Estatística para Engenharia

Prof. Charles Casimiro Cavalcante

charles@gtel.ufc.br

Departamento de Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://charlescasimiro.github.io

UNIVERSIDADE FEDERAL DO CEARÁ

“A teoria da aleatoriedade é, fundamentalmente, uma codificação do bom senso.”

Leonard Mlodinow (O Andar do Bêbado, 2008)

Conteúdo do curso

1. Introdução à Probabilidade
2. Variáveis Aleatórias
3. Geração de Números Aleatórios
4. Inferência Estatística

Introdução à Probabilidade

Conceitos básicos

Incerteza

- Probabilístico/Estatístico $\rightarrow$ "Incertezas" objetivas são mensuráveis
- Lógica nebulosa (*fuzzy*) $\rightarrow$ difíceis de quantizar

Conceitos básicos - cont.

Experimento aleatório

É um experimento cujo resultado não pode ser *predito/antecipado* com certeza antes do experimento ser realizado, muito embora os possíveis resultados possam eventualmente ser enumerados.

Exemplos típicos de experimentos aleatórios

- Jogar uma moeda ou um dado.
- Retirar uma carta de um baralho.
- **Realizar uma medição de uma variável**.

Conceitos básicos - cont.

Observações importantes

- Assume-se que o experimento pode ser repetido infinitas vezes sob as mesmas condições.
- Importante pois teoria de probabilidades está interessada no comportamento dos resultados de um experimento que é repetido muitas vezes.
- À medida que o número de repetições aumenta, **emergem** certos padrões na freqüência de ocorrência dos resultados.
- Uma definição completa de um experimento aleatório requer um definição cuidadosa sobre o que chamamos de *resultados do experimento*.

Conceitos básicos - cont.

Espaço amostral e eventos

- O **espaço amostral**, $\mathcal{S}$, associado com um experimento é a coleção de todos os resultados possíveis do experimento.
- Um **evento** é o conjunto de pontos amostrais com uma dada característica particular.
- Eventos são simbolizados por letras maiúsculas em itálico, por exemplo: $A, B, C, ...$ ou $A_1, A_2, ...$.
- Cada ponto amostral é chamado evento simples ou um elemento do espaço amostral.
- Necessidade de operações em conjuntos!

Conceitos básicos - cont.

$\mathcal{S}$ = espaço amostral

Conjunto dos resultados possíveis de um determinado experimento

	moeda:	$\mathcal{S} = \{\text{cara}, \text{coroa}\}$	(*espaço discreto*)
Exs:	dado:	$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$	(*espaço discreto*)
		$\mathcal{S} = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	(*espaço contínuo*)

Conceitos básicos - cont.

$\mathcal{F}$ = campo (classe) de Borel

Deve satisfazer

1. Fechado em relação à complemento

$$A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F} \tag{1}$$

2. Fechado quanto à adição

$$\left.\begin{array}{l} A \in \mathcal{F} \\ B \in \mathcal{F} \end{array}\right\} (A + B) \in \mathcal{F} \tag{2}$$

Ex: moeda: $\mathcal{F} = \{\{\text{cara}\}, \{\text{coroa}\}, \mathcal{S}, \emptyset\}$
dado: $\mathcal{F} = \{\{\text{par}\}, \{\text{ímpar}\}, \mathcal{S}, \emptyset\}$

Conceitos básicos - cont.

Propriedades de $\mathcal{F}$

❶ $\mathcal{S} \in \mathcal{F}$ pois

$$A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$$

$$\mathcal{S} = (A + \overline{A}) \in \mathcal{F}$$

❷ Se $A \in \mathcal{F}$ e $B \in \mathcal{F} \Rightarrow A, B \in \mathcal{F}$

Teorema de *De Morgan*

$$A \cdot B = \overline{\overline{A} + \overline{B}}$$

$$\overline{A} \in \mathcal{F}, \overline{B} \in \mathcal{F} \Rightarrow (\overline{A} + \overline{B}) \in \mathcal{F} \Rightarrow \overline{\overline{A} + \overline{B}} \in \mathcal{F} \Rightarrow A \cdot B \in \mathcal{F}$$

❸ $\emptyset \in \mathcal{F}$

$$A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$$

$$A, \overline{A} \in \mathcal{F} \Rightarrow \emptyset \in \mathcal{F}$$

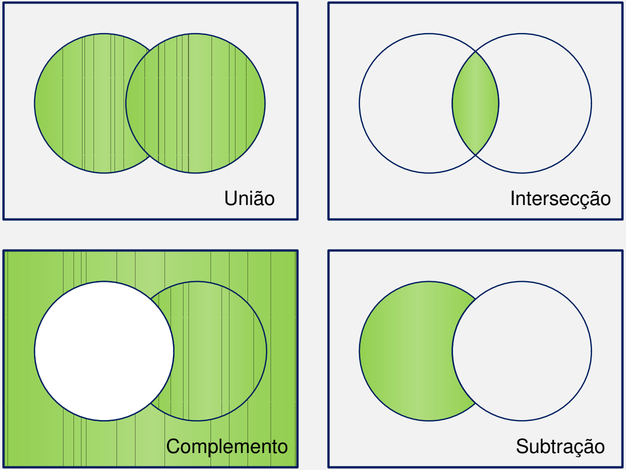
Conceitos básicos - cont.

Conceitos essenciais em conjuntos

- ✎ Conjunto: representa uma coleção de objetos;
- ✎ Elemento: qualquer um dos componentes de um conjunto;
- ✎ Pertinência: é a característica associada a um elemento que faz parte de um conjunto - $a \in A$ e $a \notin A$;
- ✎ Cardinalidade: é número de elementos de um conjunto finito;
- ✎ Operações: união ($\cup$ ou $+$), intersecção ($\cap$ ou $\cdot$), complementaridade e diferença.

Conceitos básicos - cont.

Operações em conjuntos



Eventos probabilísticos

Evento

Definição: Qualquer subconjunto de $\mathcal{S}$ que constitui um $\mathcal{F}$

Eventos mutuamente exclusivos (disjuntos)

- É natural pensar que os elementos de espaço amostral devem ser *mutualmente exclusivos* ou *disjuntos*.
- Ou seja, para um dado experimento, a ocorrência de um evento exclui a ocorrência de um outro.

Exemplo: Tirar a sorte no cara ou coroa

$$\left.\begin{array}{c} A = \{\text{cara}\} \\ B = \{\text{coroa}\} \end{array}\right\rangle A{\cdot}B = \emptyset \quad \textbf{(eventos mutuamente exclusivos)}$$

Definição de probabilidade

Probabilidade: definição axiomática

Axioma 1 - A probabilidade de um evento A não pode ser negativa: $\Pr(A) \geq 0$.

Axioma 2 - O evento $\mathcal{S}$, que inclui todos os resultados possíveis, deve ter probabilidade unitária: $\Pr(\mathcal{S}) = 1$

Axioma 3 - Para eventos *mutuamente exclusivos*, a probabilidade da união destes eventos é a soma das probabilidades dos eventos individuais:

$$\Pr(A_1 \cup A_2 \cup \cdots A_n) = \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n) \quad (3)$$

Definição de probabilidade - cont.

Probabilidade: definição axiomática (cont.)

Logo, esta definição axiomática, em termos de um campo de Borel é uma função que mapeia o campo em um número real para mensurar a incerteza. Ou seja, temos

$$\Pr(\cdot) : \mathcal{F} \to \mathbb{R} \tag{4}$$

Probabilidade: definição axiomática (cont.)

Com base nos três axiomas, pode-se fazer a seguinte definição:

> *Probabilidades são quantidades numéricas que quantificam a chance de ocorrência de eventos, e que são não-negativas (axioma 1), aditivas sobre eventos mutuamente exclusivos (axioma 3) e cujo valor máximo é igual a 1 sobre todos os resultados mutuamente exclusivos possíveis (axioma 2).*

Definição Clássica (a priori)

- A probabilidade de um certo evento A ocorrer, é definida como:

$$\Pr(A) = \frac{\text{Possíveis resultados favorecendo o evento } A}{\text{Número total de resultados possíveis}}$$

- "Probabilidade de um evento é uma medida numérica da chance de ocorrência de um evento de interesse".
- **Exemplo 1**: Qual a probabilidade de se tirar o número 5 ao jogar um dado?
- **Exemplo 2**: Qual a probabilidade de se tirar uma quadra no pôquer? E um *straight flush*?

Definição Clássica (a priori) - cont.

- Em outras palavras:

  "A probabilidade de um evento é o número de vezes em que o resultado desejado pode ocorrer, dividido pelo total de resultados possíveis".

- **Importante** - Na definição clássica de probabilidade assumimos **implicitamente** que os resultados possíveis têm todos a mesma chance de ocorrer.

Definições alternativas de probabilidade - cont.

Definição Freqüentista (a posteriori)

- Suponha que alguém realize o experimento de jogar uma moeda quatro vezes, obtendo 3 caras e 1 coroa.
- Sendo a *freqüência* de um evento A definida como o número de observações ou ocorrências de cada evento, pode-se definir a freqüência relativa, $f(A)$, deste evento como:

$$\boxed{f(A) = \frac{\text{Número de ocorrências do evento } A}{\text{Número total de repetições do experimento}}}$$

Definições alternativas de probabilidade - cont.

Definição Freqüentista (a posteriori) - cont.

- Assim,

$$f(\text{cara}) = \frac{3}{4} = 0,75 \qquad \text{e} \qquad f(\text{coroa}) = \frac{1}{4} = 0,25$$

- **Questão**: A probabilidade teórica de se tirar cara ou coroa não é $\frac{1}{2}$ para cada um dos eventos?
- Na verdade, define-se probabilidade no sentido freqüentista como o valor limite de $f(a)$, para um número infinito de repetições do experimento.
- Na prática, "infinito" é substituído por grande!

Definições alternativas de probabilidade - cont.

Resumo

- *Clássica* ("preditiva"):

$$\Pr(A) = \frac{n_A}{N}$$

em que n_A é o número de possibilidades de ocorrências de A e N é o número total de ocorrências

- *Freqüencial*

$$\Pr(A) = \lim_{N \to \infty} \frac{n_A}{N}$$

(aplicação maior em eventos discretos)

Propriedades de probabilidade

❶ $\Pr(\emptyset) = 0$

$\Pr(A) = \Pr(A + \emptyset) = \Pr(A) + \Pr(\emptyset) \Rightarrow \Pr(\emptyset) = 0$

❷ $\Pr(\overline{A}) = 1 - \Pr(A)$

$\Pr(\mathcal{S}) = \Pr(A + \overline{A}) = 1 \Rightarrow \Pr(A) + \Pr(\overline{A}) = 1 \Rightarrow \Pr(\overline{A}) = 1 - \Pr(A)$

❸ $\Pr(A + B) = \Pr(A) + \Pr(B) - \Pr(AB)$

$A + B = A + AB + \overline{A}B = A + \overline{A}B$

mas A e $\overline{A}B$ são mutuamente exclusivos

❹ Se $A \subseteq B \Rightarrow \Pr(A) \leq P(B)$

$B = A + \overline{A}B, A \cdot (\overline{A}B) = \emptyset$

$\Pr(B) = \Pr(A) + \underbrace{\Pr(\overline{A}B)}_{\geq 0}$

Definição de probabilidade conjunta

Probabilidade conjunta

É a probabilidade de se obter simultaneamente 2 ou mais eventos específicos como resultado de um experimento aleatório.

Considerando dois eventos

- Seja $\mathcal{S}$ o espaço amostral associado a um certo experimento aleatório. Sejam A e B dois eventos de $\mathcal{S}$.
- A probabilidade de se obter o evento $A \cap B$ na realização do experimento é denotada por

$$\Pr(A \cap B) = \Pr(A \text{ e } B) = \Pr(A\,, B) = \Pr(AB) \qquad (5)$$

Definição de probabilidade conjunta - cont.

Considerando n eventos

- Considerando $\mathcal{S}$ o espaço amostral associado a um certo experimento aleatório;
- Sendo $A_1, A_2, \ldots, A_n$, n eventos de $\mathcal{S}$,
- Assim, a probabilidade de se obter o evento $A_1 \cap A_2 \cap \cdots \cap A_n$ quando da realização do experimento é denotada por

$$\begin{aligned} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n) &= \Pr(A_1 \text{ e } A_2 \text{ e } \cdots \text{ e } A_n) \\ &= \Pr(A_1, A_2, \ldots, A_n) \\ &= \Pr(A_1 A_2 \ldots A_n) \end{aligned} \tag{6}$$

Definição de probabilidade conjunta - cont.

Exemplo

Considere o experimento de rolar um par de dados sendo um verde (A) e outro vermelho (B). Considere que $A_i,\ i = 1, \ldots, 6$, simboliza a ocorrência da face i no dado verde e que $B_j,\ j = 1, \ldots, 6$, denota a ocorrência da face j no dado vermelho. Determinar as probabilidades conjuntas $\Pr(A_i, B_j),\ i, j = 1, 2, \ldots, 6$.

Solução

- **Espaço amostral** $(\mathcal{S})$: $\mathcal{S} = \{(A_i, B_j),\ i, j = 1, 2, \ldots, 6\}$.
- **Cardinalidade de** $\mathcal{S}$ $(\sharp\mathcal{S})$: $\sharp\mathcal{S} = 36$.
- **Evento de interesse** (E_{ij}): $E_{ij} = A_i \cap B_j$, logo $\sharp E_{ij} = 1$
- Assim, $\Pr(E_{ij}) = \Pr(A_i \cap B_j) = \frac{\sharp E_{ij}}{\sharp\mathcal{S}} = \frac{1}{36}$

Definição de probabilidade conjunta - cont.

Para o caso em que se deseja calcular a probabilidade conjunta de dois eventos A_i e B_j, é comum organizar os valores das probabilidades $\Pr(A_i \cap B_j)$ em uma matriz (ou tabela) de probabilidades conjuntas.

➡ Para o exemplo anterior, em que $i, j = 1, \ldots, 6$, tal tabela é apresentada a seguir.

Definição de probabilidade conjunta - cont.

	B_1	B_2	B_3	B_4	B_5	B_6	$\Pr(A_i)$
A_1	$\Pr(A_1, B_1)$	$\Pr(A_1, B_2)$	$\Pr(A_1, B_3)$	$\Pr(A_1, B_4)$	$\Pr(A_1, B_5)$	$\Pr(A_1, B_6)$	$\Pr(A_1)$
A_2	$\Pr(A_2, B_1)$	$\Pr(A_2, B_2)$	$\Pr(A_2, B_3)$	$\Pr(A_2, B_4)$	$\Pr(A_2, B_5)$	$\Pr(A_2, B_6)$	$\Pr(A_2)$
A_3	$\Pr(A_3, B_1)$	$\Pr(A_3, B_2)$	$\Pr(A_3, B_3)$	$\Pr(A_3, B_4)$	$\Pr(A_3, B_5)$	$\Pr(A_3, B_6)$	$\Pr(A_3)$
A_4	$\Pr(A_4, B_1)$	$\Pr(A_4, B_2)$	$\Pr(A_4, B_3)$	$\Pr(A_4, B_4)$	$\Pr(A_4, B_5)$	$\Pr(A_4, B_6)$	$\Pr(A_4)$
A_5	$\Pr(A_5, B_1)$	$\Pr(A_5, B_2)$	$\Pr(A_5, B_3)$	$\Pr(A_5, B_4)$	$\Pr(A_5, B_5)$	$\Pr(A_5, B_6)$	$\Pr(A_5)$
A_6	$\Pr(A_6, B_1)$	$\Pr(A_6, B_2)$	$\Pr(A_6, B_3)$	$\Pr(A_6, B_4)$	$\Pr(A_6, B_5)$	$\Pr(A_6, B_6)$	$\Pr(A_6)$
$\Pr(B_j)$	$\Pr(B_1)$	$\Pr(B_2)$	$\Pr(B_3)$	$\Pr(B_4)$	$\Pr(B_5)$	$\Pr(B_6)$	1,00

Tabela 1: Tabela de probabilidades conjuntas.

Definição de probabilidade conjunta - cont.

No caso do exemplo, dois dados justos, temos que todos os eventos são equiprováveis. Assim, a tabela fica como a seguir.

	B_1	B_2	B_3	B_4	B_5	B_6	$\Pr(A_i)$
A_1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
A_2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
A_3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
A_4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
A_5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
A_6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$\Pr(B_j)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1,00

Tabela 2: Tabela de probabilidades conjuntas (eventos equiprováveis).

Definição de probabilidade conjunta - cont.

Exemplo hipotético

- ✍ Imagine o experimento de lançar os dois dados do exemplo anterior 100 vezes e determinar as probabilidades conjuntas $\Pr(A_i, B_j)$.
- ✍ Podemos então anotar a frequência relativa de cada evento e calcular as probabilidades conjuntas. Como é esperado, os valores das probabilidades neste caso podem ser diferentes.
- ✍ A tabela de probabilidades conjuntas é mostrada a seguir. Os valores entre parênteses se referem ao número de ocorrências do evento.

Definição de probabilidade conjunta - cont.

	B_1	B_2	B_3	B_4	B_5	B_6	$\Pr(A_i)$
A_1	0,03 (3)	0,02 (2)	0,01 (1)	0,01 (1)	0,02 (2)	0,01 (1)	0,10
A_2	0,01 (1)	0,01 (1)	0,02 (2)	0,03 (3)	0,03 (3)	0,02 (2)	0,12
A_3	0,02 (2)	0,01 (1)	0,02 (2)	0,01 (1)	0,04 (4)	0,03 (3)	0,13
A_4	0,01 (1)	0,01 (1)	0,03 (3)	0,02 (2)	0,05 (5)	0,06 (6)	0,18
A_5	0,01 (1)	0,02 (2)	0,03 (3)	0,04 (4)	0,06 (6)	0,07 (7)	0,23
A_6	0,02 (2)	0,02 (2)	0,02 (2)	0,04 (4)	0,07 (7)	0,07 (7)	0,24
$\Pr(B_j)$	0,10	0,09	0,13	0,15	0,27	0,26	1,00

Tabela 3: Tabela de ocorrência e probabilidade relativa do exemplo hipotético.

Problema Proposto: Calcular a probabilidade de ocorrer o número 1 no dado verde (evento A_1) e o número 3 no dado vermelho (evento B_3), ou seja, determinar $\Pr(A_1, B_3)$.

Definição de probabilidade marginal

Probabilidade marginal

- É a probabilidade de ocorrência do evento A_i, $\Pr(A_i)$, independentemente dos resultados dos eventos B_j, $j = 1, ..., N$.

$$\Pr(A_i) = \sum_{j=1}^{N} \Pr(A_i, B_j) \tag{7}$$

- De modo similar, a probabilidade marginal de B_j, $\Pr(B_j)$, é dada por:

$$\Pr(B_j) = \sum_{i=1}^{N} \Pr(A_i, B_j) \tag{8}$$

Probabilidades condicionais

- A probabilidade de ocorrência de A dado que ocorreu B é dada por:

$$\boxed{\Pr(A|B) \triangleq \frac{\Pr(AB)}{\Pr(B)}, \quad \Pr(B) > 0} \tag{9}$$

- Por sua vez, a probabilidade de ocorrência de B dado que ocorreu A é dada por:

$$\boxed{\Pr(B|A) \triangleq \frac{\Pr(AB)}{\Pr(A)}, \quad \Pr(A) > 0}$$

Probabilidade condicional - cont.

É uma probabilidade?

$\Pr(A|B)$ **é probabilidade**, pois

- ✓ $\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)} \geq 0$
- ✓ $\Pr(\mathcal{S}|B) = 1$
- ✓ Para
  $A \cdot C = \emptyset \Rightarrow \Pr\left[(A+C)|B\right] = \Pr(A|B) + \Pr(C|B)$

Probabilidades de Bayes

- Notamos ainda que temos:

$$\boxed{\Pr(A|B)P(B) = \Pr(AB) = \Pr(B|A)\Pr(A)} \quad (10)$$

- Assim, podemos escrever:

$$\boxed{\Pr(A|B) \triangleq \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}, \quad \Pr(B) > 0} \quad (11)$$

- Esta expressão é conhecida como **Regra (da Probabilidade) de Bayes**.

Definição de eventos independentes

Eventos independentes

- Dois eventos são ditos independentes se a ocorrência de um deles não afeta a ocorrência do outro.
- Formalmente, os eventos A e B são ditos $independentes$ se

$$\boxed{\Pr(A \cap B) = \Pr(A \cdot B) = \Pr(A) \cdot \Pr(B)} \tag{12}$$

Eventos independentes - cont.

