E[X]E[Z]}$$
$$= E[XY] - \mathrm{E[X]E[Y]} + \mathrm{E[XZ]} - \mathrm{E[X]E[Z]}$$
$$= Cov(X,Y) + Cov(X,Z)$$

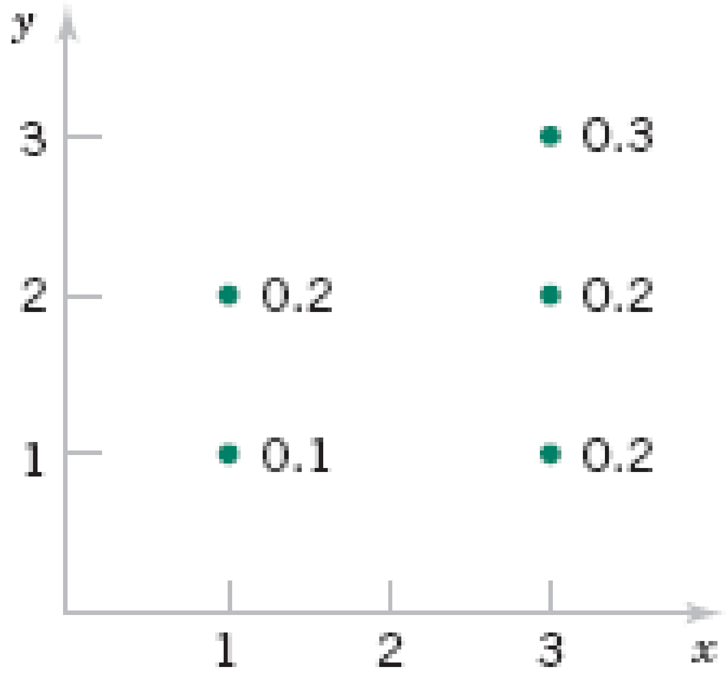
- Generalização: $\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$

Covariância de *n* variáveis

- Se tivermos um vector de *n* variáveis aleatórias $Y = (Y_1, Y_2, \ldots, Y_n)$
- $Cov(Y) = \begin{bmatrix} Cov(Y_1, Y_1) & \cdots & Cov(Y_1, Y_n) \\ \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & \cdots & Cov(Y_n, Y_n) \end{bmatrix}$
- $= \begin{bmatrix} Var(Y_1) & \cdots & Cov(Y_1, Y_n) \\ \vdots & \ddots & \vdots \\ Cov(Y_1, Y_n) & \cdots & Var(Y_n) \end{bmatrix}$

Exemplo

- Considere a seguinte distribuição conjunta de X e Y e calcule $Cov(X, Y)$



Cov(X,Y)= ?

- E(X) = ?
  = 1 x 0,3 + 3 x 0,7 = 2,4
- E(Y)= ?
- = 1 x 0,3 + 2 x 0,4 + 3 x 0,3 = 2,0

- Cov(X,Y) = E[(X-E[X]) (Y-E[Y])]
- = (1-2,4)(1-2,0) x 0,1 + (1-2,4)(2-2,0) x 0,2
- + (3-2,4)(1-2,0) x 0,2 + (3-2,4)(2-2,0) x 0.2
- + (3-2,4)(3-2,0) x 0,3 = 0,2

Coeficiente de correlação

- A **coeficiente de correlação** de duas variáveis X e Y é:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \ \sigma_Y}$$

