Cálculo de Probabilidades II

para Estatística, Ciências Atuariais, Engenharia Matemática e Matemática Aplicada

Código: MAD352

Oferecido pelo:

DEPARTAMENTO DE MÉTODOS ESTATÍSTICOS - DME

INSTITUTO DE MATEMÁTICA - 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**: $\Omega = \{\omega \in \mathbb{R} \mid 0 \leq \omega \leq 24\} = [0, 24]$

Teoria de conjuntos

Objetos de interesse: Espaço amostral Ω e certos sub-conjuntos $A, B, C, \cdots \subset \Omega$

1) $\emptyset$, *conjunto vazio*
2) A^c, *complementar*
3) $A_1 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i$, *união*
4) $A_1 \cap \ldots \cap A_n = A_1 A_2 \ldots A_n = \bigcap_{i=1}^{n} A_i$, *interseção*

   Obs.: Também podemos considerar uniões e interseções infinitas
5) $A - B = A \cap B^c$, *diferença*
6) $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$, *diferença simétrica*

Teoria de conjuntos

1) A e B são ditos *disjuntos* se $A \cap B = \emptyset$
2) Dizemos que A está *contido* em B, denotado por $A \subset B$ se todo elementos de A também é um elemento de B
3) O conjunto de todos os sub-conjuntos de Ω é chamado de *conjunto das partes* e é denotado por 2^{Ω}. Porque?!

Teorema:

1) O conjunto vazio esta contido em todo conjunto
2) $A = B$ se e somente se $A \subset B$ e $B \subset A$
3) *Propriedades distributivas da união e interseção:*

$$A \cap \left(\bigcup_i A_i \right) = \bigcup_i (A \cap A_i) \quad \text{e} \quad A \cup \left(\bigcap_i A_i \right) = \bigcap_i (A \cup A_i)$$

Exemplo

Lançamento de 10 moedas: Se $Cara \leftrightarrow 0$ e $Coroa \leftrightarrow 1$, então

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

Alguns eventos de interesse:

1) "o j-ésimo lançamento foi coroa": $A_j = \{\omega \in \{0,1\}^{10} : \omega_j = 1\}$
2) "pelo menos um lançamento foi coroa": $B = \bigcup_{j=1}^{10} A_j$
3) "todos lançamentos foram cara": $C = B^c = \{(0, 0, \ldots, 0)\}$

Teoria de conjuntos - Leis de de Morgan

Teorema: Seja $A_1, A_2, \ldots$ uma família de sub-conjuntos de Ω. Então vale que

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

Nem todos os subconjuntos de Ω podem ser eventos – 1

- 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 $\mathbb{P}(\Omega) = 1$.

Nem todos os subconjuntos de Ω podem ser eventos – 2

- 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.

Nem todos os subconjuntos de Ω podem ser eventos – 3

- 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!

- **Experimento**: Lançar um dado honesto
- **Espaço amostral**: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Obviamente sabemos calcular $\mathbb{P}(\emptyset) = 0$ e $\mathbb{P}(\Omega) = 1$
- Se sabemos que $\mathbb{P}(\{1, 2\}) = 2/6$, então obviamente $\mathbb{P}(\{1, 2\}^c) = \mathbb{P}(\{3, 4, 5, 6\}) = 1 - 2/6 = 4/6$
- Finalmente, se sabemos que $\mathbb{P}(\{1\}) = 1/6$ e $\mathbb{P}(\{2\}) = 1/6$, devemos saber que $\mathbb{P}(\{1\} \cup \{2\}) = \mathbb{P}(\{1, 2\}) = 2/6$

σ-álgebra - definição

Seja Ω o espaço amostral de um determinado experimento. Uma coleção $\mathcal{F}$ de sub-conjuntos de Ω é dita uma *σ-álgebra* se satisfaz as seguintes propriedades:

- $\Omega \in \mathcal{F}$
- Se $A \in \mathcal{F}$ então $A^c \in \mathcal{F}$
- Se $A_i \in \mathcal{F}$, para $i = 1, 2, 3, \ldots$, então $\displaystyle\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Dado um espaço amostral Ω de um determinado experimento, uma σ-álgebra conterá os conjuntos aos quais saberemos associar uma probabilidade. Chamaremos tais conjuntos, elementos de uma σ-álgebra, de *eventos*.