- Neste caso, eventos independentes, tem-se que as probabilidades condicionais reduzem-se às probabilidades marginais:

$$\boxed{\begin{aligned} \Pr(A_i|B_j) = \Pr(A_i) \quad \text{e} \\ \Pr(B_j|A_i) = \Pr(B_j) \end{aligned}} \begin{aligned} (13) \\ (14) \end{aligned}$$

Eventos independentes - cont.

Dois eventos A e B são independentes se

$$\Pr(A \cdot B) = \Pr(A) \cdot \Pr(B)$$

Generalizando (para três eventos): A, B e C

$$\left.\begin{aligned} \Pr(AB) &= \Pr(A) \cdot \Pr(B) \\ \Pr(AC) &= \Pr(A) \cdot \Pr(C) \\ \Pr(BC) &= \Pr(B) \cdot \Pr(C) \end{aligned}\right\} \Pr(ABC) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C)$$

Independência probabilística - cont.

Propriedades de eventos independentes

1. $\Pr(A|B) = \Pr(A)$
2. $\Pr(\overline{A}B) = \Pr(\overline{A}) \cdot \Pr(B)$
3. $\Pr(A\overline{B}) = \Pr(A) \cdot \Pr(\overline{B})$ e $\Pr(\overline{AB}) = \Pr(\overline{A}) \cdot \Pr(\overline{B})$

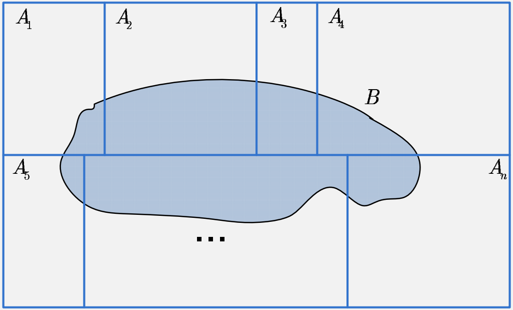
⚠ Se A e B são independentes, A e $\overline{B}$ são independentes e $\overline{A}$ e $\overline{B}$ também o são.

Probabilidade total

Teorema da probabilidade total

Sejam $A_1, A_2, A_3, \ldots, A_n$ eventos mutuamente exclusivos

- $\Pr(A_i) > 0, \quad i = 1, 2, \ldots, n$
- $B \subset \{A_1 + A_2 + \cdots + A_n\}$



Probabilidade total - cont.

Teorema da probabilidade total - cont.

- $\Pr(B) = \Pr(BA_1) + \Pr(BA_2) + \cdots + \Pr(BA_n)$ pois $B = \underbrace{BA_1 + BA_2 + \cdots + BA_n}_{\text{mutuamente exclusivos}}$
- $\Pr(B) = \Pr(B|A_1) \cdot \Pr(A_1) + \cdots + \Pr(B|A_n) \cdot \Pr(A_n)$

Probabilidade total

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B|A_i) \cdot \Pr(A_i) \qquad (15)$$

Regra de Bayes

Regra de Bayes

A regra de Bayes é o princípio inverso do conceito da probabilidade total. Em termos matemáticos tem-se

$$\boxed{\Pr(A_j|B) = \frac{\Pr(B|A_j) \cdot \Pr(A_j)}{\sum_{i=1}^{n} \Pr(B|A_i) \cdot \Pr(A_i)}} \tag{16}$$

Também chamada de probabilidade *a posteriori*

Eventos conjuntos

Dado $\mathcal{S}$ (espaço amostral), podemos atribuir diferentes atributos aos eventos pertencentes a diferentes classes de Borel

$$\mathcal{S} = \{x_1, x_2, \dots, x_n\}$$

$$\begin{cases} A_1, A_2, \cdots, A_n \in \mathcal{F}_1 \\ B_1, B_2, \cdots, B_n \in \mathcal{F}_2 \end{cases}$$

Exemplo:

$\mathcal{S} = \{$João, José, Maria $\}$

(idade,altura)

Rescrevendo:

$$\mathcal{S} = \{(10, 1.50), (30, 1.80), (32, 1.65)\}$$

Probabilidade marginal - eventos conjuntos

Probabilidade marginal - eventos conjuntos

$$A_1 + A_2 + \cdots + A_n = \mathcal{S}_1$$

$$B_1 + B_2 + \cdots + B_n = \mathcal{S}_2$$

$$\left.\begin{aligned} P(A_i) &= \sum_{j=1}^{n} P(A_i, B_j) \\ P(B_j) &= \sum_{i=1}^{n} P(A_i, B_j) \end{aligned}\right\} \sum_{i=1}^{n}\sum_{j=1}^{n} P(A_i, B_j) = 1 \tag{17}$$

Comportamento de um usuário de celular (dispositivo móvel)

É possível caracterizar (modelar) o comportamento de um usuário de telefone celular através da probabilidade condicional?

Definição

Por **comportamento** entende-se o perfil de utilização do telefone celular por um dado usuário ao longo de um período de tempo (e.g. 1 dia).

Referência:

J. Hollmén and V. Tresp (1999). "Call-based fraud detection in mobile communications networks using a hierarchical regime-switching model", In:*Advances in Neural Information Processing Systems*, M. Kearns and S. Solla and D. Cone (Eds.).

Aplicações - cont.

Caracterização do problema

Seja $X_n \in \{0,1\}$ o evento indicando se uma determinada chamada está em curso no instante n, para um dado usuário.

- $X_n = 0$ (não há chamada em curso no instante n)
- $X_n = 1$ (chamada em curso no instante n)

Interesse nas **transições** da condição de uma chamada em diferentes instantes de tempo. Neste modelo, 4 situações são possíveis:

① $X_{n-1} = 0 \to X_n = 1$ (Início de uma chamada)

② $X_{n-1} = 1 \to X_n = 0$ (Término de uma chamada)

③ $X_{n-1} = 1 \to X_n = 1$ (Chamada em andamento)

④ $X_{n-1} = 0 \to X_n = 0$ (Não-ocorrência de chamada)

Aplicações - cont.

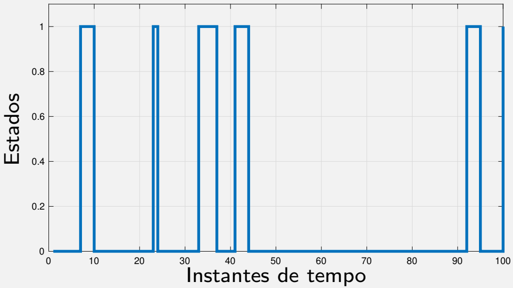
Caracterização do problema - cont.

Assim, deseja-se para um dado conjunto de medidas de X_n, $n = 1, \ldots, N$, determinar as seguintes probabilidades condicionais:

- $p_{01} = \Pr(X_n = 1 | X_{n-1} = 0)$ (Prob. iniciar chamada)
- $p_{10} = \Pr(X_n = 0 | X_{n-1} = 1)$ (Prob. terminar chamada)
- $p_{11} = \Pr(X_n = 1 | X_{n-1} = 1)$ (Prob. continuar chamada)
- $p_{00} = \Pr(X_n = 0 | X_{n-1} = 0)$ (Prob. continuar sem chamada)

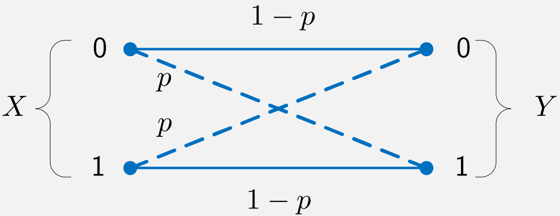
Caracterização do problema - cont.

Um usuário hipotético poderia ter o comportamento representado na forma de um sinal binário, conforme mostra figura abaixo.



Canal binário simétrico

Em sistemas de comunicação, um importante modelo de canal de comunicação conhecido como **canal binário simétrico**, ilustrado na figura abaixo.



Aplicações - cont.

Canal binário simétrico - cont.

- O canal binário simétrico só pode transmitir um dentre 2 símbolos (geralmente convencionados como 0 e 1).
- Um canal não-binário pode transmitir mais de 2 símbolos, possivelmente até infinitos símbolos.
- **A transmissão não é perfeita**, e ocasionalmente o receptor recebe o bit errado.
- A probabilidade de erro de um bit é p.
- Este canal é freqüentemente usado em telecomunicações por ser um dos modelos de **canal ruidoso** mais simples de analisar.

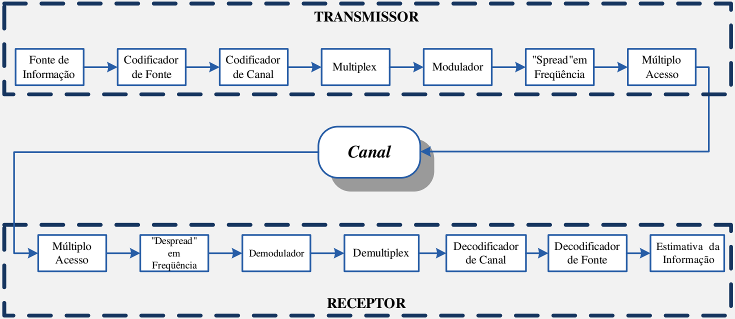
Aplicações - cont.

Canal binário simétrico - cont.

- Seja $X \in \{0, 1\}$ um evento probabilístico referente à transmissão de um bit por um canal ruidoso do tipo binário simétrico.
- Seja $Y \in \{0, 1\}$ o evento correspondente à recepção do bit transmitido.

Canal binário simétrico - cont.

Esquematicamente, o sistema completo seria representado como na figura abaixo



Canal binário simétrico - cont.

O canal binário simétrico é caracterizado por probabilidades condicionais:

- $\Pr(Y = 0|X = 0) = 1 - p$
- $\Pr(Y = 0|X = 1) = p$
- $\Pr(Y = 1|X = 0) = p$
- $\Pr(Y = 1|X = 1) = 1 - p$

em que $0 \leq p \leq 1/2$.

Aplicações - cont.

Canal binário simétrico - cont.

Uma questão de interesse para o caso de um canal binário simétrico (CBS) é verificar a probabilidade de erro do sistema.

Esta probabilidade de erro é determinada pela combinação das probabilidade de erro de cada um dos bits, o que é dado pela expressão abaixo:

$$\begin{aligned} P_e &= \Pr(Y = 1, X = 0) + \Pr(Y = 0, X = 1) \\ &= \Pr(Y = 1 | X = 0) \cdot \Pr(X = 0) + \\ &\qquad + \Pr(Y = 0 | X = 1) \cdot \Pr(X = 1) \end{aligned} \tag{18}$$

Canal binário simétrico - cont.

Assumindo as probabilidades de transmissão para os bits '0' e '1' como sendo $\mathrm{Pr}(X = 0) = \pi_0$ e $\mathrm{Pr}(X = 1) = \pi_1$, teremos a seguinte expressão para a probabilidade de erro:

$$P_e = p \cdot \pi_0 + p \cdot \pi_1 = p \cdot \underbrace{(\pi_0 + \pi_1)}_{=1} = p \qquad (19)$$

Canal binário simétrico - cont.

Uma outra questão importante em sistemas de comunicação é fazer uma *inferência* sobre o bit transmitido, a partir da observação do bit recebido.

Neste quesito, por exemplo, se for recebido um bit ‘1’, dadas as probabilidades condicionais das transições, podemos estar interessados (também como exemplo) na probabilidade de ter sido enviado um bit ‘0’, neste caso.

Uma forma de responder esta questão é através da **regra de Bayes**.

Canal binário simétrico - cont.

Neste caso mencionado teríamos a seguinte expressão:

$$\Pr(X = 0|Y = 1) = \frac{\Pr(Y = 1|X = 0) \cdot \Pr(X = 1)}{\Pr(Y = 1)} \tag{20}$$

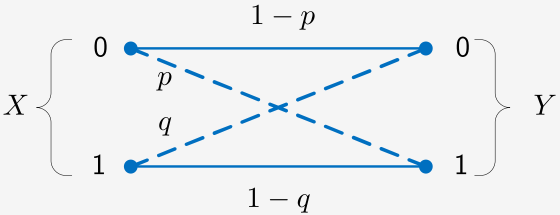
em que temos, pela regra da probabilidade total

$$\Pr(Y = 1) = \sum_{i=0}^{1} \Pr(Y = 1|X = i) \cdot \Pr(X = i) \tag{21}$$

Canal binário não-simétrico

E no caso do canal não ser simétrico e ter as probabilidades condicionais de erro para cada um dos bits diferentes?

O que muda?



Variáveis Aleatórias

A primeira pergunta importante: o que são variáveis?

Temos uma boa noção sobre isso no estudo de funções. Por exemplo, em

$$f(x) = x^2 + 2x - 1$$

temos que x é normalmente o que chamamos de *variável*.

? Outra questão importante

Mas o que significa uma variável **aleatória**?

Espaço amostral e resultados

Descrição

O espaço amostral $\mathcal{S}$ correspondente ao lançamento de uma moeda honesta três vezes seguidas é definido como

$$\mathcal{S} = \{CCC, CCK, CKC, CKK, KCC, KCK, KKC, KKK\}$$

em que $C =$ coroa e $K =$ cara.

- **Problema:** a representação dos resultados é *simbólica*;
- **Mais problema:** uma representação é dita simbólica quando **não** podemos utilizá-la para realizar operações matemáticas;
- ⚠ Por exemplo, a soma do resultado CCC com KKK não faz sentido algum.

Espaço amostral e resultados - cont.

Importante!

✎ Na prática, entretanto, muitos problemas de interesse para engenharia requerem que os resultados sejam representados em forma númerica.

✎ A representação de um resultado é dita **numérica** quando podemos utilizá-la para realizar operações matemáticas.

Pergunta

? É possível sair de uma representação simbólica de eventos probabilísticos para uma representação numérica mais adequada com a realidade da engenharia?

Variável aleatória

Descrição

Variável aleatória pode ser entendida como o resultado numérico da operação de um mecanismo não-determinístico ou execução de um experimento não-determinístico para gerar um valor aleatório. Ao contrário da prática comum com outras variáveis matemáticas, uma variável aleatória não pode ter um valor associado; uma variável aleatória não descreve o valor atual de uma realização de um evento particular, mas descreve o possível, ainda que indeterminado, resultado em termos de números reais.

Variável aleatória - cont.

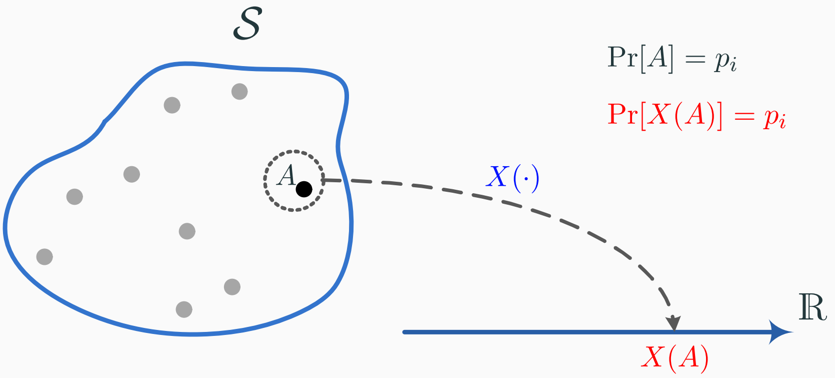


Figura 1: Definição de variável aleatória.

Variável aleatória - cont.

Exemplo

Lançando uma moeda:

$$\mathcal{S} = \{\text{cara}, \text{coroa}\}$$

$$X(\text{cara}) = 0 \;\Rightarrow\; \Pr(0) = \Pr(\text{cara})$$

$$X(\text{coroa}) = 1 \;\Rightarrow\; \Pr(1) = \Pr(\text{coroa})$$

Variável aleatória - cont.

Definição

Variável aleatória (v.a.) é qualquer função definida no espaço amostral $\mathcal{S}$ tal que:

$$\{X(\cdot) : \mathcal{S} \to \mathbb{R} \mid X(A) \in (-\infty, x], A \in \mathcal{S}\} \in \mathcal{F} \tag{22}$$

Deve-se notar que

1. Domínio da função X é $\mathcal{S}$.
2. Imagem da função X é a reta dos números reais.
3. Como toda função, a cada resultado $A \in \mathcal{S}$ corresponde exatamente um valor $x = X(A)$. Ou seja, $X(\cdot)$ é uma função sobrejetora.
4. Porém, valores diferentes de A podem levar ao mesmo x.

Variável aleatória - cont.

Exemplo 1: Lançamento de uma moeda

- Representação simbólica: $\mathcal{S} = \{\text{cara}, \text{coroa}\}$.
- Representação numérica: $X(\text{cara}) = 0$ e $X(\text{coroa}) = 1$. Logo, o conjunto-imagem de X é $R_X = \{x : x = 0, 1\}$.

Exemplo 2: Lançamento de uma moeda três vezes

☛ $\mathcal{S} = \{CCC, CCK, CKC, CKK, KCC, KCK, KKC, KKK\}$.

☛ Seja X o número de caras, então:

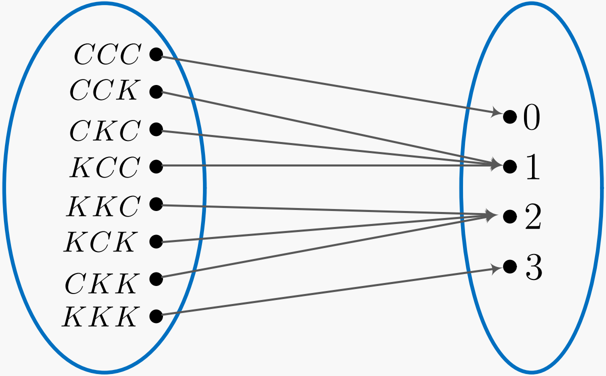
$$X(CCC) = 0, \quad X(CCK) = 1, \quad X(CKC) = 1$$

$$X(CKK) = 2, \quad X(KCC) = 1, \quad X(KCK) = 2$$

$$X(KKC) = 2, \quad X(KKK) = 3$$

Variável aleatória - cont.

Exemplo 2: Lançamento de uma moeda três vezes - cont.



Variável aleatória - cont.

Atenção!

Quando o resultado em $\mathcal{S}$ já é a característica numérica desejada, então

$$x = X(A) = A \quad \text{(função identidade)} \tag{23}$$

Exemplo 3: Tempo até primeira falha de um equipamento

➙ $\mathcal{S} = \{t : t \in \mathbb{R}, t \geq 0\}$.

➙ Se X é o tempo até a (primeira) falha do equipamento, então

$$x = X(t) = t \tag{24}$$

Observações importantes

- Pela definição dada, uma VA não é uma *variável* de fato, mas sim uma *função*.
- Na prática, consideramos como VA qualquer grandeza cujo valor está sujeito a variações (flutuações) aleatórias.
- Essas flutuações aleatórias recebem o nome genérico de **ruído aleatório**.
- Por exemplo, o valor teórico (nominal) da tensão em uma tomada da sua casa é 220 Volts.
- Contudo, se formos medir com o instrumento adequado, pode ser que não observemos exatamente este valor. **Por quê?**

Observações importantes - cont.

- Por causa das flutuações aleatórias, que podem ter uma série de causas, tais como
  1. Carga elétrica pendurada na rede de energia.
  2. Aferição inadequada do equipamento de medição.
  3. Localização de sua casa ao longo da linha de transmissão.
  4. Falha na rede de distribuição.
  5. Erro do operador, dentre outras causas.
- O melhor que podemos fazer é tentar diminuir a influência que os fenômenos físicos que causam as flutuações exercem sobre o processo de medição da variável.

Observações importantes - cont.

É útil entender a definição de VA através das seguintes associações:

- A variável a ser medida é representada por X. Esta variável é aleatória devido a flutuações que distorcem o seu valor.
- As flutuações, na verdade, são resultantes de um acontecimento aleatório que ocorrem em $\mathcal{S}$, cujo resultado não sabemos de antemão.
- Podemos afirmar que $\mathcal{S}$ é o espaço do eventos **não-observáveis**.
- Um evento qualquer em $\mathcal{S}$ é simbolizado por A. E o efeito da ocorrência de A no **mundo observável** é representado pelo valor medido $X(A)$, que é o valor mostrado pelo equipamento.