- Demonstra-se que $-1 \leq \rho_{XY} \leq 1$

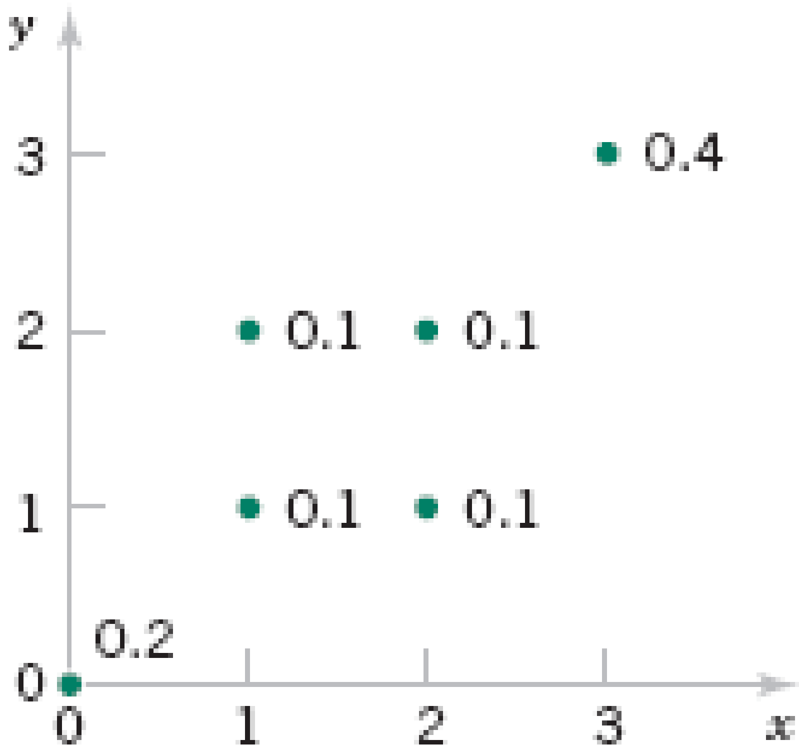
- E que os valores extremos (1 e -1) se obtém para a relação linear $Y = a\,X + b$ com a> 0 ou a <0, respectivamente

Coeficiente de correlação

- Se $\rho_{XY} = 0$ as variáveis dizem-se **descorrelacionadas**

- Como se viu, se X e Y são independentes, a sua covariância é nula e portanto são descorrelacionadas
  - Mas o contrário não é verdadeiro

Exemplo de cálculo de ρ_{XY}



x	y	P(x,y)
0	0	0,2
1	1	0,1
1	2	0,1
2	1	0,1
2	2	0,1
3	3	0,4
	SOMA	1,0

Cálculo de E[XY], E[X] e E[Y]

x	y	P(x,y)	xy P(x,y)	x P(x)	y P(y)	$x^2\ P(x)$
0	0	0,2	0x0x0,2=0	0	0	0
1	1	0,1	1x1x0,1=0,1	0,1	0,1	0,1
1	2	0,1	0,2	0,1	0,2	0,1
2	1	0,1	0,2	0,2	0,1	0,4
2	2	0,1	0,4	0,2	0,2	0,4
3	3	0,4	3,6	1,2	1,2	3,6
	SOMA	1,0	4,5	1,8	1,8	4,6

Exemplo de cálculo de ρ_{XY}

- Var(*X*) = $E[X^2] - (E[X])^2$ = 4,6 – 3,24 = 1,36
- Var(*Y*) é igual à de *X*

- $Cov(X, Y) = E[XY] - E[X]E[Y]$
- = 4,5 – (1,8) (1,8)=1,26

- Finalmente:
- $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \ \sigma_Y} = \frac{1,26}{(\sqrt{1,36})(\sqrt{1,36})} = 0,926$

MPEI

Soma e Combinação Linear de Variáveis Aleatórias

Funções de Variáveis Aleaórias

Motivação

- Se somarmos duas variáveis aleatórias X_1 e X_2 quais as características da variável aleatória $S = X_1 + X_2$?

  - Em termos de momentos ?
    - Em especial média e variância

  - Em termos de função de distribuição ?

- E no caso geral $S_n = X_1 + X_2 + \cdots + X_n$?

Média da soma de n variáveis

- Sejam $X_1, X_2, \dots, X_n$, n variáveis aleatórias e $S_n = X_1 + X_2 + \cdots + X_n$ a sua soma
- Teorema: **A média da soma de *n* variáveis é igual à soma das médias**
- Demonstração

$$
\begin{aligned}
E\left[S_n\right] &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\sum_{j=1}^{n} x_j\right) f_{X_1 \dots X_n}(x_1, \dots, x_n)\, dx_1 \dots dx_n = \\
&= \sum_{j=1}^{n} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_j f_{X_1 \dots X_n}(x_1, \dots, x_n)\, dx_1 \dots dx_n = \\
&= \sum_{j=1}^{n} \int_{-\infty}^{+\infty} x_j f_{X_j}(x_j)\, dx_j = \sum_{j=1}^{n} E\left[X_j\right]
\end{aligned}
$$

