Estatística e Probabilidade

para Ciência da Computação

Código: MAD243

Oferecido por:

Prof. Hugo Carvalho (hugo@dme.ufrj.br)

Departamento de Métodos Estatísticos - DME

Instituto de Matemática - IM/UFRJ

História

Recomendação de leitura: L. Mlodinow - *O andar do bêbado: Como o acaso determina nossas vidas*

- Por séculos, acreditou-se que deuses regiam eventos incertos, que estavam além da compreensão humana
- Gerolamo Cardano (1501–1575), físico e viciado em apostas italiano: primeiro estudo sistemático de probabilidades, inspirado por jogos de azar
- Definiu probabilidade como "eventos favoráveis"/"eventos totais", óbvio hoje em dia porém um grande avanço na sua época
- 1654, início do estudo de Probabilidade, segundo historiadores: correspondências entre Pierre de Fermat (1601–1665) e Blaise Pascal (1623–1662). Abordagem mais sistemática, ainda inspirada em jogos de azar!

História

- 1657, astrônomo holandês Christiaan Huygens (1629–1695) tomou ciência dessas correspondências, introduziu o conceito de valor esperado e aprimorou a abordagem deles
- 1713, matemático suíço Jakob Bernoulli (1654–1705) publica *Ars Conjecturandi*, primeira teoria geral para cálculo de probabilidades
- 1812, matemático francês Pierre Simon Laplace (1749–1827) publica *Théorie Analytique des Probabilités*. Aplica ideias de probabilidade a problemas práticos. Grande avanço na Probabilidade e Estatística.
- Porém, até então sabia-se *calcular* probabilidades, e não o que *é* probabilidade!
- Definição aceitável de Probabilidade, 1933 com o matemático russo Andrey Nikolaevich Kolmogorov (1903–1987). Seus axiomas serão o ponto de partida para nosso curso!

Aplicações

Quantificação de incertezas!

- Mercado financeiro
- Companhia de seguros
- Epidemiologia
- Até juízes e médicos podem se valer de Probabilidade para tomar melhores decisões!
  - C. Colmez, L. Schneps - *A matemática nos tribunais: Uso e abuso dos números em julgamentos*
  - S. Senn - *Dicing with Death: Chance, Risk And Health*

Modelos determinísticos e probabilísticos

- **Modelo determinístico**: Pode ser descrito facilmente através de uma fórmula
  Exemplo: Corpo em queda livre no vácuo. A velocidade final, em cm/s, é dada por $v = \sqrt{2gh}$, onde g é a aceleração da gravidade (em cm/s^2) e h é a altura inicial (em cm).
- **Modelo probabilístico**: Sabemos os possíveis resultados do experimento, mas não sabemos (ou é extremamente difícil) precisá-lo.
  Exemplo: Lançamento de uma moeda ou lançamento de um dado.

Interpretações de probabilidade

1) **Interpretação frequentista**: *"Ao se jogar uma moeda honesta um grande número de vezes, em condições similares, observaremos aproximadamente metade cara e metade coroa"*
2) **Interpretação clássica**: *"Se um processo tem n possíveis resultados, então podemos assumir que todos são igualmente prováveis"*
3) **Interpretação subjetiva**: *"A chance de chover amanhã é de 70%. Para mim, isso é o suficiente para que eu saia de casa com um guarda-chuvas"*

A Teoria de Probabilidades **não** depende da escolha de interpretação!

Conceitos básicos

- **Experimento**: Qualquer processo, real ou hipotético, no qual os resultados podem ser identificados ao longo do tempo
- **Espaço amostral**: É a coleção de todos os possíveis resultados de um experimento. Usualmente denotado pela letra Ω. Elementos do espaço amostral são usualmente denotados pela letra ω
- **Eventos**: São sub-conjuntos do espaço amostral

Conceitos básicos

Exemplo: Uma rede de computadores está em operação contínua, mas pode sofrer avaria a qualquer momento. Na ocorrência de falha, o tempo de colocar a rede novamente em operação depende de vários fatores envolvendo a extensão e a causa da falha, entre outras.

- **Experimento**: Observar número de falhas em um dia
- **Espaço amostral**: $\Omega = \{0, 1, 2, 3, \dots\}$
- **Experimento**: Observar a hora do dia na qual a primeira falha ocorre
- **Espaço amostral**: $\Omega = \{\omega \in \mathbb{R} \mid 0 \leq \omega \leq 24\} = [0, 24]$

Exemplo

Experimento: Selecione uma carta de um baralho de 52 cartas

Espaço amostral:

$$\Omega = \{\mathrm{A}\spadesuit, \ldots, K\spadesuit, \mathrm{A}\heartsuit, \ldots, K\heartsuit, \mathrm{A}\diamondsuit, \ldots, K\diamondsuit, \mathrm{A}\clubsuit, \ldots, K\clubsuit\}$$

Eventos:

“A carta é Ás”	$A = \{\mathrm{A}\spadesuit, \mathrm{A}\heartsuit, \mathrm{A}\diamondsuit, \mathrm{A}\clubsuit\}$
“A carta é de cor escura”	$B = \{\mathrm{A}\spadesuit, \ldots, K\spadesuit, \mathrm{A}\clubsuit, \ldots, K\clubsuit\}$
“A carta é de copas”	$C = \{\mathrm{A}\heartsuit, \ldots, K\heartsuit\}$
“A carta é de cor branca”	$E = B^c$ ⚠

Quantos eventos podemos construir para esse espaço amostral?

$$2^{52} \approx 4.5 \times 10^{15} \text{ eventos!!}$$

Exemplo

Experimento: Lançamento de 10 moedas

$$Cara \leftrightarrow 0 \text{ e } Coroa \leftrightarrow 1$$

Espaço amostral:

$$\begin{aligned}\Omega &= \{\omega = (\omega_1, \omega_2, \ldots, \omega_{10}) \mid \omega_i \in \{0,1\} \text{ para } 1 \leq i \leq 10\} \\ &= \{0,1\}^{10}\end{aligned}$$

Eventos:

“o j-ésimo lançamento foi coroa” $A_j = \{\omega \in \{0,1\}^{10} : \omega_j = 1\}$

“pelo menos um lançamento foi coroa” $B = \bigcup_{j=1}^{10} A_j$ ⚠

“todos lançamentos foram cara” $C = \{(0,0,\ldots,0)\} = B^c$

Operação em conjuntos

Sejam Ω espaço amostral e $A, B \subseteq \Omega$ eventos em Ω.

- *Conjunto vazio*: $\emptyset$
- *Complementar de A*: $A^c = \{\omega \in \Omega \mid \omega \notin A\}$
- *União de A e B*: $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ ou } \omega \in B\}$
- *Interseção de A e B*: $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ e } \omega \in B\}$
- *Diferença de A e B*: $A - B = A \setminus B = A \cap B^c$
- *Diferença simétrica entre A e B*: $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$
- A e B são *disjuntos* se $A \cap B = \emptyset$
- Dizemos que A está *contido* em B, denotado por $A \subset B$ se todo elementos de A também é um elemento de B
- O conjunto de todos os sub-conjuntos de Ω é chamado de *conjunto das partes* e é denotado por 2^{Ω}. Porque?!

Operação em conjuntos

Se $A_1, \ldots, A_n$ são eventos em Ω:

- *União de $A_1, \ldots, A_n$*: $A_1 \cup \cdots \cup A_n = \displaystyle\bigcup_{i=1}^{n} A_i$
- *Interseção de $A_1, \ldots, A_n$*: $A_1 \cap \cdots \cap A_n = \displaystyle\bigcap_{i=1}^{n} A_i$
- $A_1, \ldots, A_n$ formam uma *partição* de Ω se são disjuntos 2 a 2 e a união deles é Ω, ou seja, $A_i \cap A_j = \emptyset$, se $i \neq j$, e $\displaystyle\bigcup_{i=1}^{n} A_i = \Omega$

Diagramas de Venn

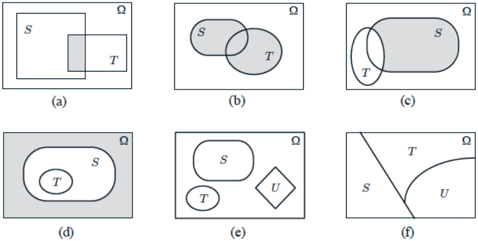


Figure 1: (Bertsekas, D.P. e Tsitsiklis, J.N. Introduction to probability, p.5.)

Em todos os diagramas Ω é o espaço amostral, e S, T e U são eventos.

a) A região sombreada é $S \cap T$
b) A região sombreada é $S \cup T$
c) A região sombreada é $S \cap T^c$
d) Aqui $T \subset S$. A região sombreada é S^c
e) Os conjuntos T, S e U são *disjuntos* (ou *mutuamente exclusivos*)
f) Os conjuntos T, S e U formam uma *partição* de Ω

Algumas propriedades úteis

- Para subconjuntos $A, A_1, A_2, \dots \subseteq \Omega$ vale que:
  - $A^c \cup A = \Omega$
  - $A \cap A^c = \emptyset$
  - $(A^c)^c = A$
  - $A \cup \Omega = \Omega$
  - $\Omega \cup A = \Omega$
  - $A \cap \left(\bigcup_i A_i\right) = \bigcup_i (A \cap A_i)$
  - $A \cup \left(\bigcap_i A_i\right) = \bigcap_i (A \cup A_i)$
- Leis de de Morgan: se $A_1, A_2, \dots, A_n, \dots$ são eventos em Ω, então

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c \quad \text{e} \quad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$$

⚠ [1]:Nem todos os subconjuntos de Ω podem ser eventos

- Lançamento de uma moeda
- Espaço amostral $\Omega = \{K, C\}$
- Se a moeda é honesta, sabemos calcular todas essas probabilidades:
  - $\mathbb{P}(\emptyset) = 0$
  - $\mathbb{P}(\{K\}) = \mathbb{P}(\{C\}) = \frac{1}{2}$
  - $\mathbb{P}(\{K, C\}) = 1$
- Ou seja, sabemos calcular $\mathbb{P}(A)$, para todo $A \subset 2^{\Omega}$.
- Porém, se não sabemos se a moeda é honesta, não sabemos calcular $\mathbb{P}(\{K\})$ nem $\mathbb{P}(\{C\})$!
- Nesse cenário, só sabemos calcular $\mathbb{P}(\emptyset) = 0$ e $\mathbb{P}(\Omega) = 1$.

⚠ [2]:Nem todos os subconjuntos de Ω podem ser eventos

- Questionário para saber idade de indivíduos
- Espaço amostral $\Omega = [0, \infty)$
- Resposta da forma: "entre 0 e 18 anos", "entre 18 e 25 anos", "entre 25 e 34 anos", etc...
- Matematicamente, $[0, 18)$, $[18, 25)$, $[25, 34)$, etc...
- Nos permite inferir informações sobre intervalos da forma $[18, 25)$, $[0, 18) \cup [25, 34)$, etc...
- Porém, nada podemos dizer sobre o intervalo $[20, 30)$, por exemplo!

$\Rightarrow$ Sub-conjuntos de 2^{Ω} que são considerados eventos devem codificar informações conhecidas sobre o experimento em questão.

⚠ [3]:Nem todos os subconjuntos de Ω podem ser eventos

- Razão bastante técnica...
- Nem todos os sub-conjuntos de $\mathbb{R}^n$ têm "volume" bem definido!
- **Paradoxo (Teorema) de Banach-Tarski**: Dada uma bola sólida de raio 1 em $\mathbb{R}^3$, existe uma decomposição dela em um número finito de conjuntos disjuntos que, se adequadamente reordenados, geram duas bolas sólidas de raio 1!

$\Rightarrow$ Sub-conjuntos de 2^Ω que são considerados eventos devem ser minimamente razoáveis!

Solução: σ-álgebra de eventos. Ponto extremamente sutil e técnico da Teoria de Probabilidades. Não iremos mais à fundo nisso.

Noções primitivas de probabilidade

- Ω finito com elementos equiprováveis (e.g., lançamento de dado honesto), $A \subset \Omega$ evento $\Rightarrow$

$$\mathbb{P}(A) = \frac{\text{número de elementos em } A}{\text{número de elementos em } \Omega} = \frac{\text{“eventos favoráveis”}}{\text{“eventos totais”}}$$

  Técnicas de combinatória e contagem

- Ω intervalo de $\mathbb{R}$, distribuição uniforme, $A \subset \Omega$ intervalo $\Rightarrow$

$$\mathbb{P}(A) = \frac{\text{comprimento de } A}{\text{comprimento de } \Omega}$$

- Seja n_A número de ocorrência de A em n repetições independentes do experimento em questão $\Rightarrow$

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

Apesar de intuitivas, essas definições nos dizem como *calcular* probabilidades, não o que *é* probabilidade!

Exemplo

Se jogarmos dois dados honestos, qual dos resultados é mais provável: a soma dê 11 ou a soma dê 12?

Solução: O espaço amostral do lançamento de dois dados é

$$\begin{aligned}\Omega &= \{\omega = (\omega_1, \omega_2) \mid \omega_i \in \{1, \dots, 6\} \text{ para } 1 \leq i \leq 2\} \\ &= \{1, \dots, 6\}^2\end{aligned}$$

Para cada resultado individual possível $\omega \in \Omega$:

$$\mathbb{P}(\omega) = \mathbb{P}(\{\omega\}) = \frac{1}{36}$$

Os eventos de interesse são:

- "A soma de 11": $A = \{(5, 6), (6, 5)\}$
- "A soma de 12": $B = \{(6, 6)\}$

Logo, $\mathbb{P}(A) = 2/36 > 1/36 = \mathbb{P}(B)$

Exemplo

Full House em uma mão de Poker: Uma mão de 5 cartas é extraída de um baralho de 52 cartas. Se o baralho foi embaralhado de forma que cada carta tenha a mesma chance de aparecimento, qual é a probabilidade de se obter um Full House?
Obs.: Uma mão é chamada Full House se há 3 cartas do mesmo valor e 2 outras cartas do mesmo valor.

Solução: Seja A o evento "a mão é um Full House". Temos que

$$|A| = 13\binom{4}{3}12\binom{4}{2}$$

logo,

$$\mathbb{P}(A) = \frac{13\binom{4}{3}12\binom{4}{2}}{\binom{52}{5}} \approx 0.00144$$

Exemplo

O problema do aniversário: Existem k pessoas numa sala. Supondo que o aniversário de cada pessoa é igualmente provável e que não há gêmeos na sala, qual a probabilidade de que duas ou mais pessoas celebrem o aniversário no mesmo dia?

Solução: Se A denota o evento "nenhuma das k pessoas fazem aniversário no mesmo dia", então queremos calcular $\mathbb{P}(A^c)$. Nesse exemplo,

$$\Omega = \{1, \ldots, 365\}^k$$

Note que $|\Omega| = 365^k$ e

$$|A| = \frac{365!}{(365-k)!} = 365 \cdot 364 \cdots (365 - k + 1)$$

Logo,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - \frac{|A|}{|\Omega|} = 1 - \frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k}$$

Exemplo (continuação)

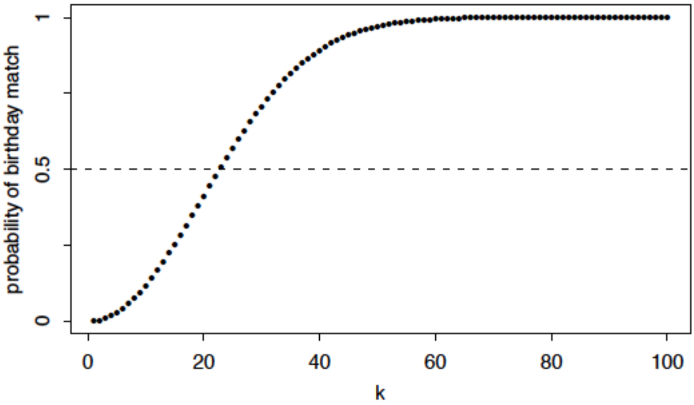


Figure 2: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.12.) Probabilidade de que num grupo de k pessoas duas ou mais celebrem aniversário no mesmo dia. Essa probabilidade excede $1/2$ para $k = 23$

Observação

Note que para Ω infinito, a definição clássica não se aplica

Exemplos de espaços amostrais Ω infinitos incluem:

- tempo de vida útil de um aparelho eletrônico: $\Omega = [0, \infty)$
- o número de clientes atendidos em uma agência bancária durante um período de tempo fixado: $\Omega = \{0, 1, 2, 3, \dots\}$

Como definir probabilidade para espaço amostral Ω infinito?

Axiomas de Kolmogorov

Seja Ω um espaço amostral. Uma probabilidade é uma função $\mathbb{P}$ que atribui valores reais entre 0 e 1 a eventos $A \subset \Omega$, satisfazendo:

1) $\mathbb{P}(\Omega) = 1$
2) Para todo evento $A \subset \Omega$, vale que $\mathbb{P}(A) \geq 0$
3) Para toda sequência $A_1, A_2, \cdots \subset \Omega$ de eventos disjuntos temos

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Um exemplo simples

Um dado honesto é lançado duas vezes. Considere os eventos:

$$A = \{\text{“a soma dos resultados é ímpar”}\}$$
$$B = \{\text{“o resultado do primeiro lançamento é ímpar”}\}$$
$$C = \{\text{“o produto dos resultados é ímpar”}\}.$$

Construa um espaço de probabilidade adequado para esse problema e calcule as probabilidades dos eventos acima.

Um exemplo não tão simples...

O "Paradoxo" de Bertrand: Considere um triângulo equilátero inscrito em um círculo unitário. Suponha que uma corda é escolhida ao acaso. Qual é a probabilidade de que a corda escolhida seja maior que o lado do triângulo ($\sqrt{3}$)?

- Para cada solução apresentada, temos um espaço amostral Ω diferente
- Não há paradoxo algum! A noção de "escolhida ao acaso", por si só, é mal formulada e ambígua!

Propriedades da probabilidade - 1

Relembremos os três axiomas de Probabilidade:

1) $\mathbb{P}(\Omega) = 1$
2) $A \subset \Omega \Rightarrow \mathbb{P}(A) \geq 0$
3) $A_1, A_2, \cdots \subset \Omega$ disjuntos $\implies \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Teorema: Seja $\mathbb{P}$ uma medida de probabilidade em um espaço amostral Ω e sejam $A, B, A_1, A_2, \cdots \subset \Omega$ eventos. Temos que:

1) $\mathbb{P}(\emptyset) = 0$
2) Se $A_1, \ldots, A_n$ são disjuntos, então $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$
3) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
4) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
5) $0 \leq \mathbb{P}(A) \leq 1$
6) $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$
7) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Propriedades da probabilidade - 2

Teorema (Desigualdades de Bonferoni): Se $A_1, A_2, \dots \subset \Omega$ não são necessariamente disjuntos, então

- $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
- $\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) \geq 1 - \sum_{i=1}^{\infty} \mathbb{P}(A_i^c)$

Probabilidade condicional

Intuição:

- Ao acordar, antes de olhar pela janela, a probabilidade de chover no dia é de 50%
- Porém, após ouvir um trovão, provavelmente essa probabilidade será modificada

$\Rightarrow$ Como que informações preliminares podem alterar a probabilidade de eventos de interesse?

- Experimento associado ao espaço amostral Ω e à medida de probabilidade $\mathbb{P}$
- É sabida a ocorrência de um elemento $\omega \in B \subset \Omega$
- Como incorporar essa informação na medida de probabilidade de modo adequado?

Probabilidade condicional

Considere uma probabilidade $\mathbb{P}$ definida em um espaço amostral Ω. Para eventos A e B com $\mathbb{P}(B) > 0$, a *probabilidade condicional* de A dado B é definida como

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

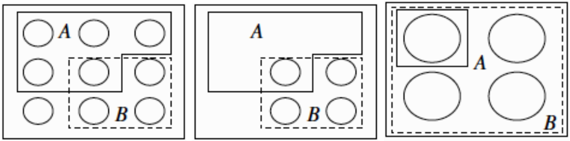


Figure 3: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.44.) Intuição para $\mathbb{P}(A|B)$. Da esquerda para direita: a) Eventos A e B do espaço amostral. b) Como sabemos que B ocorreu, descartamos todos resultados em B^c. c) No espaço amostral restrito, renormalizamos de modo que a massa total seja 1

Exemplo

Duas cartas são escolhidas aleatoriamente, uma de cada vez e sem reposição, de um baralho de 52 cartas (dos tipos α, β, γ e δ; cada tipo numeradas de 1 até 13). Seja A o evento "a primeira carta é de copas (do tipo α)" e B o evento "a segunda carta é vermelha (do tipo α ou β)". Determine $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$

Solução: Note que:

- $\mathbb{P}(A \cap B) = \dfrac{13 \times 25}{52 \times 51}$
- $\mathbb{P}(A) = \dfrac{1}{4}$
- $\mathbb{P}(B) = \dfrac{26}{52}$

Logo, temos $\mathbb{P}(A|B) = 25/102$ e $\mathbb{P}(B|A) = 25/51$

Note que $\mathbb{P}(B|A)$ é $2 \times \mathbb{P}(A|B)$

Exemplo

Um paciente é testado para uma doença que atinge 1% da população. O resultado do teste é positivo. Seja D o evento "o paciente testado tem a doença" e T o evento "o resultado do teste para esse paciente é positivo". Suponha que a precisão do teste é 95%, ou seja, $\mathbb{P}(T|D) = 0.95 = \mathbb{P}(T^c|D^c)$. Calcule $\mathbb{P}(D|T)$.