A motivação anterior nos mostra que tais propriedades são razoáveis de se esperar!

σ-álgebra - exemplos

1 Considere $\Omega = \{1, 2, 3\}$ e as coleções de sub-conjuntos abaixo:

$$\mathcal{F}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$$
$$\mathcal{F}_2 = \{\emptyset, \Omega, \{1\}, \{2\}, \{1, 3\}, \{2, 3\}\}.$$

Ambas são σ-álgebras?

2 Para qualquer conjunto Ω, 2^{Ω} é sempre uma σ-álgebra.

3 Seja $A \subset \Omega$. Quem é uma σ-álgebra que contenha o conjunto A?

σ-álgebra - exemplos

4 Considere o experimento "escolher um número real ao acaso, segundo uma distribuição normal de média 0 e variância 1". Quem é uma σ-álgebra razoável?

Teorema: Seja Ω um espaço amostral e $\mathcal{F}$ uma σ-álgebra de eventos. Sejam também $A_1, A_2, A_3, \ldots$ eventos em $\mathcal{F}$. Então:

1. $\emptyset \in \mathcal{F}$
2. $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

Sequências de eventos

1) $A_1, A_2, A_3, \ldots$ é dita uma *sequência* de eventos em $\mathcal{F}$, e também a denotaremos por $(A_n)_{n\in\mathbb{N}}$.
2) A sequência de eventos $(A_n)_{n\in\mathbb{N}}$ é dita *crescente* se $A_n \subset A_{n+1}$, para todo $n \in \mathbb{N}$. Notação: $A_n \uparrow$
3) A sequência de eventos $(A_n)_{n\in\mathbb{N}}$ é dita *decrescente* se $A_{n+1} \subset A_n$, para todo $n \in \mathbb{N}$. Notação: $A_n \downarrow$

Sequências de eventos crescentes ou decrescentes

1) Sequência $(A_n)_{n\in\mathbb{N}}$ crescente $\Rightarrow \lim_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} A_n$

2) Sequência $(A_n)_{n\in\mathbb{N}}$ decrescente $\Rightarrow \lim_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} A_n$

Sequências de eventos mais gerais

- Se a sequência $(A_n)_{n\in\mathbb{N}}$ não é crescente nem decrescente não temos uma intuição clara do seu limite
- Porém, podemos fazer duas definições análogas:
- **Definição**: Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de eventos. Definimos os *limite superior* e *limite inferior*, respectivamente, como:

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad \text{e} \quad \liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

- Mas o que essas definições significam?

Interpretação dos limites superior e inferior

Teorema: Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de eventos em Ω.

1) O evento $\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ representa "ocorrência de infinitos eventos A_n"

2) O evento $\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$ representa que "todos os eventos A_n ocorrem, para n suficientemente grande"

Um pouco mais sobre os limites superior e inferior

1) **Teorema**: Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de eventos em uma σ-álgebra $\mathcal{F}$. Temos que:
   a) Os eventos $\liminf A_n$ e $\limsup A_n$ também pertencem à $\mathcal{F}$.
   b) Se $(A_n)_{n\in\mathbb{N}}$ é uma sequência crescente, então
   $$\limsup A_n = \liminf A_n = \lim A_n = \bigcup_{n=1}^{\infty} A_n$$
   c) Se $(A_n)_{n\in\mathbb{N}}$ é uma sequência decrescente, então
   $$\limsup A_n = \liminf A_n = \lim A_n = \bigcap_{n=1}^{\infty} A_n$$
2) **Atenção**: NÃO é verdade que $\liminf A_n$ e $\limsup A_n$ são eventos complementares
3) **Proposição**: Vale que $\liminf A_n \subset \limsup A_n$

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!

Axiomas de Kolmogorov

Seja Ω um espaço amostral, munido de uma σ-álgebra $\mathcal{F}$. Uma *probabilidade* é uma função $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ satisfazendo:

1) $\mathbb{P}(\Omega) = 1$
2) Para todo evento $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$
3) Para toda sequência $A_1, A_2, \dots \in \mathcal{F}$ de eventos disjuntos temos

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

- A trinca $(\Omega, \mathcal{F}, \mathbb{P})$ é dita um *espaço de probabilidade*.
- $\mathbb{P}$ também é chamada de *medida de probabilidade*.

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 ($\sqrt{3}$)?

Um exemplo não tão simples...