Variância da soma de n variáveis

- Considerando da mesma forma $S_n = X_1 + X_2 + \cdots + X_n$:
- Teorema: **A variância da soma de *n* variáveis aleatórias é dada pela soma de todas as variâncias e covariâncias**

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{\substack{j=1 \\ j \neq k}}^{n} \sum_{k=1}^{n} Cov(X_j, X_k)$$

- Demonstração:

$$\begin{aligned} Var(S_n) &= E\left[\sum_{j=1}^{n}(X_j - E[X_j]) \sum_{k=1}^{n}(X_k - E[X_k])\right] = \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} E[(X_j - E[X_j])] \, E[(X_k - E[X_k])] \end{aligned}$$

Variância da soma de *n* variáveis

- Se as variáveis **são independentes**, $Cov(X_j, X_k) = 0$, para todo o $j \neq k$, pelo que:

- $\boldsymbol{Var(S_n) = \sum_{i=1}^{n} Var(X_i)}$
  - **Variância da soma igual a soma das variâncias**

- Se para além de independentes forem **identicamente distribuídas (IID)**
  e tivermos $E[X_i] = \mu$ e $Var(X_i) = \sigma^2, i = 1,2, \dots, n$
  a média e variância da soma são dadas por:

- $\mathrm{E}[S_n] = n\,\mu$ e $Var(S_n) = n\,\sigma^2$

Função de distribuição da soma de 2 variáveis aleatórias independentes

- Caso discreto (2 v.a. Discretas X e Y)
- Fazendo $\mathrm{Z} = X + Y$
- $p_Z(z) = P(X + Y = z)$

$=\sum_{\{(x,y)|x+y=z\}} P(X = x, Y = y)$

$$= \sum_{x} P(X = x, Y = z - x)$$

$= \sum_x p_X(x) \quad p_Y(z - x)$; devido à indep.