Ainda não sabemos responder a tal pergunta!
Em breve saberemos.

Teorema: (Regra da multiplicação) Se A e B são eventos com probabilidade positiva, então

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

Mais geralmente, se $A_1, \ldots, A_n$ são eventos tais que $\left(\bigcap_{i=1}^{k} A_i\right)$ tem probabilidade positiva para todo $1 \leq k \leq (n-1)$, então

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\ldots\mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$$

Exemplo

Três cartas são extraídas, sem reposição, de um baralho com 52 cartas (dos tipos α, β, γ e δ; cada tipo numeradas de 1 até 13). Assumindo que em cada extração, cada uma das cartas restantes têm a mesma chance de ser escolhida, determine a probabilidade de nenhuma das três cartas seja de copas (do tipo α)

Solução: Defina para cada $i = 1, 2, 3$, o evento

$$A_i = \text{“a } i\text{-ésima carta não é de copas (do tipo } \alpha\text{)”}$$

Note que $\mathbb{P}(A_1) = 39/52$, $\mathbb{P}(A_2|A_1) = 38/51$ e $\mathbb{P}(A_3|A_1 \cap A_2) = 37/50$. Logo, pela regra da multiplicação,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) = \frac{39}{52}\frac{38}{51}\frac{37}{50} \approx 0{,}41.$$

Lei da Probabilidade Total

Teorema: (Lei da Probabilidade Total) Considere $A_1, \ldots, A_n$ formam uma partição de Ω tal que $\mathbb{P}(A_i) > 0$ para todo $1 \leq i \leq n$. Para todo evento B, temos que

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)$$

Lembrete: Os eventos $A_1, \ldots, A_n$ formam uma *partição* de Ω se são disjuntos 2 a 2 e a união deles é Ω, ou seja, $A_i \cap A_j = \emptyset$, se $i \neq j$, e $\bigcup_{i=1}^{n} A_i = \Omega$

Exemplo

João decide participar de um torneio de xadrez, onde a metade dos jogadores é profissional (tipo 1), um quarto é semi-profissional (tipo 2) e um quarto é amador (tipo 3). Suponha que as probabilidades do João vencer uma partida sejam 0,3, 0,4 e 0,5, se o adversário for respectivamente do tipo 1, tipo 2 e tipo 3. Se um adversário é escolhido ao acaso, determine a probabilidade do João sair vitorioso

Solução: Defina para cada $i = 1, 2, 3$, o evento

$$A_i = \text{“o adversário escolhido é do tipo } i\text{”}.$$

Temos que $\mathbb{P}(A_1) = 0{,}5$, $\mathbb{P}(A_2) = 0{,}25$ e $\mathbb{P}(A_3) = 0{,}25$. Defina B o evento “João saiu vitorioso”. Temos que $\mathbb{P}(B|A_1) = 0{,}3$, $\mathbb{P}(B|A_2) = 0{,}4$ e $\mathbb{P}(B|A_3) = 0{,}5$ Assim, pela lei da probabilidade total

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) + \mathbb{P}(A_3)\mathbb{P}(B|A_3) = 0{,}375.$$

Exemplo

Uma moeda honesta é lançada 100 vezes. Qual a probabilidade de observarmos exatamente 50 caras?

Exemplo

Alguém rola dois dados honestos sem que você veja. Você pergunta a esta pessoa se houve um seis entre os dois lançamentos e ela responde, honestamente, “sim”. Qual é a probabilidade que os dois dados rolaram seis?

Exemplo

Em um jogo, dois jogadores, A e B, alternam rolagens de um dado honesto, de acordo com as seguintes regras:

1) O jogador A rola primeiro, e ganha se cair um seis.
2) Se para A não cair um seis, então B rola o dado e ganha se cair cinco ou seis.
3) Se B não ganhar (ou seja, não cair cinco nem seis), A joga o dado novamente e ganha ao observar quatro, cinco ou seis.
4) O jogo continua com a mesma lógica, até alguém ganhar.

Qual é a probabilidade do jogador A ganhar?

Exemplo

Um baralho de 52 cartas (com cartas dos tipos α, β, γ e δ; cada tipo numeradas de 1 até 13) é embaralhado aleatoriamente, e cartas são retiradas sequencialmente. Qual a probabilidade da primeira carta ser um Ás de copas (a carta $\alpha 1$)? E a probabilidade dessa carta ser a segunda? E a última?

⚠ Para casa

Considere o seguinte jogo com uma moeda viciada, cuja probabilidade de cair “cara” é um número p satisfazendo $0 < p < 1$: Ana, Bia e Carlos lançam a moeda nessa ordem (primeiro Ana, depois Bia e finalmente Carlos), até alguém obter “cara”, sendo essa pessoa a vencedora. Determine em função de p a probabilidade de cada um dos participantes ganhar o jogo. Quem tem mais chances de ganhar o jogo? Porque?

Teorema de Bayes

Teorema: (Teorema de Bayes) Se A e B são eventos com probabilidade positiva, então

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

Além disso, se $B_1, \ldots, B_n$ formam uma partição de Ω tal que $\mathbb{P}(B_j) > 0$ para todo $1 \leq j \leq n$, então

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \ldots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)}.$$

Interpretação do Teorema de Bayes

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k=1}^{n} \mathbb{P}(A|B_k)\mathbb{P}(B_k)}$$

- Cada B_j representa uma causa (fator de influência) no resultado de um experimento aleatório
- Probabilidades *a priori* de ocorrências $\mathbb{P}(B_j)$, para $j = 1, \ldots n$
- Evento A é observado na realização do experimento
- Reavaliar quanto cada fator B_j é responsável pela ocorrência do evento A
- Probabilidades *a posteriori* $\mathbb{P}(B_j|A)$, para $j = 1, \ldots, n$

Retornando ao exemplo do paciente testado...

Um paciente é testado para uma doença que atinge 1% da população. O resultado do teste é positivo. Seja D o evento "o paciente testado tem a doença" e T o evento "o resultado do teste para esse paciente é positivo". Suponha que a precisão do teste é 95%, ou seja, $\mathbb{P}(T|D) = 0.95 = \mathbb{P}(T^c|D^c)$. Calcule $\mathbb{P}(D|T)$.

Solução: Pelo Teorema de Bayes e a Lei da Probabilidade total:

$$\mathbb{P}(D|T) = \frac{\mathbb{P}(D)\mathbb{P}(T|D)}{\mathbb{P}(T)} = \frac{\mathbb{P}(D)\mathbb{P}(T|D)}{\mathbb{P}(D)\mathbb{P}(T|D) + \mathbb{P}(D^c)\mathbb{P}(T|D^c)} \approx 0{,}16.$$

Exemplo: A Falácia do Procurador

Em 1999, Sally Clark foi condenada pelo assassinato dos seus dois filhos, ambos recém nascidos. A decisão foi tomada com base no depoimento de um especialista. Segundo ele, a probabilidade de uma bebê recém nascido morrer da síndrome de morte súbita infantil (SMSI) era de 1/8500, de forma que a probabilidade de duas mortes devida a SMSI na mesma família era $(1/8500)^2 \approx 1/(73 \times 10^6)$. Com base nisso, o especialista conclui que a probabilidade da inocência de Clark era de $1/(73 \times 10^6)$. Qual é o problema com a linha de raciocínio do especialista?

- Hipótese de independência não justificada
- Usou incorretamente $\mathbb{P}(\text{“evidência”}|\text{“inocência”})$ ao invés de $\mathbb{P}(\text{“inocência”}|\text{“evidência”})$.

Mensagem: não confunda $\mathbb{P}(A|B)$ com $\mathbb{P}(B|A)$!

Fonte: Coralie Colmez & Leila Schneps: “A matemática nos tribunais: Uso e abuso dos números em julgamentos”

Exemplo

Considere duas moedas, sendo uma honesta e a outra viciada, cuja probabilidade de cara é $3/4$. Escolhemos uma das moedas ao acaso e então efetuamos três lançamentos. Em todos os lançamentos o resultado foi cara. Dado essa informação, qual é a probabilidade de a moeda honesta ter sido escolhida?

Solução: Sejam A o evento "em todos os lançamentos o resultado foi cara" e F o evento "a moeda honesta foi escolhida". Queremos calcular $\mathbb{P}$. É imediato calcular $\mathbb{P}(A|F)$ e $\mathbb{P}(A|F^c)$

$$\mathbb{P}(A|F) = 1/8 = (1/2)^3 \text{ e } \mathbb{P}(A|F^c) = (3/4)^3$$

Também temos que $\mathbb{P}(F) = \mathbb{P}(F^c) = 1/2$. Logo, pelo Teorema de Bayes e a Lei da Probabilidade total:

$$\mathbb{P}(F|A) = \frac{\mathbb{P}(A|F)\mathbb{P}(F)}{\mathbb{P}(A|F)\mathbb{P}(F) + \mathbb{P}(A|F^c)\mathbb{P}(F^c)} \approx 0{,}23.$$

Exemplo

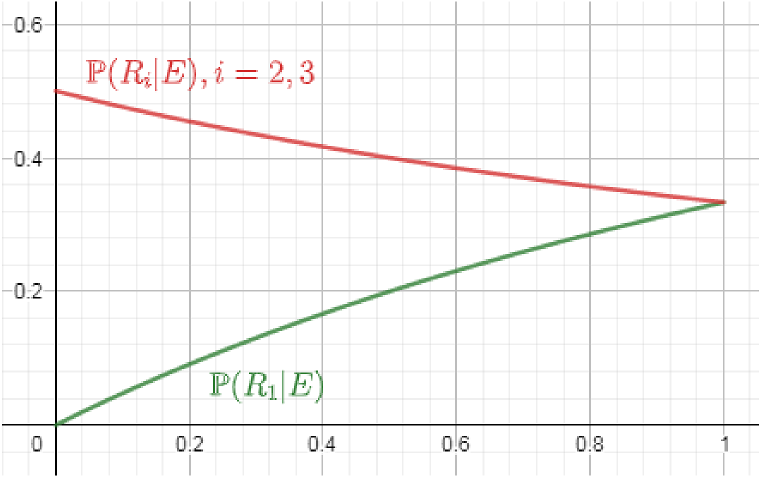
Uma caixa contém cinco moedas, com respectivas probabilidades p_i de dar cara, a saber: $p_1 = 0, p_2 = 1/4, p_3 = 1/2, p_4 = 3/4$ e $p_5 = 1$. Ao escolher aleatoriamente uma moeda da caixa e lançá-la, observa-se uma cara. Qual é a probabilidade de que a i-ésima moeda foi selecionada? Se a mesma moeda for lançada novamente, qual é a probabilidade de se observar outra cara? E se o primeiro lançamento fosse coroa, qual seria a probabilidade do segundo lançamento (da mesma moeda) ser cara?

Solução: Denotando por q_i a probabilidade da i-ésima moeda ser escolhida, temos que $q_i = \dfrac{\frac{1}{5}p_i}{\sum_{j=1}^{5} \frac{1}{5}p_j}$, de modo que temos $q_1 = 0$, $q_2 = 0{,}1$, $q_3 = 0{,}2$, $q_4 = 0{,}3$ e $q_5 = 0{,}4$. Portanto, a probabilidade de se obter uma nova cara é dada por $\displaystyle\sum_{i=1}^{5} q_i p_i = \frac{3}{4}$. Um raciocínio análogo mostra que se o primeiro lançamento fosse coroa, a probabilidade do segundo lançamento da mesma moeda ser cara é de $1/4$.

Exemplo

Um avião desapareceu e presume-se que seja igualmente provável que ele tenha caído em qualquer uma das três regiões possíveis. Denote por $1 - \beta_i$, $i = 1, 2, 3$, a probabilidade de que o avião seja encontrado após uma busca na região i quando ele de fato está nessa região (as constantes β_i são ditas *probabilidades de negligência*, pois representam a probabilidade de não encontrar o avião; em geral são atribuídas às condições climáticas e geográficas da região). Qual é a probabilidade de que o avião esteja na região i dado que a busca na região 1 tenha sido mal-sucedida?

Exemplo (continuação)



Exemplo

No campeonato mundial de *bridge* que ocorreu em Buenos Aires em maio de 1965, os famosos parceiros de *bridge* Terrence Reese e Boris Schapiro foram acusados de trapacear utilizando um sistema de sinais com os dedos que poderia indicar o número de cartas do naipe de copas que eles tinham. Reese e Schapiro negaram a acusação e o caso acabou em uma audiência realizada pela liga inglesa de *bridge*. A audiência foi feita na forma de um julgamento com grupos de acusação e defesa, ambos com o poder de chamar e arguir as testemunhas. Durante o procedimento em curso, o promotor examinou mãos específicas jogadas por Reese e Schapiro e afirmou que suas jogadas eram consistentes com a hipótese de eles terem tido conhecimento ilícito acerca das cartas do naipe de copas. Neste ponto, o advogado de defesa disse que as jogadas que eles fizeram eram consistentes com a sua linha de jogo usual. Entretanto, o promotor afirmou em seguida que, desde que suas jogadas fossem consistentes com a hipótese de culpa, isso deveria ser contado como evidência em direção a essa hipótese. O que você pensa do raciocínio da acusação?

Exemplo

Um crime foi cometido por um único indivíduo, que deixou algumas amostras de seu DNA na cena do crime. Peritos que estudaram o DNA recuperado notaram que apenas 5 cadeias puderam ser identificadas, e que uma pessoa inocente teria uma probabilidade de 10^{-5} de ter seu DNA compatível com todas as cadeias recuperadas. O promotor do distrito imagina que o criminoso pode ser qualquer um dos 1 milhão de habitantes da cidade. Dez mil desses residentes foram libertados da prisão ao longo dos últimos 10 anos; consequentemente, tem-se arquivada uma amostra de seu DNA. Antes de verificar os arquivos de DNA, o promotor do distrito acha que cada um dos dez mil ex-criminosos tem probabilidade α de ter cometido o novo crime, enquanto cada um dos 990.000 moradores restantes tem probabilidade β, onde $\alpha = c\beta$ (isto é, o promotor supõe que cada condenado recém-libertado é c vezes mais propenso a cometer um crime do que um morador da cidade que não seja um condenado recém-libertado). Quando o DNA analisado é comparado com a base de dados dos dez mil ex-condenados, descobre-se que A. J. é o único cujo DNA é compatível com a amostra coletada. Supondo que a estimativa entre α e β feita pelo promotor da cidade seja precisa, qual é a probabilidade de que A. J. seja o culpado?

Independência entre eventos

Considere uma probabilidade $\mathbb{P}$ definida em um espaço amostral Ω. Lembre que para quaisquer eventos A e B, vale que

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

Definição: Dois eventos A e B são *independentes* se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Teorema: Seja $\mathbb{P}$ uma probabilidade definida em um espaço amostral Ω. Se A e B são eventos independentes, então também o são os eventos A^c e B, A e B^c, e A^c e B^c.

Exemplo

Em um jogo de dados você ganha se sair 5 ou 6 em uma rolagem. Alguém que observou o resultado da rolagem antes de você quer te vender uma informação, com o argumento que o conhecimento dela te ajudará a tomar uma decisão em relação à sua aposta. A informação é "o valor rolado é um número primo". Vale a pena pagar por essa informação? E se a informação fosse "o valor rolado é um numero primo e ímpar"?

Independência entre mais de dois eventos

Definição: Os eventos $A_1, \ldots, A_n$ são *independentes dois a dois*, se os eventos A_i e A_j são independentes para todos $i \neq j$.

Definição: Os eventos $A_1, \ldots, A_n$ são *independentes* se

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i),$$

para qualquer subconjunto $J \subseteq \{1, \ldots, n\}$.

Independência entre eventos

⚠ Independência dois a dois não implica independência ⚠

De fato, no lançamento de 2 moedas honestas, os eventos

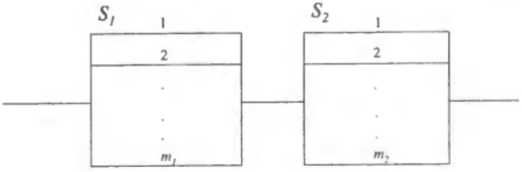
$$A = \text{“primeira lançamento é cara”}$$
$$B = \text{“segundo laçamento é cara”}$$
$$C = \text{“os dois lançamentos são iguais”}$$

são independentes dois a dois mas não são independentes.

É realmente necessário verificar independência para toda sub-coleção de eventos de interesse!

Exemplo

- A *confiabilidade* de um sistema ou componente é a probabilidade que ele funcione. Considere um sistema com dois sub-sistemas em série S_1 e S_2 com, respectivamente m_1 e m_2 componentes idênticos em paralelo.



- O evento em que o componente j do sub-sistema S_i funciona é representado por A_{ij}, para $i = 1, 2$ e $j = 1, 2, \ldots, m_i$.
- Suponha que a probabilidade de funcionamento de cada componente dentro de um mesmo sub-sistema seja igual, ou seja, $\mathbb{P}(A_{ij}) = \alpha_i$, para todo j.
- Quantos componentes devemos ter em cada sub-sistema para garantir uma confiabilidade de pelo menos γ?

⚠ Atenção!

- Independência: assumida ou deduzida?
- Eventos independentes $\neq$ eventos disjuntos!

Exemplo

Realizam-se tentativas independentes que consistem no lançamento de um par de dados honestos, e considera-se a soma dos valores observados. Qual é a probabilidade de que um resultado igual a 5 apareça antes de um resultado igual a 7?

Variáveis aleatórias - motivação

Às vezes, nosso espaço amostral pode ser bem complicado...

- Gás com 10^{28} partículas movendo-se livremente em $\mathbb{R}^3$
- Ω = posição e momento de todas as partículas = $\mathbb{R}^{6\times 10^{28}}$!
- Impossível medir posição e momento de todas as partículas!
- Além disso, fazer contas nesse espaço é virtualmente inviável.
- Porém, é fácil observarmos, por exemplo, pressão e temperatura desse gás
- Para cada configuração $\omega \in \Omega$, temos uma pressão e temperatura diferentes, ou seja, temos funções

$$
\begin{aligned}
X :&\Omega \to \mathbb{R} \\
&\omega \mapsto X(\omega) = \text{temperatura associada à configuração } \omega \\
Y :&\Omega \to \mathbb{R} \\
&\omega \mapsto Y(\omega) = \text{pressão associada à configuração } \omega
\end{aligned}
$$

- X e Y são *resumos numéricos* de Ω. Tornemos isso mais rigoroso!