- Para cada solução apresentada, temos um espaço amostral Ω diferente
- Consequentemente, temos σ-álgebras diferentes para cada resolução
- Não há paradoxo algum! A noção de “escolhida ao acaso”, por si só, é mal formulada e ambígua!

O óbvio acontece!

Teorema: Se Ω é finito, munido da σ-álgebra das partes, então $\mathbb{P}(A) = \dfrac{\text{quantidade de elementos em } A}{\text{quantidade de elementos em } \Omega}$, para $A \subset \Omega$ é uma medida de probabilidade.

Teorema: Se $\Omega = [a, b] \subset \mathbb{R}$, munido da σ-álgebra dos borelianos $\mathcal{B}([a, b])$, então $\mathbb{P}(A) = \dfrac{\text{“tamanho” de } A}{b - a}$, para $A \in \mathcal{B}([a, b])$ é uma medida de probabilidade.

Propriedades da probabilidade - 1

Relembremos os três axiomas de Probabilidade:

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

Teorema: Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e sejam $A, B, A_1, A_2, \cdots \in \mathcal{F}$ 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 $A_1, A_2, \dots \in \mathcal{F}$ não são necessariamente disjuntos, então $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ e

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

Teorema: (Continuidade da medida de probabilidade) Sejam $A, A_1, A_2, \dots, \in \mathcal{F}$ eventos. Temos que:

1) Se $A_n \uparrow A$, então $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$
2) Se $A_n \downarrow A$, então $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$

Probabilidade condicional – motivação

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 de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$
- É sabida a ocorrência de um elemento $\omega \in B \in \mathcal{F}$
- Como incorporar essa informação na σ-álgebra e na medida de probabilidade de modo adequado?

Probabilidade condicional e regra do produto

Definição: Seja $B \in \mathcal{F}$ um evento tal que $\mathbb{P}(B) > 0$. Definimos a *probabilidade condicional de A dado B* como

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

Note que $\mathbb{P}(A|B)$, considerando $A \in \mathcal{F}$ também é uma medida de probabilidade.

Proposição (Regra do Produto): Para eventos $A_1, A_2, \ldots, A_n \in \mathcal{F}$ com $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) > 0$, vale que

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\ldots\mathbb{P}(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

Lei da Probabilidade Total

Definição (Partição de um espaço amostral): Os eventos $C_1, \ldots, C_n \in \mathcal{F}$ formam uma *partição* de Ω se são dois-a-dois disjuntos e sua união é Ω, ou seja, $C_i \cap C_j = \emptyset$, se $i \neq j$, e $\displaystyle\bigcup_{i=1}^{n} C_i = \Omega$.

Teorema (Lei da Probabilidade Total): Suponha que os eventos $C_1, \ldots, C_n \in \mathcal{F}$ formem uma partição de Ω e todos têm probabilidade positiva. Então para qualquer evento $A \in \mathcal{F}$ vale que

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|C_i)\mathbb{P}(C_i).$$

Exemplos – 1

Uma caixa contém três cartas: uma é vermelha em ambos os lados, uma é azul em ambos os lados e a outra é vermelha de um lado e azul do outro. Uma carta é escolhida ao acaso de tal caixa, e um lado de tal carta é observado ao acaso. Se tal lado da carta é azul, qual a probabilidade do outro lado da carta também ser azul?

Exemplos – 2

Uma caixa contém três moedas com "cara" em ambos os lados, quatro moedas com "coroa" em ambos os lados e duas moedas honestas. Se uma dessas nove moedas é escolhida ao acaso e lançada uma vez, qual é a probabilidade de que uma "cara" seja observada?

Teorema de Bayes

Teorema (Teorema de Bayes): Suponha que os eventos $C_1, \ldots, C_n \in \mathcal{F}$ formem uma partição de Ω e todos têm probabilidade positiva. Então para qualquer evento $A \in \mathcal{F}$ com $\mathbb{P}(A) > 0$ vale que

$$\mathbb{P}(C_j|A) = \frac{\mathbb{P}(A|C_j)\mathbb{P}(C_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|C_j)\mathbb{P}(C_j)}{\sum_{i=1}^{n} \mathbb{P}(A|C_i)\mathbb{P}(C_i)}$$

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

Exemplos – 1

Considere um teste clínico para diagnosticar determinada doença que seja 99% sensível e 99% específico (ou seja, o teste produz 99% de verdadeiros positivos e 99% de verdadeiros negativos). Suponha que 0,5% da população seja portadora de tal doença.