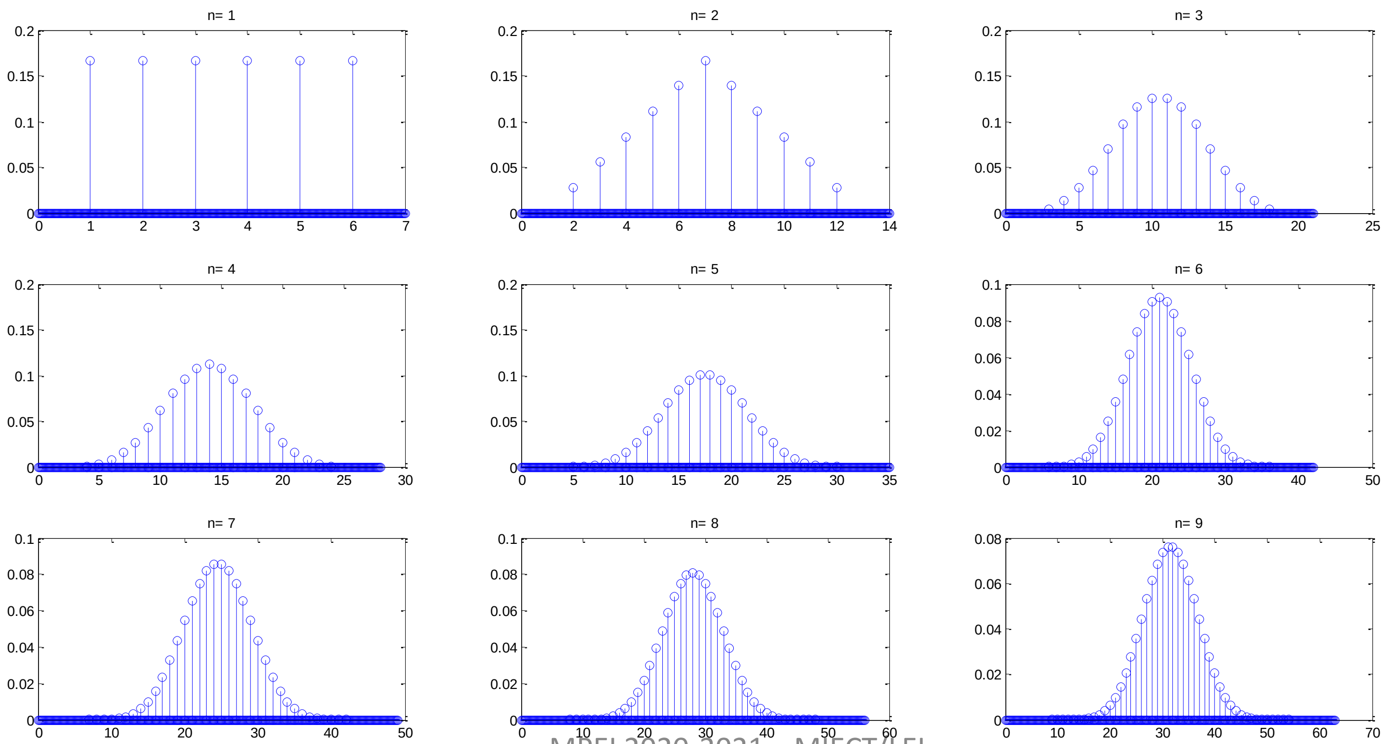
$= p_X(x) * p_Y(z)$

demoConvolucao.m

- Que é a **convolução** discreta de p_X e p_Y

Exemplo (em Matlab)

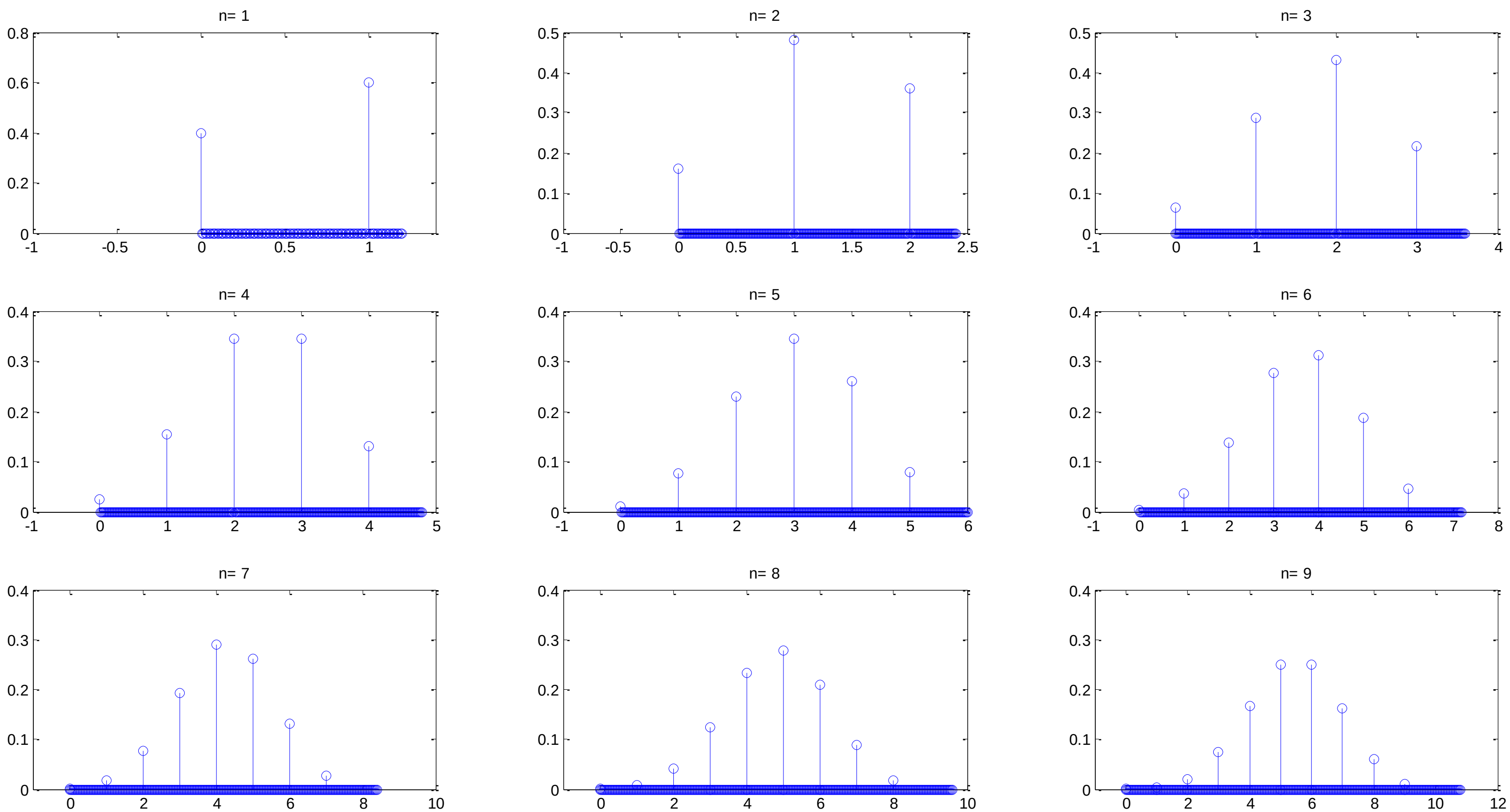
- Usando conv() e a pmf relativa à variável X correspondente ao lançamento de um dado honesto (n=1, 2, ..., 9)