Variável aleatória

Seja Ω um espaço amostral. Uma *variável aleatória* (abreviada por v.a.) é uma função de Ω assumindo valores em $\mathbb{R}$

Fornecem "resumos numéricos" de um experimento aleatório

Variáveis aleatórias serão denotadas em geral por letras maiúsculas do final do alfabeto:

$$X, Y, Z, \ldots$$

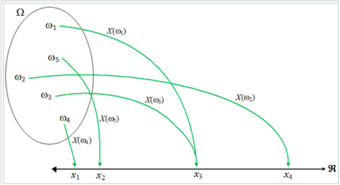


Figure 4: (Variável Aleatória. In: Wikipédia, a enciclopédia livre.) Visualização de uma variável aleatória X. A v.a. X atribuí um valor numérico $X(\omega)$ para cada resultado possível $\omega \in \Omega$

Exemplos de variáveis aleatórias

1) Em um experimento envolvendo 10 lançamentos de uma moeda, o número de caras obtidos é uma variável aleatória
2) Em um experimento envolvendo 5 lançamentos de um dado, os seguintes são variáveis aleatórias:
    a) A soma das 5 faces
    b) O número de faces 4 nos 5 lançamentos
    c) O valor da maior face obtida
    d) O valor da menor face obtida
3) Em um experimento envolvendo a transmissão de uma mensagem, o tempo necessário para transmitir a mensagem, o número de símbolos recebidos com erro e o tempo que leva para a mensagem ser recebida são variáveis aleatórias

Eventos e variáveis aleatórias

Para todo $x \in \mathbb{R}$, escrevemos $\{X = x\}$, $\{X \leq x\}$ e $\{X \geq x\}$ para denotar respectivamente os eventos

$$\{\omega \in \Omega \mid X(\omega) = x\}, \ \{\omega \in \Omega \mid X(\omega) \leq x\} \text{ e } \{\omega \in \Omega \mid X(\omega) \geq x\}$$

Se a v.a. X conta o número de coroas em 10 lançamentos de uma moeda honesta, então usando $Cara \leftrightarrow 0$ e $Coroa \leftrightarrow 1$

- $\{X = x\} = \{\omega \in \{0,1\}^{10} : \omega_1 + \ldots + \omega_{10} = x\}$
- $\{X \leq x\} = \{\omega \in \{0,1\}^{10} : \omega_1 + \ldots + \omega_{10} \leq x\}$
- $\{X \geq x\} = \{\omega \in \{0,1\}^{10} : \omega_1 + \ldots + \omega_{10} \geq x\}$

Variáveis aleatórias discretas

Uma variável aleatória X é dita *discreta* se existe um conjunto discreto $\Omega_X \subset \mathbb{R}$ tal que

$$X(\omega) \in \Omega_X \text{ para todo } \omega \in \Omega$$

A *função massa de probabilidade* (f.m.p.) de uma v.a. X discreta é a função p_X dada por

$$p_X(x) = \mathbb{P}(X = x), \quad x \in \Omega_X$$

Note que a f.m.p. p_X de uma v.a. X satisfaz

$$\begin{cases} p_X(x) \geq 0 \text{ para todo } x \in \Omega_X \\ \displaystyle\sum_{x \in \Omega_X} p_X(x) = 1 \end{cases}$$

Exemplo

Ao ser executado, um determinado algoritmo tem uma probabilidade p de dar erro, a depender do estado de certas variáveis internas. Assumindo que o valor de p é conhecido, qual a probabilidade de observarmos k erros em n execuções do algoritmo? Assuma que cada tentativa é independente.

Exemplo

Ao ser lançada, uma moeda tem probabilidade p de dar cara. Assumindo que o valor de p é conhecido, qual a probabilidade de observarmos k caras em n lançamentos? Assuma que cada lançamento é independente.

⚠ Atenção!

- Não é mera coincidência que o raciocínio dos dois *slides* anteriores tenha sido o mesmo
- O conceito de variável aleatória nasce com dois objetivos:
  - Incorporar a utilização do Cálculo Diferencial e Integral no cálculo de probabilidades
  - Criar um arcabouço unificado para tratar de problemas semelhantes

$\Rightarrow$ Variáveis aleatórias "iguais" podem modelar problemas de naturezas distintas!

Exemplo

Tentativas independentes que consistem em jogar uma moeda com probabilidade p de dar cara são realizadas continuamente até que dê cara ou que um total de n jogadas tenha sido realizado. Explicite o espaço amostral e expresse esse cenário através de uma variável aleatória, encontrando a sua fmp. Repita o raciocínio permitindo que a moeda seja jogada até que cara seja observada mas sem limitar o número máximo de lançamentos possíveis.

Variáveis aleatórias contínuas

Enquanto v.a.'s discretas *contam* fenômenos de interesse, v.a.'s contínuas vêm de um processo de *medição*. Alguns exemplos são:

- Tempo que leva para atendimento em uma fila
- Tempo de vida de um equipamento eletrônico
- Tamanho de uma peça em uma fábrica

Nesses exemplos temos $\Omega_X = [0, \infty)$, um conjunto *não-enumerável*

Dizemos que X é uma v.a. *contínua* se existe uma função $f_X : \mathbb{R} \to \mathbb{R}$, denominada *função densidade de probabilidade* (f.d.p.) satisfazendo:

1) $f_X(x) \geq 0, \forall x \in \mathbb{R}$
2) $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$

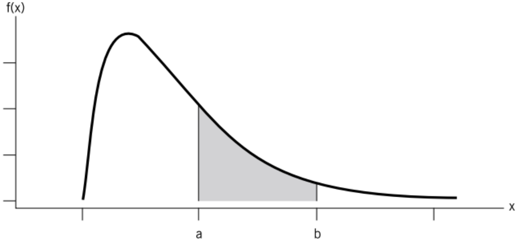


Figure 5: (Pinheiro, J. I. D., et. al. Probabilidade e Estatística, p. 83) Função densidade de probabilidade de uma variável aleatória contínua

⚠ Note que $f_X(x)$ NÃO é $\mathbb{P}(X = x)$! Esse último valor é sempre zero:

$$\mathbb{P}(X = x) = \int_x^x f_X(t)\, dt = 0$$

Porém, se $a < b$ é verdade que

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$$

Um exemplo

O tempo de vida útil de determinado componente eletrônico, quando medido em dias, é descrito pela variável aleatória X, contínua, cuja função densidade de probabilidade é dada por

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

a) Qual a probabilidade de que tal componente funcione por mais de três dias?

b) Sabendo que tal componente já funcionou por três dias, qual a probabilidade de que ele funcione por mais três dias?

Outro exemplo

Para fazer uma poção mágica, uma maga coloca uma substância secreta na mistura, cuja concentração é desconhecida, porém seguindo a função densidade de probabilidade abaixo:

$$g(z) = \begin{cases} \frac{3}{2}(1 - z^2), & \text{para } 0 \leq z \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Sejam X e Y as concentrações da substância em dois preparos da poção, feitos de modo independente um do outro. Calcule a probabilidade de que em ambas as poções a concentração da substância secreta seja superior a 0,5.

Mais um exemplo

O tempo de vida, em horas, de uma válvula de rádio é uma variável aleatória com função densidade de probabilidade dada por

$$f(x) = \begin{cases} 0, & \text{se } x \leq 100 \\ \frac{100}{x^2}, & \text{se } x > 100. \end{cases}$$

Qual é a probabilidade de que exatamente 2 de 5 válvulas no circuito de um aparelho de rádio tenham que ser trocadas nas primeiras 150 horas de operação? Suponha que os eventos E_i, para $i = 1, \ldots, 5$, em que a i-ésima válvula tem que ser substituída dentro deste intervalo de tempo sejam independentes.

Função de probabilidade acumulada

Definição: Seja X uma variável aleatória em um espaço amostral Ω com uma medida de probabilidade $\mathbb{P}$. Definimos sua *função de probabilidade acumulada* (também chamada de *função de distribuição*) como

$$F_X(x) = \mathbb{P}(X \in (-\infty, x]) = \mathbb{P}(X \leq x), \text{ para } x \in \mathbb{R}.$$

"Teorema": A função de probabilidade acumulada nos permite obter qualquer informação probabilística sobre X, ou seja, $\mathbb{P}(X \in I)$ pode ser obtida a partir de F_X, para todo intervalo $I \subset \mathbb{R}$.

Obs.: Aspas acima pois o enunciado não está no máximo de seu rigor, mas juro que é tudo verdade :-)

Veremos mais à frente outras formas interessantes de caracterizar uma variável aleatória.

Exemplo

Encontre a fpa das variáveis aleatórias X e Y descritas abaixo:

- X é discreta tal que:

$$\mathbb{P}(X = 1) = 0{,}5 \quad \mathbb{P}(X = 2) = 0{,}2 \quad \mathbb{P}(X = 4) = 0{,}3$$

- Y é contínua com fdp

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

Propriedades da fpa

Teorema: Se $F_X(x)$ é fpa de alguma variável aleatória X, então:

- $a \leq b \Rightarrow F_X(a) \leq F_X(b)$, ou seja, F_X é uma função não-decrescente
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- $\lim_{x \to b^+} F_X(x) = F_X(b)$, ou seja, F_X é uma função contínua à direita

Observação: $\lim_{x \to b^+}$ denota o limite à direita, ou seja, x se aproxima de b por valores maiores que b.

Relação da fpa com a fmp e a fdp

- Se X é va discreta e $\mathbb{P}(X = a) > 0$, então:

$$\mathbb{P}(X = a) = F_X(a) - F_X(a^-)$$

Observação: $\lim\limits_{x \to a^-}$ denota o limite à esquerda, ou seja, x se aproxima de a por valores menores a.

- Se Y é va contínua com fdp $f_Y(y)$, então:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\, dt,$$

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

Variável aleatória Bernoulli

Uma v.a. X é *Bernoulli* de parâmetro $p \in [0, 1]$, se:

- X assume valores 0 ou 1, isto é $\Omega_X = \{0, 1\}$
- $p_X(1) = \mathbb{P}(X = 1) = p$ e $p_X(0) = \mathbb{P}(X = 0) = 1 - p$
- Sinteticamente: $p_X(x) = p^x(1-p)^{1-x}$, para $x = 0, 1$

Notação: $X \sim \mathrm{Ber}(p)$

Observações:

- Codificação: $0 \leftrightarrow$ “fracasso” e $1 \leftrightarrow$ “sucesso”
- O espaço amostral Ω não aparece explicitamente na definição
- *Ensaio de Bernoulli*: experimento aleatório que resulta em “sucesso” ou “fracasso”, mas NÃO ambos
- Parâmetro p: probabilidade de sucesso

Variável aleatória Binomial

Uma v.a. X é *Binomial* de parâmetros n e p, se:

- X assume valores $0, 1, \dots, n$, isto é $\Omega_X = \{0, 1, \dots, n\}$
- Para cada $x \in \{0, 1, \dots, n\}$:

$$p_X(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Notação: $X \sim \text{Bin}(n, p)$

Observação: se X é o número de sucessos em n ensaios independentes de Bernoulli com probabilidade p de sucesso $\Rightarrow X \sim \text{Bin}(n, p)$

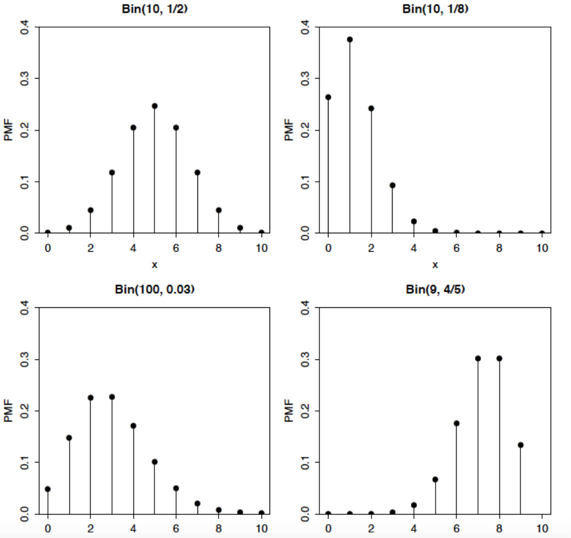


Figure 6: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.102.) Gráfico da f.m.p. para quatro variáveis aleatórias Binomiais

Exemplo

Um determinado sistema eletrônico contém 10 componentes. A probabilidade de determinado componente falhar é de 0,2, e assuma que os componentes sejam independentes. Qual é a probabilidade de pelo menos dois componentes terem falhado? Agora, sabendo que ao menos um dos componentes falhou, qual é a probabilidade de pelo menos dois componentes terem falhado?

Exemplo (para motivar a próxima)

Uma loja recebe em média 4,5 consumidores por hora. Qual é a distribuição de probabilidade do número de consumidores que chega à loja em determinado intervalo particular de uma hora?

Variável aleatória Poisson

Uma v.a. X é *Poisson* de parâmetro $\lambda > 0$, se:

- X assume valores $0, 1, 2, \ldots$, isto é $\Omega_X = \mathbb{N}$
- Para cada $x \in \mathbb{N}$:

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

Notação: $X \sim \text{Poi}(\lambda)$

Observações:

- p_X é uma f.m.p. válida: $e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$
- **Relação Binomial e Poisson**: fixe $\lambda > 0$ e denote $p = \lambda/n$. Se n for grande o suficiente, então o número de sucessos em n ensaios de Bernoulli independentes com probabilidade p de sucesso pode ser bem aproximada por uma va $\text{Poi}(\lambda)$

Variável aleatória Poisson

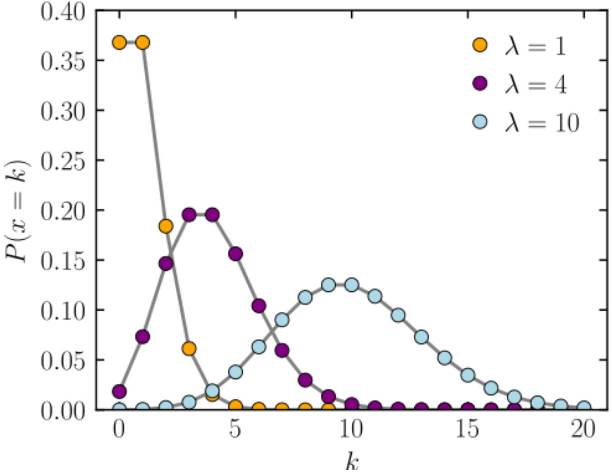


Figure 7: (Wikipedia) Gráfico da f.m.p. para três variáveis aleatórias Poisson

Variável aleatória Poisson

(Trecho extraído de [SR], p. 182) Alguns exemplos de dados que geralmente obedecem à lei de probabilidades de Poisson são:

- Número de erros de impressão em uma página (ou em um grupo de páginas) de um livro
- Número de pessoas em determinada comunidade que vivem mais de 100 anos
- Quantidade de números de telefone discados incorretamente em um dia
- Número de pacotes de biscoitos vendidos em uma determinada loja em um dia
- Número de clientes que entram em uma agência dos correios em um dia
- Número de partículas descarregadas por um material radioativo em um período de tempo fixo

Exemplo

Continuando no exemplo inicial, onde uma loja recebe em média 4,5 consumidores por hora.

- Qual a probabilidade da loja receber mais de 4 consumidores em um intervalo de uma hora?
- Em um intervalo de uma hora, dado que a loja já recebeu 3 consumidores, qual a probabilidade dela receber pelo menos mais dois?
- Com 99% de certeza, qual o número máximo de consumidores que entrarão na loja em um intervalo de uma hora?

Variável aleatória Geométrica

Uma v.a. X é *Geométrica* de parâmetro $p \in (0,1]$, se:

- X assume valores $0, 1, 2, \ldots$, isto é $\Omega_X = \mathbb{N}$
- Para cada $x \in \mathbb{N}$:
$$p_X(x) = p(1-p)^x$$

Notação: $X \sim \text{Geo}(p)$

Observações:

- p_X é uma f.m.p. válida: $p\sum_{x=0}^{\infty}(1-p)^x = p\frac{1}{1-(1-p)} = 1$
- **Perda de memória**: Para todos $k, t \geq 0$ vale que $\mathbb{P}(X = k + t | X \geq k) = \mathbb{P}(X = t)$. A geométrica é a única distribuição discreta com essa propriedade
- **Uma fórmula útil**: $\sum_{k=m}^{n} ar^k = \frac{a(r^m - r^{n+1})}{1-r}$

Exemplo

Um inteiro positivo é escolhido ao acaso, segundo uma distribuição geométrica de parâmetro p conhecido. Determine a probabilidade que o valor escolhido seja par. É possível fazer com que essa probabilidade seja de 50%?

Exemplo

Um sistema eletrônico contém n componentes que funcionam de forma independente e estão conectados em série. Suponha que cada componente funciona propriamente por certo número de ciclos e eventualmente falha. Assuma também que para $i = 1, \ldots, n$ o número de ciclos no qual o componente i funciona tem distribuição geométrica com parâmetro p_i. Encontre a distribuição de probabilidade do número de ciclos no qual o sistema completo funciona.

Um exemplo

Um aluno de educação física, conhecido na escola por ser "bom de pontaria", é desafiado pelos amigos num jogo de basquete. O desafio consiste em fazer 20 lances livres de forma independente ao cesto de basquete e avaliar sua performance nessas jogadas. Sabe-se que em média a cada 20 lançamentos ele acerta 15. Com base nesta informação, responda as perguntas a seguir.

a) Calcule a probabilidade do aluno acertar ao menos 17 das 20 jogadas

b) ⚠ O professor de basquete da escola decide entrar na brincadeira e fazer o desafio. Sabe-se que a probabilidade do professor errar o arremesso ao cesto é de 0.01. Usando a aproximação da Binomial pela Poisson, calcule a probabilidade do professor errar no máximo 2 vezes em 300 lançamentos ao cesto

Solução:

a) Seja X a v.a. que representa o número de acertos ao cesto em 20 jogadas. Temos que $X \sim \text{Bin}(20, 0.75)$ e queremos calcular $\mathbb{P}(X \geq 17)$. Para isso, basta lembrar

$$\mathbb{P}(X \geq 17) = \sum_{x=17}^{20} \binom{20}{x} (0.75)^x (0.25)^{20-x} \approx 0.225$$

b) Seja Y o número de erros ao cesto em 300 jogadas, temos que $Y \sim \text{Bin}(300, 0.01)$, que pode ser aproximada por uma v.a. Poisson de parâmetro 3. Logo

$$\mathbb{P}(Y \leq 2) \approx \sum_{x=0}^{2} e^{-3} \frac{(3)^x}{x!} \approx 0.423$$

Outro exemplo ⚠

Uma empresa faz transporte de material por meio de caminhões e o volume de encomendas que ela recebe oscila ao longo do tempo. Admita que, escolhendo ao acaso um dia de trabalho:

- O número de entregas a serem feitas segue uma distribuição de Poisson com "média" de 5 entregas/dia
- A empresa pode contratar, por empreitada, trabalhadores autônomos que constam de uma lista de 6 nomes. Todos eles tem a mesma chance p de virem a ser contratados para servir à empresa naquele dia, e há independência entre as decisões de se contratar ou não relativas aos diversos trabalhadores da lista

Pergunta-se:

a) Qual é a probabilidade de que nesse dia sejam feitas no máximo 4 entregas, dado que ocorrerão pelo menos 3 entregas?

b) Quais são os valores possíveis de p, se a probabilidade de serem recrutados exatamente 3 empregados da lista de 6 é igual a 0.27648?

Solução:

a) Seja E o número de entregas do dia. Queremos calcular $\mathbb{P}(E \leq 4|E \geq 3)$. Como $E \sim \text{Poi}(5)$, temos que

$$\mathbb{P}(E \leq 4|E \geq 3) = \frac{\mathbb{P}(3 \leq E \leq 4)}{\mathbb{P}(E \geq 3)} = \frac{\mathbb{P}(E = 3) + \mathbb{P}(E = 4)}{1 - \mathbb{P}(E \leq 2)} \approx 0.361$$

b) Seja T o número de trabalhadores autônomos contratados no dia. Então, $T \sim \text{Bin}(6, p)$ e queremos achar p tal que

$$\mathbb{P}(T = 3) = 0.27648$$

Como, $\mathbb{P}(T = 3) = 20p^3(1-p)^3 = 20[p(1-p)]^3$ temos que

$$p(1-p) = (0.2764820)^{1/3} = 0.24$$

o que nos leva à equação do 2^o grau: $p^2 - p + 0.24 = 0$, cujas soluções são $p = 0.6$ e $p = 0.4$, ambas válidas

Exemplo

O seguinte procedimento é realizado com 100 moedas honestas: as 100 são lançadas e as que caíram “cara” são separadas; as moedas que caíram “coroa” são lançadas novamente e as que caíram “cara” são separadas; finalmente, as moedas que caíram “coroa” nas duas primeiras tentativas são lançadas novamente e as que caíram “cara” são também separadas. Qual é a função massa de probabilidade do número de moedas que são separadas?

Exemplo

Uma determinada cidade promove um jogo de loteria no estilo 6/40: o vencedor deve acertar exatamente os seis números sorteados entre 1 e 40; caso haja mais de um vencedor, o prêmio é dividido igualmente. Um dia, onde o prêmio é de 12,5 milhões de dinheiros, um milionário está na cidade. Ele decide então comprar 5 milhões de jogos, onde os 6 números são escolhidos aleatoriamente com igual probabilidade. Assuma que os moradores locais compraram 2 milhões de jogos, e para simplificar, assuma também que os respectivos números também são escolhidos aleatoriamente com igual probabilidade. Cada jogo custa 1 dinheiro. Qual é a probabilidade que o milionário esteja entre os vencedores e qual é a probabilidade que ele seja o único vencedor? Qual é a probabilidade de que o milionário consiga recuperar o seu investimento inicial?