Variável aleatória - cont.

Eventos definidos em termos de variáveis aleatórias

- Se X é uma VA e x é um número real fixo, podemos definir o evento $(X = x)$ como

$$(X = x) = \{A : X(A) \;=\; x\}. \tag{25}$$

- De modo semelhante, para números fixos x, x_1 e x_2, tal que $x_1 < x_2$, podemos definir os seguintes eventos:

$$\begin{aligned}
(X \leq x) &= \{A : X(A) \leq x\} && (26)\\
(X > x) &= \{A : X(A) > x\} && (27)\\
(x_1 < X \leq x_2) &= \{A : x_1 < X(A) \leq x_2\} && (28)
\end{aligned}$$

Probabilidades em termos de variáveis aleatórias

Os eventos anteriores têm probabilidades que são denotadas da seguinte forma:

$$\Pr(X = x) = \Pr\{A : X(A) = x\} \quad (29)$$

$$\Pr(X \leq x) = \Pr\{A : X(A) \leq x\} \quad (30)$$

$$\Pr(X > x) = \Pr\{A : X(A) > x\} \quad (31)$$

$$\Pr(x_1 < X \leq x_2) = \Pr\{A : x_1 < X(A) \leq x_2\} \quad (32)$$

Variável aleatória - cont.

Variáveis aleatórias contínuas e discretas

- Uma VA é dita **discreta** se ela puder assumir apenas valores (níveis) específicos.
- Uma forma de identificar uma variável aleatória discreta é através do conjunto universo de X. Se este conjunto for *enumerável*[1], finito ou infinito, então X é uma VA discreta.
- Uma VA é dita **contínua** se ela puder assumir *qualquer* valor dentro de seu conjunto universo.
- Uma forma de identificar uma variável aleatória discreta é através do conjunto universo de X. Se este conjunto for *não-enumerável*, então X é uma VA contínua.

[1] Um conjunto é dito enumerável se os seus elementos puderem ser colocados em correspondência um-para-um com os números naturais.

Exemplo resolvido

- Seja X a soma dos números das faces resultantes do lançamento de dois dados honestos.
- Se A_i é a face resultante de um dos dados e B_j é a face resultante para o outro, então $X = A_i + B_j$.
- Assim, as probabilidades para os possíveis valores de X são:

$$\begin{aligned}
\Pr(X=2) &= \Pr\{(1,1)\} = 1/36 \\
\Pr(X=3) &= \Pr\{(1,2),(2,1)\} = 2/36 \\
\Pr(X=4) &= \Pr\{(1,3),(3,1),(2,2)\} = 3/36 \\
\Pr(X=5) &= \Pr\{(1,4),(2,3),(3,2),(4,1)\} = 4/36 \\
\Pr(X=6) &= \Pr\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = 5/36 \\
\Pr(X=7) &= \Pr\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = 6/36 \\
\Pr(X=8) &= \Pr\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = 5/36 \\
\Pr(X=9) &= \Pr\{(3,6),(4,5),(5,4),(6,3)\} = 4/36 \\
\Pr(X=10) &= \Pr\{(4,6),(5,5),(6,4)\} = 3/36 \\
\Pr(X=11) &= \Pr\{(5,6),(6,5)\} = 2/36 \\
\Pr(X=12) &= \Pr\{(6,6)\} = 1/36
\end{aligned}$$

Variável aleatória - cont.

O que podemos dizer sobre uma variável aleatória?

1. Só porque uma VA tem um comportamento **pontualmente** imprevisível, isto não quer dizer que seu comportamento global não obedeça certos **padrões numéricos**.
2. Estes padrões refletem como o valor de uma certa grandeza física se distribui na reta dos números reais.
3. Há diferentes padrões de distribuição de valores de uma VA.
4. Estes padrões são formalizados matematicamente na forma de leis probabilísticas, que quantificam as probabilidades de ocorrência de valores de uma certa VA:
   - ➠ função de distribuição cumulativa de probabilidade (cdf);
   - ➠ função de densidade de probabilidade (pdf).

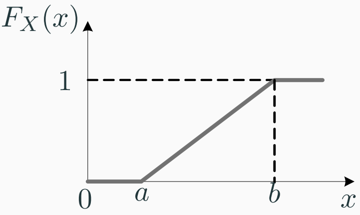
Função que mede a probabilidade **acumulada** até certo valor.

Definição da cdf

$$F_X(x) \triangleq \Pr\{X \leq x\} \tag{33}$$

Exemplo: *Distribuição uniforme*

$$F_X(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{x-a}{b-a}, & \text{se } a < x \leq b \\ 1, & \text{se } x > b \end{cases}$$



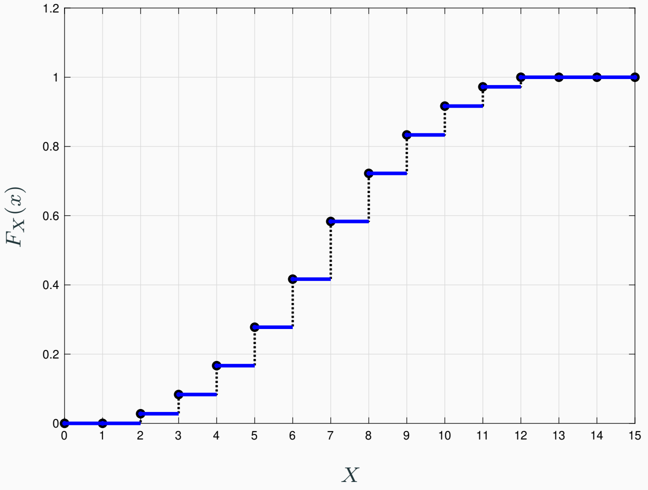
Função distribuição cumulativa de probabilidade - cont.

Propriedades da cdf

1. $F_x(x) \geq 0, \quad \forall x$
2. $F_X(-\infty) = \Pr(X \leq -\infty) = \lim_{x \to -\infty} \Pr(X \leq x) = 0$
3. $F_X(+\infty) = \Pr(X \leq +\infty) = \lim_{x \to +\infty} \Pr(X \leq x) = 1$
4. $\Pr\{x_1 \leq X \leq x_2\} = F_X(x_2) - F_X(x_1)$
5. $\Pr(X > x_1) = 1 - F_X(x_1)$
6. Se $x_1 \leq x_2$, então $F_X(x_1) \leq F_X(x_2)$, ou seja, a cdf é monotônica não-decrescente
7. A FDA é contínua à direita. Isto é, para todo x e todo $\delta < 0$, temos $\lim_{\delta \to 0} [F_X(x + \delta) - F_X(x)] = 0$.

Função distribuição cumulativa de probabilidade - cont.

Exemplo: Soma das faces no lançamento de dois dados

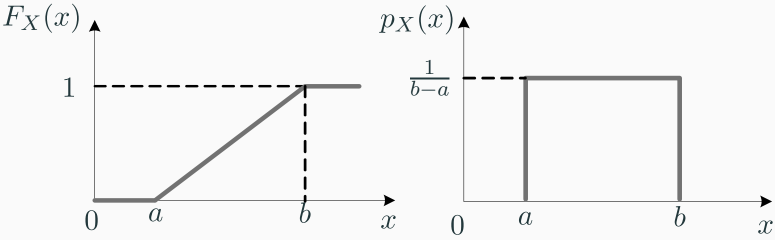


Função que mede a **distribuição** (variação) da probabilidade no espaço de ocorrência da v.a.

Definição da pdf

$$p_X(x) \triangleq \frac{d}{dx} F_X(x) \tag{34}$$

Exemplo: *Distribuição uniforme*



Função densidade de probabilidade - cont.

Propriedades da pdf

1. $F_X(x) = \int\limits_{-\infty}^{x} p_X(\xi)\, d\xi$

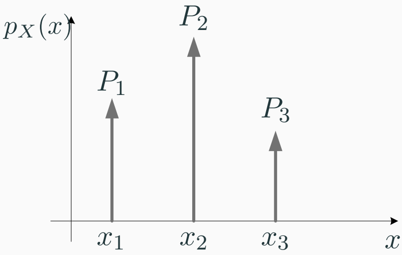
2. $F_X(x)$ é monotônico não-decrescente $\Rightarrow p_X(x) \geq 0$

3. $\Pr\{X > x\} = 1 - F_X(x) = 1 - \Pr\{X \leq x\}$

4. $\Pr\{x_1 < X \leq x_2\} = F_X(x_2) - F_X(x_1) = \int\limits_{x_1}^{x_2} p_X(\xi)\, d\xi$

Dificuldade

Neste caso, as variáveis admitem valores somente em determinados instantes de tempo. Mas o que ocorre com as probabilidades?



Função $\delta(\cdot)$ de Dirac (impulso)

$$\delta(t)\begin{cases}\int_{-\infty}^{\infty} f(t)\delta(t-t_0)\,dt = f(t_0)\\ \int_{-\infty}^{\infty} \delta(t)\,dt = 1\end{cases}$$

$\delta(t) = \frac{d}{dt}u(t)$, em que $u(t)$é a função degrau unitário

Logo, teremos

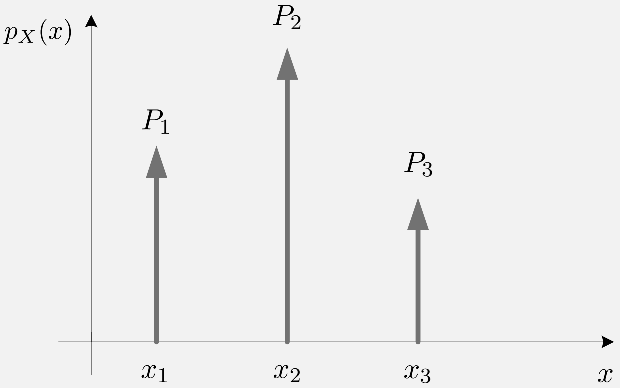
$$p_X(x) = \sum_{i=1}^{N} \Pr\{X = x_i\} \cdot \delta(x - x_i)$$

$$F_X(x) = \int_{-\infty}^{x} p_X(\xi)\, d\xi = \sum_{i=1}^{N} \Pr\{X = x_i\} \cdot \int_{-\infty}^{x} \delta(\xi - x_i)\, d\xi$$

Mas sabe-se que

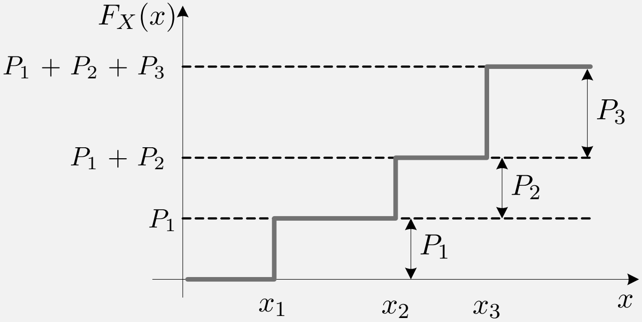
$$\int_{-\infty}^{x} \delta(\xi - x_i)\, d\xi = \begin{cases} 0, & \text{se } x < x_i \\ 1, & \text{se } x \geq x_i \end{cases}$$

pdf



Variáveis aleatórias discretas - cont.

cdf



Função de densidade de probabilidade gaussiana

Definição

Seja X uma v.a., X é dito ter distribuição de probabilidade gaussiana, ou *normal*, se sua densidade de probabilidade pode ser escrita da seguinte forma

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{35}$$

Notação usual:

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Parâmetros $\begin{cases} \mu \rightarrow & \text{média} \\ \sigma^2 \rightarrow & \text{variância} \end{cases}$

pdf gaussiana - cont.

Normalização

$$Z \sim \mathcal{N}(0,1)$$

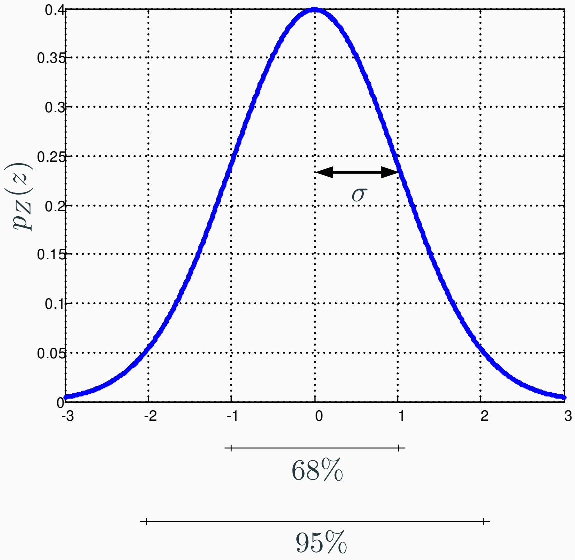
$$p_Z(z) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{z^2}{2}\right) \quad (36)$$

Função erro

cdf da distribuição gaussiana normalizada

$$\begin{aligned} \text{erf}(x) &= F_Z(x) = \Pr\{Z \le x\} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\right) \, dz \end{aligned} \quad (37)$$

pdf gaussiana - cont.

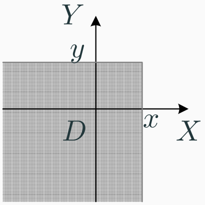


Função distribuição de probabilidade conjunta

$$F_{X,Y}(x,y) = \Pr \underbrace{\{X \leq x, Y \leq y\}}_{\text{intersecção}} \tag{38}$$

$$\underbrace{\{X \leq x, \quad Y \leq y\}}_{\text{evento}} = \{\ w \in \mathcal{S} | [X(w), Y(w)] \in D\ \}$$

em que $D = \{\ (X,Y) | X \in (-\infty, x], Y \in (-\infty, y]\ \}$



$$\boxed{p_{X,Y}(x,y) = \frac{\partial^2}{\partial x\ \partial y} F_{X,Y}(x,y)} \tag{39}$$

Variáveis aleatórias bidimensionais - cont.

Propriedades distribuições conjuntas

❶ $F_{X,Y}(-\infty, y) = 0$

❷ $F_{X,Y}(x, -\infty) = 0$

❸ $F_{X,Y}(\infty, \infty) = 1$

❹ $$\left.\begin{aligned} F_{X,Y}(x, \infty) &= F_X(x) \\ F_{X,Y}(\infty, y) &= F_Y(y) \end{aligned}\right\}\text{distribuições marginais}$$

❺ $$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{X,Y}(\alpha, \beta)\ d\alpha\ d\beta$$

❻ $$\begin{aligned} p_X(x) &= \int_{-\infty}^{\infty} p_{X,Y}(x, y)\ dy \\ p_Y(y) &= \int_{-\infty}^{\infty} p_{X,Y}(x, y)\ dx \end{aligned}$$

Variáveis aleatórias independentes

Definição

Sejam X e Y v.a.'s. Elas são independentes se:

$$\boxed{F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)} \tag{40}$$

$$\begin{aligned} F_{X,Y}(x,y) &= \Pr\{\ X \leq x, Y \leq y\ \} \\ &= \Pr\{X \leq x\} \cdot \Pr\{Y \leq y\} \end{aligned}$$

Variáveis aleatórias independentes - cont.

Logo, para as densidades conjuntas temos:

$$\boxed{p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)} \tag{41}$$

pois

$$\begin{aligned}
p_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \, \partial y} F_{X,Y}(x,y) \\
&= \frac{\partial^2}{\partial x \, \partial y} \left(F_X(x) \cdot F_Y(y) \right) = \frac{\partial}{\partial x} F_X \cdot \frac{\partial}{\partial y} F_Y \\
&= p_X(x) \cdot \; p_Y(y)
\end{aligned}$$

Representação de variáveis aleatórias

Uma vez que, como já discutido, as variáveis aleatórias não podem ter um valor associado por se tratarem de representações de experimentos não-determinísticos, faz-se necessário métodos de representação das v.a.

Algumas destas representações:

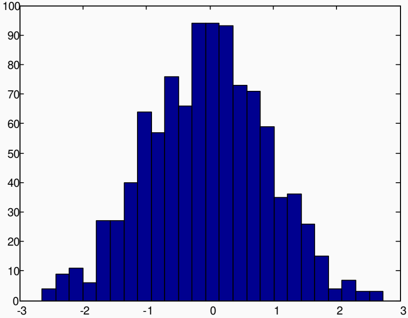
① Função densidade de probabilidade (pdf)

② Função de distribuição cumulativa (cdf)

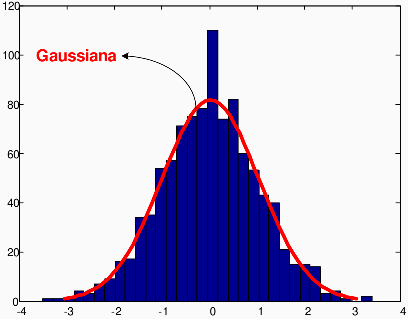
③ **Histograma**

Histograma

É uma representação da *distribuição* das variáveis no espaço amostral em relação ao número de ocorrências.



O histograma serve então para comparar com determinadas distribuições conhecidas e verificar quão próxima (ou quão diferente) é a distribuição dos dados com uma distribuição qualquer.



Representação de variáveis aleatórias - cont.

Construindo o histograma...

Passo 1 - Calcular, a partir do conjunto de dados, o *range* (domínio) da variável $[\texttt{menor valor}, \texttt{maior valor}]$;

Passo 2 - Definir em quantos M intervalos (*bins*) será observada a variável;

Passo 3 - Para cada intervalo Δx_i, $i = 1, ..., M$, determinar:

$$C(\Delta x_i) = \text{No. de observações de } X \in \Delta x_i$$

Passo 4 - Desenhar o gráfico de barras $C(\Delta x_i) \times \Delta x_i$.

Medidas estatísticas

Momentos

São *estatísticas* de uma variável aleatória capazes de representar seu comportamento probabilístico.

Momentos são as mensurações para quantificar o comportamento de uma variável aleatória.

Os infinitos momentos estatísticos definem a função de densidade de probabilidade.

Média

Também chamada de esperança matemática, valor esperado, momento de 1^a ordem, é definido como

$$\boxed{\mu = \mathbb{E}\{X\} \triangleq \int_{-\infty}^{\infty} x \cdot p_X(x)\ dx} \tag{42}$$

para $X \sim$ v.a. contínua

Medidas estatísticas - cont.

Média

Para X discreta temos

$$\boxed{\mu = \mathbb{E}\{X\} \triangleq \sum_{i=-\infty}^{\infty} x_i \cdot \Pr\{X = x_i\}} \tag{43}$$

? Pergunta: Num jogo de moeda, qual a média? E no caso de uma distribuição uniforme entre $[0, 1]$?

Medidas estatísticas - cont.

Interpretações da média

- A média é o ponto de equilíbrio da curva da densidade de probabilidade: 50% dos eventos possíveis é menor que a média e os 50% restantes são maiores que a média.
- É uma aproximação (grosseira) do valor a ser obtido quando não se conhece nada da distribuição.
- Se $p_X(x)$ for simétrico em relação a um valor $x = a \Rightarrow \mathbb{E}\{X\} = a$.

Medidas estatísticas - cont.

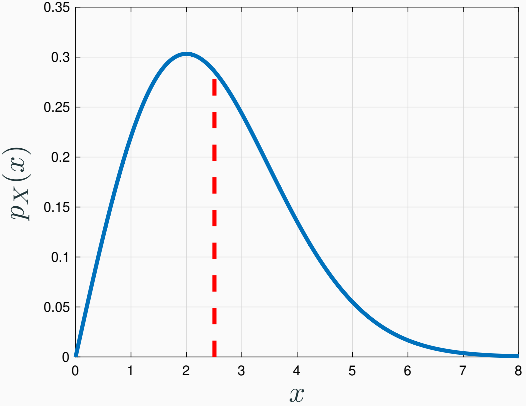


Figura 2: Distribuição de Rayleigh e sua média.

Medidas estatísticas - cont.

Propriedades da média

1. $\mathbb{E}\{X+Y\} = \mathbb{E}\{X\} + \mathbb{E}\{Y\}$
2. $\mathbb{E}\{X \cdot Y\} = \mathbb{E}\{X\} \cdot \mathbb{E}\{Y\}$ se X e Y são v.a.'s independentes
3. $\mathbb{E}\{f(X,Y)\} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x,y) p_{X,Y}(x,y)\ dxdy$
4. Se X e Y são independentes
   $\mathbb{E}\{f(X) \cdot g(Y)\} = \mathbb{E}\{f(X)\} \cdot \mathbb{E}\{g(Y)\}$

Medidas estatísticas - cont.