Outro exemplo

demoConvolucaoMoeda.m

- Sendo X relativa ao número de caras num lançamento de moeda não honesta
  - com probabilidade de cara = 0,6



Caso contínuo

- Sendo X e Y independentes e contínuas
- Fazendo novamente $\mathrm{Z} = X + Y$
- Para obter a função densidade prob. de Z , primeiro obtém-se a f. densidade conjunta de X e Z e depois integra-se

$$
\begin{aligned}
F_{Z|X}(z \mid x) &= \mathbf{P}(Z \leq z \mid X = x) = \mathbf{P}(X + Y \leq z \mid X = x) = \mathbf{P}(x + Y \leq z \mid X = x) = \mathbf{P}(Y \leq z - x \mid X = x) \\
&= \mathbf{P}(Y \leq z - x) \qquad \text{using the independence of } X \text{ and } Y \\
&= F_Y(z - x)
\end{aligned}
$$

$$
f_{Z|X}(z \mid x) = \frac{d}{dz} F_{Z|X}(z \mid x) = \frac{d}{dz} F_Y(z - x) = f_Y(z - x)
$$

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Z}(x,z)\,dx = \int_{-\infty}^{\infty} f_X(x) f_{Z|X}(z \mid x)\,dx \\
&= \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,dx \equiv f_X * f_Y(z)
\end{aligned}
$$

Obtém-se através da convolução, agora contínua

- Para mais informação, ver, por exemplo:
- https://engineering.purdue.edu/~ipollak/ece302/SPRING12/notes/23_GeneralRVs-8_Sums.pdf

Combinações lineares de variáveis aleatórias

- Os **resultados anteriores generalizam-se** facilmente para o caso de termos uma soma pesada (combinação linear) $Y_n = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$
  - Em que $c_1, c_2, \ldots, c_n$ são constantes

- $E[Y_n] = c_1 E[X_1] + c_2 E[X_2] + \cdots + c_n E[X_n]$
- $Var(Y_n) =$
  $\sum_{i=1}^{n} {c_i}^2 \, Var(X_i) + \sum_i \sum_{j\ (\neq i)} c_i c_j Cov(X_i, X_j)$

- Se independentes $Var(Y_n) = \sum_{i=1}^{n} {c_i}^2 \, Var(X_i)$

Exemplo de aplicação

- Um semicondutor tem 3 camadas.
- Se a espessura das várias camadas tiver uma variância de 25, 40 e 30 nanómetros quadrados, qual a variância da espessura das 3 camadas ?

- Considerando X_1, X_2, X_3 como as espessuras das camadas e independência
- $Var(X_1 + X_2 + X_3) = Var(X_1) + var(X_2) + var(X_3) = 95$
- Mostrando a propagação das variâncias de todas as camadas para o resultado final
  - Os problemas somam-se ☹

Funções de variáveis aleatórias

A soma (simples ou pesada) de variáveis aleatórias é um caso particular

Funções de v. a. múltiplas

- Muitas vezes encontramos **problemas em que temos uma transformação das v. a.** $X_1, X_2, \ldots, X_n$ que produz variáveis aleatórias $Y_1, Y_2, \ldots, Y_m$
- O caso mais simples é termos uma função escalar de várias variáveis aleatórias

$$Z = g(X_1, X_2, \ldots, X_n)$$

- A função de distribuição acumulada de Z é determinada como sabemos calculando a probabilidade do conjunto $\{Z \leq z\}$

  i.e. A região R_z do espaço n-dimensional tal que

  $R_z = \{\mathbf{x}: g(\mathbf{x}) \leq \mathbf{z}\}$ $\quad\quad$ $\mathbf{x} = (\boldsymbol{x_1}\ \boldsymbol{x_2}\ \ldots\ \boldsymbol{x_n})$
- Logo
  - $F_Z(z) = \int_{\mathrm{x} \in R_Z} \ldots \int \; f_X(\mathrm{x})\, d\mathrm{x}$