Exemplo

Um baralho de 52 cartas é “bem embaralhado“, ou seja, toda carta tem igual probabilidade de estar em qualquer posição. Usando a distribuição de Poisson, mostre que a probabilidade de que não tenhamos duas cartas consecutivas com o mesmo valor é de aproximadamente e^{-3}.

Variável aleatória uniforme

Uma v.a. contínua é dita *uniforme* no intervalo $[a, b]$ se a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \leq x \leq b \\ 0, & \text{caso contrário} \end{cases}$$

Notação: $X \sim \text{Unif}([a, b])$

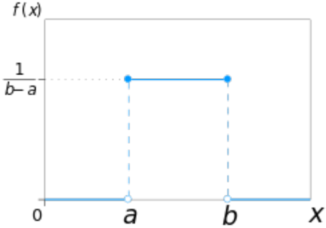


Figure 8: Distribuição Uniforme. In: Wikipédia, a enciclopédia livre

Exemplo 1

Ao arredondar números reais para os inteiros mais próximos, os erros de arredondamento assumem valores entre −0,5 e 0,5. É razoável modelá-los como uma v.a. uniforme nesse intervalo, de modo que a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} 1, & \text{se } -0{,}5 \leq x \leq 0{,}5 \\ 0, & \text{caso contrário} \end{cases}$$

Dessa forma, a probabilidade de que o erro de arredondamento, em módulo, seja menor que ou igual a $0,2$ é dada por

$$P(|X| \leq 0{,}2) = \int_{-0{,}2}^{0{,}2} f_X(x)\, dx = 0{,}4$$

Uma v.a. contínua é dita *Exponencial* de parâmetro λ se a sua f.d.p. é dada por

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

Notação: $X \sim \text{Exp}(\lambda)$

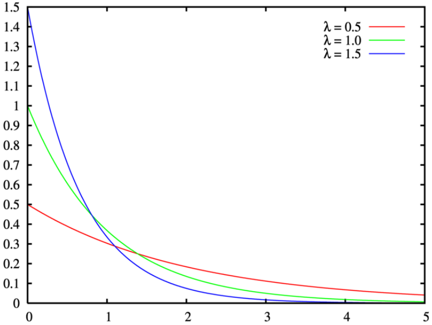


Figure 9: Distribuição Exponencial. In: Wikipédia, a enciclopédia livre

Variável aleatória exponencial - aplicações

- Tempo até uma partícula radioativa decair
- Tempo entre chamadas telefônicas em uma central de auto-atendimento
- Distância entre mutações em uma fita de DNA
- Altura de moléculas de gás em um campo gravitacional uniforme a temperatura e pressão fixos
- Em Meteorologia, precipitação máxima em um determinado intervalo de tempo
- Tempo de vida útil de aparelhos eletrônicos

Exemplo 2

Seja T o tempo entre duas falhas consecutivas de um aparelho eletrônico, medido em dias, e suponha que $T \sim \text{Exp}(2)$

a) Qual a probabilidade de termos pelo menos dois dias entre duas falhas consecutivas?

b) Sabendo que não ocorre uma falha há pelo menos cinco dias, qual a probabilidade de que precisemos esperar pelo menos mais dois dias até a próxima falha?

Solução:

a) Sendo $f_T(t) = 2e^{-2t}$, para $t \geq 0$, temos que:

$$\mathbb{P}(T > 2) = \int_2^\infty 2e^{-2t}\ dt = \frac{1}{e^4} \approx 0{,}0183$$

b)

$$\mathbb{P}(T > 7 \,|\, T > 5) = \frac{\mathbb{P}(T > 7)}{\mathbb{P}(T > 5)} = \frac{\int_7^\infty 2e^{-2t}\ dt}{\int_5^\infty 2e^{-2t}\ dt} = \frac{1/e^{14}}{1/e^{10}} = \frac{1}{e^4} \approx 0{,}0183$$

Variável aleatória exponencial - perda de memória

Teorema: Uma variável aleatória contínua X satisfaz a propriedade da *perda de memória*

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s), \forall s, t > 0$$

se e somente se $X \sim \text{Exp}(\lambda)$

Prova: Provemos somente que se $X \sim \text{Exp}(\lambda)$ então X tem a propriedade da perda de memória. De fato,

$$\begin{aligned}\mathbb{P}(X > s + t \mid X > t) &= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s)\end{aligned}$$

A prova da implicação contrária foge do escopo do curso

Exemplo

Evidências a respeito da culpa ou inocência de um réu em uma investigação criminal podem ser resumidas pelo valor de uma variável aleatória X com distribuição exponencial, cujo parâmetro λ depende da culpa do réu. Se ele é inocente, então $\lambda = 1$; se ele é culpado, $\lambda = 1/2$. O juiz considerará o réu culpado caso $X > c$ para algum valor de c adequadamente escolhido.

a) Se o juiz quer estar 95% certo de que uma pessoa inocente não seja condenada, qual deve ser o valor de c?

b) Usando o valor de c obtido no item a), qual é a probabilidade de que um réu culpado seja condenado?

c) Se uma pessoa qualquer tem 90% de chance de ser inocente, também usando o valor de c obtido no item a), qual é a probabilidade de que ela seja condenada?

Exemplo

Dois clientes A e B, não relacionados entre si, estão em uma loja. Eles permanecem na loja em média por 10 e 20 minutos, respectivamente, com distribuição exponencial. Qual é a probabilidade do cliente que demorar mais tempo para sair da loja demore mais que 20 minutos?

Exemplo

Para amaldiçoar alguém, uma bruxa escolhe somente uma entre duas maldições, cada uma com respectiva probabilidade p_1 e p_2 de ser escolhida, onde $p_1 + p_2 = 1$. A duração de ambas as maldições, medida em dias, são dadas por variáveis aleatórias exponenciais, com respectivos parâmetros λ_1 e λ_2. Se uma maldição escolhida aleatoriamente é lançada e continua a ter efeito após x dias, qual é a probabilidade de que ela continue a ter efeito após y dias adicionais?

Dica: Denote por X o tempo de duração da maldição. NÃO use diretamente a perda de memória em X, pois ela não é exponencial! Use primeiro a definição de probabilidade condicional para expressar a quantidade desejada.

Relação entre a Poisson e a Exponencial

- Poisson modela (sob hipóteses adequadas) ocorrências de "algo" por unidade de tempo
- Parâmetro λ: taxa média de ocorrência por unidade de tempo
- **Pergunta**: Tal "algo" acabou de ocorrer. Qual a distribuição de probabilidade do tempo até a próxima ocorrência?
- **Resposta**: $\text{Exp}(\lambda)$

Exemplo: A quantidade de carros do seu ônibus que passa no seu ponto a cada hora segue uma distribuição de Poisson. Você chega no ponto, atrasado, e pergunta para alguém se faz muito tempo que o último carro passou. A pessoa te responde: "faz um bom tempo já, então daqui a pouco outro está chegando". Essa afirmação está correta? Por quê?

Variável aleatória normal

Uma variável aleatória X é dita *normal* ou *gaussiana* de parâmetros μ e σ^2 se a sua função densidade de probabilidade é dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \text{ para } x \in \mathbb{R}$$

Notação: $X \sim N(\mu, \sigma^2)$

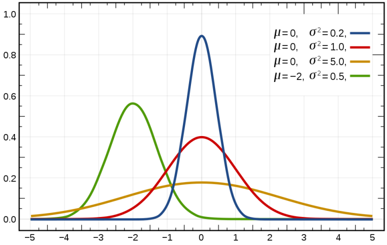


Figure 10: (Distribuição Normal. In: Wikipédia, a enciclopédia livre.) Visualização de uma variável aleatória normal, com diversos valores de μ e σ^2

Variável aleatória normal

- Uma das distribuições mais importantes da Probabilidade e Estatística
- Teorema Central do Limite: "Se $E_1, \ldots, E_n$ são variáveis aleatórias independentes, então $E_1 + \cdots + E_n$ tem distribuição aproximadamente normal, se n é suficientemente grande". Voltaremos a isso em breve!
- Pode ser usada para modelar fenômenos que são descritos pela sobreposição de diversos fenômenos independentes de origem desconhecida (e.g., interferências e ruído de medição de equipamentos)
- É impossível calcular $\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)\, dx$ analiticamente $\Rightarrow$ uso de tabelas para o cálculo de probabilidades

Variável aleatória normal padrão

Dizemos que Z tem distribuição *normal padrão* se $Z \sim N(0,1)$. Denotamos por φ e Φ as suas fdp e fpa, respectivamente:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$
$$\Phi(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)\ dz$$

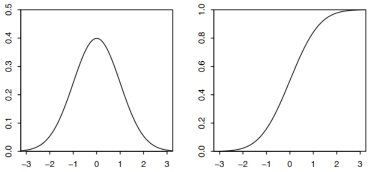
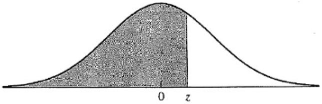


Figure 11: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.212.) Gráficos das funções φ e Φ, respectivamente

A tabela da normal padrão



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319

Figure 12: Excerto de uma tabela de probabilidades da normal padrão.

Variável aleatória normal - propriedades

Se $X \sim N(\mu, \sigma^2)$ então vale que:

- A f.d.p. de X é simétrica com respeito a μ, ou seja,

$$f_X(\mu - x) = f_X(\mu + x) \quad \forall x \in \mathbb{R}$$

- Em particular, para $Z \sim N(0, 1)$

$$\mathbb{P}(Z \leq -z) = \mathbb{P}(Z \geq z) \quad \forall z \in \mathbb{R}$$

ou seja

$$\Phi(-z) = 1 - \Phi(z) \quad \forall z \in \mathbb{R}$$

Padronização da distribuição Normal

Se $X \sim N(\mu, \sigma^2)$, então a *padronização* de X é a v.a. definida por

$$\frac{X-\mu}{\sigma}$$

Temos que $\dfrac{X-\mu}{\sigma}$ tem distribuição normal padrão

Essa última propriedade nos permite calcular qualquer probabilidade usando uma tabela da normal padrão Z:

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\mathbb{P}(X \leq a) = \mathbb{P}\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\mathbb{P}(X \geq a) = \mathbb{P}\left(Z \geq \frac{a-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Teorema: ("Regra" do 68 – 95 – 99.7%) Se $X \sim N(\mu, \sigma^2)$, então

$$\begin{aligned} \mathbb{P}(|X-\mu| < \sigma) &\approx 0.68 \\ \mathbb{P}(|X-\mu| < 2\sigma) &\approx 0.95 \\ \mathbb{P}(|X-\mu| < 3\sigma) &\approx 0.997 \end{aligned}$$

Variável aleatória normal

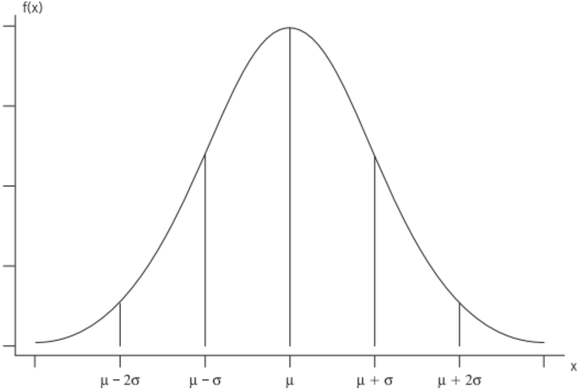


Figure 13: (Pinheiro, J. I. D., et. al. Probabilidade e Estatística, p. 99) Função densidade de probabilidade de uma variável aleatória normal, exibindo alguns valores importantes: 68,3% de probabilidade encontra-se entre $\mu - \sigma$ e $\mu + \sigma$; 95,4% de probabilidade encontra-se entre $\mu - 2\sigma$ e $\mu + 2\sigma$

Exemplo

Seja $X \sim N(\mu, \sigma^2)$. Verifique que

$$\mathbb{P}(|X - \mu| > k\sigma) = 2[1 - \Phi(k)],$$

para todo $k > 0$.

Exemplo

O tempo de vida útil de dois componentes eletrônicos são variáveis aleatórias X_1 e X_2, independentes e com distribuição normal, de respectivos parâmetros μ_i e σ_i^2. Encontre a probabilidade de que ambas deixem de funcionar com uma distância de no máximo a unidades de tempo.

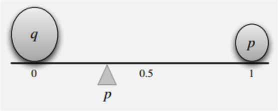
Média, valor esperado, esperança, etc...

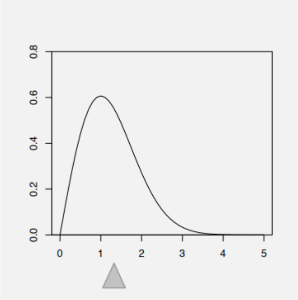
- Jogo no qual n resultados distintos podem ser obtidos $\Rightarrow$ cada com retorno x_i ao jogador, $i = 1, \ldots, n$
- Mesa cobra C para participar de uma rodada $\Rightarrow$ a mesa está levando vantagem?
- Jogar N vezes, resultado x_i tem probabilidade p_i de ocorrer $\Rightarrow$ x_i acontece aproximadamente Np_i vezes
- $\displaystyle\sum_{i=1}^{n} x_i p_i N - NC = N\left(\sum_{i=1}^{n} x_i p_i - C\right) \approx$ lucro aproximado do jogador após jogar N vezes
- $\displaystyle\sum_{i=1}^{n} x_i p_i N \approx$ quanto a mesa dá ao jogador apos N rodadas
- $N = 1 \Rightarrow \displaystyle\sum_{i=1}^{n} x_i p_i \approx$ quanto a mesa dá em uma única rodada
- $\Rightarrow \displaystyle\sum_{i=1}^{n} x_i p_i$ será chamado o *retorno esperado* do jogo

Definição: Definimos a *média*, *esperança* ou *valor esperado* de uma variável aleatória X como

$$\mathbb{E}[X] = \underbrace{\sum_{n=-\infty}^{\infty} x_n \mathbb{P}(X = x_n)}_{\text{se } X \text{ é discreta}} = \underbrace{\int_{-\infty}^{+\infty} x f_X(x)\, dx}_{\text{se } X \text{ é contínua}},$$

quando o somatório ou a integral existirem.





Um exemplo

Seja X a variável aleatória que codifica o número observado em um lançamento de dado honesto. Calcule e interprete $\mathbb{E}[X]$.

Exemplo tranquilo

Determine $\mathbb{E}[X]$ quando a fdp de X é dada por

$$f(x) = \begin{cases} 2x, & \text{se } 0 \leq x \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Outro exemplo tranquilo

Lembremos que o tempo de vida, em horas, de uma válvula de rádio é uma variável aleatória com função densidade de probabilidade dada por

$$f(x) = \begin{cases} 0, & \text{se } x \leq 100 \\ \frac{100}{x^2}, & \text{se } x > 100. \end{cases}$$

Em média, quanto tempo irá funcionar uma determinada válvula?

Um exemplo mais emocionante

Um competidor em um jogo de perguntas e respostas recebe duas questões, numeradas 1 e 2, às quais tentará responder na ordem que preferir. Se ele decidir tentar a questão i primeiro, então ele poderá passar à questão j somente se a sua resposta para a questão i estiver correta. Se a sua resposta inicial estiver incorreta, ele não poderá responder à outra questão. O competidor receberá V_i reais se responder à questão i corretamente, para $i = 1, 2$. Se a probabilidade de que ele saiba responder a questão i é igual a p_i, para $i = 1, 2$, que questão ele deveria tentar responder primeiro de forma a maximizar seu prêmio esperado? O que ocorre no cenário particular em que $p_1 = 0{,}6, V_1 = 200, p_2 = 0{,}8$ e $V_2 = 100$?

Exemplo

Mostre os seguintes resultados referentes aos modelos probabilísticos previamente apresentados:

- $X \sim \text{Bern}(p) \Rightarrow \mathbb{E}[X] = p$
- $X \sim \text{Bin}(n, p) \Rightarrow \mathbb{E}[X] = np$
- $X \sim \text{Poi}(\lambda) \Rightarrow \mathbb{E}[X] = \lambda$
- $X \sim \text{Geo}(p) \Rightarrow \mathbb{E}[X] = \dfrac{1-p}{p}$
- $X \sim \text{Unif}([a, b]) \Rightarrow \mathbb{E}[X] = \dfrac{a+b}{2}$
- $X \sim \text{Exp}(\lambda) \Rightarrow \mathbb{E}[X] = \dfrac{1}{\lambda}$
- $X \sim \text{N}(\mu, \sigma^2) \Rightarrow \mathbb{E}[X] = \mu$

Exemplo para motivar

Seja X uma variável aleatória que assume os valores $-1, 0$ e 1 com probabilidades 0,2, 0,5 e 0,3, respectivamente, e seja $Y = X^2$. Calcule $\mathbb{E}[Y]$.

Resposta: $\mathbb{E}[Y] = 0{,}5$

⚠ **Atenção:** Note que

$$0{,}01 = \mathbb{E}[X]^2 \neq \mathbb{E}[X^2] = \mathbb{E}[Y] = 0{,}5.$$

Exemplo para confundir

Paradoxo de São Petersburgo: Suponha que uma moeda honesta é jogada repetidamente até que a primeira cara apareça. O jogo paga 2^n reais se a primeira cara aparecer na n-ésima jogada. Qual o lucro médio ao jogar esse jogo? Qual o preço que um indivíduo pagaria para entrar neste jogo?

Lei do Estatístico Preguiçoso

Teorema: Seja X uma variável aleatória. Temos que

$$\mathbb{E}[g(X)] = \underbrace{\sum_{n=-\infty}^{\infty} g(x_n)\mathbb{P}(X = x_n)}_{\text{se } X \text{ é discreta}} = \underbrace{\int_{-\infty}^{+\infty} g(x) f_X(x)\; dx}_{\text{se } g(X) \text{ é contínua}},$$

caso $\mathbb{E}[g(X)]$ exista.

- Tal resultado nos diz que não é necessário encontrar a distribuição de $g(X)$ para calcularmos seu valor esperado!
- Em geral, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$!

Um exemplo simples

A função densidade de probabilidade de X é dada por

$$f_X(x) = \begin{cases} 1, & \text{se } 0 \leq x \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Calcule $\mathbb{E}[e^X]$.

Um exemplo mais emocionante

Uma vareta de comprimento 1 é dividida em um ponto U que é uniformemente distribuído ao longo do intervalo $(0, 1)$. Determine o comprimento esperado do pedaço que contém o ponto z, para um valor $0 < z < 1$ fixado.

Propriedades do valor esperado

1) Se $\mathbb{P}(X = c) = 1$, então $\mathbb{E}[X] = c$

2) Se existe uma constante a tal que $\mathbb{P}(X \geq a) = 1$ (ou $\mathbb{P}(X \leq a) = 1$), então $\mathbb{E}[X] \geq a$ (ou $\mathbb{E}[X] \leq a$)

3) Se $\mathbb{E}[X] = a$ e $\mathbb{P}(X \geq a) = 1$ (ou $\mathbb{P}(X \leq a) = 1$), então $\mathbb{P}(X = a) = 1$

4) $\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i] + b$

5) Se $X_1, \ldots, X_n$ são independentes, então $\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i]$

Interlúdio – vetores aleatórios

Definição: Um vetor aleatório de dimensão n é uma função $\mathbf{X} : \Omega \to \mathbb{R}^n$. Ou seja, cada componente de $\mathbf{X} = (X_1, \ldots, X_n)^T$ é uma variável aleatória em Ω.

Definição: A função de probabilidade acumulada conjunta do vetor aleatório $\mathbf{X}$ é definida por

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, \ldots, x_n \leq X_n),$$

para todo $\mathbf{x} \in \mathbb{R}^n$.

Definição: Um vetor aleatório $\mathbf{X}$ é dito discreto se assume seus valores em um subconjunto discreto de $\mathbb{R}^n$. Nesse caso, ele terá uma função massa de probabilidade conjunta:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).$$

Interlúdio – vetores aleatórios

Definição: Um vetor aleatório $\mathbf{X}$ é dito contínuo se tem uma função densidade de probabilidade conjunta, ou seja, se existe uma função $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ tal que:

- $f_{\mathbf{X}} \geq 0$
- $\mathbb{P}(X \in C) = \int_C f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$, para $C \subset \mathbb{R}^n$

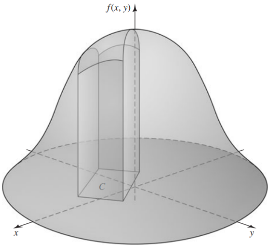


Figure 14: (deGroot & Schervish. Probability and Statistics, p. 121) Função densidade de probabilidade conjunta, e visualização da probabilidade do vetor aleatório estar na região C.