Momentos de ordem k

$$\mu_k = \mathbb{E}\left\{X^k\right\} = \int_{-\infty}^{\infty} x^k \cdot p_X(x)\, dx \qquad (44)$$

- Os momentos de uma variável aleatória são uma representação da pdf da variável
- A coletânea dos *infinitos* momentos da v.a. definem sua pdf
- Algumas distribuições possuem alguns momentos nulos
- A estimativa de momentos cresce em complexidade e decresce em precisão com o aumento direto de k

Momentos centrados

Uma importante medida estatística é avaliar o comportamento da v.a. *em torno* da média. Assim, define-se o *momento centrado de ordem k* como sendo

$$c_k = \mathbb{E}\left\{(X-\mu)^k\right\} = \int_{-\infty}^{\infty} (x-\mu)^k \cdot p_X(x)\; dx \tag{45}$$

De particular interesse: variância ($\sigma^2 = \mathbb{E}\left\{(X-\mu)^2\right\} \geq 0$)

Se $p_X(x)$ é simétrica em relação a média, temos que $c_k = 0$ para $\forall k$ **ímpar**!

Medidas estatísticas - cont.

✎ É possível escrever os momentos centrados em função dos momentos!

Exemplo

$$\begin{aligned}
\sigma^2 &= \mathbb{E}\left\{(X-\mu)^2\right\} \\
&= \mathbb{E}\left\{X^2 - 2X\mu + \mu^2\right\} \\
&= \mathbb{E}\left\{X^2\right\} - 2\cdot\mathbb{E}\left\{X\right\}\cdot\mu + \mu^2 \\
&= \mathbb{E}\left\{X^2\right\} - \left(\mathbb{E}\left\{X\right\}\right)^2 \\
&= \mu_2 - \mu^2
\end{aligned} \tag{46}$$

Fórmula para achar outros momentos?

Exemplo

Conhecendo-se a média e a variância de uma v.a. X $(\mu = \mathbb{E}\{X\}, \sigma^2 = \mathbb{E}\left\{(X-\mu)^2\right\})$ deseja-se achar $\mathbb{E}\left\{X^3\right\}$ em função de μ e σ^2.

Aproximando X^3 em série de Taylor

$$\begin{aligned} X^3 &= \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \left.\frac{d^k X^3}{dX^k}\right|_{X=\mu} (X-\mu)^3 \\ &= \mu^3 + 3\mu^2 \cdot (X-\mu) + 3\mu \cdot (X-\mu)^2 + (X-\mu)^3 \end{aligned}$$

Aplicando-se o operador esperança

$$\begin{aligned} \mathbb{E}\{X^3\} &= \mu^3 + 3\mu \cdot \cancelto{0}{\mathbb{E}\{(X-\mu)\}} + 3\mu \cdot \mathbb{E}\{(X-\mu)^2\} \\ &= \mu^3 + 3\mu\sigma^2 \end{aligned}$$

Medidas estatísticas - cont.

Resultado importante

$$\boxed{\mu_{2k} \geq \mu_k^2, \quad k = 1, 2, 3, \ldots} \tag{47}$$

Prova: Como $\mathbb{E}\left\{\left(X^k - \alpha\right)^2\right\} \geq 0, \quad \forall\alpha$, temos

$$\mathbb{E}\left\{X^{2k}\right\} - 2 \cdot \alpha\mathbb{E}\left\{X^k\right\} + \mathbb{E}\left\{\alpha^2\right\} \geq 0$$

$$\mu_{2k} - 2 \cdot \alpha\mu_k + \alpha^2 \geq 0 \quad \forall\alpha$$

$$\alpha^2 - 2 \cdot \alpha\mu_k + \mu_{2k} \geq 0 \quad \forall\alpha \quad \text{(equação de 2a ordem)}$$

$$\Rightarrow \text{discriminante } \leq 0$$

$$4\mu_k^2 - 4\mu_{2k} \leq 0$$

$$\mu_k^2 \leq \mu_{2k}$$

Desigualdade de Markov

Seja X uma v.a. tal que $p_X(x) = 0$ se $x < 0$

$$\boxed{\Pr\{X \geq \alpha\mu\} \leq \frac{1}{\alpha}, \quad \alpha > 0} \tag{48}$$

Mensuração da probabilidade da v.a. assumir valores que são relacionados à média daquela variável, que só possui valores positivos.

Desigualdades - cont.

Prova - Desigualdade de Markov

$$\Pr\{X \geq \alpha\mu\} = \int_{\alpha\mu}^{\infty} p_X(x)\ dx$$

$$\mu = \mathbb{E}\{X\} = \int_{-\infty}^{\infty} x \cdot p_X(x)\ dx = \int_{0}^{\infty} x \cdot p_X(x)\ dx$$

$$\mu \geq \int_{\alpha\mu}^{\infty} x \cdot p_X(x)\ dx \geq \int_{\alpha\mu}^{\infty} \alpha\mu \cdot p_X(x)\ dx$$

$$\mu \geq \alpha\mu \underbrace{\int_{\alpha\mu}^{\infty} p_X(x)\ dx}_{\Pr\{X \geq \alpha\mu\}} \Rightarrow \Pr\{X \geq \alpha\mu\} \leq \frac{1}{\alpha}$$

Desigualdade de Chebyshev

Seja X uma v.a. com *distribuição qualquer*, então

$$\boxed{\Pr\{|X-\mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}, \quad \epsilon > 0} \tag{49}$$

Mensuração da probabilidade de uma v.a. em torno de sua média, como função de sua variância.

Prova - Desigualdade de Chebyshev

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot p_X(x)\ dx$$

$$= \int_{-\infty}^{\mu-\epsilon} (x-\mu)^2 \cdot p_X(x)\ dx + \int_{\mu-\epsilon}^{\mu+\epsilon} (x-\mu)^2 \cdot p_X(x)\ dx + \int_{\mu+\epsilon}^{\infty} (x-\mu)^2 \cdot p_X(x)\ dx$$

$$\sigma^2 \geq \int_{-\infty}^{\mu-\epsilon} (x-\mu)^2 \cdot p_X(x)\ dx + \int_{\mu+\epsilon}^{\infty} (x-\mu)^2 \cdot p_X(x)\ dx$$

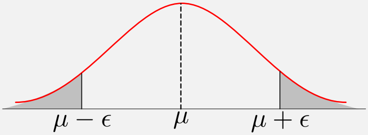
$$\sigma^2 \geq \int_{-\infty}^{\mu-\epsilon} \epsilon^2 \cdot p_X(x)\ dx + \int_{\mu+\epsilon}^{\infty} \epsilon^2 \cdot p_X(x)\ dx$$

Prova - Desigualdade de Chebyshev - cont.

$$\frac{\sigma^2}{\epsilon^2} \geq \underbrace{\int_{-\infty}^{\mu-\epsilon} p_X(x)\ dx}_{\Pr\{X\leq\mu-\epsilon\}} + \underbrace{\int_{\mu+\epsilon}^{\infty} p_X(x)\ dx}_{\Pr\{X\geq\mu+\epsilon\}}$$

$$\frac{\sigma^2}{\epsilon^2} \geq \Pr\{X-\mu \leq -\epsilon\} + \Pr\{X-\mu \geq \epsilon\}$$

$$\frac{\sigma^2}{\epsilon^2} \geq \Pr\{|X-\mu| \geq \epsilon\}$$



Primeira função característica

$$C(\omega) \triangleq \int_{-\infty}^{\infty} p_X(x) \cdot \exp(j\omega x)\ dx \tag{50}$$

Representação da v.a. no domínio da frequência.

Funções característica - cont.

Importante

$$C(-\omega) \triangleq \int_{-\infty}^{\infty} p_X(x) \cdot \exp(-j\omega x)\, dx$$

$C(-\omega) = \mathfrak{F}\{p_X(x)\}$ (transformada de Fourier de $p_X(x)$)

Funções característica - cont.

Importante - cont.

Notação

$$P_X(\omega) = \mathfrak{F}\{p_X(x)\} = \int_{-\infty}^{\infty} p_X(x) \cdot \exp(-j\omega x)\ dx \tag{51}$$
$$= \mathbb{E}\{\exp(-j\omega X)\}$$

Assim

$$\boxed{p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_X(\omega) \cdot \exp(j\omega x)\ d\omega} \tag{52}$$

Funções característica - cont.

No caso de termos variáveis aleatórias discretas, a definição é essecialmente a mesma, levando em conta apenas a existência de probabilidade para certos pontos no espaço.

Primeira função característica - Variáveis discretas

$$\begin{aligned} P_X(w) &= \sum_{i=-\infty}^{\infty} \Pr\{X = x_i\} \cdot \exp\left(-j\omega x_i\right) \\ &= \mathbb{E}\{\exp(-j\omega X)\} \end{aligned} \tag{53}$$

Propriedades $P_X(\omega)$

1. $|P_X(\omega)| \leq 1$
   Prova:
$$|P_X(\omega)| = \left| \int_{-\infty}^{\infty} p_X(x) \cdot \exp(-j\omega x)\ dx \right|$$
$$\leq \int_{-\infty}^{\infty} |p_X(x)| \cdot |\exp(-j\omega x)|\ dx = \int_{-\infty}^{\infty} p_X(x)\ dx$$
   (desigualdade de Schwartz)

2. $P_X(0) = 1$

Função geradora de momentos

$$\exp(-j\omega x) = \sum_{k=-\infty}^{\infty} \frac{(-j\omega X)^k}{k!} \quad \text{(Série de Taylor em torno de } X = 0\text{)}$$

$$P_X(\omega) = \mathbb{E}\{\exp(-j\omega x)\} = \mathbb{E}\left\{\sum_{k=-\infty}^{\infty} \frac{(-j\omega)^k \cdot X^k}{k!}\right\}$$

$$\mathbb{E}\{\exp(-j\omega x)\} = \sum_{k=-\infty}^{\infty} \frac{(-j\omega)^k}{k!} \cdot \underbrace{\mathbb{E}\left\{X^k\right\}}_{\mu_k}$$

$$\boxed{\Rightarrow P_X(\omega) = \sum_{k=-\infty}^{\infty} \frac{(-j\omega)^k}{k!} \cdot \mu_k} \qquad (54)$$

Função geradora de momentos - cont.

Ainda

$$P_X(\omega) = \sum_{k=0}^{\infty} \left.\frac{d^k P_X(\omega)}{d\omega^k}\right|_{\omega=0} \cdot \frac{1}{k!} \cdot \omega^k$$

Logo

$$\sum_{k=0}^{\infty} \left.\frac{d^k P_X(\omega)}{dw^k}\right|_{\omega=0} \cdot \frac{\omega^k}{k!} = \sum_{k=0}^{\infty} \mu_k \cdot (-j)^k \cdot \frac{\omega^k}{k!}$$

e assim

$$\boxed{\mu_k \cdot (-j)^k = \left.\frac{d^k P_X(\omega)}{dw^k}\right|_{\omega=0}} \tag{55}$$

Funções característica - cont.

Exemplo

Sejam X e Y v.a.s independentes com $p_X(x)$ e $p_Y(y)$ conhecidas. Se $Z = X + Y$, $p_Z(z) =?$

Solução

$$P_Z(\omega) = \mathbb{E}\{\exp(-j\omega Z)\} = \mathbb{E}\{\exp[-j\omega(X+Y)]\}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-j\omega X)\cdot\exp(-j\omega Y)\cdot p_{X,Y}(x,y)\ dx\ dy$$

$$\underbrace{\int_{-\infty}^{\infty} \exp(-j\omega X)\cdot p_X(x)\ dx}_{\mathbb{E}\{\exp(-j\omega X)\}}\cdot\underbrace{\int_{-\infty}^{\infty} \exp(-j\omega Y)\cdot p_Y(y)\ dy}_{\mathbb{E}\{\exp(-j\omega Y)\}}$$

$$P_Z(\omega) = P_X(\omega)\cdot P_Y(\omega)$$

$$p_Z(z) = p_X(x)\star p_Y(y)$$

Funções característica - cont.

Segunda função característica

$$\Psi(\omega) = \ln[P_X(\omega)] \tag{56}$$

Importante

- A segunda função característica é também chamada de **função geradora de cumulantes**
- Os cumulantes são de extrema importância na caracterização estatística de uma v.a.

História

Os cumulantes foram inicialmente introduzidos pelo astrônomo, contador, matemático e estaticista dinamarquês Thorvald N. Thiele (1838-1910) que os denominou *semi-invariantes*.

O termo *cumulante* surgiu pela primeira vez em 1931 no artigo "The Derivation of the Pattern Formulæ of Two-Way Partitions from Those of Simpler Patterns", Proceedings of the London Mathematical Society, Series 2, vol. 33, pp. 195-208, publicado pelo geneticista e estaticista Sir Ronald Fisher e o estaticista John Wishart, epônimo da distribuição de Wishart.

O historiador Stephen Stigler comenta que o termo cumulante foi sugerido a Fisher numa carta de Harold Hotelling. Em um outro artigo publicado em 1929, Fisher chamou-os de funções de momentos cumulativos.

Cumulantes - cont.

Definição

O cumulante de ordem k é definido como

$$\boxed{\kappa_k = \frac{\partial^k \Psi(\omega)}{\partial \omega^k}} \quad (57)$$

Propriedades dos cumulantes

1. **Invariância e equivariância**

$$\kappa_1(Y+\alpha) = \kappa_1(Y) + \alpha$$
$$\kappa_k(Y+\alpha) = \kappa_k(Y)$$

sendo α uma constante qualquer.

2. **Homogeneidade** (ou multilinearidade)

$$\kappa_k(\alpha Y) = \alpha^k \cdot \kappa_k(Y)$$

Cumulantes - cont.

Propriedades dos cumulantes - cont.

3. **Aditividade**

$$\kappa_k(X+Y) = \kappa_k(X) + \kappa_k(Y)$$

se X e Y são v.a.s independentes

Cumulantes e momentos

Os cumulantes são relacionados com os momentos através da seguinte recursão:

$$\boxed{\kappa_k = \mu_k - \sum_{i=1}^{k-1} \binom{k-1}{i-1} \kappa_i \cdot \mu_{k-i}} \tag{58}$$

Cumulantes e momentos - cont.

Desta forma, o $k-$ésimo momento é um polinômio de grau k dos k primeiros cumulantes, dados, para o caso em que $k = 1, \ldots, 6$, na seguinte forma:

$$\mu_1 = \kappa_1$$

$$\mu_2 = \kappa_2 + \kappa_1^2$$

$$\mu_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3$$

$$\mu_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4$$

$$\mu_5 = \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3$$

$$\mu_6 = \kappa_6 + 6\kappa_5\kappa_1 + 15\kappa_4\kappa_2 + 15\kappa_4\kappa_1^2 + 10\kappa_3^2 + 60\kappa_3\kappa_2\kappa_1 + 20\kappa_3\kappa_1^3 + 15\kappa_2^3 + 45\kappa_2^2\kappa_1^2 + 15\kappa_2\kappa_1^4 + \kappa_1^6.$$

Distribuição uniforme

$$\mathcal{S} = [a, b]$$

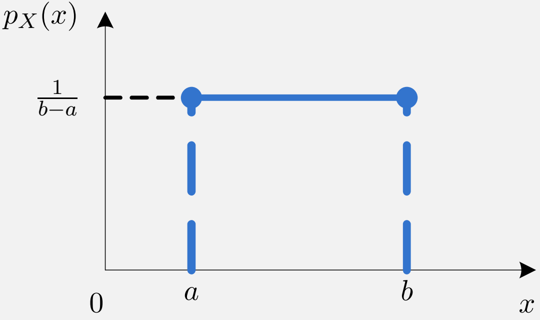
$$p_X(x) = \frac{1}{b-a},$$

$$a \leq x \leq b$$

$$\mathbb{E}\{X\} = \frac{a+b}{2} \qquad \sigma_X^2 = \frac{(b-a)^2}{12}$$

$$P_X(\omega) = \frac{\exp(-j\omega b) - \exp(-j\omega a)}{-j\omega(a-b)}$$

Distribuição uniforme - cont.



Distribuição exponencial

$$\mathcal{S} = [0, \infty)$$

$$p_X(x) = \lambda \exp(-\lambda x),$$

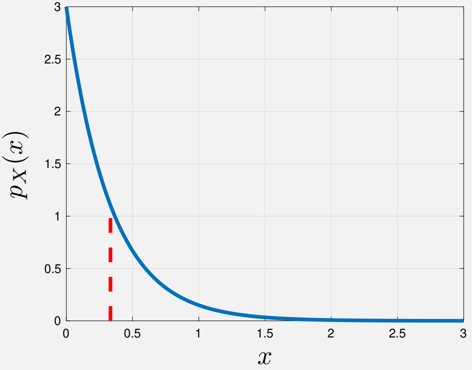
$$x \geq 0, \lambda \geq 0$$

$$\mathbb{E}\{X\} = \frac{1}{\lambda} \qquad \sigma_X^2 = \frac{1}{\lambda^2}$$

$$P_X(\omega) = \frac{\lambda}{\lambda + j\omega}$$

Distribuição exponencial - cont.

$\lambda = 3$



Distribuições importantes (contínuas) - cont.

Distribuição gaussiana

$$\mathcal{S} = (-\infty, \infty)$$

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

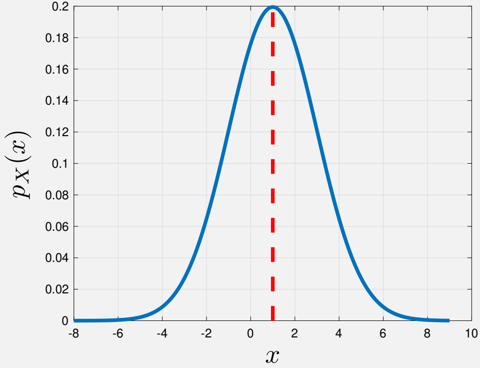
$$-\infty < x < \infty, \sigma > 0$$

$$\mathbb{E}\{X\} = \mu \qquad \sigma_X^2 = \sigma^2$$

$$P_X(\omega) = \exp\left(\frac{-j\mu\omega - \sigma^2\omega^2}{2}\right)$$

Distribuição gaussiana - cont.

$\mu = 1, \quad \sigma = 2$



Distribuição Gamma

$$\mathcal{S} = (0, \infty)$$

$$p_X(x) = \frac{1}{\lambda^a \Gamma(\alpha)} \cdot x^{\alpha-1} \exp\left(-\frac{x}{\lambda}\right),$$

$$x > 0, \quad \alpha > 0, \quad \lambda > 0$$

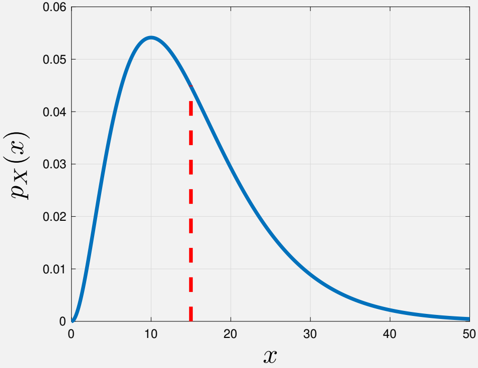
$$\mathbb{E}\{X\} = \alpha \cdot \lambda \qquad \sigma_X^2 = \alpha \cdot \lambda^2$$

$$P_X(\omega) = \exp\left(\frac{1}{(1 + j\omega\lambda)^\alpha}\right)$$

Distribuições importantes (contínuas) - cont.

Distribuição Gamma - cont.

$\alpha = 3, \quad \lambda = 5$



Distribuição de Rayleigh

$$\mathcal{S} = [0, \infty)$$

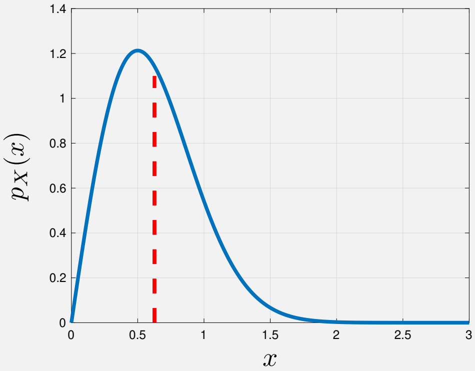
$$p_X(x) = \frac{x}{\sigma^2} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

$$x \geq 0, \quad \sigma > 0$$

$$\mathbb{E}\{X\} = \sigma\sqrt{\pi/2} \qquad \sigma_X^2 = \left(\frac{4-\pi}{2}\right)\sigma^2$$

Distribuição de Rayleigh - cont.