Expectância de funções de v. aleatórias

- Expectância de uma função de 2 var. aleatórias

  $Z = g(X, Y)$

  $$E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dx\, dy$$

- Para o caso de variáveis discretas :
  - $E[Z] = \sum_i \sum_n g(x_i, y_n)\, p_{X,Y}(x_i, y_n)$

- Resultado generalizável para uma função com um número arbitrário de variáveis aleatórias

  $Z = g(\mathbf{X})$ em que $\mathbf{X}$ (bold) é um vector

  $$E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

Exemplo

- $Z = g(X, Y) = X + Y \quad i.e.$ Soma de 2 v.a.
- $E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y) f_{X,Y}(x,y)\,dx\,dy$
- $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{X,Y}(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\,dx\,dy$
- $= \int_{-\infty}^{\infty} x\, f_X(x)\,dx + \int_{-\infty}^{\infty} y f_Y(y)\,dy$
- $= \boldsymbol{E(X) + E(Y)}$
- Resultado válido quer as variáveis **sejam independentes ou não**
- Mostra (conjuntamente com a propriedade de escala, $E(aX) = aE(X)$) que a esperança é um **operador linear**

Momentos de funções de variáveis aleatórias

- Momento de ordem n de uma função escalar de um vector aleatório:

  $Z = g(\mathbf{X})$ em que $\mathbf{X}$ (bold) é um vector

  $$E[\boldsymbol{Z^n}] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \boldsymbol{g^n}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

- Aplicando para obter a variância temos:
- $\mathrm{Var}[Z] = E[Z^2] - E^2[Z]$
- $= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \boldsymbol{g^2}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} - \left(\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \boldsymbol{g}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}\right)^{\mathbf{2}}$

Média de variáveis aleatórias

- E se a função aplicada a um vetor de n variáveis aleatórias for a média dessas variáveis?

- Interessa-nos em especial o caso em que essas *n* variáveis são independentes e identicamente distribuídas (IID)

Valor esperado da Média

- Se criarmos a variável aleatória relativa à média de *n* variáveis IID X_i,

$$\boldsymbol{M_n = \frac{S_n}{n}},$$

- e assumindo $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$
- teremos :
- $\boldsymbol{E[M_n]} = E\left[\frac{S_n}{n}\right]$
- $= \frac{\sum_i E[X_i]}{n}$
- $= E[X_i] = \boldsymbol{\mu}$

Variância da Média

- **Var($\boldsymbol{M_n}$)= ?**
- **Var[$\boldsymbol{M_n}$]** $= Var\left(\frac{S_n}{n}\right)$
- $= \frac{1}{n^2}\frac{\sum_i Var[X_i]}{1}$
- $= \frac{Var(X_i)}{n} = \boldsymbol{\frac{\sigma^2}{n}}$

- À medida que se aumenta o número de experiências vai diminuindo a variância da estimativa da média

Aplicação interessante de soma de variáveis aleatórias

Contadores estocásticos

Motivação

- Evitar contadores grandes quando o volume de dados é grande
  - Por exemplo: na contagem de células

- Como um contador de n bits contará no máximo até 2^n eventos, será este o limite a ultrapassar

Primeira solução

- Para duplicar o número de eventos que se podem contar, incrementa-se o contador com probabilidade 1/2 cada vez que ocorre um evento

- A ideia *é incrementar o contador metade das vezes (em média)*

…

- Com base na função rand() podemos agora tomar decisões aleatórias com probabilidade 1/2 e portanto construir uma função para incrementar (ou não) o contador:

```
if (rand() < 0.5) then
    incrementar_contador
endif
```

Em Matlab

- Podemos facilmente simular o resultado após 100 eventos:

```
% gera 100 var aleatórias indep em [0,1]
x = rand(1, 100);

% calcular quantas são < 0.5
n = sum(x < 0.5);
```

- n representará o valor do contador após os 100 eventos

Qual é o valor médio do contador após k eventos?