Interlúdio – variáveis aleatórias independentes

Definição: Dizemos que um vetor aleatório discreto ou contínuo $\mathbf{X}$ tem componentes independentes ou que as variáveis aleatórias $X_1, \ldots, X_n$ são independentes se podemos fazer as seguintes fatorações:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1}(x_1) \ldots p_{X_n}(x_n), \text{ se } \mathbf{X} \text{ é discreto,}$$
$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) \ldots f_{X_n}(x_n), \text{ se } \mathbf{X} \text{ é contínuo.}$$

Observação: Note as semelhanças/diferenças com a independência de eventos. Ademais, aqui a independência também pode ser assumida na construção do modelo ou deduzida.

Variância – motivação

A média é um resumo numérico de uma variável aleatória, informando sobre a sua tendência central. Porém, nada sabemos sobre a sua *dispersão*:

$$X = 0 \text{ com probabilidade } 1$$

$$Y = \begin{cases} -1 & \text{com probabilidade } 1/2 \\ 1 & \text{com probabilidade } 1/2 \end{cases}$$

$$Z = \begin{cases} -100 & \text{com probabilidade } 1/2 \\ 100 & \text{com probabilidade } 1/2 \end{cases}$$

Note que temos $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$, porém obviamente a dispersão de Z é maior que a de Y que é maior que a de X!

Variância

Definição: Seja X uma variável aleatória com média μ finita. A sua *variância* é definida como

$$\mathbb{V}(X) = \mathbb{E}[(X - \mu)^2].$$

O seu desvio padrão é definido como

$$\mathrm{dp}(X) = \sqrt{\mathbb{V}(X)}.$$

Teorema: A variância de X também pode ser calculada como

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Exemplo

Mostre os seguintes resultados referentes aos modelos probabilísticos previamente apresentados:

- $X \sim \text{Bern}(p) \Rightarrow \mathbb{V}(X) = p(1-p)$
- $X \sim \text{Bin}(n,p) \Rightarrow \mathbb{V}(X) = np(1-p)$
- $X \sim \text{Poi}(\lambda) \Rightarrow \mathbb{V}(X) = \lambda$
- $X \sim \text{Geo}(p) \Rightarrow \mathbb{V}(X) = \dfrac{1-p}{p^2}$
- $X \sim \text{Unif}([a,b]) \Rightarrow \mathbb{V}(X) = \dfrac{(b-a)^2}{12}$
- $X \sim \text{Exp}(\lambda) \Rightarrow \mathbb{V}(X) = \dfrac{1}{\lambda^2}$
- $X \sim \text{N}(\mu,\sigma^2) \Rightarrow \mathbb{V}(X) = \sigma^2$

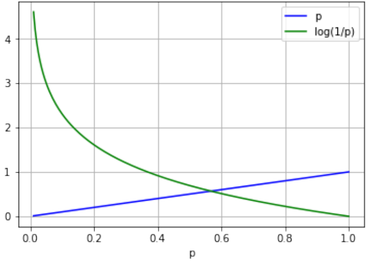
Propriedades da variância

1) $\mathbb{V}(X) \geq 0$, sendo que $\mathbb{V}(X) = 0$ se e somente se existe $c \in \mathbb{R}$ tal que $\mathbb{P}(X = c) = 1$.

4) $\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i)$, se as variáveis aleatórias $X_1, \ldots, X_n$ são independentes.

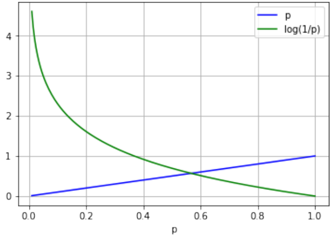
Aplicação: Entropia de variáveis aleatórias

- Seja A um evento em um espaço de probabilidade $(\Omega, \mathbb{P})$
- Queremos medir quão "surpresos" ficamos ao ver a ocorrência de A
- Será que $S(A) = \log\left(\frac{1}{\mathbb{P}(A)}\right)$ é uma boa quantidade? Vejamos um gráfico:



- Ou seja, se $\mathbb{P}(A) = 0$ ficamos bastante surpreso de ver sua ocorrência; e se $\mathbb{P}(A) = 1$ nossa surpresa é nula.

Entropia de variáveis aleatórias



- Além disso, valem as propriedades abaixo:
  - Se $\mathbb{P}(A) \leq \mathbb{P}(B)$ então $S(A) \geq S(B)$
  - Se A e B são independentes, então $S(A \cap B) = S(A) + S(B)$
- Dada uma variável aleatória (discreta ou contínua) X com função de probabilidade ou função densidade de probabilidade f_X, definimos a sua *entropia* como a surpresa média das suas observações, ou seja,

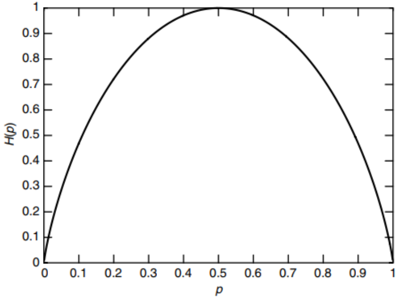
$$\mathcal{H}(X) = \mathbb{E}\left[\log\left(\frac{1}{f_X(X)}\right)\right]$$

Entropia de variáveis aleatórias

Assumindo X discreta e abreviando $f_X(x_n) = \mathbb{P}(X = x_n) = p_n$, podemos reescrever tal quantidade como

$$\mathcal{H}(X) = \sum_n p_n \log\left(\frac{1}{p_n}\right) = -\sum_n p_n \log(p_n)$$

Exemplo: Se X assume os valores 1 ou 0 com probabilidades p e $1 - p$, respectivamente, temos que:



Aplicações do conceito de entropia

- Introduzido por Claude Shannon em 1948, no trabalho *A Mathematical Theory of Communication*
- **Teorema da codificação**: N variáveis aleatórias i.i.d., cada com entropia $\mathcal{H}(X)$ podem ser comprimidas em pelo menos $N\mathcal{H}(X)$ bits com risco essencialmente nulo de perda de informação, se $N \to \infty$
- Nos permite estudar distribuições de probabilidade mais "aleatórias" possíveis, dadas algumas restrições.
  Princípio da máxima entropia: Suponha que X assume os valores $1, \ldots, n$ com probabilidades $p_1, \ldots, p_n$, respectivamente. Então $\mathcal{H}(X)$ é máxima quando $p_1 = \cdots = p_n = 1/n$

E no caso contínuo?!

Aplicações do conceito de entropia

- Finalmente... quantas perguntas o gênio precisa fazer para descobrir em quem você está pensando?

Matematicamente, podemos formular: Um valor X é sorteado de acordo com uma variável aleatória que pode assumir qualquer um de n valores possíveis $1, \ldots, n$ com respectivas probabilidades $p_1, \ldots, p_n$. Através de perguntas do tipo "sim" ou "não" (por exemplo, "X é igual a x?" ou "X é igual a x_1, x_2 ou x_3?") qual é o número médio de questões que você precisará fazer para determinar o valor de X? $\Rightarrow \mathcal{H}(X)$, calculada com $\log_2$!

Exemplos atrasados

Primeiro, os exemplos que deixamos para trás até agora! :-)

Exemplo

Suponha que se você estiver adiantado s minutos para um compromisso, então você tem que arcar com o custo cs. Caso contrário, se estiver atrasado s minutos, você incorre no custo ks. Assuma que o tempo de viagem do seu ponto de partida até o local do compromisso é uma variável aleatória contínua com fdp f. Visando minimizar o custo médio, determine o momento em que você deve sair para o compromisso. Encontre e interprete o tempo ótimo se $c = 2$, $k = 5$ e o tempo de viagem é exponencialmente distribuído com média de meia hora.

Exemplo

Um grande grupo de pessoas fará exames de sangue para uma doença específica, que contamina pessoas ao acaso com probabilidade $p = 0{,}005$. Para diminuir os custos da testagem, decide-se dividir o grande grupo em grupos menores de r pessoas, de modo que as amostras de sangue das r pessoas são misturadas e testadas de uma vez só. Dessa forma, a mistura testará negativo se todos os indivíduos no grupo estiverem negativados, e nesse caso um teste é suficiente para o grupo todo. Caso contrário, se o teste der positivo, r testes adicionais serão necessários para determinar quem está com a doença. Qual o valor de r que minimiza o número médio de testes por pessoa?

Exemplo

Um investidor tem \$100.000 para investir nas seguintes possibilidades, durante um período de um ano: ações com retorno aleatório R_1, ações com retorno aleatório R_2 e um investimento com retorno fixo de 3,6% ao ano. Assuma que R_1 tem média 6 e variância 55, enquanto que R_2 tem média 4 e variância 28. Assuma também que uma ação com retorno R_1 custe \$60 e que uma ação com retorno R_2 custe \$48. O investidor irá comprar s_1 ações com retorno R_1, s_2 ações com retorno R_2 e o valor restante s_3 irá ser investido na taxa fixa. Estude boas estratégias de alocação do dinheiro nesses formatos de investimento.

Exemplo

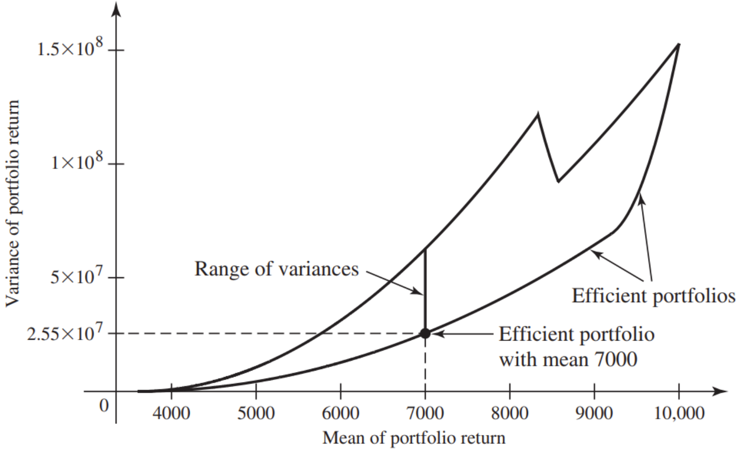


Figure 15: (deGroot & Schervish. Probability and Statistics, p. 231) Todas as médias e variâncias das possíveis alocações de investimento. Para cada tripla (s_1, s_2, s_3) possível, há um ponto correspondente dentro da região.

Exemplo

Seja Z uma variável aleatória com distribuição normal padrão. Verifique que a função densidade de probabilidade de $|Z|$ é dada por

$$\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}z^2}, \text{ para } z > 0.$$

Adicionalmente, verifique que $\mathbb{E}[|Z|] = \sqrt{2/\pi}$ e que $\mathbb{V}(|Z|) = 1 - 2/\pi$.

Função geradora de momentos

Definição: A função geradora de momentos (FGM) de uma variável aleatória X é definida como

$$\psi_X(t) = \mathbb{E}[e^{tX}], \text{ para } t \in (-\varepsilon, \varepsilon) \text{ onde tal quantidade seja finita.}$$

- **Caso discreto**: $\psi_X(t) = \sum_x e^{tx}\mathbb{P}(X = x)$
- **Caso contínuo**: $\psi_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx$

FGM e a transformada de Laplace

Definição: A transformada de laplace bilateral da função f é definida como:

$$\mathcal{L}\{f\}(s) = \int_{-\infty}^{+\infty} e^{-sx} f(x)\, dx,$$

para os valores de $s \in \mathbb{R}$ onde a integral existe e é finita.

Relação da função geradora de momentos com a transformada de Laplace (caso contínuo somente):

$$\psi_X(t) = \mathcal{L}\{f_X\}(-t).$$

Juro que isso será importante em outras disciplinas :-)

Função geradora de momentos

Uma das principais importâncias da função geradora de momentos é que ela, tal como a função de probabilidade acumulada, caracteriza a distribuição de uma variável aleatória de modo único. Mais precisamente:

Teorema: Se duas variáveis aleatórias X e Y têm funções geradoras de momento satisfazendo $\psi_X(t) = \psi_Y(t)$, para $t \in (-\varepsilon, \varepsilon)$, para algum $\varepsilon > 0$, então a distribuição de X e Y é a mesma.

O que significa “gerar momentos”?

Teorema: Suponha que a função geradora de momentos de X exista para $t \in (-\varepsilon, \varepsilon)$, para algum $\varepsilon > 0$. Então $\mathbb{E}[X^k]$ existe, para $k = 0, 1, 2, \ldots$ e temos que

$$\mathbb{E}[X^k] = \psi_X^{(k)}(t)\Big|_{t=0} = \frac{d^k}{dt^k}\psi_X(t)\Big|_{t=0}.$$

Algumas funções geradoras de momentos

Vamos calcular algumas funções geradoras de momentos:

- $X \sim \text{Bern}(p) \implies \psi_X(t) = pe^t + 1 - p,\ \text{para } t \in \mathbb{R}$
- $X \sim \text{Bin}(p) \implies \psi_X(t) = (pe^t + 1 - p)^n,\ \text{para } t \in \mathbb{R}$
- $X \sim \text{Poi}(p) \implies \psi_X(t) = e^{-\lambda(1-e^t)},\ \text{para } t \in \mathbb{R}$
- $X \sim \text{Geo}(p) \implies \psi_X(t) = \frac{p}{1-(1-p)e^t},\ \text{para } t < -\ln(1-p)$
- $X \sim \text{Unif}([a,b]) \implies \psi_X(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)}, & \text{se } t \neq 0 \\ 1, & \text{se } t = 0 \end{cases}$
- $X \sim \text{Exp}(\lambda) \implies \psi_X(t) = \dfrac{\lambda}{\lambda - t},\ \text{para } t < \lambda$
- $X \sim \text{N}(\mu, \sigma^2) \implies \psi_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2},\ \text{para } t \in \mathbb{R}$
- $X \sim \text{Cauchy} \implies \psi_X(t)$ não existe para $t \neq 0$

Também podemos calcular as médias e variâncias (quando existirem) a partir das FGM acima!

Propriedades das FGM

1) $\psi_X(0)$ sempre está definida e vale 1, para qualquer variável aleatória X.
2) Se X tem FGM $\psi_X(t)$ e $Y = aX + b$, então $\psi_Y(t) = e^{bt}\psi_X(at)$.
3) Se X tem FGM $\psi_X(t)$ definida para t em torno de zero, então

$$\mathbb{E}[X^k] = \psi_X^{(k)}(t)\Big|_{t=0} = \frac{d^k}{dt^k}\psi_X(t)\Big|_{t=0}.$$

4) Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes e $X = X_1 + \cdots + X_n$, então vale que

$$\psi_X(t) = \prod_{i=1}^{n} \psi_{X_i}(t).$$

Exemplo

Mostre que a soma de binomiais independentes (com o mesmo parâmetro p) ainda é binomial, que a soma de Poisson independentes ainda é Poisson e que combinação linear de normais independentes ainda é normal. Em todos os casos encontre os respectivos parâmetros.

Exemplo

Uma distribuição extremamente importante em Estatística é a distribuição chi-quadrado com n graus de liberdade. Essa distribuição surge naturalmente em diversas aplicações, e é usualmente descrita na forma

$$\chi_n = Z_1^2 + \cdots + Z_n^2,$$

onde $Z_1, \ldots, Z_n$ são variáveis aleatórias normais padrão independentes. Encontre a função geradora de momentos de tal distribuição.

Exemplo

Seja X uma variável aleatória com FGM dada por

$$\psi_X(t) = \frac{1}{2}\frac{e^t - 1}{t} + \frac{1}{2},$$

para $t \in \mathbb{R}$. Argumente que X não é contínua nem discreta, e encontre a sua distribuição de probabilidade.

Teorema Central do Limite (TCL)

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média μ e variância σ^2, e seja $S_n := \sum_{i=1}^{n} X_i$

Consideramos uma nova variável aleatória

$$Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{V}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

Temos que $\mathbb{E}[Z_n] = 0$ e $\mathbb{V}(Z_n) = 1$. O TCL permite conhecer a distribuição limite de Z_n (quando $n \to \infty$):

Teorema Central do Limite: A distribuição de Z_n se aproxima de uma normal padrão $Z \sim N(0,1)$, ou seja, para todos $x \in \mathbb{R}$ vale que:

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \underset{n\to\infty}{\longrightarrow} \mathbb{P}\left(Z \leq x\right) = \Phi(x)$$

Teorema Central do Limite

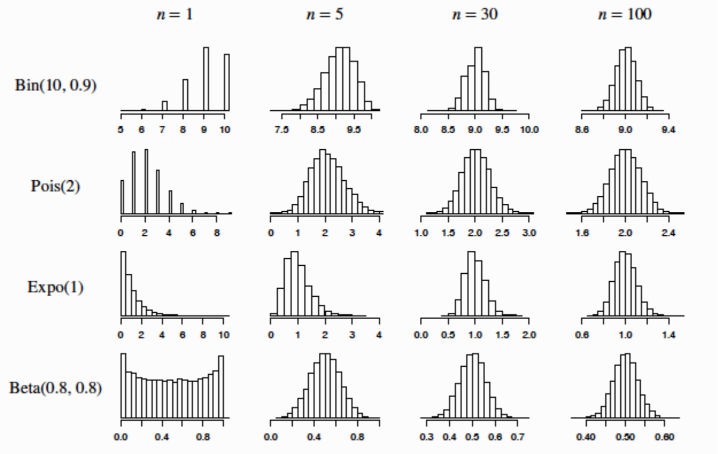


Figure 16: Histogramas da distribuição de S_n/n para diferentes distribuições das v.a.s X_j. Cada histograma foi construído com base em 10.000 simulações de S_n/n. (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.474)

Observação: $X_1, X_2, \ldots$ i.i.d. com média μ e variância σ^2, $S_n = \sum_{i=1}^{n} X_i$

$$\mathbb{P}(S_n \leq c) = \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{c - n\mu}{\sqrt{n}\sigma}\right) \underset{TCL}{\approx} \mathbb{P}\left(Z \leq \frac{c - n\mu}{\sqrt{n}\sigma}\right) = \Phi\left(\frac{c - n\mu}{\sqrt{n}\sigma}\right),$$

onde a aproximação melhora a medida que n cresce.

Exemplo

Carregamos um avião com 100 caixas cujos pesos são modelados por uma variável aleatória uniforme entre 5 e 50 kg. Qual é a probabilidade que o peso total exceda 3.000 kg?

Solução: X_i: peso da i-ésima caixa; $\mu = 27{,}5$ e $\sigma^2 = \frac{(50-5)^2}{12} = 168{,}75$.

$$\mathbb{P}(S_{100} \leq 3.000) = \mathbb{P}\left(\frac{S_{100} - 100(27{,}5)}{10\sqrt{168{,}75}} \leq \frac{3.000 - 100(27{,}5)}{10\sqrt{168{,}75}}\right) \approx \Phi(1{,}92) = 0{,}97.$$

Então $\mathbb{P}(S_{100} > 3.000) = 1 - \mathbb{P}(S_{100} \leq 3.000) = 0{,}03$.

Exemplo

Parece plausível observar 5.250 caras em 10.000 lançamentos de uma moeda honesta?

Exemplo

Uma partícula caminha em uma reta, saltando, a partir de um ponto inicial, uma unidade para a direita ou uma unidade para a esquerda, com igual probabilidade e de modo independente dos passos anteriores. Seja D_n a distância da partícula ao ponto inicial após n passos. Verifique que $\mathbb{E}[D_n] \approx \sqrt{(2/\pi)n}$ e que $\mathbb{V}(D_n) \approx (1 - 2/\pi)n$, para n suficientemente grande. Verifique também que $\mathbb{P}(D_n \leq x) \approx \Phi(x/\sqrt{n}) - \Phi(-x/\sqrt{n})$, também quando n é suficientemente grande.

Exemplo

Verifique que a probabilidade de obter r ou mais vezes o número seis em $6r$ lançamentos de um dado honesto tende a $1/2$ a medida que $r \to \infty$.

Exemplo

Verifique que a soma

$$e^{-n}\left(1+\frac{n}{1!}+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!}\right)$$

tende a $1/2$ a medida que $n \to \infty$.

Exemplo

Seja $X_1, X_2, X_3, \dots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média μ e desvio padrão σ. Mostre que

$$\mathbb{P}\left(\left|\frac{1}{n}(X_1 + \cdots + X_n) - \mu\right| > c\right) \approx 2\left[1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right],$$

para n suficientemente grande.