$\sigma = 0.5$



Distribuições importantes (contínuas) - cont.

Distribuição de Cauchy

$$\mathcal{S} = (-\infty, \infty)$$

$$p_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2},$$

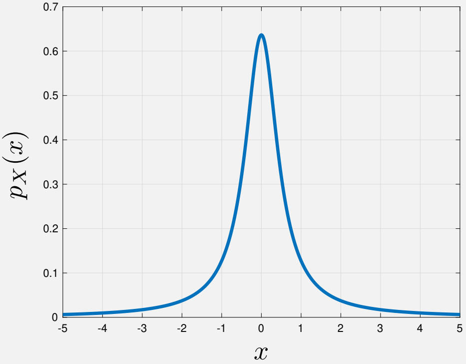
$$-\infty < x < \infty, \alpha > 0$$

Média e variância não são definidas

$$P_X(\omega) = \exp(-\alpha|\omega|)$$

Distribuição de Cauchy - cont.

$\alpha = 0.5$



Distribuição laplaciana

$$\mathcal{S} = (-\infty, \infty)$$

$$p_X(x) = \frac{\alpha \exp(-\alpha|x|)}{2},$$

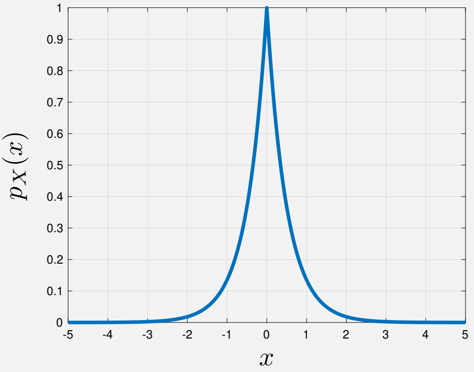
$$-\infty < x < \infty, \quad \alpha > 0$$

$$\mathbb{E}\{X\} = 0 \qquad \sigma_X^2 = \frac{2}{\alpha^2}$$

$$P_X(\omega) = \frac{\alpha^2}{\omega^2 + \alpha^2}$$

Distribuição laplaciana - cont.

$\alpha = 2$

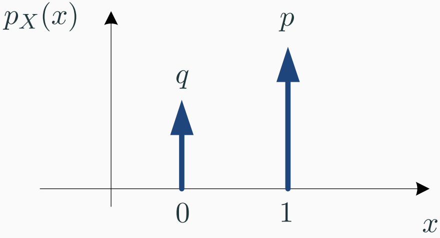


Distribuição de Bernoulli

$$\mathcal{S} = 0, 1$$

$$\Pr(0) = q = 1 - p, \quad \Pr(1) = p, \quad 0 \leq p \leq 1$$

$$\mathbb{E}\{X\} = p \qquad \sigma_X^2 = p(1 - p)$$



Distribuições importantes (discretas) - cont.

Distribuição binomial

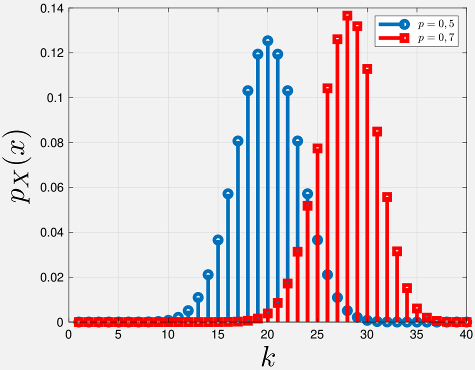
$$\mathcal{S} = 0, 1, \ldots, n$$

$$\Pr(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

$$\mathbb{E}\{X\} = np \qquad \sigma_X^2 = np(1-p)$$

Distribuição binomial - cont.

$n = 40$



Distribuições importantes (discretas) - cont.

Distribuição binomial negativa

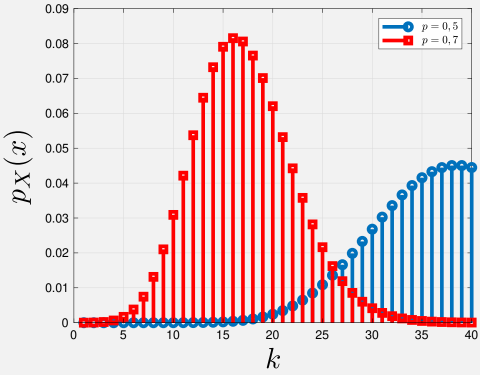
$$\mathcal{S} = r, r+1, \ldots, \quad \text{em que } r \text{ é um inteiro positivo}$$

$$\Pr(k) = \begin{pmatrix} k-1 \\ r-1 \end{pmatrix} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

$$\mathbb{E}\{X\} = \frac{r}{p} \qquad \sigma_X^2 = \frac{r(1-p)}{p^2}$$

Distribuição binomial negativa - cont.

$n = 40$



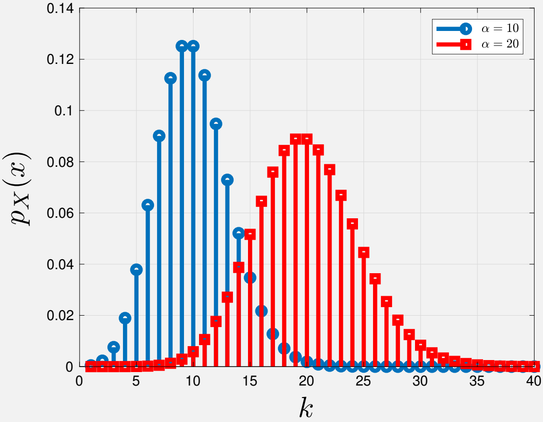
Distribuição de Poisson

$$\mathcal{S} = 0, 1, 2, \ldots$$

$$\Pr(k) = \frac{\alpha^k}{k!} \exp(-\alpha), \quad k = 0, 1, 2, \ldots, \quad \alpha > 0$$

$$\mathbb{E}\{X\} = \alpha \qquad \sigma_X^2 = \alpha$$

Distribuição de Poisson - cont.



Correlação

$$
\begin{aligned}
R(X, Y) &\triangleq \mathbb{E}\{X \cdot Y\} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot p_{X,Y}(x, y) \; dxdy
\end{aligned} \tag{59}
$$

ⓘ Correlação é uma medida de quanto de informação (relação) uma variável aleatória tem de outra. A correlação assume valores na faixa $(-\infty, \infty)$.

Covariância

$$
\text{cov}(X, Y) \triangleq \mathbb{E}\left\{[X - \mu(X)] \cdot [Y - \mu(Y)]\right\} \tag{60}
$$

Correlação e covariância - cont.

Propriedades

1. $\text{cov}(X, Y) = \mathbb{E}\{X \cdot Y\} - \mu(X)\mu(Y)$

$$\begin{aligned}\text{cov}(X, Y) &= \mathbb{E}\left\{[X - \mu(X)] \cdot [Y - \mu(Y)]\right\} \\ &= \mathbb{E}\left\{XY - X\mu(Y) - \mu(X)Y + \mu(X)\mu(Y)\right\} \\ &= \mathbb{E}\{XY\} - \mathbb{E}\{X\}\mu(Y) - \mathbb{E}\{Y\}\mu(X) + \mu(X)\mu(Y) \\ &= \mathbb{E}\{XY\} - \mu(X)\mu(Y) - \mu(Y)\mu(X) + \mu(X)\mu(Y) \\ &= R(X, Y) - \mu(X)\mu(Y)\end{aligned}$$

2. Se X e Y forem independentes

$$\mathbb{E}\{XY\} = \mathbb{E}\{X\} \cdot \mathbb{E}\{Y\} = \mu(X)\mu(Y)$$

$$\Rightarrow \text{cov}(X, Y) = 0$$

Propriedades - cont.

3. Dizemos que X e Y são **descorrelacionados** se e somente se

$$\text{cov}(X,Y) = 0$$

4. Quando $X = Y$

$$\begin{aligned}\text{cov}(X,X) &= \mathbb{E}\{X \cdot X\} - \mu(X)\mu(X) \\ &= \mathbb{E}\left\{X^2\right\} - \mathbb{E}^2\{X\} \qquad = \sigma_X^2\end{aligned}$$

5. Se $\text{cov}(X,Y) = 0$ **não implica** que X e Y são independentes

Exemplo

$\theta \sim$ v.a. uniforme entre $[0, 2\pi]$

$$\begin{cases} X = \sin(\theta) \\ Y = \cos(\theta) \end{cases}$$

$$\text{cov}(X, Y) = \mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}$$

$$\mathbb{E}\{X\} = 0 \quad \text{e} \quad \mathbb{E}\{Y\} = 0$$

$$\mathbb{E}\{XY\} = \mathbb{E}\{\sin(\theta) \cdot \cos(\theta)\} = \int_0^{2\pi} \frac{1}{2\pi} \sin(\theta) \cdot \cos(\theta)\ d\theta = 0$$

cov$(X, Y) = 0$ não implica X e Y independentes!

Coeficiente de correlação

$$\rho_{XY} \triangleq \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \tag{61}$$

⚠ Limitar os valores da correlação/covariância para facilitar a comparação.

Propriedade

$$|\rho_{XY}| \leq 1$$

Correlação e covariância - cont.

Propriedade (prova)

$$\mathbb{E}\left\{[\alpha(X-\mu(X))-(Y-\mu(Y))]^2\right\} \geq 0, \quad \forall\alpha$$

$$\mathbb{E}\left\{\alpha^2(X-\mu(X))^2+(Y-\mu(Y))^2 \right.$$
$$\left. -2\cdot\alpha(X-\mu(X))(Y-\mu(Y))\right\} \geq 0, \quad \forall\alpha$$

$$\alpha^2\sigma_X^2+\sigma_Y^2-2\alpha\mathsf{cov}(X,Y) \geq 0, \quad \forall\alpha$$

(equação do 2o grau para $\Delta \leq 0$)

$$\Delta = 4\cdot\mathsf{cov}^2(X,Y) - 4\cdot\sigma_X^2\cdot\sigma_Y^2 \leq 0$$

$$\mathsf{cov}^2(X,Y) \leq \sigma_X^2\cdot\sigma_Y^2$$

$$\Rightarrow \rho_{XY}^2 = \frac{\mathsf{cov}^2(X,Y)}{\sigma_X^2\cdot\sigma_Y^2} \leq 1$$

$$|\rho_{XY}| \leq 1 \qquad \text{(c.q.d)}$$

Distribuições e densidades condicionais

Meta

Caracterização da distribuição de probabilidade de uma variável aleatória condicionada a ocorrência de outra variável aleatória ou evento

Definição distribuição cumulativa condicionada

$$\begin{aligned} F_X(x|A) &\triangleq \Pr\{X \leq x|A\} \\ &= \frac{\Pr\{X \leq x, A\}}{\Pr\{A\}} \end{aligned} \tag{62}$$

$$\boxed{p_X(x|A) \triangleq \frac{d}{dx} F_X(x|A)}$$

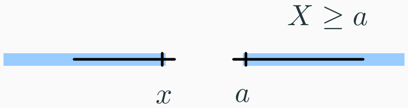
Distribuições e densidades condicionais - cont.

Analisando...

Seja $A = \{x \geq a\}$

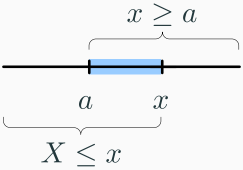
$$F_X(x|A) = F_X(x|X \geq a)$$
$$= \Pr\{X \leq x|X \geq a\}$$

(a) $x < a$



$$\Rightarrow F_X(x|x \geq a) = 0$$

(b) $x \geq a$



$$F_X(x|x \geq a) = \frac{\Pr\{X \leq x, X \geq a\}}{\Pr\{X \geq a\}} = \frac{\int_a^x p_X(\alpha)\ d\alpha}{\int_a^\infty p_X(\beta)\ d\beta}$$

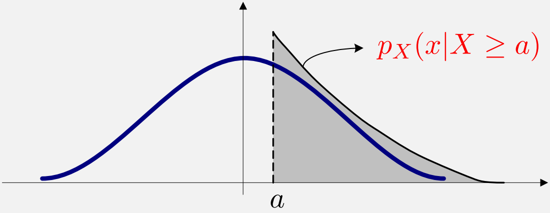
$$= \frac{F_X(x) - F_X(a)}{1 - F_X(a)}$$

$$p_X(x|X \geq a) = \frac{d}{dx} F_X(x|X \geq a) = \frac{\frac{d}{dx} F_X(x)}{1 - F_X(a)} = \frac{p_X(x)}{1 - F_X(a)}$$

Distribuições e densidades condicionais - cont.

Resumindo

$$p_X(x|X \geq a) = \frac{p_X(x)}{\int\limits_0^{\infty} p_X(\beta)\, d\beta} U(x-a)$$



Observações

1.

$$\int_{-\infty}^{\infty} p_X(x|X \geq a)\ dx = \int_{-\infty}^{\infty} \left[\frac{p_X(x)}{\int_a^{\infty} p_X(\beta)\ d\beta} \right] \cdot U(x-a)\ dx$$

$$= \frac{\int_a^{\infty} p_X(x)\ dx}{\int_a^{\infty} p_X(\beta)\ d\beta} = 1$$

Observações - cont.

❷ $\mathbb{E}\{X|A\} = ?$ para $A = \{X \geq a\}$

$$
\begin{aligned}
\mathbb{E}\{X|X \geq a\} &= \int_{-\infty}^{\infty} x \cdot p_X(x|X \geq a)\ dx \\
&= \int_{\infty}^{\infty} \frac{x p_X(x)}{\int_a^{\infty} p_X(\beta)\ d\beta} \cdot U(x-a)\ dx
\end{aligned}
$$

$$
\boxed{\mathbb{E}\{X|X \geq a\} = \frac{\int_a^{\infty} x \cdot p_X(x)\ dx}{\int_a^{\infty} p_X(\beta)\ d\beta}}
$$

Observações - cont.

❸ Caso em que $A = \{X = a\}$

$$F_X(x|X=a) = \frac{\Pr\{X \leq x, X = a\}}{\underbrace{\Pr\{X = a\}}_{\text{v.a. contínua} \Rightarrow \Pr\{X=a\}=0}}$$

Relaxando "um pouco" $X = a$ para

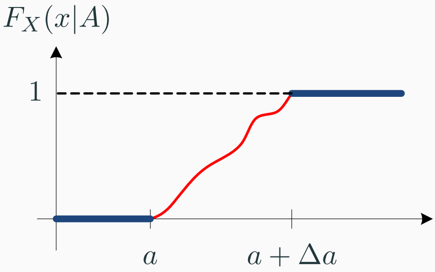
$$a \leq X \leq a + \Delta a, \quad \Delta a \to 0 \quad \text{(depois)}$$

$$\begin{aligned} F_X(x|A) &= F_X(x|a \leq X \leq a + \Delta a) \\ &= \frac{\Pr\{X \leq x, a \leq X \leq a + \Delta a\}}{\Pr\{a \leq X \leq a + \Delta a\}} \end{aligned}$$

Observações - cont.

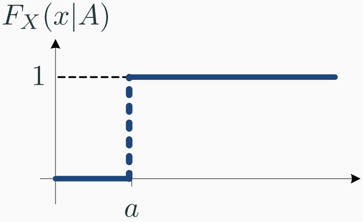
Temos 3 situações

- $X < a \Rightarrow F_X(x|A) = 0$
- $a \leq X \leq a + \Delta a \Rightarrow F_X(x|A) = \frac{\Pr\{a \leq X \leq x\}}{\Pr\{a \leq X \leq a+\Delta a\}}$
- $X > a + \Delta a \Rightarrow F_X(x|A) = 1$



Observações - cont.

Fazendo $\Delta a \to 0$:



$$F_X(x|A) = \underbrace{U(x-a)}_{\text{função degrau unitário}}$$

ou seja

$$\Rightarrow F_X(x|X=a) = U(x-a)$$

$$\boxed{p_X(x|X=a) = \frac{d}{dx} F_X(x|X=a) = \delta(x-a)}$$

Distribuições e densidades condicionais - cont.

Densidade condicional de duas variáveis aleatórias

❓ $p_Y(y|x) = ?$

$$F_Y(y|x \leq X \leq x + \Delta x) = \frac{\Pr\{Y \leq y, x \leq X \leq x + \Delta\}}{\Pr\{x \leq X \leq x + \Delta a\}}$$

Ainda

$$\begin{aligned}
\Pr\{Y \leq y, x \leq X \leq x + \Delta x\} &= \int_{-\infty}^{y} \int_{x}^{x+\Delta x} p_{X,Y}(\alpha, \beta)\ d\alpha d\beta \\
&\cong p_{X,Y}(x, \beta)\Delta x \quad \text{(método de Euler)} \\
&= \int_{-\infty}^{y} p_{X,Y}(x, \beta)\ d\beta \cdot \Delta x
\end{aligned}$$

Densidade condicional de duas variáveis aleatórias - cont.

$$\Pr\{x \leq X \leq x + \Delta x\} = \int_{x}^{x+\Delta x} p_X(\gamma)\ d\gamma \cong p_X(x) \cdot \Delta x$$

$$F_Y(y|x) \cong \frac{\int_{-\infty}^{y} p_{X,Y}(x,\beta)\ d\beta \cdot \Delta x}{p_X(x) \cdot \Delta x} = \frac{\int_{-\infty}^{y} p_{X,Y}(x,\beta\ d\beta}{p_X(x)}$$

$$\boxed{p_Y(y|x) = \frac{dF_Y(y|x)}{dy} = \frac{p_{X,Y}(x,y)}{p_X(x)}}$$

Se X e Y forem independentes

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y) \quad \Rightarrow \quad p_Y(y|x) = p_Y(y) \qquad (63)$$

Teorema central do limite

- ✔ Muitos problemas em v.a. envolvem uma composição de eventos
- ✔ Tais fenômenos resultants podem ser reduzidos, de forma exata ou aproximada, por uma soma de variáveis independentes
- ✔ O estudo deste tipo de fenômenos resulta no mais importante teorema relativo a varáveis aleatórias: o **Teorema Central do Limite**
- ✔ O estudo dele no entanto repousa sobre à chamada *Lei dos Grande Números*

Soma de variáveis aleatórias

Seja $X_1, X_2, \ldots, X_n$ uma seqüência de variáveis aleatórias e sua soma é representada por $S_n = X_1 + X_2 + \ldots + X_n$

Então temos:

$$\boxed{\mathbb{E}\{S_n\} = \sum_{i=1}^{n} \mathbb{E}\{X_i\}} \tag{64}$$

Ou seja, a média de uma soma de v.a.s é a soma das médias das v.a.s

Entretanto, no caso da variância isso é mais complexo.

Teorema central do limite - cont.

Soma de variáveis aleatórias - cont.

Para chegar ao resultado da variância da soma de n variáveis aleatórias, vamos considerar inicialmente duas delas, ou seja, $S_2 = X_1 + X_2$. Então

$$\text{var}(S_2) = \mathbb{E}\left\{(S_2 - \mathbb{E}\{S_2\})^2\right\} = \mathbb{E}\left\{[X_1 + X_2 - \mathbb{E}\{X_1\} - \mathbb{E}\{X_2\}]^2\right\}$$
$$= \mathbb{E}\left\{[(X_1 - \mathbb{E}\{X_1\}) + (X_2 - \mathbb{E}\{X_2\})]^2\right\}$$

Então, após simplificações chegamos ao seguinte resultado

$$\boxed{\text{var}(S_2) = \text{var}(X_1) + \text{var}(X_2) + 2\text{cov}(X_1, X_2)}$$

Ou seja, as covariâncias são levadas em conta no cálculo da variância da soma de v.a.

Soma de variáveis aleatórias - cont.

Ao fazermos as contas para a soma de n v.a.s $(S_n = X_1 + X_2 + \ldots + X_n)$ temos

$$\boxed{\mathsf{var}(S_n) = \sum_{i=1}^{n} \mathsf{var}(X_i) + \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{i=1}^{n} \mathsf{cov}(X_i, X_j)}$$

Entretanto, para **v.a. independentes** temos

$$\mathsf{var}(S_n) = \sum_{i=1}^{n} \mathsf{var}(X_i)$$

uma vez que $\mathsf{cov}(X_i, X_j) = 0 \quad \forall i \neq j$

Soma de variáveis aleatórias - cont.