- O contador é uma variável aleatória, determinada por uma sucessão de experiências aleatórias

- Associando uma variável aleatória a cada evento, de forma a representá-lo probabilisticamente
- Seja X_i a variável aleatória que representa o incremento i, com valor 1 se o contador foi incrementado, e valor zero caso contrário.
- Como $P(X_i = 0)$ e $P(X_i = 1)$ são iguais a 1/2, tem-se

- $E[X_i] = 0 \times P(X_i = 0) + 1 \times P(X_i = 1) = \frac{1}{2}$

Valor médio

- O valor do contador após k eventos é a soma dos k incrementos, $S = X_1 + X_2 + \cdots + X_k$

- E o valor médio:
- $E[S] = E[X_1 + X_2 + \cdots + X_k]$
- $= E[X_1] + E[X_2] + \cdots + E[X_k]$
- $= \frac{1}{2} + \cdots + \frac{1}{2} = \frac{k}{2}$

- Como o valor médio do contador após k eventos é $k/2$, o número de eventos pode ser estimado através do dobro do número registado pelo contador

Variância

- A variância de um qualquer dos X_i é
- $Var(X_i) = E\left[{X_i}^2\right] - \left(E[X_i]\right)^2$

- $E\left[{X_i}^2\right] = 0^2 \times P(X_i = 0) + 1^2 \times P(X_i = 1)$
- $= \frac{1}{2}$

- $Var(X_i) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

Variância (continuação)

- Como as variáveis X_i são independentes, a variância de S é
- $Var(S) = Var(X_1 + X_2 + \cdots + X_k)$
- $= Var(X_1) + Var(X_2) + \cdots + Var(X_k)$
- $= \frac{1}{4} + \cdots + \frac{1}{4} = \frac{k}{4}$
- O que implica $\sigma = \frac{\sqrt{k}}{2}$
- Para n=10000 teremos:
  - média 5000
  - desvio padrão 50

Comparar com simulação (contadores1.m)

Distribuição de probabilidade

- Pode calcular-se a probabilidade de, após k eventos, o valor do contador ser n.

- Fixemos $k = 4$:
- Teremos $X_1, X_2, X_3\ e\ X_4$
  - Variáveis binárias que descrevem se o contador é incrementado ou não após o evento 1,2,3 e 4
- O que nos dá 16 possibilidades (2^4)

X_1	X_2	X_3	X_4	valor do contador
0	0	0	0	0
0	0	0	1	1
0	0	1	0	1
0	0	1	1	2
0	1	0	0	1
0	1	0	1	2
0	1	1	0	2
0	1	1	1	3
1	0	0	0	1
1	0	0	1	2
1	0	1	0	2
1	0	1	1	3
1	1	0	0	2
1	1	0	1	3
1	1	1	0	3
1	1	1	1	4

- É agora fácil determinar as probabilidades, por contagem:
- $p(0) = \frac{1}{16}$
- $p(1) = \frac{4}{16}$
- $p(2) = \frac{6}{16}$
- $p(3) = \frac{4}{16}$
- $p(4) = \frac{1}{16}$

Generalizando

- Sendo p a probabilidade de incrementar e $1 - p$ a probabilidade de não incrementar …
- A probabilidade de observar uma soma igual a n após k experiências é:

$$p(n) = \binom{k}{n} p^n \ (1-p)^{k-n}$$

Variante 1

- Como proceder para alargar mais a gama do contador?

- Imaginemos, por exemplo, que se quer multiplicar por 64 essa gama. A solução natural é incrementar com probabilidade $1/64$ em vez de ½

- O valor médio de X_i será agora $\frac{1}{64}$
- E $E[S] = \quad \ldots . = \frac{k}{64}$
- Neste caso, o número de eventos pode ser estimado por $64\, n$, sendo n o valor do contador

Segunda solução

- Neste caso o contador é incrementado com probabilidade cada vez menor à medida que o seu valor aumenta:
- quando o contador contém *n*, a probabilidade de um incremento é 2^{-n}

Eventos	*Valor do contador*	*Número de eventos*
x	1	1
x		
x	2	3
x		
x		
x		
x	3	7
x		
x		
x		
x		
x		
x		
x		
x	4	15