Transformada z: motivação

- Caso particular da distribuição zeta (generalização de Lei de Zipf):

$$\mathbb{P}(X = x) = \frac{6}{\pi^2}\frac{1}{x^2}, x = 1, 2, 3, \ldots$$

- Tentemos calcular a sua FGM:

$$\psi_X(t) = \mathbb{E}[e^{tX}] = \frac{6}{\pi^2}\sum_{x=1}^{\infty}\frac{e^{tx}}{x^2} < \infty \iff t = 0$$

- Existem generalizações da FGM para os casos onde ela não existe para $t \neq 0$?
- Sim! Especialmente a função característica (para todo tipo de variável aleatória) e a função geradora de probabilidades (ou transformada z – apenas para variáveis aleatórias discretas)

Transformada z

Definição: Seja X uma variável aleatória discreta assumindo valores nos inteiros positivos. Definimos sua transformada z ou função geradora de probabilidades como

$$G_X(z) = \mathbb{E}[z^X] = \sum_{x=0}^{\infty} z^x \mathbb{P}(X = x)$$

Tal nome se justifica pois

$$\mathbb{P}(X = k) = \frac{1}{k!} \frac{d^k}{dz^k} G_X(z) \bigg|_{z=0}, k = 1, 2, \ldots$$

Mais diretamente, o coeficiente de z^k é exatamente $\mathbb{P}(X = k)$

Propriedades da transformada z

- $\mathbb{P}(X = k) = \left. \dfrac{1}{k!} \dfrac{d^k}{dz^k} G_X(z) \right|_{z=0}, k = 1, 2, \ldots$
- $\mathbb{E}[X(X-1)(X-2)\ldots(X-k+1)] = \left. \dfrac{d^k}{dz^k} G_X(z) \right|_{z=1}, k = 1, 2, \ldots$
- $\mathbb{E}[X] = G'_X(1)$ e $\mathbb{E}[X^2] = G''_X(1) + G'_X(1)$
- $\mathbb{V}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$
- $G_{X+Y}(z) = G_X(z) G_Y(z)$, se X e Y são independentes e assumem valores inteiros não-negativos
- Se X tem FGM, então vale que $\psi_X(t) = G_X(e^t)$
- Sejam $X_1, \ldots, X_n$ iid com transformada z dada por $G_X(z)$, independentes de N, com transformada z dada por $G_N(z)$. Então a transformada z de $\displaystyle\sum_{i=1}^{N} X_i$ é dada por $G_N(G_X(z))$

Exemplo

Calcule a transformada z da fmp de uma distribuição de Poisson de parâmetro λ. Use esse resultado para encontrar a média e a variância dessa distribuição.

Exemplo

Três dados de seis faces são rolados. Encontre a fmp da soma das faces voltadas para cima.

Exemplo

A fmp de X não é conhecida, mas a sua transformada z é. Mostre que a probabilidade de X assumir um valor par é dada por $\frac{1}{2}[G_X(-1) + 1]$.

Exemplo

Um experimento que dá sucesso (com probabilidade p) ou fracasso (com probabilidade $1-p$) é repetido N vezes, onde $N \sim \text{Poi}(\lambda)$. Considere que as realizações do experimento são independentes entre si e de N. Encontre a fmp da quantidade de sucessos observadas.

Exemplo

Uma moeda honesta é lançada sucessivamente. Seja X o número de tentativas necessárias até observarmos r caras seguidas. Qual a média e variância de X?

Exemplo

O número de partículas emitidas por uma fonte radioativa durante um período de uma hora é uma variável aleatória de Poisson com parâmetro λ, a princípio desconhecido. É sabido que a probabilidade de nenhuma partícula ser emitida durante esse intervalo de tempo é de 0,673%. Além disso, um dispositivo contador é utilizado para registrar a quantidade de partículas emitidas em um determinado período de uma hora. Faça o que se pede abaixo:

a) Se mais de 10 partículas são emitidas nesse período, o contador é incapaz de registrar o excesso, e registra simplesmente 10. Qual é a probabilidade do limite do contador ser excedido em um período de uma hora, na precisão de duas casas decimais?

b) Seja Y a variável aleatória que representa o número de partículas registradas pelo dispositivo contador. Qual é a distribuição de probabilidade de Y? *Basta indicar as fórmulas, não é necessário colocar a resposta em valores numéricos.*

c) Sabemos que o tempo T entre duas emissões consecutivas, medido em horas, é exponencialmente distribuído, com parâmetro λ. Se em 20 minutos após uma emissão nenhuma outra partícula foi emitida, qual a probabilidade de que precisamos esperar mais meia hora pela próxima emissão?

Exemplo

Faça o que se pede abaixo:

a) Mostre que
$$\frac{\mathbb{P}(B|A)}{\mathbb{P}(C|A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(C)} \frac{\mathbb{P}(A|B)}{\mathbb{P}(A|C)}.$$

b) Explique tal resultado intuitivamente.

c) Suponha que, antes que uma nova evidência seja observada, a hipótese B tenha três vezes mais probabilidade de ser verdadeira do que a hipótese C. Se a nova evidência é duas vezes mais provável quando C é verdade do que quando B é verdade, qual hipótese é mais provável após a observação da nova evidência?

Exemplo

Considere n jogadas independentes de uma moeda com probabilidade p de dar cara. Dizemos que uma *inversão* ocorre sempre que um resultado difira daquele que o precede. Por exemplo, se $n = 5$ e a sequência de lançamentos é KKCKC (sendo K uma cara e C uma coroa), então há 3 inversões. Faça o que se pede abaixo:

a) Determine o número esperado de inversões nos n lançamentos. *Dica: Expresse o número de inversões como a soma de* $n - 1$ *variáveis aleatórias de Bernoulli adequadas.*

b) Qual o valor de p que maximiza o número esperado de inversões?

c) Justifique intuitivamente o resultado obtido no item b).

Exemplo

A quantidade de óleo contida em cada lata fabricada por uma indústria tem peso que segue uma distribuição normal com média 990g e variância 100g^2. Uma lata é imprópria para venda se o seu peso está um desvio padrão acima ou abaixo da média.

a) Qual é a probabilidade de uma lata ser imprópria para venda?

b) Em uma amostra de 5 latas, qual é a probabilidade de que tenhamos 2 latas impróprias para venda?

c) Qual deve ser a nova distância até a média para que tenhamos apenas 5% das latas impróprias para venda?

Exemplo

Uma companhia de cigarros alega que a quantidade de nicotina em um de seus cigarros é uma variável aleatória com média 2,2mg e variância 0,3mg^2. Entretanto, verificou-se um conteúdo médio de 3,1mg de nicotina em 100 cigarros escolhidos aleatoriamente. Faça o que se pede abaixo:

a) Calcule uma aproximação para a probabilidade de um evento pelo menos tão discrepante quanto o observado, supondo que a alegação da companhia é verdadeira.

b) O que você pode dizer sobre a veracidade da afirmativa da empresa de cigarros? Justifique com base na resposta obtida no item a).

Exemplo (2023-02 - Avaliação presencial 02)

No sistema eletrônico de um ônibus espacial há um determinado componente, fundamental e extremamente frágil, que tem tempo de funcionamento exponencialmente distribuído com média de dois dias. Só há um de tal componente no sistema, de modo que ele precisa ser trocado assim que dá defeito. Assuma que a substituição de um componente estragado por um novo é instantânea.

a) Como cada componente funciona, em média, por dois dias, em uma missão espacial de 100 dias é sugerido que sejam levados apenas 50 componentes, assumidos serem independentes. Denote por $T_1, \ldots, T_{50}$ o tempo de vida de cada um desses componentes. Usando o Teorema Central do Limite calcule a probabilidade de que essa quantidade de componentes seja insuficiente para os 100 dias da missão.

b) No contexto do item a), quantos dias deveria ter a missão para garantir que 50 componentes sejam suficientes, com uma probabilidade de 99,9%?

c) Assumindo que o ônibus espacial possa carregar uma quantidade infinita de tais componentes, seja X o número desses componentes utilizados por dia no ônibus espacial. Argumente que X tem distribuição de Poisson e encontre o seu parâmetro.

d) Ainda no contexto do item c), seja agora Y o número de tais componentes que serão utilizados ao longo de toda a missão de 100 dias. Argumente que Y também tem distribuição de Poisson e encontre o seu parâmetro.

Exemplo (2022-02 - Avaliação presencial 02)

Seja X uma variável aleatória codificando uma característica de uma população cuja média μ é um valor desconhecido e queremos estimar. Por estudos anteriores, a variância de X é conhecida e denotada por σ^2. Para se estimar o valor de μ, coleta-se então dados, representados na forma de variáveis aleatórias $X_1, \ldots, X_n$, independentes e identicamente distribuídas, com a mesma distribuição de X.

a) Uma boa forma de se estimar μ é através da *média amostral*, definida por

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Note que $\overline{X}_n$ também é uma variável aleatória, pois depende do sorteio de uma amostra da população. Portanto, calcule os valores de $\mathbb{E}[\overline{X}_n]$ e $\mathbb{V}(\overline{X}_n)$ e com base nisso justifique porque $\overline{X}_n$ pode ser considerado um "bom estimador" para μ.

Exemplo (2022-02 - Avaliação presencial 02 – cont.)

b) Agora, vamos tentar aferir algo mais quantitativo sobre a "qualidade" do estimador $\overline{X}_n$. Mostre que a probabilidade do estimador $\overline{X}_n$ se desviar do valor verdadeiro μ por mais do que c unidades é dada, aproximadamente, por $2\left[1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right]$, ou seja, mostre que

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| > c\right) \approx 2\left[1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right],$$

onde Φ denota a função de probabilidade acumulada de uma distribuição normal padrão.

c) Deseja-se estimar o salário médio de uma população, na unidade monetária de salários mínimos. Pesquisas anteriores afirmam que o desvio padrão dos salários é dado por 1,1 salário mínimo. Qual deve ser o tamanho mínimo da amostra para garantir que tenhamos, com 99% de confiança, que o salário médio estimado não se afastará do salário médio "verdadeiro" por mais do que 1% de um salário mínimo.

Exemplo ([SR] Exemplo 3a, Sec. 8.3)

Um astrônomo está interessado em medir a distância, em anos-luz, entre o seu observatório e uma estrela. Embora o astrônomo disponha de uma boa técnica de medição, ele sabe que, em função da variação das condições climáticas e de erros normais, cada vez que faz uma medição ele não obtém a distância exata, mas sim uma estimativa deste parâmetro. Como resultado, o astrônomo planeja fazer uma série de medições e então usar o valor médio dessas medições como seu valor estimado da distância real. Se o astrônomo acredita que os valores das medições são variáveis aleatórias independentes e identicamente distribuídas com média comum d anos-luz (a distância real) e variância comum 4 anos-luz^2, quantas medições precisam ser feitas para que se garanta que a distância estimada tenha uma precisão de 0,5 anos-luz?

Exemplo ([SR] Exemplo 3b, Sec. 8.3)

O número de estudantes que se matriculam em um curso de Ciência da Computação é uma variável aleatória de Poisson com média 100. O professor encarregado do curso decidiu que, se o número de matrículas for maior ou igual a 120, ele dará aulas para duas turmas separadas. Por outro lado, se esse número for menor que 120, ele dará as aulas para todos os estudantes juntos em uma única turma. Qual é a probabilidade de que o professor tenha que dar aulas para duas turmas?

Exemplo ([SR] Exemplo 2h, Sec. 7.2)

Suponha que N pessoas joguem os seus chapéus no centro de uma sala. Os chapéus são misturados e cada pessoa seleciona um deles aleatoriamente. Determine o número esperado de pessoas que selecionam o próprio chapéu.

Exemplo ([SR] Exemplo 2j, Sec. 7.2)

Dez caçadores estão esperando uma revoada de patos. Quando aparece um bando de patos, os caçadores atiram simultaneamente, mas cada um deles escolhe o seu alvo aleatoriamente, independentemente dos demais. Se cada caçador atinge o seu alvo de maneira independente com probabilidade p, calcule o número esperado de patos que escapam ilesos quando um bando com 10 deles passa voando na frente dos caçadores.

Exemplo ([SR] Exemplo 2l, Sec. 7.2)

Considere uma partícula localizada inicialmente em um ponto dado no plano, e suponha que ela siga uma sequência de passos de tamanho fixo, mas em direção completamente aleatória. Especificamente, suponha que, após cada passo, a nova posição esteja a uma unidade de distância da posição anterior, em um ângulo de orientação uniformemente distribuído ao longo de $(0, 2\pi)$ (veja a figura abaixo). Calcule o valor esperado do quadrado da distância em relação à origem após n passos.

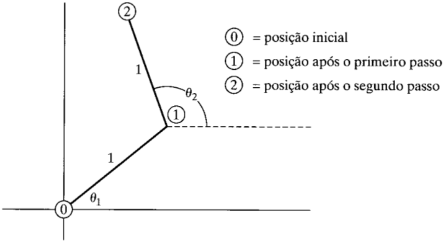


Figure 17: [SR], pág. 365.

Exemplo ([SR] Exemplo 2m, Sec. 7.2)

Suponha que nos deparemos com um conjunto de n valores distintos $x_1, x_2, \ldots, x_n$, e que queiramos ordená-los. Um procedimento eficiente para realizar essa tarefa é o algoritmo de ordenação rápida, que é definido da seguinte maneira. Quando $n = 2$, o algoritmo compara os dois valores e coloca-os na ordem apropriada. Quando $n > 2$, um dos elementos é escolhido aleatoriamente – digamos que seja o elemento x_i – e então os demais valores são comparados com x_i. Aqueles que são menores que x_i, são colocados em uma chave à sua esquerda, e aqueles que são maiores são colocados em uma chave à sua direita. O algoritmo então se repete no interior das chaves e continua até que todos os valores tenham sido ordenados. Em média, quantas comparações precisam ser feitas?

Variáveis aleatórias (discretas) bidimensionais

Em diversas ocasiões é interessante estudar a relação entre múltiplas variáveis aleatórias de um mesmo experimento. Por exemplo:

- *Medicina*: Para avaliar a efetividade de um tratamento, pode-se tomar várias medidas por paciente; a pressão arterial, a frequência cardíaca e as taxas de colesterol, conjuntamente podem ser mais informativas do que qualquer uma dessas medidas separadamente.
- *Finanças:* Para estudar a evolução do preço das ações de uma empresa, coletamos uma série de medidas ao longo do tempo, e as estudamos conjuntamente. Essa série de medidas quando consideradas conjuntamente podem ser usadas para deduzir a tendência do preço das ações da empresa.

Distribuição conjunta de um vetor aleatório discreto

Sejam X e Y v.a.s definidas no mesmo espaço amostral. O par (X, Y) é chamado *vetor aleatório bidimensional*

O vetor aleatório bidimensional (X, Y) é chamado *discreto* se X e Y são v.a.s discretas

A *função massa de probabilidade conjunta* do v.a. (X, Y) discreto é a função $p_{X,Y}(x, y)$ definida por

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) \text{ para todo } (x, y) \in \Omega_X \times \Omega_Y$$

Note que a f.m.p. conjunta $p_{X,Y}$ satisfaz

$$\begin{cases} p_{X,Y}(x, y) \geq 0 \text{ para todo } (x, y) \in \Omega_X \times \Omega_Y \\ \displaystyle\sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} p_{X,Y}(x, y) = 1 \end{cases}$$

Distribuições marginais de um vetor aleatório discreto

Considere um v.a. bidimensional (X, Y) discreto. A *função massa de probabilidade marginal* de X é a função $p_X(x)$ definida por

$$p_X(x) = \sum_{y \in \Omega_Y} \mathbb{P}(X = x, Y = y) \text{ para todo } x \in \Omega_X$$

De modo similar, a *função massa de probabilidade marginal* de Y é a função $p_Y(y)$ definida por

$$p_Y(y) = \sum_{x \in \Omega_X} \mathbb{P}(X = x, Y = y) \text{ para todo } y \in \Omega_Y$$

Um exemplo

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x,y) = \begin{cases} 1/2, & \text{se } (x,y) = (0,0) \\ 0, & \text{se } (x,y) = (0,1) \\ 1/4, & \text{se } (x,y) = (1,0) \\ 1/4, & \text{se } (x,y) = (1,1) \end{cases}$$

Marginal de X: $p_X(0) = 1/2 = p_X(1) = 1/2$

Marginal de Y : $p_Y(0) = 3/4$ e $p_Y(1) = 1/4$

Logo, $X \sim \text{Ber}(1/2)$ e $Y \sim \text{Ber}(1/4)$

Independência de variáveis aleatórias discretas

Duas v.a.s discretas X e Y são *independentes* se

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

para todo $x \in \Omega_X$ e $y \in \Omega_Y$

Em outras palavras, as v.a.s aleatórias X e Y são independentes se a f.m.p. conjunta $p_{X,Y}$ é o produto das maginais p_X e p_Y:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

para todo $x \in \Omega_X$ e $y \in \Omega_Y$

De volta ao exemplo anterior

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x,y) = \begin{cases} 1/2, & \text{se } (x,y) = (0,0) \\ 0, & \text{se } (x,y) = (0,1) \\ 1/4, & \text{se } (x,y) = (1,0) \\ 1/4, & \text{se } (x,y) = (1,1) \end{cases}$$

Vimos que a marginal de X é: $p_X(0) = 1/2 = p_X(1) = 1/2$

Vimos também que a marginal de Y é: $p_Y(0) = 3/4$ e $p_Y(1) = 1/4$

As v.a.s X e Y são independentes? Não, já que

$$p_{X,Y}(0,1) = 0 \neq \frac{1}{2}\frac{1}{4} = p_X(0)p_Y(1)$$

Outro exemplo

A tabela abaixo mostra a função de probabilidade conjunta entre os números diários de crianças com alergia (X) e com pneumonia (Y) atendidos diariamente em um determinado posto de saúde

$X \backslash Y$	0	1	2	$\mathbb{P}(X = x)$
0	1/16	1/16	1/8	1/4
1	1/8	1/8	0	1/4
2	1/16	1/8	1/8	5/16
3	0	1/8	1/16	3/16
$\mathbb{P}(Y = y)$	1/4	7/16	5/16	1

São X e Y independentes?

Solução: X e Y não são independentes, pois

$$\mathbb{P}(X = 1, Y = 2) = 0 \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 2) = 5/64$$

Mais um exemplo

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x,y) = \begin{cases} 2/6, & \text{se } (x,y) = (0,0) \\ 1/6, & \text{se } (x,y) = (0,1) \\ 2/6, & \text{se } (x,y) = (1,0) \\ 1/6, & \text{se } (x,y) = (1,1) \end{cases}$$

As v.a.s X e Y são independentes?

Solução: Marginal de X:

$$p_X(0) = p_{X,Y}(0,0) + p_{X,Y}(0,1) = 1/2 \text{ e } p_X(1) = 1/2$$

Marginal de Y:

$$p_Y(0) = p_{X,Y}(0,0) + p_{X,Y}(1,0) = 2/3 \text{ e } p_Y(1) = 1/3$$

Logo, X e Y são independentes!