Se as v.a.s X_i forem independentes e todas tiverem média μ e variância σ^2 teremos então para $S_n = X_1 + X_2 + \cdots + X_n$:

$$\boxed{\begin{aligned} \mathbb{E}\{S_n\} &= n \cdot \mu \\ \text{var}(S_n) &= n \cdot \sigma^2 \end{aligned}}$$

Soma de variáveis aleatórias - cont.

Para calcular a pdf de uma soma de n v.a.s independentes, podemos usar a *função característica*. Iniciando com o caso para $n = 2$ ($S_2 = X_1 + X_2$):

$$\begin{aligned}
P_{S_2}(\omega) &= \mathbb{E}\left\{\exp(-j\omega S_2)\right\} \\
&= \mathbb{E}\left\{\exp[-j\omega(X_1 + X_2)]\right\} \\
&= \mathbb{E}\left\{\exp(-j\omega X_1)\exp(-j\omega X_2)\right\} \\
&= \mathbb{E}\left\{\exp(-j\omega X_1)\right\}\mathbb{E}\left\{\exp(-j\omega X_2)\right\} \\
&= P_{X_1}(\omega)P_{X_2}(\omega)
\end{aligned}$$

Assim, para o caso de duas variáveis

$$p_{S_2}(s) = p_{X_1}(x) \star p_{X_2}(x)$$

Soma de variáveis aleatórias - cont.

Para n variáveis $(S_n = X_1 + X_2 + \ldots + X_n)$ temos:

$$\begin{aligned}
P_{S_n}(\omega) &= \mathbb{E}\left\{\exp(-j\omega S_n)\right\} \\
&= \mathbb{E}\left\{\exp[-j\omega(X_1 + X_2 + \ldots + X_n)]\right\} \\
&= \mathbb{E}\left\{\exp(-j\omega X_1)\exp(-j\omega X_2)\cdots\exp(-j\omega X_n)\right\} \\
&= \mathbb{E}\left\{\exp(-j\omega X_1)\right\}\mathbb{E}\left\{\exp(-j\omega X_2)\right\}\cdots\mathbb{E}\left\{\exp(-j\omega X_n)\right\} \\
&= P_{X_1}(\omega)P_{X_2}(\omega)\cdots P_{X_n}(\omega)
\end{aligned}$$

Logo, a densidade de probabilidade será dada por:

$$\boxed{p_{S_n}(X) = \mathfrak{F}^{-1}\{P_{X_1}(\omega)P_{X_2}(\omega)\cdots P_{X_n}(\omega)\}}$$

Soma de variáveis aleatórias - cont.

Seja $S_n = X_1 + X_2 + \ldots + X_n$ a soma de n v.a. independentes e gaussianas com médias $\mu_1, \mu_2, \ldots, \mu_n$ e variâncias $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. Temos que a função característica de X_k é:

$$P_{X_k} = \exp\left(-j\omega\mu_k - \frac{\omega^2\sigma_k^2}{2}\right)$$

Então, pelo resultado anterior

$$\begin{aligned} P_{S_n}(\omega) &= P_{X_1}(\omega)P_{X_2}(\omega)\cdots P_{X_n}(\omega) \\ &= \exp\left[-j\omega(\mu_1 + \ldots + \mu_n) - \frac{\omega^2(\sigma_1^2 + \ldots + \sigma_n^2)}{2}\right] \end{aligned}$$

que é a função característica de uma variável aleatória gaussiana com média $\mu = \mu_1 + \ldots + \mu_n$ e variância $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$

Soma de variáveis aleatórias - cont.

Se as variáveis X_i que compõem a soma forem independentes e identicamente distribuídas (iid), ou seja, com as mesmas estatísticas, cada uma com função característica

$$P_{X_k}(\omega) = P_X(\omega)$$

então, a função característica da soma será

$$\boxed{P_{S_n}(\omega) = [P_X(\omega)]^n}$$

⚠ Isso nos leva a poder enunciar o teorema central do limite.

Teorema

Seja S_n uma soma de n v.a. i.i.d. com média μ e variância σ^2 finitas, e seja Z_n uma v.a. de média nula e variância unitária, definida por

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

então

$$\boxed{\lim_{n\to\infty} \Pr\{Z_n \leq z\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{x^2}{2}\right)\, dx} \quad (65)$$

Teorema central do limite - cont.

Prova do Teorema Central do Limite

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu)$$

A função característica de Z_n é dada por

$$\begin{aligned}
P_{Z_n}(\omega) &= \mathbb{E}\{\exp(-j\omega Z_n)\} \\
&= \mathbb{E}\left\{\exp\left[\frac{-j\omega}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu)\right]\right\} \\
&= \mathbb{E}\left\{\prod_{k=1}^{n}\exp\left(\frac{-j\omega(X_k - \mu)}{\sigma\sqrt{n}}\right)\right\} \\
&= \prod_{k=1}^{n}\mathbb{E}\left\{\exp\left(\frac{-j\omega(X_k - \mu)}{\sigma\sqrt{n}}\right)\right\} \\
&= \left[\mathbb{E}\left\{\exp\left(\frac{-j\omega(X_k - \mu)}{\sigma\sqrt{n}}\right)\right\}\right]^n
\end{aligned}$$

Teorema central do limite - cont.

Prova do Teorema Central do Limite - cont.

Expandindo a exponencial em série de Taylor, temos

$$\begin{aligned}
&\mathbb{E}\left\{\exp\left(\frac{-j\omega(X-\mu)}{\sigma\sqrt{n}}\right)\right\} \\
&= \mathbb{E}\left\{1 + \frac{-j\omega}{\sigma\sqrt{n}}(X-\mu) + \frac{(j\omega)^2}{2!\sigma^2 n}(X-\mu)^2 + R(\omega)\right\} \\
&= 1 + \frac{-j\omega}{\sigma\sqrt{n}}\mathbb{E}\{(X-\mu)\} + \frac{(j\omega)^2}{2!\sigma^2 n}\mathbb{E}\{(X-\mu)^2\} + \mathbb{E}\{R(\omega)\}
\end{aligned}$$

Notando que $\mathbb{E}\{(X-\mu)\} = 0$ e $\mathbb{E}\{(X-\mu)^2\} = \sigma^2$, temos

$$\mathbb{E}\left\{\exp\left(\frac{-j\omega(X-\mu)}{\sigma\sqrt{n}}\right)\right\} = 1 - \frac{\omega^2}{2n} + \mathbb{E}\{R(\omega)\}$$

O termo $\mathbb{E}\{R(\omega)\}$ pode ser desprezado em relação a $\frac{\omega^2}{2n}$ quando n torna-se grande.

Prova do Teorema Central do Limite - cont.

Ao substituirmos o resultado anterior na fórmula da função característica de Z_n, tem-se

$$P_{Z_n}(\omega) = \left[1 - \frac{\omega^2}{2n}\right]^n$$

E sabe-se que

$$P_{Z_n}(\omega) \to \exp\left(-\frac{\omega^2}{2}\right), \quad \text{quando} \quad n \to \infty$$

O resultado acima é a função característica de uma v.a. gaussiana de média zero e variância unitária. (c.q.d.) $\square$

Média das Amostras e Lei dos Grandes Números

Seja X uma v.a. cuja média $\mathbb{E}\{X\} = \mu$ é desconhecida. Sejam $X_1, \ldots, X_n$ n repetições independentes medidas de X, ou seja, os X_i são v.a. iid com a mesma pdf de X. A média das amostras da seqüência usada para estimar $\mathbb{E}\{X\}$ é dada por:

$$\boxed{\widehat{M}_n = \frac{1}{n}\sum_{j=1}^{n} X_j} \tag{66}$$

Perguntas

Qual a média e a variância do estimador $\widehat{M}_n$ de tal forma que ele é um bom estimador de $\mathbb{E}\{X\}$? E qual a influência do n no estimador?

Média das Amostras e Lei dos Grandes Números -cont.

A média das amostras é por si só uma v.a. e desta maneira irá exibir características de variação aleatória. Um bom estimador deve possuir duas propriedades:

1. Deve ser não-polarizado, ou seja, na média seu valor deve tender para o valor da variável;
2. Não deve variar muito em torno do valor correto, ou seja, $\mathbb{E}\left\{\left(\widehat{M_n} - \mu\right)^2\right\}$ deve ser pequeno.

Média das Amostras e Lei dos Grandes Números -cont.

Valor esperado do estimador:

$$\begin{aligned}\mathbb{E}\left\{\widehat{M_n}\right\} = \mathbb{E}\left\{\frac{1}{n}\sum_{j=1}^{n} X_j\right\} &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left\{X_j\right\} \\ &= \frac{1}{n}\cdot n\mu \\ &= \mu\end{aligned}$$

Assim, temos um estimador não-polarizado.

Teorema central do limite - cont.

Média das Amostras e Lei dos Grandes Números -cont.

Já a variância temos:

$$\mathsf{var}(\widehat{M_n}) = \mathbb{E}\left\{\left(\widehat{M_n} - \mu\right)^2\right\} = \mathbb{E}\left\{\left(\widehat{M_n} - \mathbb{E}\{\widehat{M_n}\}\right)^2\right\}$$

Deve-se notar que $\widehat{M_n} = \frac{S_n}{n}$, e $\mathsf{var}(S_n) = n \cdot \mathsf{var}(X_j) = n \cdot \sigma^2$, uma vez que X_i são iid. Logo

$$\mathsf{var}(\widehat{M_n}) = \frac{1}{n^2}\mathsf{var}(S_n) = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Assim, a variância tende a zero quando o número de amostras aumenta! Isto implica que a probabilidade da média amostral ser próxima do valor correto tende a um quando n assume um valor grande.

Média das Amostras e Lei dos Grandes Números -cont.

A afirmação anterior pode ser então formalizada por meio do uso da desigualdade de Tchebchev:

$$\Pr\left\{\left|\widehat{M_n} - \mathbb{E}\left\{\widehat{M_n}\right\}\right| \geq \epsilon\right\} \leq \frac{\mathsf{var}(\widehat{M_n})}{\epsilon^2}$$

Então, substituindo os valores, temos:

$$\Pr\left\{\left|\widehat{M_n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

Considerando o complemento:

$$\boxed{\Pr\left\{\left|\widehat{M_n} - \mu\right| < \epsilon\right\} \geq 1 - \frac{\sigma^2}{n\epsilon^2}} \qquad (67)$$

Média das Amostras e Lei dos Grandes Números -cont.

Note que se $n \to \infty$ então

$$\lim_{n \to \infty} \Pr\left\{\left|\widehat{M_n} - \mu\right| < \epsilon\right\} = 1$$

Este resultado nos remete então para duas importantes leis em estatística

1. **Lei fraca dos grandes números**;
2. **Lei forte dos grandes números**

Lei Fraca dos Grandes Números

Seja $X_1, X_2, \ldots$ uma seqüência de v.a. i.i.d. com média $\mathbb{E}\{X\} = \mu$, então para $\epsilon > 0$

$$\lim_{n \to \infty} \Pr\left\{ \left| \widehat{M_n} - \mu \right| < \epsilon \right\} = 1 \qquad (68)$$

A lei fraca dos grandes números afirma que para um valor fixo n suficientemente grande , a média amostral usando n valores vai ser próxima do valor correto com alta probabilidade.

Lei Forte dos Grandes Números

Seja $X_1, X_2, \ldots$ uma seqüência de v.a. i.i.d. com média finita $\mathbb{E}\{X\} = \mu$ e variância finita, então

$$\Pr\left\{\lim_{n\to\infty} \widehat{M_n} = \mu\right\} = 1 \qquad (69)$$

Embora semelhante a lei fraca dos grandes números, há uma diferença crucial. Ela afirma que com probabilidade um, toda seqüência usada para cálculo da média amostral vai eventualmente se aproximar e permanecer próxima a $\mathbb{E}\{X\} = \mu$. Este é o tipo de convergência que se espera em situações físicas nas quais a regularidade estatística se mantém.

Problema geral

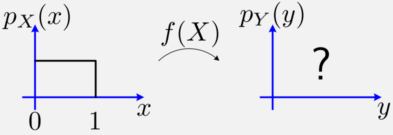
Dados $X \sim$ v.a. e uma função

$$Y = f(X)$$

a questão é como determinar $p_Y(y)$ conhecendo-se $p_X(x)$.

Por exemplo, seja X uma v.a. uniforme em $[0, 1]$ e

$$Y = \frac{1}{\lambda} \cdot \ln\left(\frac{1}{X}\right)$$



Vamos analisar dois casos

1o. caso

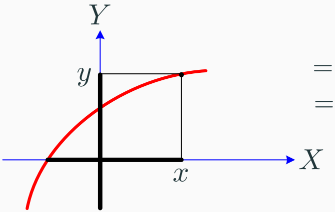
$$X \sim \text{v.a.} \quad p_X(x) \text{ conhecido}$$

$$\begin{cases} y = f(x) \\ \frac{df(x)}{dx} > 0 \Rightarrow f(x) \text{ é monotônico crescente} \end{cases}$$

$$\rightarrow f \text{ é biunívoca } [p_Y(y), F_Y(y)]$$

$$F_Y(y) = \Pr\{Y \leq y\} = \Pr\{f(X) \leq y\}$$

Como $\frac{df(x)}{dx} > 0 \Rightarrow$



$$= \Pr\{X \leq f^{-1}(y)\}$$
$$= F_X\left[f^{-1}(y)\right]$$

1o. caso - cont.

$$
\begin{aligned}
p_Y(y) = \frac{dF_Y(y)}{dy} &= \frac{dF_X\left[f^{-1}(y)\right]}{dy} \\
&= \frac{d}{dy}\int_{-\infty}^{f^{-1}(y)} p_X(x)\ dx \\
&= p_X\left[f^{-1}(y)\right]\cdot\frac{df^{-1}(y)}{dy}
\end{aligned}
$$

Mas: $\frac{df^{-1}(y)}{dy} = \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{\frac{df(x)}{dx}}$. Logo,

$$
p_Y(y) = p_X(x)\cdot\left.\frac{1}{\frac{df(x)}{dx}}\right|_{x=f^{-1}(y)} \qquad \text{para} \quad \frac{df(x)}{dx} > 0
$$

2o. caso

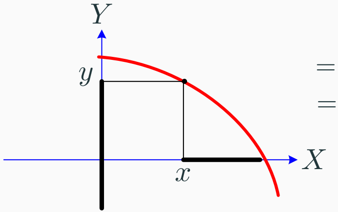
$$X \sim \text{v.a.} \quad p_X(x) \text{ conhecido}$$

$$\begin{cases} y = f(x) \\ \frac{df(x)}{dx} < 0 \Rightarrow f(x) \text{ é monotônico decrescente} \end{cases}$$

$$\rightarrow f \text{ é biunívoca } [p_Y(y), F_Y(y)]$$

$$F_Y(y) = \Pr\{Y \leq y\}$$

$$F_Y(y) = \Pr\{f(X) \leq y\}$$



$$= \Pr\{X > f^{-1}(y)\}$$
$$= 1 - F_X\left[f^{-1}(y)\right]$$

2o. caso - cont.

$$p_Y(y) = \frac{F_Y(y)}{dy} = \frac{-p_X(x)}{\frac{df(x)}{dx}}\Bigg|_{x=f^{-1}(y)} \qquad \text{com } \frac{df(x)}{dx} < 0$$

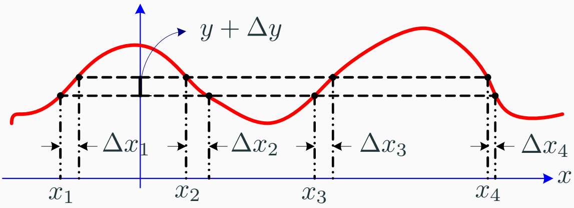
$$= \frac{p_X(x)}{\frac{df(x)}{dx}}$$

O sinal desaparece pois $\frac{df(x)}{dx}$ também é negativo.

Resumindo, para funções de uma variável aleatória temos:

$$\boxed{p_Y(y) = \left.\frac{p_X(x)}{\left|\frac{df(x)}{dx}\right|}\right|_{x=f^{-1}(y)}} \tag{70}$$

Funções não-biunívocas



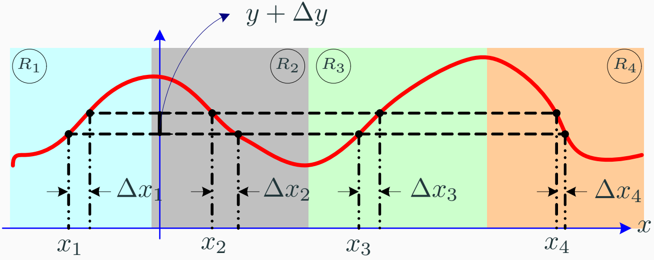
$$
\begin{aligned}
F_Y(y) &= \Pr\{Y \leq y\} \\
&= \Pr\{y \leq Y \leq y + \Delta y\}
\end{aligned}
$$

Como há contribuições de diferentes intervalos de x, então

$$
\begin{aligned}
F_Y(y) = &\Pr\{x_1 \leq X_1 \leq x_1 + \Delta x_1\} + \Pr\{x_2 - \Delta x_2 \leq X_2 \leq x_2\} \\
&+ \Pr\{x_3 \leq X_3 \leq x_3 + \Delta x_3\} + \Pr\{x_4 - \Delta x_4 \leq X_4 \leq x_4\}
\end{aligned}
$$

Funções não-biunívocas - cont.

Nestes casos, dividimos o espaço amostral em regiões biunívocas. Ou seja, teremos algo como



Funções não-biunívocas - cont.

Nestes casos, chamando de $y = f_i(x_i)$ em cada região R_i, temos então a seguinte expressão para encontrar a densidade de probabilidade de Y:

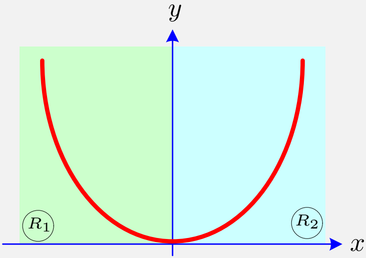
$$\boxed{p_Y(y) = \sum_{R_i} \frac{p_X(x_i)}{\left|\frac{df_i(x_i)}{dx_i}\right|}\Bigg|_{x_i=f^{-1}(y)}} \tag{71}$$

Funções de variáveis aleatórias - cont.

Exemplo

$$X \sim N(0, \sigma^2)$$

$$Y = X^2$$



$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Exemplo - cont.

$$p_Y(y) = \underbrace{\left.\frac{p_X(x)}{\left|\frac{df_1(x)}{dx}\right|}\right|_{x=f_1^{-1}(y)}}_{R_1} + \underbrace{\left.\frac{p_X(x)}{\left|\frac{df_2(x)}{dx}\right|}\right|_{x=f_2^{-1}(y)}}_{R_2}$$

$$\frac{df_1(x)}{dx} = \frac{dx^2}{dx} = 2x \quad x = -\sqrt{y}$$

$$\frac{df_2(x)}{dx} = \frac{dx^2}{dx} = 2x \quad x = \sqrt{y}$$

$$p_Y(y) = p_X(x)|_{x=-\sqrt{y}} \cdot \frac{1}{2\sqrt{y}} + p_X(x)|_{x=\sqrt{y}} \cdot \frac{1}{2\sqrt{y}}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{y}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{y}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Funções de várias variáveis aleatórias

Problema

$$X \sim \text{v.a.} \qquad Y \sim \text{v.a.}$$

$$\begin{cases} U = f(X,Y) \\ V = g(X,Y) \end{cases}$$

Conhecido $p_{X,Y}(x,y)$, como achar $p_{U,V}(u,v)$?

Funções de várias variáveis aleatórias

Em regiões biunívocas:

$$\boxed{p_{U,V}(u,v) = \frac{p_{X,Y}(x,y)}{\left|J\left(\frac{u,v}{x,y}\right)\right|}\Bigg|_{\substack{x=\mathfrak{f}(u,v)\\ y=\mathfrak{g}(u,v)}}} \tag{72}$$

em que $\mathfrak{f}(\cdot)$ e $\mathfrak{g}(\cdot)$ são as funções inversas de f e g, respectivamente. E

$$J\left(\frac{u,v}{x,y}\right) = \det\begin{bmatrix} \frac{du}{dx} & \frac{du}{dy} \\ \frac{dv}{dx} & \frac{dv}{dy} \end{bmatrix} = \frac{1}{\det\begin{bmatrix} \frac{d\mathfrak{f}}{du} & \frac{d\mathfrak{f}}{dv} \\ \frac{d\mathfrak{g}}{du} & \frac{d\mathfrak{g}}{dv} \end{bmatrix}}$$

é chamado *Jacobiano de* u, v *em relação a* x, y

Funções de várias variáveis aleatórias

Caso particular

$$Z = X + Y,$$

$$p_{X,Y}(x,y) \text{ conhecido}$$

$$p_Z(z) = ?$$

Definir então

$$\begin{cases} z = x + y \\ w = x \end{cases} \Rightarrow p_{Z,W}(z,w)$$

$$J\left(\frac{z,w}{x,y}\right) = \det \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = -1$$

Funções de várias variáveis aleatórias - cont.

Caso particular - cont.

Logo,

$$p_{Z,W}(z,w) = \left.\frac{p_{X,Y}(x,y)}{|-1|}\right|_{\substack{x=\mathfrak{f}(z,w)\\ y=\mathfrak{g}(z,w)}}$$

$$x = w, \qquad y = z - w$$

$$p_{Z,W}(z,w) = p_{X,Y}(w, z-w)$$

Funções de várias variáveis aleatórias - cont.

Caso particular - cont.

Então, para achar a densidade marginal $p_Z(z)$, temos

$$p_Z(z) = \int_{-\infty}^{\infty} p_{X,Y}(w, z - w) \, dw$$

Se supormos que X e Y são independentes

$$p_{X,Y}(w, z - w) = p_X(w) \cdot p_Y(z - w)$$
$$\Rightarrow p_{Z,W}(z, w) = p_X(w) \cdot p_Y(z - w)$$

Funções de várias variáveis aleatórias - cont.

Caso particular - cont.

$$p_Z(z) = \int_{-\infty}^{\infty} p_{Z,W}(z,w)\ dw = \underbrace{\int_{-\infty}^{\infty} p_X(w) \cdot p_Y(z-w)\ dw}_{\text{convolução}}$$

Assim:

$$\boxed{p_Z(z) = p_x(x) \star p_Y(y)} \tag{73}$$

Geração de Números Aleatórios

Aleatoriedade × computadores

Simular processos ou eventos probabilísticos é importante em várias áreas da engenharia. Alguns exemplos:

1. Modelagem de sistemas de telecomunicações (tráfego, usuários, propagação, etc.).
2. Geração de seqüências de treinamento de filtros adaptativos.
3. Simulação de tráfego de pacotes em redes de computadores.
4. Métodos de otimização estocástica (e.g. algoritmos genéticos).
5. Navegação de robôs móveis.

? Mas como fazer um computador gerar um número aleatório?

Aleatoriedade × computadores - cont.

Na verdade, o que se consegue fazer é o computador gerar um número que **parece** aleatório.

Alguns métodos são então utilizados para gerar números que possuem padrão que se assemelha a um valor aleatório, ou seja, emulamos um número aleatório com um padrão complexo que não parece ser determinístico.

Métodos congruenciais

Métodos congruenciais usam algoritmos determinísticos recursivos, implementados na forma de curtos programas computacionais, com o objetivo de gerar uma seqüência de números toda vez que tal programa for executado.

A idéia básica por trás deste tipo de método consiste em usar o número *aleatório* gerado na iteração anterior para produzir o número *aleatório* da iteração atual.

Existem diversas implementações de métodos congruenciais. Aqui nos restringiremos ao método conhecido como "**congruencial misto**".

Métodos congruenciais - cont.

O método congruencial misto usa a seguinte equação recursiva:

$$\boxed{x(n) = [a \cdot x(n-1) + c] \;\; \text{modulo} \;\; m} \tag{74}$$

A operação "**modulo** m" significa que $x(n)$ é o resto da divisão da quantidade $ax(n-1) + c$ pelo número m.

1. $x(n)$ e $x(n-1)$ são os números (pseudo)-aleatórios gerados na iteração atual (n) e na iteração anterior $(n-1)$, respectivamente.
2. a, c e m são *constantes* positivas escolhidas adequadamente, de forma a obedecer as seguintes relações:

$$a < m \qquad \text{e} \qquad c < m \tag{75}$$

Métodos congruenciais - cont.

Pergunta

Como funciona a equação (74)?

Um passo - Se $a = 5$, $c = 3$ e $m = 16$ e $x(0) = 7$, teremos que $a \cdot x(0) + c = 38$ e assim $x(1) = 38 \mod 16 = 6$.

Vários passos - Por exemplo, para $a = 5$, $c = 3$ e $m = 16$ e $x(0) = 7$, iremos obter **sempre** a seguinte seqüência de números: $x(1) = 6$, $x(2) = 1$, $x(3) = 8$, $x(4) = 11$, $x(5) = 10$, $x(6) = 5$, $x(7) = 12$, $x(8) = 15$, $x(9) = 14$, $x(10) = 9$, $x(11) = 0$, $x(12) = 3$, $x(13) = 2$, $x(14) = 13$ e $x(15) = 4$; ou seja, todos os números entre 0 e 15 (são 16 números).

Métodos congruenciais - cont.

Assim, teríamos um algoritmo que, para gerar N amostras de uma variável (pseudo)-aleatória, poderia ser escrito da seguinte maneira:

Algoritmo

Input $a, c, m, x(0)$

for $i = 1 : N$

$$x(i) = [a \cdot x(i-1) + c] \mod m$$

end

Como $x(0)$ é o ponto de partida da sequência, geralmente é denominado de *semente* (ou *seed*, do inglês).

Pergunta

Para um certo valor de m, quantos números (pseudo) aleatórios diferentes podem ser obtidos a partir da Equação (74)?

Resposta

No máximo m números, ou seja, todos os números entre 0 e $m-1$.

Qual seria o valor de $x(16)$ nesta seqüência? Fazendo as contas, temos $x(16) = a \cdot x(15) + c = 23 \mod 16 = 7$. Desta forma, a seqüência passará a se repetir com $x(17) = 6$, $x(18) = 1$, e assim por diante.

Métodos congruenciais - cont.

- ✔ O exemplo anterior ilustra que qualquer seqüência de números aleatórios gerada por meio da Equação (74) é **cíclica** (motivo do termo $congruencial$).
- ✔ Cada ciclo consiste necessariamente de uma quantidade finita de números distintos (no máximo m). Após um ciclo ser completado, um novo ciclo idêntico ao anterior é começado. Em nosso exemplo, cada ciclo consiste de 16 números distintos.
- ✔ Um comprimento de ciclo pequeno pode causar problemas em uma simulação, visto que uma sucessão de ciclos pequenos e idênticos não se comporta como uma seqüência de variáveis aleatórias independentes.
- ✔ Entretanto, se m for grande e as constantes a e c forem escolhidas adequadamente para tornar o comprimento de cada ciclo semelhante ao tamanho de m, o caráter finito dos ciclos não terá importância prática.

Métodos congruenciais - cont.

- Considere o caso em que m é feito igual a 2^b, em que b é o número de bits do processador de um computador hipotético.
- Este é o valor de m usado em simulações computacionais, visto que 2^b é o total de números inteiros que podem ser expressos em forma binária com o números de bits disponíveis.
- Quando $b = 32$, tem-se que $m = 2^{32} \approx$ 4,3 bilhões de números diferentes. Com um ciclo (período) desta ordem de magnitude, a maioria das simulações computacionais produzirão resultados confiáveis. Se $b = 64$, fica ainda mais longo o comprimento máximo da sequência.

Métodos congruenciais - cont.

- ✔ Vale notar que o significado da palavra *aleatório* aplicada aos números produzidos pela Equação (74) não é preciso.
- ✔ Obviamente, se nós conhecemos a, c e m, então podemos prever perfeitamente a seqüência completa de números que sucederão o número inicial $x(0)$.
- ✔ Assim, as seqüências de números produzidas pela Equação (74), assim como por outros métodos congruenciais, são chamados de números **pseudo-aleatórios**.
- ✔ Este fato, contudo, é apenas de importância teórica, desde que as seqüências geradas imitem o comportamento de amostras independentes de uma variável aleatória com distribuição uniforme (distribuída entre 0 e m).

Métodos congruenciais - cont.

- ✔ Na prática, podemos resolver a questão de gerar sequências diferentes escolhendo-se uma semente diferente cada vez que seqüências de números aleatórios uniformes precisarem ser geradas.
- ✔ Em geral, com o intuito de automatizar o processo de escolha da semente, escolhe-se o número associado ao tempo (relógio) do computador em que a simulação está sendo efetuada.

Deve-se notar também que como a sequência apresenta números uniformemente distribuídos entre 0 e $m-1$ e para obter a distribuição uniforme entre 0 e 1 deve-se dividir todos os elementos por $m-1$.

Comentários sobre escolha das constantes a e c

- ✎ Uma escolha infeliz de a e c pode levar a ciclos muito curtos, mesmo para valores elevados de m.
- Orientações gerais para uma boa escolha de a e c: para $m = 2^b$, a constante a deve ser escolhida tal que ela seja igual a 1 em aritmética de módulo 4, ou seja, $a = 1, 5, 9, 13, \ldots$ e que a constante c seja um número ímpar.
- ✎ Para estes valores de m, a e c, a escolha de $x(0)$ torna-se irrelevante no que concerne o comprimento do ciclo.

Propriedades desejadas na geração de sequências aleatórias (método congruencial)

1. Os números devem parecer ser estatisticamente independentes uns dos outros. Embora, estritamente falando, eles detenham um certo grau de correlação.
2. Os números devem estar uniformemente distribuídos sobre uma certa faixa de valores.
3. A seqüência de números não deve se repetir, qualquer que seja o comprimento desejado pelo usuário. Para o caso do método congruencial apresentado, esta propriedade só pode ser aproximada, escolhendo-se um valor elevado para m.

Propriedades desejadas na geração de sequências aleatórias (método congruencial) - cont.

❹ Os números devem ser gerados rapidamente. Em outras palavras, o método não deve ser "pesado" computacionalmente.

❺ O método também não deve exigir muito espaço em memória.

❻ Qualquer seqüência de números gerados durante uma certa simulação deve ser *reproduzível*, posteriormente. Para o caso do método congruencial, basta usar a mesma semente para reproduzir a mesma seqüência de números.

Do ponto de vista prático, o método da equação (74) atende apenas de forma aproximada todas estas propriedades.

Exercício computacional

Gerar uma seqüência de 1000 números aleatórios uniformes entre 0 e 1. Plotar o histograma e calcular os valores da média e variância, comparando com os valores teóricos esperados. Especificar os parâmetros a, c e m escolhidos, bem como foi determinada a semente.

Geração de números aleatórios gaussianos

O teorema central do limite (TCL) pode ser utilizado para gerar números (pseudo-)aleatórios gaussianos, a partir de números aleatórios uniformes gerados via métodos congruenciais.

Mas como fazer isso?

Geração de números aleatórios gaussianos - cont.

Do TCL sabemos que

$$\boxed{Z_n \sim N(0,1), \qquad \text{à medida que} \qquad n \to \infty} \tag{76}$$

em que

$$Z_n = \frac{S_n - n\mu}{\sigma/\sqrt{n}} \tag{77}$$

$$= \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}, \tag{78}$$

assumindo que $\{X_1, X_2 \ldots, X_n\}$ são v.a.'s i.i.d.

Geração de números aleatórios gaussianos - cont.

Sem perda de generalidade, podemos considerar que X_i, $i = 1, 2 \ldots, n$, são v.a.'s uniformes no intervalo (a, b). Sendo $a = 0$ e $b = 1$, temos que:

$$\begin{aligned} \mu &= \frac{b-a}{2} &&= \frac{1}{2} \\ \sigma^2 &= \frac{(b-a)^2}{12} &&= \frac{1}{12}. \end{aligned} \tag{79}$$

Neste caso, esses valores nos levam a ter a seguinte expressão para Z_n:

$$\boxed{Z_n = \frac{\sum_{i=1}^{n} X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}}} \tag{80}$$

Geração de números aleatórios gaussianos - cont.

Podemos simplificar a expressão de Z_n se adotarmos $n = 12$, o que nos leva à seguinte expressão:

$$\boxed{Z_{12} = \sum_{i=1}^{12} X_i - 6} \tag{81}$$

A equação acima nos mostra que para gerar uma v.a. *aproximadamente* gaussiana basta somar 12 v.a.'s uniformemente distribuídas entre 0 e 1, subtraindo o valor 6 em seguida.

Esta fórmula exprime um dos métodos mais simples e bastante usado para se gerar v.a.'s gaussianas a partir de v.a.'s uniformes.

Geração de números aleatórios gaussianos - cont.

Dica importante

A fórmula da Eq. (81) sempre gera números gaussianos de média zero ($\mu = 0$) e variância unitária ($\sigma^2 = 1$).

Para gerar variáveis aleatórias gaussianas Z_{12}^* de qualquer média μ_d e variância σ_d^2, basta implementar a seguinte operação:

$$Z_{12}^* = \mu_d + \sigma_d \cdot Z_{12} \sim N\left(\mu_d, \sigma_d^2\right) \tag{82}$$

ou seja, basta multiplicar Z_{12} pelo desvio-padrão desejado (σ_d) e adicionar a média desejada μ_d.

Geração de números aleatórios gaussianos - cont.

Algoritmo (pseudo-código)

$z = 0;$

for $j = 1 : 12$

  Input $a, c, m, x_j(0)$

  for $i = 1 : N$

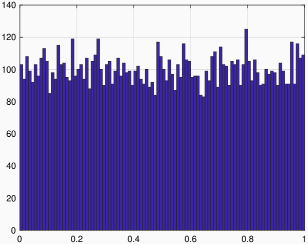
$$x_j(i) = [a \cdot x_j(i-1) + c] \mod m$$

  end

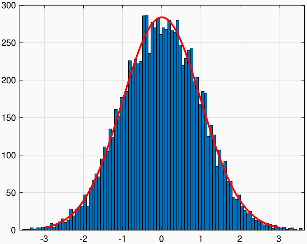
  $z = z + x_j;$

end

$z = z - 6;$



(a)



(b)

Figura 3: (a) Histograma de uma das 12 v.a.'s uniformes geradas. (b) Histograma da v.a. gaussiana resultante da soma de 12 v.a.'s uniformes.

Método da cdf inversa

Já foi vimos que métodos congruenciais geram números (pseudo-)aleatórios uniformemente distribuídos. Também foi visto que somando tais números é possível gerar números com distribuição gaussiana.

O método da cdf inversa usa números aleatórios uniformes gerados via métodos congruenciais para gerar números aleatórios distribuídos segundo uma outra densidade de probabilidade qualquer.

Mas como isso funciona?

Método da cdf inversa - cont.

Este método baseia-se em um resultado bastante importante da teoria de funções de variáveis aleatórias. Vamos apresentá-lo na forma de um exercício resolvido.

Problema

Seja uma variável aleatória contínua X, cuja pdf é conhecida e dada por $p_X(x)$. Considere também a variável aleatória $Y = F_X(x)$, em que $F_X(x)$ é a cdf da variável aleatória X. Pede-se determinar $p_Y(y)$.

Método da cdf inversa - cont.

Solução

Sabe-se que a pdf de uma variável $Y = f(X)$, que é função de uma outra variável X de pdf $p_X(x)$ conhecida, é dada por

$$p_Y(y) = \left.\frac{p_X(x)}{\left|\frac{dy}{dx}\right|}\right|_{x=f^{-1}(y)} \tag{83}$$

Assim, aplicando a Equação (83) à solução do exercício, chega-se a

$$p_Y(y) = \frac{p_X(x)}{\left|\frac{dy}{dx}\right|} = \frac{p_X(x)}{\left|\frac{dF_X(x)}{dx}\right|} = \frac{p_X(x)}{p_X(x)} = 1 \tag{84}$$

O resultado da resolução nos diz que, independente da pdf da variável X, os valores de $Y = F_X(x)$ (ou seja, as probabilidades associadas a X) estão distribuídas ao longo do intervalo $[0, 1]$ segundo uma pdf uniforme.

? Pergunta importante

Como o resultado mostrado na Eq. (84) pode ser usado para gerar números aleatórios não-uniformes ou não-gaussianos?

Vamos fornecer a resposta em forma de um algoritmo, apresentado a seguir.

Método da cdf inversa - cont.

Passo 1 - Seja X uma variável aleatória cuja pdf $p_X(x)$ é conhecida, obter a fórmula da cdf por integração:

$$F_X(x) = \int_{-\infty}^{x} p_X(u)du. \quad (85)$$

Passo 2 - Determinar a expressão analítica da cdf inversa de X:

$$x = F_X^{-1}(y) \quad (86)$$

Passo 3 - Gerar uma seqüência de N números aleatórios uniformes $\{y_i\}$, $i = 1, ..., N$, na faixa de 0 a 1. Para cada número aleatório y_i, $i = 1, \ldots, N$, calcular o valor correspondente de x_i usando a fórmula obtida da Eq. (86):

$$x_i = F_X^{-1}(y_i), \qquad \forall i = 1, \ldots, N \quad (87)$$

Passo 4 - Verificar se a seqüência de números aleatórios $\{x_i\}$ realmente é consistente com hipótese de ser equivalente a uma amostra de X.

Observações

Uma forma bem prática de se fazer a avaliação exigida no Passo 4 é através do histograma da distribuição $\{x_i\}$.

Um procedimento quantitativo útil consiste na comparação dos valores da média e variância da distribuição $\{x_i\}$, chamados aqui de valores estimados, com a média e variância teóricas da pdf de X.

O exemplo a seguir ilustra todos os passos do algoritmo de geração de números aleatórios não-uniformes pelo método da cdf inversa.

Método da cdf inversa - cont.

Exercício

Gerar uma seqüência de 1000 números aleatórios obedecendo a uma pdf exponencial, de parâmetro $\lambda = 3$.

Solução

Passo 1 - A pdf de uma variável aleatória X exponencialmente distribuída é dada por $p_X(x) = \lambda \exp(-\lambda x)$, $x \geq 0$. Assim, a cdf correspondente pode ser obtida por integração:

$$\begin{aligned} y &= F_X(x) = \int_0^x \lambda \exp(-\lambda x)\, du \\ &= 1 - \exp(-\lambda x) \end{aligned} \qquad (88)$$

Solução - cont.

Passo 2 - A expressão da cdf inversa é obtida a partir da Eq. (88), resultando em :

$$x = F_X^{-1}(y) = -\frac{1}{\lambda}\log(1-y) \tag{89}$$

em que $\log(z)$ denota o logaritmo natural de z.

Solução - cont.

Passo 3 - De posse de $N = 10000$ números aleatórios uniformemente distribuídos, y_i, gerados pelo método congruencial misto, são calculados 10000 números aleatórios exponencialmente distribuídos x_i por meio da Eq. (89):

$$x_i = F_X^{-1}(y_i) = -\frac{1}{\lambda}\log(1-y_i),\ i = 1, \ldots, 10000 \tag{90}$$

Método da cdf inversa - cont.

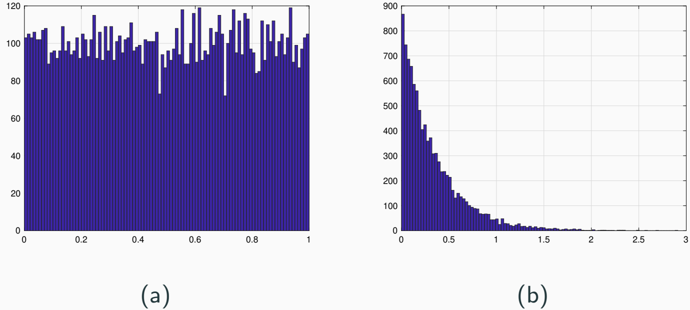


Figura 4: Histogramas da amostra de números aleatórios uniformes $\{y_i\}$ e da amostra de números aleatórios exponencialmente distribuídos $\{x_i\}$.

Em termos númericos, os valores estimados e teóricos da média e da variância de $\{x_i\}$ foram os seguintes:

$$\textbf{Teóricos}: \mu = \frac{1}{\lambda} = \frac{1}{3} = 0,3333 \text{ e } \sigma^2 = \frac{1}{\lambda^2} = \frac{1}{9} = 0,1111$$

$$\begin{aligned} \textbf{Amostrais}: \bar{x} &= \frac{\sum_{i=1}^{10000} x_i}{10000} = 0,3303 \\ s^2 &= \frac{\sum_{i=1}^{10000} (x_i - \bar{x})^2}{10000} = 0,1056 \end{aligned} \tag{91}$$

Nota-se que os valores estimados e teóricos são bem próximos, como era de se esperar. Quanto maior a quantidade N de números uniformes gerados, mais similares serão os valores estimados e os teóricos.