Um último exemplo

Em um lançamento de dois dados honestos, se X é a face do primeiro dado e Y a face do segundo, então as v.a.s $X+Y$ e $X-Y$ não são independentes

Por um lado temos que

$$\mathbb{P}(X+Y=12, X-Y=0) = \mathbb{P}(X=6, Y=6) = 1/36$$

Por outro lado, $\mathbb{P}(X+Y=12) = \mathbb{P}(X=6, Y=6) = 1/36$ e

$$\mathbb{P}(X-Y=0) = \mathbb{P}(X=1, Y=1) + \ldots + \mathbb{P}(X=6, Y=6) = 1/6$$

de modo que

$$\mathbb{P}(X+Y=12, X-Y=0) = \frac{1}{36} \neq \frac{1}{36}\frac{1}{6} = \mathbb{P}(X+Y=12)\mathbb{P}(X-Y=0)$$

Logo as v.a.s $X+Y$ e $X-Y$ são dependentes

Distribuição condicional

Seja (X, Y) um v.a. com f.m.p. conjunta $p_{X,Y}(x, y)$. A *função massa de probabilidade condicional* de X dado $Y = y$, com $y \in \Omega_Y$ tal que $p_Y(y) > 0$, é a função $p_{X|Y}(x|y)$ definida por

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \text{ para todo } x \in \Omega_X$$

Note que a f.m.p. condicional $p_{X|Y}$ satisfaz

$$\begin{cases} p_{X|Y}(x|y) \geq 0 \text{ para todo } (x, y) \in \Omega_X \times \Omega_Y \\ \displaystyle\sum_{x \in \Omega_X} p_{X|Y}(x|y) = 1 \text{ para cada } y \in \Omega_Y \end{cases}$$

Um exemplo

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x, y) = \begin{cases} 1/2, & \text{se } (x, y) = (0, 0) \\ 0, & \text{se } (x, y) = (0, 1) \\ 1/4, & \text{se } (x, y) = (1, 0) \\ 1/4, & \text{se } (x, y) = (1, 1) \end{cases}$$

A f.m.p condicional de X dado $Y = 0$:

$$p_{X|Y}(x|0) = \frac{p_{X,Y}(x, 0)}{p_Y(0)} = \begin{cases} 2/3, & \text{se } x = 0 \\ 1/3, & \text{se } x = 1 \end{cases}$$

A f.m.p. condicional de X dado $Y = 1$:

$$p_{X|Y}(x|1) = \frac{p_{X,Y}(x, 1)}{p_Y(1)} = \begin{cases} 0, & \text{se } x = 0 \\ 1, & \text{se } x = 1 \end{cases}$$

Outro exemplo

Foi feito o loteamento de uma área rural em terrenos retangulares. Para cada terreno, seu comprimento e sua largura, ambos em decâmetros, podem ser iguais a 1 dam, 2 dam ou 3 dam. Denote por X o comprimento e Y a largura de um terreno sorteado ao acaso. A tabela a seguir fornece (apenas parcialmente) a distribuição conjunta de X e Y, variáveis aleatórias supostas independentes:

	Y			
X	1	2	3	$p_X(x)$
1	0,35		0,14	
2				0,20
3	0,05			
$p_Y(y)$		0,30		

Considere as v.a.s $V = 2X + 2Y$ = perímetro do terreno e $W = XY$ = área do terreno (em dam^2) e calcule:

a) O valor de cada probabilidade, conjunta ou marginal, omitido na tabela acima

b) A probabilidade condicional de que a área seja igual a 4 dam^2, dado que o perímetro é igual a 8 dam

Solução:

a) Denote $\mathbb{P}(X = 1) = a$ e $\mathbb{P}(Y = 1) = b$. Em seguida, use a independência para concluir que $a = 0.7$ e $b = 0.5$. Com essas informações, podemos deduzir que

$$\begin{cases} \mathbb{P}(X = 3) = 1 - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) = 1 - 0.7 - 0.2 = 0.1 \\ \mathbb{P}(Y = 3) = 1 - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 2) = 1 - 0.5 - 0.3 = 0.2 \\ \mathbb{P}(X = 1, Y = 2) = (0.7)(0.3) = 0.21 \\ \mathbb{P}(X = 2, Y = 1) = (0.2)(0.5) = 0.10 \\ \mathbb{P}(X = 2, Y = 3) = (0.2)(0.2) = 0.04 \\ \mathbb{P}(X = 3, Y = 2) = (0.1)(0.3) = 0.03 \\ \mathbb{P}(X = 3, Y = 3) = (0.1)(0.2) = 0.02 \end{cases}$$

Continuação da solução:

b) Queremos calcular $\mathbb{P}(W=4|V=8)$. Ora,

$$\mathbb{P}(W=4|V=8) = \frac{P(V=8, W=4)}{P(V=8)}$$
$$= \frac{\mathbb{P}(2X+2Y=8, XY=4)}{\mathbb{P}(2X+2Y=8)}$$

Agora, $\mathbb{P}(2X+2Y=8)$ pode ser reescrita como

$$\mathbb{P}(X=1, Y=3) + \mathbb{P}(X=2, Y=2) + \mathbb{P}(X=3, Y=1) = 0.25$$

e $\mathbb{P}(2X+2Y=8, XY=4) = \mathbb{P}(X=2, Y=2) = 0.06$

Logo,

$$\mathbb{P}(W=4|V=8) = \frac{0.06}{0.25} = 0.24$$

Covariância

A *Covariância* entre v.a.s X e Y é definida por

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))] \\ &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]\ (\textit{Linearidade da esperança})\end{aligned}$$

As seguintes propriedades são imediatas da definição:

1. $\mathrm{Cov}(X,X)=\mathbb{V}(X)$
2. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$
3. $\mathrm{Cov}(X,Z)=0$, se Z é v.a. constante com probabilidade 1
4. $\mathrm{Cov}(aX,Y)=a\mathrm{Cov}(X,Y)$, para $a\in\mathbb{R}$
5. Para quaisquer números reais a,b,c e d

$$\begin{aligned}\mathrm{Cov}\left(aX+bY,cX+dY\right) = ac\mathbb{V}(X)+bd\mathbb{V}(Y) \\ +(ad+bc)\mathrm{Cov}(X,Y)\end{aligned}$$

Exemplo 1

Consideramos n lançamentos de uma moeda cuja probabilidade de obter Coroa é p e Cara $1-p$. Seja X o número de Caras e Y o número de Coroas. Calcular o covariância entre X e Y.

Solução:

$$X+Y=n \implies \mathbb{E}[X]+\mathbb{E}[Y]=n$$

O que implica

$$X-\mathbb{E}[X]=-(Y-\mathbb{E}[Y])$$

Então temos que

$$\operatorname{Cov}(X,Y)=-\mathbb{V}(X)=-np(1-p)$$

ATENÇÃO

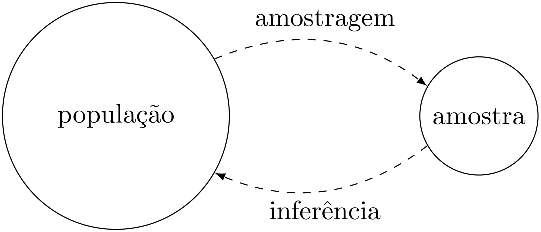
Daqui pra frente nem tudo que está nos slides será apresentado em sala de aula. Eu usualmente falo um pouco de Inferência Estatística porém mais voltado para aplicações em Aprendizagem de Máquina. Porém, acho importante ter esse conteúdo nos slides para que vocês possam consultar em momento oportuno.

Inferência estatística

A inferência estatística tem como objetivo inferir "alguma coisa" sobre uma população a partir da amostra (sub-conjunto da população)

Exemplo: *Idade média dos brasileiros, sabendo a idade dos cariocas*

Observação: Caso fosse possível obter a idade de todos brasileiros poderíamos saber a distribuição exata da idade e daí calcular a média. A situação onde é possível observar a população toda é rara (*custos elevados, tempo para a coleta muito longo, etc...*)



Perguntas naturais sobre inferência estatística

Sempre que estudamos uma população a partir de uma amostra existe a possibilidade de cometermos algum tipo de erro

Exemplo: população dos pesos (kg) de 6 pessoas: 64, 75, 68, 80, 72, 70. A média *populacional* é $\mu = 71,5$. Consideramos as duas amostras

$$a) \quad 64, 75, 72, 80 \qquad\qquad b) \quad 75, 80, 72, 70$$

As médias *amostrais* são $72,75$ e $74,25$, claramente diferentes de μ

Algumas perguntas naturais:

1) Como é que escolhemos uma amostra?
2) É possível medir o erro da estimativa obtida a partir da amostra?

Amostragem

Existem várias formas de selecionar uma amostra de uma população:

Amostragem aleatória simples: sucessivos sorteios aleatórios até completar-se o tamanho da amostra n (*mais utilizado*)

Amostragem sistemática: útil quando a população é ordenada segundo algum critério; aumenta a representatividade da amostra

Exemplo: população $= \{1, 2, \ldots, N\}$ e $n =$ tamanho da amostra. Define-se $b = N/n$ e escolhe-se aleatoriamente um elemento x da população entre o primeiro e o b-ésimo. A amostra é dada por $\{x, x+b, x+2b, \ldots, x+(n-1)b\}$

Como medir o erro de estimação?

Para responder essa pergunta, a Estatística serve-se da Teoria de Probabilidades

A Teoria de Probabilidades é fundamental na Inferência Estatística, pois permite definir um modelo matemático para relacionar as amostras com as populações

estatística descritiva	*população = conjunto de elementos*
inferência estatística	*população = distribuição de uma variável aleatória X*

A distribuição de X é desconhecida e a Inferência Estatística quer utilizar amostras de X para estimar essa distribuição

Conhecimento parcial sobre a população

Para simplificar a análise, nós assumimos que a distribuição da população pertence a uma família de distribuições indexada por um ou mais parâmetros, porém o verdadeiro valor do parâmetro é desconhecido (*conhecimento parcial sobre a população*)

Usamos a notação $f(x|\theta)$, $\theta \in \Theta$ para a família de distribuições, onde θ pode também ser um vetor de mais de um parâmetro

Exemplo: Bernoulli com parâmetro $\theta = p$
Normal com média e variança $(\theta_1 = \mu, \theta_2 = \sigma^2)$

Nesse arcabouço, o objetivo da Inferência Estatística é estimar o verdadeiro parâmetro usando a informação contida na amostra

Definições

População: conjunto de todos os possível resultados de um observação; é modelada por uma variável aleatória X (discreta ou contínua) com uma distribuição $f(x|\theta)$, onde θ é desconhecido

Amostra aleatória (tamanho n): coleção de n variáveis aleatórias $X_1, X_2, \ldots, X_n$ independentes e com a mesma distribuição de X (população)

Intuitivamente, X_i representa a i-ésima observação. Os valores observados correspondentes a amostra $X_1, X_2, \ldots, X_n$ são denotados $x_1, x_2, \ldots, x_n$

Observação: A amostra é um vetor aleatório com distribuição conjunta

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Conceitos básicos da inferência estatística

Parâmetro: medida numérica (em geral desconhecida) que descreve uma caraterística de interesse da população (da distribuição)
É possível resumir a informação contida na amostra em algumas caraterísticas chave da amostra. Isso é o que se chama *estatística*

Estatística: Uma função $T(\boldsymbol{X}) = T(X_1, \ldots, X_n)$ de uma amostra aleatória $X_1, \ldots, X_n$

Estimador: Uma estatística destinada a estimar um parâmetro de uma população

Como os estimadores, e mais em geral as estatísticas, são funções de variáveis aleatórias, elas mesmas são variáveis aleatórias. Podemos então falar de distribuição de um estimador (*distribuição amostral*)

Exemplos de estatísticas

Seja $X_1, \dots X_n$ uma amostra de tamanho n tirada de uma certa população. Algumas estatísticas importante são:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{Média da amostra}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \text{Variância da amostra}$$

$$X_{(1)} = \min(X_1, \dots, X_n) \qquad \text{menor valor da amostra}$$

$$X_{(n)} = \max(X_1, \dots, X_n) \qquad \text{maior valor da amostra}$$

Função de verossimilhança

Uma estatística importante é a *função de verossimilhança*

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n tirada de uma população X com distribuição $f(x|\theta)$. Dada uma observação da amostra $\boldsymbol{x} = (x_1, \ldots, x_n)$ a função de θ definida por

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

é chamada *função de verossimilhança*

Observação: Quando consideramos $f(\boldsymbol{x}|\theta)$ assumimos que θ seja fixo (desconhecido), e $\boldsymbol{x}$ a variável. Quando consideramos $L(\theta|\boldsymbol{x})$ assumimos que $\boldsymbol{x}$ seja fixo, e θ a variável

Exemplo: função de verossimilhança

Seja $X_1, \ldots, X_n$ uma amostra de tamanho $n = 5$ tirada de uma população $X \sim \text{Ber}(p)$. Dada uma observação da amostra $\boldsymbol{x} = (x_1, \ldots, x_5)$, onde cada $x_i = 1$ or $x_i = 0$, a função de verossimilhança é

$$L(p|\boldsymbol{x}) = \prod_{i=1}^{5} p^{x_i}(1-p)^{1-x_i}$$

Procedimentos básicos de inferência estatística

Estimação pontual (por ponto): se preocupa com achar "bons" estimadores para o parâmetro de uma população

Estimação intervalar (por intervalo): se preocupa com achar intervalos que contêm o parâmetro da população

Teste de hipótese: se preocupa de como escolher entre duas hipóteses complementares a respeito do parâmetro da população

Em todos os três procedimentos os dados extraídos de uma amostra são usado para inferir algo sobre a população

Estimadores pontuais

Estimadores pontuais visam estimar o(s) parâmetro(s) de uma população a partir de uma amostra

Lembrete: dada uma população com distribuição $f(x|\theta)$ o conhecimento do parâmetro θ leva conhecimento sobre a população

Lembrete: um *estimador pontual* é uma estatística, ou seja uma função $T(X_1, \ldots, X_n)$ da amostra $X_1, \ldots, X_n$

1) Como achamos estimadores dos parâmetros?
2) Como podemos avaliar a qualidade da estimativa?

Os métodos comumente usados para achar estimadores são:

- método dos momentos
- método de máxima verossimilhança

O estimador média amostral

A *média amostral* de uma amostra aleatória $X_1, \ldots, X_n$ é definida como

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Sendo uma função de v.a. ela é também uma v.a., é podemos falar da sua distribuição, esperança e variância

- Se $X_1, \ldots, X_n$ é uma amostra aleatória de uma população X com distribuição $N(\mu, \sigma^2)$, é possível provar que

$$S_n = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \implies \overline{X} = \frac{S_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Se $X_1, \ldots, X_n$ é uma amostra de uma população qualquer, temos

$$\text{TCL} \implies \overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right), \text{ para } n \text{ grande}$$

<u>Em geral:</u> $n \geq 30$ é suficiente para usar tal aproximação

Exemplo: média amostral

As especificações de uma característica de qualidade estabelecem um limite máximo de $150,6$ unidades. A medição dessa característica comporta-se como uma v.a. $X \sim N(150; 2,1)$. Determine a probabilidade de que a média amostral de uma amostra aleatória de tamanho $n = 49$ ultrapasse a especificação limite de 150,6.

Solução: Temos que a média amostral $\overline{X} \sim N(150; \frac{2,1}{49} = 0,04)$. Então

$$\mathbb{P}\left(\overline{X} > 150,6\right) = \mathbb{P}\left(\frac{\overline{X} - 150}{0,2} > \frac{150,6 - 150}{0,2}\right) = 1 - \Phi(3) = 0,0013$$

Observação: nesse caso não precisamos do TCL!

O estimador proporção amostral

Consideramos uma população cujos elementos podem ser classificados em dois tipos: *sucesso* e *fracasso*. Nesse caso, modelamos a população como uma variável aleatória $X \sim \text{Ber}(p)$

Dada uma amostra aleatória $X_1, \ldots, X_n$, um estimador do parâmetro p é a *proporção amostral* definida por

$$\overline{p} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

O Teorema central do limite, nos garante que se n é "grande", então vale que

$$\frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

<u>Em geral:</u> essa aproximação vale quando $np(1-p) \geq 3$

Exemplo: proporção amostral

Uma empresa fabrica diodos usados em placas de circuito impresso. Digamos que tenha sido coletada uma amostra aleatória da linha de produção com $n = 50$ diodos e que exatamente um número Y deles esteja fora das especificações

A cada elemento da amostra corresponde uma v.a. $X_i \sim \text{Ber}(p)$; $X_i = 1$ se o diodo analisado está fora das especificações e $X_i = 0$, caso contrário

Y: número de diodos na amostra que estão fora das especificações $\implies Y \sim \text{Bin}(50; p)$

$$\overline{p} := \frac{Y}{50} \approx N\left(p, \frac{p(1-p)}{50}\right)$$

Os estimadores variância/desvio padrão amostral

A *variância amostral* de uma amostra aleatória $X_1, \ldots, X_n$ é definida como

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

É uma estatística da dispersão de uma amostra aleatória

O *desvio padrão amostral* de uma amostra aleatória $X_1, \ldots, X_n$ é definido como $S = \sqrt{S^2}$

Sendo S^2 e S v.a., elas têm uma distribuição. Em, geral não é fácil conhecer a distribuição delas

Porém, se $X_1, \ldots, X_n$ é uma amostra aleatória de uma população $X \sim N(\mu, \sigma^2)$, é possível provar que

$$S^2 \text{ tem uma distribuição } \chi^2$$

Como avaliar a qualidade dos estimadores?

Vimos que existem diferentes métodos para achar estimadores. Por isso é importante introduzir algum critério para avaliá-los

Seja X uma variável aleatória cuja distribuição de probabilidade depende de um parâmetro desconhecido θ, e seja $\hat{\theta} = T(X_1, \ldots, X_n)$ um estimador de θ baseado numa amostra aleatória de X, de tamanho n

Uma das propriedades desejáveis para o estimador $\hat{\theta}$ é que o seu valor esteja o mais "próximo" possível do verdadeiro parâmetro θ

Estimadores não tendenciosos

Lembre que um estimador é uma variável aleatória com uma determinada distribuição. Por isso não está claro o que significa para uma v.a. ser "próxima" de um determinado valor

Possível solução: Uma propriedade desejável para um estimador é que a média da sua distribuição seja igual ao parâmetro sendo estimado (em média, o estimador "acerta")

Definimos o *viés* de um estimador pontual $\hat{\theta}$ de um parâmetro θ como

$$B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

Um estimador pontual $\hat{\theta}$ é dito *não tendencioso* se

$$B(\hat{\theta}) = 0 \iff \mathbb{E}(\hat{\theta}) = \theta$$

Exemplo: estimadores não tendenciosos

Seja $X_1, \ldots, X_n$ uma amostra de uma população $X \sim N(\mu, \sigma^2)$

Consideramos os estimadores $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$

Temos que $\mathbb{E}(\overline{X}) = \mu \implies \overline{X}$ é não tendencioso para μ Para S^2 temos

$$
\begin{aligned}
\mathbb{E}(S^2) &= \frac{1}{n-1}\mathbb{E}\Big(\sum_i (X_i - \overline{X})^2\Big) = \frac{1}{n-1}\mathbb{E}\Big(\sum_i X_i^2 - n\overline{X}^2\Big) \\
&= \frac{1}{n-1}\mathbb{E}\Big(\sum_i (X_i^2 - \overline{X}^2)\Big) = \frac{1}{n-1}\mathbb{E}\Big(\sum_i \big((X_i^2 - \mu^2) - (\overline{X}^2 - \mu^2)\big)\Big) \\
&= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\mathrm{Var}(\overline{X}) = \frac{n}{n-1}\sigma^2 - \frac{1}{n-1}\sigma^2 = \sigma^2
\end{aligned}
$$

onde usamos que $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$. Então, S^2 sé não tendencioso para σ^2

Observação: Essa é a motivação para o fator de normalização $\frac{1}{n-1}$

Erro quadrático médio (EQM)

Uma outra métrica para medir se um estimador retorna estimativas que sejam próximas do valor verdadeiro do parâmetro é o *erro quadrático médio*, definido como

$$\mathrm{EQM}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2)$$

Observação: o fato que um estimador seja não tendencioso não garante que o EQM seja pequeno

Para estimadores não tendenciosos, o EQM mede quão espalhadas as estimativas dadas pelo estimador ficam em relação ao valor verdadeiro do parâmetro

Erro quadrático médio (EQM)

O EQM de um estimador pode ser reescrito em função do seu viés:

$$\mathrm{EQM}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + B^2(\hat{\theta})$$

O EQM incorpora duas componentes: uma mede a precisão (variância) e a outra mede o viés. Estimadores que têm boas propriedades de EQM têm variança e viés pequenos

Observação 1: para estimadores não tendenciosos o EQM é igual à variança do estimador

Observação 2: estimadores não tendenciosos não necessariamente têm menor EQM do que estimadores tendenciosos

Exemplo: erro quadrático médio (EQM)

Seja $X_1, \ldots, X_n$ uma amostra aleatória tirada de uma população $X \sim \text{Ber}(p)$. Consideramos o estimador $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Para calcular o $\text{EQM}(\overline{X})$ podemos calcular $B(\overline{X})$ e $\text{Var}(\overline{X})$

$B(\overline{X}) = 0 \implies \overline{X}$ é não tendencioso $\implies \text{EQM}(\overline{X}) = \text{Var}(\overline{X})$

$$\text{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}$$

Erro absoluto da estimação pontual

Mais uma propriedade desejável para um estimador é que ele retorne estimativas que sejam próximas do valor verdadeiro do parâmetro dentro de um certo erro absoluto d fixado, ou seja $|\hat{\theta} - \theta| \leq d$

Sendo $\hat{\theta}$ uma v.a., não é possível garantir que $|\hat{\theta} - \theta| \leq d$ seja sempre válido

Em algumas situações é possível garantir que a condição $|\hat{\theta} - \theta| \leq d$ seja atendida com alta probabilidade, ou seja

$$\mathbb{P}(|\hat{\theta} - \theta| \leq d) \text{ seja perto de } 1$$

Erro absoluto para a média amostral

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população X com média μ (desconhecida) e variância σ^2 (conhecida)

No caso do estimador $\overline{X}$ é possível garantir que $\mathbb{P}(|\overline{X} - \mu| \leq d) \approx 1$:

$$\mathbb{P}(|\overline{X} - \mu| \leq d) = \mathbb{P}\left(\frac{|\overline{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} \leq \frac{d}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$\stackrel{TCL}{\approx} \mathbb{P}\left(|Z| \leq \frac{d}{\frac{\sigma}{\sqrt{n}}}\right) = 2\Phi\left(\frac{d}{\frac{\sigma}{\sqrt{n}}}\right) - 1,$$

e o valor de Φ pode ser obtido pela tabela da normal padrão

Observação 1: É preciso conhecer o valor de σ!