Inferência Estatística

Teste de hipóteses

Hipótese estatística

Uma hipótese estatística é uma **afirmação** sobre os parâmetros de uma ou mais populações.

Teste de hipótese

Procedimento para **análise** da hipótese e tomada de decisão

Teste de hipóteses - cont.

Conjectura (população) $\longrightarrow$ **Hipótese estatística**

Regra para decidir sobre a hipótese (se é verdadeira ou não) $\longrightarrow$ **Teste de hipótese**

Teste de hipóteses - cont.

Seja um engenheiro de telecomunicações interessado no volume de dados que chegam por segundo em determinado roteador.

$X =$ volume de dados que chegam por segundo em determinado roteador

$X \sim$ distribuição de probabilidade (variável aleatória)

Podemos fazer uma *hipótese* a respeito do número **médio** de pacotes no roteador, expressando isto formalmente como

$$\boxed{\begin{aligned} H_0 &: \mu = 50 \text{ Mbps} \\ H_1 &: \mu \neq 50 \text{ Mbps} \end{aligned}} \tag{92}$$

Teste de hipóteses - cont.

O interesse está numa afirmação sobre taxa **média** populacional de pacotes ➜ **hipótese**

$$
\begin{aligned}
H_0 &: \quad \mu = 50 \text{ Mbps} \quad \text{hipótese nula}\\
H_1 &: \quad \mu \neq 50 \text{ Mbps} \quad \text{hipótese alternativa}
\end{aligned}
$$

E como construir H_0?

- resultado de um experimento anterior ou conhecimento do processo;
- modelo ou teoria;
- especificações ou legislação.

Teste de hipóteses - cont.

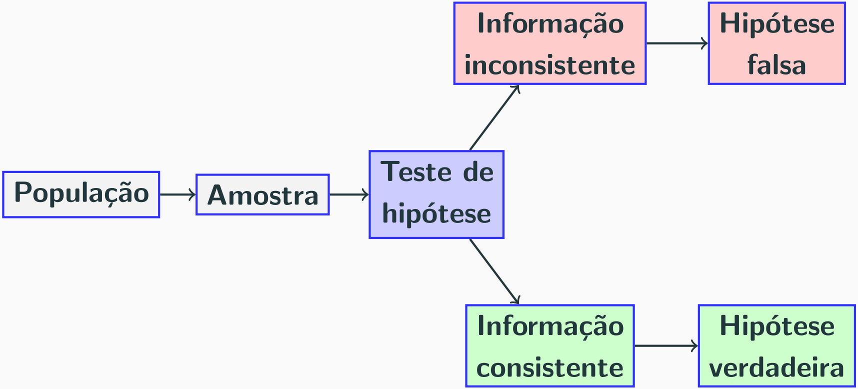
No nosso exemplo

$H_0 :$ $\mu = 50$ Mbps (hipótese nula)
$H_1 :$ $\mu \neq 50$ Mbps (hipótese alternativa bilateral)
$H_1 :$ $\mu \leq 50$ Mbps (hipótese alternativa unilateral inferior)
$H_1 :$ $\mu \geq 50$ Mbps (hipótese alternativa unilateral superior)

H_0 **é aquela que queremos testar.**

A rejeição de H_0 leva sempre à aceitação de H_1 (teste binário).

Estrutura de um teste de hípótese



Teste de hipóteses - cont.

- O teste **não permite ter certeza** se uma hipótese em particular é verdadeira ou falsa. Como são tomadas **amostras** de uma determinada distribuição, elas refletem apenas uma parcela das informações da distribuição.
- Um teste de hipótese é desenvolvido com a **probabilidade de ter uma conclusão errada** (não refletir a realidade): não é possível coletar todas as amostras da população (censo).

Teste de hipóteses - cont.

Retomando nosso exemplo...

Tomamos uma amostra de tamanho N e calculamos a média amostral $\overline{X}$.

Aqui temos duas possibilidades

❶ $\overline{X}$ próximo de 50: evidência que a média verdadeira é 50, ou seja, não rejeitamos H_0.

❷ $\overline{X}$ longe de 50: evidência que a média verdadeira não é 50, ou seja, rejeitamos H_0 (H_1 é aceita).

? **Pergunta**

Mas, o que determina **"próximo"** e **"longe"**?

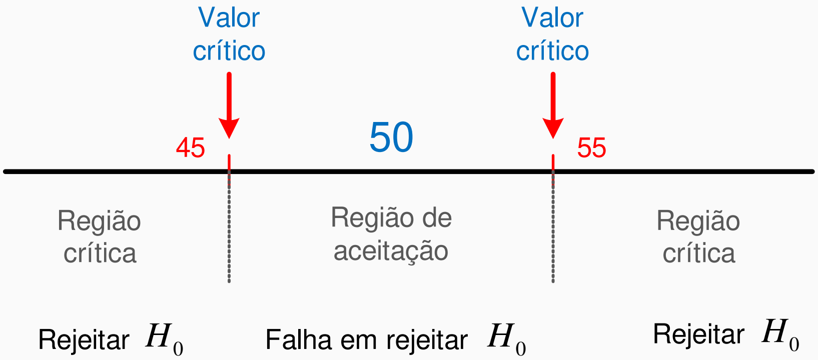
Retomando nosso exemplo... - cont

Teremos de trabalhar com **valores críticos** (limites)!

Podemos supor os seguintes limites/valores críticos:

- Se $45 < \overline{X} < 55 \rightarrow$ não rejeitamos H_0
- Se $\overline{X} < 45$ ou $\overline{X} > 55 \rightarrow$ rejeitamos H_0 em favor de H_1

Retomando nosso exemplo... - cont.



Teste de hipóteses - cont.

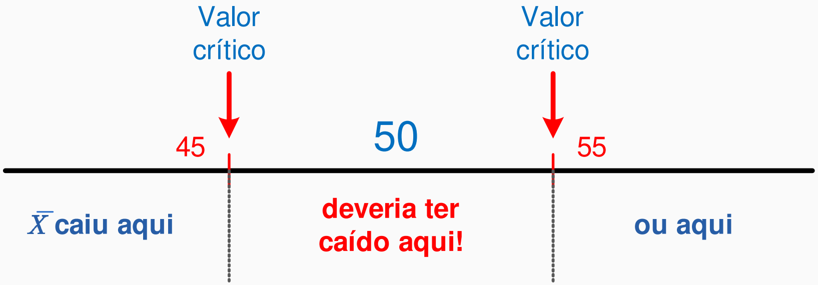
O que pode acontecer?

- ✔ O volume médio de dados **pode ser igual** a 50 Mbps, no nosso exemplo, (correspondendo a H_0), mas para a amostra, os valores médios podem estar na região crítica.
- ✔ Conclusão do teste: **Rejeitar a hipótese nula**.
- ✗ Mas: **A hipótese nula é verdadeira!**
- ✗ Logo: **Erro tipo I**.

$$\alpha = \Pr(\text{Erro I}) = \Pr(\text{Rejeitar } H_0 \text{ quando } H_0 \text{ for verdadeira})$$

⚠ α é chamado de nível de **significância**.

O que pode acontecer? - cont.

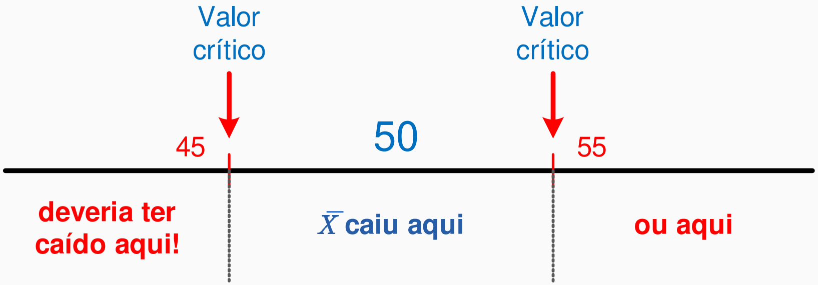


O que pode acontecer? - cont.

- ✔ O volume médio de dados **é diferente** de 50 Mbps, (correspondendo a H_0), mas para a amostra, os valores médios podem estar na região de aceitação.
- ✔ Conclusão do teste: **Não rejeita a hipótese nula**.
- ✗ Mas: **A hipótese nula é falsa!**
- ✗ Logo: **Erro tipo II**

$$\beta = \Pr(\text{Erro II}) = \Pr(\text{Não rejeitar } H_0 \text{ quando } H_0 \text{ for falsa})$$

O que pode acontecer? - cont.



Tipos de erros

Erro tipo I

Rejeitar a hipótese nula H_0 quando ela for verdadeira.

Erro tipo II

Deixar de rejeitar a hipótese nula H_0 quando ela é falsa.

Teste de hipóteses - cont.

Decisões do teste de hipótese

Decisão	H_0 é verdadeira	H_0 é falsa
Não rejeitar H_0	Não há erro	Erro tipo II
Rejeitar H_0	Erro tipo I	Não há erro

Como funciona um teste de hipótese?

1. Contexto do problema: identificar o parâmetro de interesse;
2. Estabelecer a hipótese nula H_0;
3. Especificar uma hipótese alternativa associada H_1;
4. Escolha do nível de significância α (1%, 5%, ...) e calcular o valor crítico (z_c ou t_c);
5. Estabelecer uma estatística apropriada de teste: $z, t, \ldots$;
6. Estabelecer uma região de rejeição para a estatística escolhida;
7. Calcular a grandeza amostral necessária com a estatística de teste e o seu valor associado;
8. Decidir se H_0 deve ser ou não rejeitada no contexto do problema (comparar t com t_c ou z com z_c).

Mas antes de avançarmos...

Precisamos saber quais são os valores (scores) críticos, ou seja, o que são z_c e t_c;

Os scores críticos, z_c e t_c, podem ser encontrados em suas respectivas tabelas Z (distribuição normal padrão - média 0 e variância 1) e T (t-Student), cuja consulta dependerá do teste de hipóteses que será utilizado.

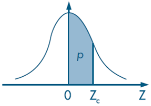
Teste de hipóteses - cont.

Tabela da distribuição normal padrão

Tabela III — Distribuição Normal Padrão

$Z \sim \mathrm{N}(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2

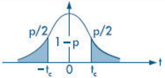
Figura 5: Tabela Z - dist. normal (Morettin e Bussab, 2014)

Tabela da distribuição normal padrão - cont.

- Na tabela, descobre-se o valor da probabilidade acumulada da distribuição normal, dado o quantil, de tal forma que a casa decimal está na coluna da esquerda da tabela e a casa centesimal está na linha superior da tabela;
- Portanto, se o teste é bilateral, por exemplo, temos os seguintes valores críticos associados a cada nível de significância (α):
  - ✔ Se $\alpha = 0,10$, então $z_c = 1,65$
  - ✔ Se $\alpha = 0,05$, então $z_c = 1,96$
  - ✔ Se $\alpha = 0,01$, então $z_c = 2,33$

Tabela da distribuição t-Student

Tabela V — Distribuição t de Student
Corpo da tabela dá os valores t_c tais que $P(-t_c < t < t_c) = 1 - p$.
Para $v > 120$, usar a aproximação normal.



Graus de liberdade v	p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%	Graus de liberdade v
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,657	318,309	636,619	1
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,598	2
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214	12,924	3
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,998	3,747	4,604	7,173	8,610	4
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,756	3,365	4,032	5,893	6,869	5
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	8
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	3,025	4,437	11
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,248	2,602	2,947	3,733	4,073	15

Figura 6: Tabela t-Student (Morettin e Bussab, 2014)

Tabela da distribuição t-Student - cont.

- Na tabela anterior, os valores da primeira coluna são os graus de liberdade e a linha representa o coeficiente de confiança;
- Portanto, se o teste é bilateral e temos 10 graus de liberdade, por exemplo, temos os seguintes valores críticos associados a cada nível de significância:
  - ✔ Se $\alpha = 0,10$, então $t_c = 1,812$
  - ✔ Se $\alpha = 0,05$, então $t_c = 2,228$
  - ✔ Se $\alpha = 0,01$, então $t_c = 3,169$

Teste de hipóteses - cont.

Entendida a parte sobre consulta às tabelas das distribuições, podemos avançar para o estudo dos casos particulares dos testes de hipóteses.

Os testes de hipóteses mais comuns

1. Teste para a média (*com variância conhecida*)
2. Teste para a média (*com variância desconhecida*)
3. Teste para a variância
4. Teste para a proporção

Caso 1: Teste para a média com variância conhecida

Contexto: neste caso estamos interessados em realizar inferência sobre a média populacional μ, com base na amostra $X_1, X_2, \ldots, X_n$ de uma população com distribuição normal com variância **conhecida**;

Caso 1 - cont.

Passo 1 - Definir as hipóteses:

- $H_0 : \mu = \mu_0$;
- $H_1 : \mu < \mu_0$ ou $H_1 : \mu \neq \mu_0$ ou $H_1 : \mu > \mu_0$;

Passo 2 - Definir o estimador e a estatística do problema:

$$\boxed{Z_{\text{obs}} = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1);} \tag{93}$$

Passo 3 - Sob H_0 como verdade, fixar o erro do tipo I e calcular a região crítica;

Caso 1 - cont.

Passo 4 - De acordo com as informações da amostra, vamos calcular o valor da estatística do teste e identificar se esse valor pertence ou não à região crítica (se pertence, rejeita-se H_0, caso contrário não rejeita-se);

- Teste bilateral: se $Z_{\text{obs}} > z_{\alpha/2}$ ou se $Z_{\text{obs}} < -z_{\alpha/2}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
- Teste unilateral à direita: se $Z_{\text{obs}} > z_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
- Teste unilateral à esquerda: se $Z_{\text{obs}} < -z_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.

Caso 2: Teste para a média com variância desconhecida

Contexto: neste caso estamos interessados em realizar inferência sobre a média populacional μ, com base na amostra $X_1, X_2, \ldots, X_n$ de uma população com distribuição normal com variância **desconhecida**.

Caso 2 - cont.

Passo 1 - Definir as hipóteses:

- $H_0 : \mu = \mu_0$;
- $H_1 : \mu < \mu_0$ ou $H_1 : \mu \neq \mu_0$ ou $H_1 : \mu > \mu_0$

Passo 2 - Definir o estimador e a estatística do teste:

$$\boxed{T_{\text{obs}} = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t(n-1)} \tag{94}$$

Passo 3 - Sob H_0 como verdade, fixar o erro do tipo I e calcular a região crítica.

Caso 2 - cont.

- **Passo 4** - De acordo com as informações da amostra, vamos calcular o valor da estatística do teste e identificar se esse valor pertence ou não à região crítica (se pertence, rejeita-se H_0, caso contrário não rejeita-se);
  - Teste bilateral: se $T_{\text{obs}} > t_{\alpha/2}$ ou se $T_{\text{obs}} < -t_{-\alpha/2}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
  - Teste unilateral à direita: se $T_{\text{obs}} > t_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
  - Teste unilateral à esquerda: se $T_{\text{obs}} < -t_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.

Caso 3: Teste para a variância

Contexto: neste caso estamos interessados em realizar inferência sobre a variância populacional σ^2, com base na amostra $X_1, X_2, \ldots, X_n$ de uma população com distribuição normal.

Caso 3 - cont.

Passo 1 - Definir as hipóteses:

- $H_0 : \sigma = \sigma_0$;
- $H_1 : \sigma < \sigma_0$ ou $H_1 : \sigma \neq \sigma_0$ ou $H_1 : \sigma > \sigma_0$;

Passo 2 - Definir o estimador e a estatística do teste:

$$\boxed{Q_{\text{obs}} = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1)} \tag{95}$$

Passo 3 - Sob H_0 como verdade, fixar o erro do tipo I e calcular a região crítica;

Teste de hipóteses - cont.

Caso 3 - cont.

Passo 4 - De acordo com as informações da amostra, vamos calcular o valor da estatística do teste e identificar se esse valor pertence ou não à região crítica (se pertence, rejeita-se H_0, caso contrário não rejeita-se);

- Teste bilateral: se $Q_{\text{obs}} > q_{\alpha/2}$ ou se $Q_{\text{obs}} < -q_{-\alpha/2}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
- Teste unilateral à direita: se $Q_{\text{obs}} > q_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
- Teste unilateral à esquerda: se $Q_{\text{obs}} < -q_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.

Caso 4: Teste para a proporção

Contexto: neste caso estamos interessados em realizar inferência sobre a proporção p de indivíduos/objetos com certa característica, com base na amostra $X_1, X_2, \ldots, X_n$ de uma população.

Caso 4 - cont.

Passo 1 - Definir as hipóteses:

- $H_0 : p = p_0$;
- $H_1 : p < p_0$ ou $H_1 : p \neq p_0$ ou $H_1 : p > p_0$;

Passo 2 - Definir o estimador e a estatística do teste (neste caso, usamos o TCL):

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right).$$

Assim:

$$\boxed{Z_{\text{obs}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim \mathcal{N}(0,1)} \qquad (96)$$

Caso 4 - cont.

Passo 3 - Sob H_0 como verdade, fixar o erro do tipo I e calcular a região crítica;

Passo 4 - De acordo com as informações da amostra, vamos calcular o valor da estatística do teste e identificar se esse valor pertence ou não à região crítica (se pertence, rejeita-se H_0, caso contrário não rejeita-se);

- Se o teste é bilateral e $Z_{\text{obs}} > Z_{\alpha/2}$ ou $Z_{\text{obs}} < -Z_{\alpha/2}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
- Se o teste é unilateral à direita e $Z_{\text{obs}} > Z_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.
- Se o teste é unilateral à esquerda e $Z_{\text{obs}} < -Z_{\alpha}$, rejeitamos H_0. Caso contrário, não rejeitamos H_0.

Teste de hipóteses - cont.

Outros testes de hipóteses comuns são:

A. Testes de aderência
B. Teste de homogeneidade
C. Teste de independência.

Caso A: Teste de aderência

Contexto: queremos testar se uma população P segue, por exemplo, uma distribuição normal (ou qualquer outra). Nesse caso, temos que testar as seguintes hipóteses:

$$\begin{aligned} H_0 &: \quad P = P_0 \\ H_1 &: \quad P \neq P_0 \end{aligned}$$

Caso A - cont.

A estatística do teste utilizada e sua distribuição é dada por:

$$\boxed{\chi^2 = \sum_{i=1}^{l} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(n-1),} \tag{97}$$

em que l é o número de categorias da variável em estudo.

Caso B: Teste de homogeneidade

Contexto: esse teste é utilizado, geralmente, quando temos o interesse em testar se dois ou mais grupos diferem ou não com relação a alguma característica.

Para um caso particular de dois grupos a serem testados, temos:

$$\begin{aligned} H_0 &: \quad P_1 = P_2 \\ H_1 &: \quad P_1 \neq P_2 \end{aligned}$$

Caso B - cont.

A estatística do teste utilizada e sua distribuição é dada por:

$$\boxed{\chi^2 = \sum_{i=1}^{l} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((l-1).(c-1))} \tag{98}$$

sendo l e c o número de categorias de cada variável em estudo.

Caso C: Teste de independência

Contexto: suponha duas variáveis aleatórias X e Y. Suponha também que seus valores amostrais podem ser classificados segundo categorias, obtendo-se uma tabela de dupla entrada. O objetivo, portanto, é testar se X e Y são independentes.

Assim, para $p_{i.} = \sum_{j=1}^{l} p_{ij}$ e $p_{.j} = \sum_{i=1}^{c} p_{ij}$, sendo l e c o número de categorias de cada variável, temos o seguinte teste:

$$\begin{aligned} H_0 : &\quad p_{ij} = p_{i.}p_{.j} \quad \text{para todo par } (i,j). \\ H_1 : &\quad p_{ij} \neq p_{i.}p_{.j} \quad \text{para algum par } (i,j) \end{aligned}$$

Caso C - cont.

A estatística do teste utilizada e sua distribuição é dada por:

$$\boxed{\chi^2 = \sum_{i=1}^{l} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((l-1).(c-1))} \tag{99}$$

sendo l e c o número de categorias de cada variável em estudo.