Observação 2: Ao aumentar o valor de n temos que $\Phi\left(\frac{d}{\frac{\sigma}{\sqrt{n}}}\right) \to 1$

Erro absoluto para a proporção amostral

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população $X \sim \text{Ber}(p)$

Também para o estimador $\overline{p}$ é possível garantir que $\mathbb{P}(|\overline{p} - p| \leq d) \approx 1$:

$$\mathbb{P}\left(|\overline{p} - p| \leq d\right) = \mathbb{P}\left(\frac{|\overline{p} - p|}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{d}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\overset{TCL}{\approx} \mathbb{P}\left(|Z| \leq \frac{d}{\sqrt{\frac{p(1-p)}{n}}}\right) \geq \mathbb{P}\left(|Z| \leq \frac{d}{\sqrt{\frac{\frac{1}{4}}{n}}}\right) = 2\Phi(2d\sqrt{n}) - 1,$$

onde usamos que $p(1-p) \leq \frac{1}{4}$ para todo $p \in [0, 1]$

Observação: $\text{Var}(\overline{p}) = \frac{\sqrt{p(1-p)}}{n}$, mas o parâmetro p é desconhecido!
Para resolver esse problema usamos $p(1-p) \leq \frac{1}{4}$

Exemplo: erro absoluto de estimação

Foi realizada uma pesquisa de opinião em uma empresa visando determinar o nível médio de satisfação dos empregados. O índice de satisfação de cada empregado varia entre 0 e 100 pontos, e o desvio padrão populacional é de 30 pontos. Se nessa pesquisa foram sorteados 324 empregados ao acaso para uma entrevista, qual a probabilidade de que o índice de satisfação médio seja estimado com erro absoluto menor que três pontos?

Solução: Seja $X_1, \ldots, X_{324}$ uma amostra aleatória e μ a média da população. Queremos calcular $\mathbb{P}(|\overline{X} - \mu| \leq 3)$

$$\begin{aligned}\mathbb{P}(|\overline{X} - \mu| \leq 3) = \mathbb{P}\left(|Z| \leq \frac{3}{30/\sqrt{324}}\right) = \mathbb{P}\left(|Z| \leq 1,8\right) &= 2\Phi(1,8) - 1 \\ &= 0,9281\end{aligned}$$

Isso quer dizer que com $92,8\%$ de chance o erro absoluto na estimação do nível médio de satisfação dos empregados será menor que três pontos

Dimensionamento da amostra

Como vimos acima, quanto maior é o tamanho da amostra em geral mais precisa é a analise estatística

Qual é o tamanho mínimo da amostra para termos estimativas aceitáveis?

Para poder responder essa pergunta, primeiramente fixamos duas constantes:

d: distância máxima considerada tolerável entre a estimativa e o valor verdadeiro do parâmetro

α: probabilidade que a distância entre estimativa e o valor verdadeiro do parâmetro ultrapasse o limite d

Dimensionando a amostra para média (σ^2 conhecido)

Nesse caso procuramos o menor valor de n tal que

$$\mathbb{P}\left(|\overline{X} - \mu| \leq d\right) = 1 - \alpha$$

Usando que

$$\mathbb{P}\left(|\overline{X} - \mu| \leq d\right) \approx \mathbb{P}\left(|Z| \leq \frac{d}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

obtemos que $z_{1-\frac{\alpha}{2}} = \frac{d}{\frac{\sigma}{\sqrt{n}}}$ o que implica

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2$$

Exemplo: dimensionamento da amostra

Foi realizada uma pesquisa de opinião em uma empresa visando determinar o nível médio de satisfação dos empregados. O índice de satisfação de cada empregado varia entre 0 e 100 pontos, e o desvio padrão populacional é de 30 pontos. Qual deveria ser o tamanho n da amostra de empregados a serem entrevistados para que o erro absoluto na estimação do índice de satisfação médio estivesse limitado por $1,5$ com uma probabilidade de $92,81\%$?

Solução: Seja $X_1, \ldots, X_n$ uma amostra aleatória e μ a média da população. Queremos o menor n tal que $\mathbb{P}(|\overline{X} - \mu| \leq 1,5) = 0,9281$

$$0,9281 = 1 - 0,0719 = \mathbb{P}(|\overline{X} - \mu| \leq 1,5) \approx \mathbb{P}\left(|Z| \leq \frac{1,5}{30/\sqrt{n}}\right)$$

$$\Longrightarrow\ n = \left(\frac{z_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2 = \left(\frac{(1,8)(30)}{1,5}\right)^2 \approx 1,296\ \text{ empregados}$$

repare que σ é assumido conhecido!

Dimensionando a amostra para estimar a proporção

Quando o objetivo é estimar a proporção populacional, o tamanho mínimo da amostra para garantir uma distância máxima d da proporção verdadeira com uma probabilidade de pelo menos $1-\alpha$ é calculado a partir da equação:

$$\mathbb{P}\left(|\overline{p}-p| \leq d\right) = 1-\alpha$$

Usando a aproximação Normal, como anteriormente, obtemos:

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d}\right)^2 p(1-p)$$

Observação: o tamanho da amostra depende do parâmetro que estamos querendo estimar (p), que de fato não conhecemos!

Usando $p(1-p) \leq 1/4$ temos que o tamanho mínimo é $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{2d}\right)^2$

Se por acaso sabemos que $p \neq 0.5$ é preciso estimar $p(1-p)$ com o valor mais alto que pode assumir no domínio dos possíveis valores de p

Exemplo: dimensionamento da amostra

Uma empresa fabrica diodos usados em placas de circuito impresso. Cada um desses diodos pode estar fora das especificações padrões com probabilidade p (desconhecida). Suponha que queremos que o erro absoluto da estimativa da fração de diodos fora das especificações não ultrapasse $0,05$ com uma probabilidade de 90%. Qual deveria ser o tamanho amostral se:

a) não tivermos qualquer informação sobre o verdadeiro valor de p?

b) sabemos que o verdadeiro valor de p é inferior a $0,2$?

Solução: Temos $d = 0,05$ e para $1 - \alpha = 0,90$ temos $z_{1-\frac{\alpha}{2}} = 1,64$

a) Como nada sabemos sobre p, na formula $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d}\right)^2 p(1-p)$

pegamos $p = 0,5 \implies n \geq \left(\frac{1{,}64}{(0{,}05)2}\right)^2 = 269$ diodos

b) Sabemos que $p \leq 0,2 \implies p(1-p) \leq 0,2(1-0,2)$. Então,

$n \geq \left(\frac{1{,}64}{(0{,}05)}\right)^2 (0,2)(0,8) = 172,13 \implies 173$ diodos

Na estimação pontual o objetivo é inferir o parâmetro da população a partir de uma amostra; um estimador pontual especifica um único valor.

Na estimativa por intervalo o objetivo é inferir a partir da amostra um *intervalo* de valores para o qual nos temos alguma *confiança* que isso contém o verdadeiro parâmetro

Exemplo: Retira-se uma amostra de 500 brasileiros e calcula-se a média das alturas encontrando-se $1,70$ metro. Logo uma estimação pontual da verdadeira altura média (μ) é dada por $\bar{x} = 1,70$m. Através do intervalo de confiança é possível encontrar um intervalo, por exemplo $[1,60\text{m}, 1,80\text{m}]$ que, em 95% das vezes, contem o valor verdadeiro μ

A lógica na construção dos intervalos de confiança

- θ: parâmetro da população (média, variância, etc)
- $\hat{\theta}$: estimador de θ (é uma variável aleatória!)
- "Conhecida" a distribuição de probabilidade de $\hat{\theta}$, é possível construir um intervalo $\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2$ e exigir que

$$\mathbb{P}(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha$$

- $(1 - \alpha)$: nível de confiança; normalmente $1 - \alpha = 0.90, 0.95, 0.99$

Intervalos de confiança

Seja $X_1, \ldots, X_n$ uma amostra de uma população com parâmetro θ

Um *intervalo de confiança* de θ é dado por um par de funções L e U da amostra (i.e., L e U são estatísticas) tais que $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ para todas possível realizações da amostra $\boldsymbol{x}$

O intervalo aleatório $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ é chamado *estimador do intervalo*

Exemplo: intervalo de confiança

X_1, X_2, X_3, X_4: amostra de uma população com distribuição $N(\theta, 1)$

Um estimador do intervalo para θ é, por exemplo, $[\overline{\boldsymbol{X}} - 1, \overline{\boldsymbol{X}} + 1]$

Na estimação pontual, nós estimamos que $\theta = \overline{\boldsymbol{X}}$, enquanto agora nós temos uma estimativa menos precisa, ou seja, $\theta \in [\overline{\boldsymbol{X}} - 1, \overline{\boldsymbol{X}} + 1]$

Na estimação pontual temos que $\mathbb{P}(\overline{\boldsymbol{X}} = \theta) = 0$, ou seja não temos um bom controle sobre o possível erro. Por outro lado, é possível calcular a *probabilidade que o intervalo contenha o verdadeiro parâmetro*:

$$\mathbb{P}(\theta \in [\overline{\boldsymbol{X}} - 1, \overline{\boldsymbol{X}} + 1]) = \mathbb{P}(\overline{\boldsymbol{X}} - 1 \leq \theta \leq \overline{\boldsymbol{X}} + 1) = \mathbb{P}(-1 \leq \overline{\boldsymbol{X}} - \theta \leq +1)$$

$$= \mathbb{P}\left(-2 \leq \frac{\overline{\boldsymbol{X}} - \theta}{\sqrt{1/4}} \leq +2\right) = \mathbb{P}(-2 \leq Z \leq +2) = 0.95 \qquad Z \sim N(0, 1)$$

Observação: a motivação para utilizar estimativas por intervalos é ter alguma garantia de capturar o verdadeiro parâmetro

Intervalo de confiança para a média populacional

Caso 1: A população X tem distribuição $N(\mu, \sigma^2)$, σ^2 **é conhecida**

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n, e $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média amostral. Sabemos que

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Então, fixado um nível de confiança $(1-\alpha)$, obtém-se o intervalo de confiança para μ usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \implies \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

Intervalo de confiança para a média populacional

Caso 2: População X têm distribuição $N(\mu, \sigma^2)$, **σ^2 é desconhecida**

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ média amostral e $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ a variância amostral (*repare que precisamos estimar a variância!*). Temos que

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tem distribuição *t de Student* com $(n-1)$ grau de liberdade

Então, fixado um nível de confiança $(1-\alpha)$, obtém-se o intervalo de confiança para μ usando

$$\mathbb{P}\left(-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{n-1;1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Longrightarrow \left[\overline{X} - t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}; \overline{X} + t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

Observação: $t_{n-1;1-\frac{\alpha}{2}}$ encontra-se na tabela da t de Student

Distribuição t de Student

Se $X_1, \ldots, X_n$ são i.i.d com distribuição $N(\mu, \sigma^2)$, a variável aleatória $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ têm uma distribuição que não depende de μ nem σ^2, e apenas depende de um parâmetro que é $(n-1)$

Tal distribuição é chamada t de Student com $(n-1)$ graus de liberdades

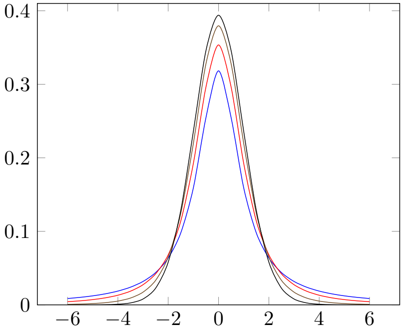
Observação: a dependência na distribuição de T do parâmetro n é devida a $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$

Observação: Se $X_1, \ldots, X_n$ não foram normais, não é mas verdade que a variável T tem distribuição t de Student!

Distribuição t de Student

A distribuição t de Student tem as seguintes propriedades:

- depende de um parâmetro chamado graus de liberdade
- como a normal é simétrica e centrada em 0
- quando os graus de liberdade tendem ao infinito a distribuição se aproxima da distribuição normal padrão
- o cálculo de probabilidades pode ser feito por meio de consulta a uma tabela de probabilidade (como no caso da normal padrão)



Exemplo: intervalo de confiança - caso 1

Os seguintes dados foram coletados para uma amostra de uma população normal $X \sim N(?, 4)$: 15, 10, 7, 13, 6, 9, 13, 7, 10

Calcule o intervalo de 90% de confiança para a média populacional

Solução: *amostra pequena, população normal e σ conhecido $\Longrightarrow$ usamos a Normal*

Para $\alpha = 0,10$ temos $z_{1-\alpha/2} = 1,65$. Então,

$$\left[\bar{x} - z_{0,95}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0,95}\frac{\sigma}{\sqrt{n}}\right] = \left[10 - 1,65\left(\frac{2}{3}\right); 10 + 1,65\left(\frac{2}{3}\right)\right]$$
$$= [8,9; 11,1]$$

Exemplo: intervalo de confiança - caso 2

Os seguintes dados foram coletados para uma amostra de uma população normal $X \sim N(?,?)$: 15, 10, 7, 13, 6, 9, 13, 7, 10.

Qual é o intervalo de 95% de confiança para a média da população?

Solução: *amostra pequena e σ desconhecido $\Longrightarrow$ usamos a t de Student*

Temos $\overline{x} = \frac{1}{n}\sum_{i=1}^{9} x_i = 10$ e $s^2 = \frac{1}{n-1}\sum_{i=1}^{9}(x_i - \overline{x})^2 = \frac{78}{8} = 9,75$ e $s = 3,12$. Usamos a distribuição t de Student com 8 graus de liberdade. Para $\alpha = 0,05$ temos $t_{8;1-\alpha/2} = 2,306$. Então,

$$\left[\bar{x} - t_{8;0,975}\frac{s}{\sqrt{n}}; \bar{x} + t_{8;0,975}\frac{s}{\sqrt{n}}\right] = \left[10 - \frac{3,12}{3}2,306; 10 + \frac{3,12}{3}2,306\right]$$
$$= [7,6; 12,4]$$

Intervalo de confiança para a média populacional

Caso 3 (Grande amostra): A população X tem uma distribuição qualquer mas a amostra é grande $(n \geq 30)$ para que o TCL se aplique

Seja $X_1, \ldots, X_n$ uma amostra e $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média amostral

σ^2 **conhecido:** usando o TCL podemos dizer que $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$; o intervalo de $(1-\alpha)$ confiança para μ é calculado usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \implies \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

σ^2 **desconhecido:** usamos $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ e, nesse caso também, o TCL nos garante que $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}} \approx N(0,1)$; o intervalo de $(1-\alpha)$ confiança para μ é calculado usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \implies \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

Exemplo: intervalo de confiança - caso 3

Deseja-se estimar a resistência média de um certo tipo de fibra usada na fabricação de um tecido. Uma amostra aleatória de 40 espécimes da fibra tem uma média amostral de $12,4$ bar. Se o desvio padrão σ da população é conhecido e igual a $2,1$ bar, determine um intervalo de confiança de 95% para o verdadeiro valor da média populacional μ

Solução: Temos $n = 40, \overline{x} = 12,4, \sigma = 2,1$, $\alpha = 0,05$ e $z_{1-\alpha/2} = 1,96$. Como $n \geq 30$ e σ é conhecido, podemos aplicar o TCL e construir um intervalo de $1-\alpha$ confiança usando $\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$:

$$\left[12,4 - (1,96)\frac{2,1}{\sqrt{40}}; 12,4 + (1,96)\frac{2,1}{\sqrt{40}}\right] = [11,75; 13,05]$$

Observação: A população pode ter qualquer distribuição, não precisa ser normal; isso porque o TCL pode ser aplicado!

Intervalo de confiança para a proporção populacional

Seja X uma população cujos elementos podem ser classificados em dois tipos: *sucesso* e *insucesso*, ou seja $X \sim \text{Ber}(p)$

Dada uma amostra $X_1, \ldots, X_n$ de tamanho n, um estimador do parâmetro p é a proporção amostral $\overline{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. O TCL $\Longrightarrow$

$$\frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1) \qquad \text{se } n \text{ é suficientemente grande, i.e., } np(1-p) \geq 3$$

Calculamos o intervalo de $(1-\alpha)$ confiança para p usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \Longrightarrow \left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} ; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]$$

Observação: O parâmetro p é de fato desconhecido, então não está claro como calcular $\sqrt{\frac{p(1-p)}{n}}$ e obter o intervalo

Dois tipos de intervalos para a proporção populacional

Para achar o intervalo de confiança para a proporção p é preciso calcular $\sqrt{\frac{p(1-p)}{n}}$. Como fazer isso se o parâmetro p é desconhecido?

Intervalo de confiança conservativo: lembrando que $p(1-p) \leq 1/4$ para todo p podemos achar o intervalo de confiança usando que

$$\left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \subset \left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}}\right]$$

Intervalo de confiança não conservativo: podemos usar na expressão da variância o valor estimado $\bar{p}$ e achar o intervalo usando

$$\left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right]$$

Observação: A escolha conservativa sempre levará a intervalo maiores

Exemplo: intervalo de confiança para proporção

Uma loja em Botafogo deseja inferir a proporção de clientes que estão satisfeitos com seu serviço. Para isto, entrevistou 30 clientes e obteve que 20 são satisfeitos e 10 são insatisfeitos. Construa um intervalo conservativo e um não conservativo de 96% confiança para a proporção de clientes satisfeitos

Solução: O intervalo não conservativo de 96% de confiança para p é $\left[\bar{p} - z_{0,98}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{0,98}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] = \left[\frac{2}{3} - 2,05(\sqrt{\frac{2/9}{30}}); \frac{2}{3} + 2,05(\sqrt{\frac{2/9}{30}})\right] = \left[\frac{2}{3} - 0,18; \frac{2}{3} + 0,18\right] = [0,49; 0,85]$

O intervalo conservativo de 96% de confiança para p é $\left[\bar{p} - z_{0,98}\sqrt{\frac{1}{4n}}; \bar{p} + z_{0,98}\sqrt{\frac{1}{4n}}\right] = \left[\frac{2}{3} - 2,05(\sqrt{\frac{1/4}{30}}); \frac{2}{3} + 2,05(\sqrt{\frac{1/4}{30}})\right] = \left[\frac{2}{3} - 0,19; \frac{2}{3} + 0,19\right] = [0,48; 0,86]$

Exemplo 6

Para estimar a média μ de uma população, foram propostos três estimadores não viesados, $\widehat{\mu}_1$, $\widehat{\mu}_2$ e $\widehat{\mu}_3$, obtidos a partir de três amostras independentes, tais que $\mathbb{V}(\widehat{\mu}_1) = 2\mathbb{V}(\widehat{\mu}_2) = 3\mathbb{V}(\widehat{\mu}_3)$. Considere os seguintes estimadores ponderados de μ:

$$T_1 = \frac{\widehat{\mu}_1 + 2\widehat{\mu}_2 + 3\widehat{\mu}_3}{6}, T_2 = \frac{3\widehat{\mu}_1 + 2\widehat{\mu}_2 - \widehat{\mu}_3}{4} \text{ e } T_3 = \frac{\widehat{\mu}_1 + \widehat{\mu}_2}{2}.$$

a) Determine se os estimadores T_1, T_2 e T_3 são ou não viesados.

b) Qual estimador possui o menor erro quadrático médio?

Exemplo 7

Sejam $\overline{X}_1$ a média de uma amostra aleatória de tamanho n extraída de uma população normal de valor esperado μ e variância σ_1^2, e $\overline{X}_2$ a média de uma amostra aleatória também de tamanho n, independente da primeira, extraída de uma população normal de valor esperado μ e variância σ_2^2. Mostre que:

a) Se $w \in [0,1]$, então $w\overline{X}_1 + (1-w)\overline{X}_2$ é um estimador não viesado de μ.

b) A variância do estimador em a) é mínima quando $w = \dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.