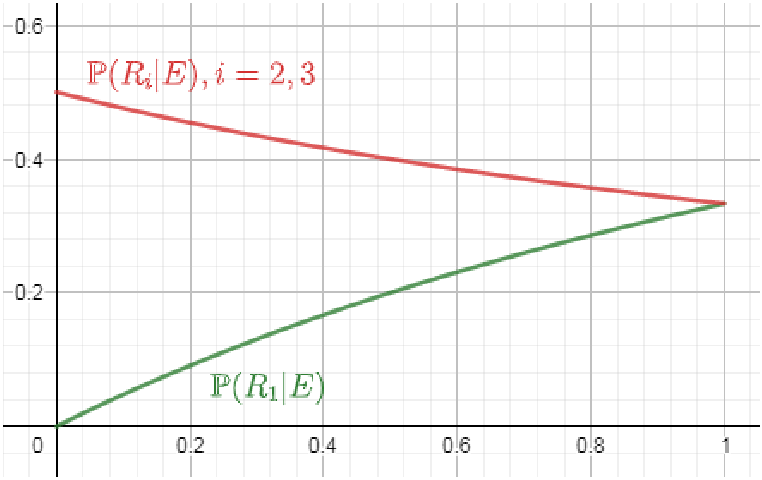
a) Se um indivíduo selecionado aleatoriamente testar positivo, qual a probabilidade de ele ter a doença, ou seja, qual a probabilidade do teste não cometer um falso positivo?

b) Se tal indivíduo realizar um novo teste, independente do primeiro, e o teste também der positivo, qual a nova probabilidade de ele ter a doença?

Exemplo – 2

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



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 $\mathbb{P}(B|A)$!

Independência entre eventos

- $(\Omega, \mathcal{F}, \mathbb{P})$ espaço de probabilidade, A e B eventos em $\mathcal{F}$
- Intuitivamente, A e B são independentes se a ocorrência de B não altera a probabilidade de ocorrência de A e vice-versa
- Matematicamente,

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

Equivalente a

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

Definição: Dizemos que os eventos A e B em $\mathcal{F}$ são *independentes* se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definição: Dizemos que os eventos $A_1, A_2, \ldots, A_n \in \mathcal{F}$ são *independentes* se para toda escolha de índices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, e $2 \leq k \leq n$ tivermos

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k}).$$

Exemplo

Considere um alfabeto que tem um total de n letras. Dentre todas as palavras formadas com 3 letras escolhemos uma delas, ao acaso. Seja s uma particular letra desse alfabeto. Defina os eventos

- $A =$ "palavra escolhida começa com a letra s
- $B =$ "palavra escolhida tem a letra s no meio"
- $C =$ "palavra escolhida tem exatamente duas letras iguais"

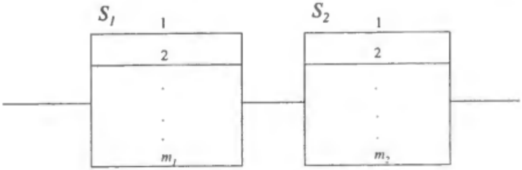
Provemos que esses eventos são *dois a dois independentes*, porém não são (coletivamente) independentes

É possível construir também um exemplo onde $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ porém eles não são independentes dois a dois (tente!)

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!

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!

Variáveis aleatórias - definição

Definição: Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Uma *variável aleatória* é uma função $X : \Omega \to \mathbb{R}$ satisfazendo

$$X^{-1}(I) := \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{F},$$

para todo intervalo $I \subset \mathbb{R}$.

Intuitivamente, a informação na σ-álgebra $\mathcal{F}$ deve ser compatível com cálculo de probabilidades sobre intervalos.

Observação: Cuidado com a nomenclatura! X não é variável (pois é função), nem aleatória (pois tem uma regra de formação bem definida)!

Exemplo

- Lançamento de um dado...
- ...porém somos informados somente a paridade do número
- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$
- Considere as funções abaixo:

$$
\begin{aligned}
X :& \Omega \to \mathbb{R} \\
& \omega \mapsto X(\omega) = 1 \text{ se o número é par e } 0 \text{ se é ímpar} \\
Y :& \Omega \to \mathbb{R} \\
& \omega \mapsto Y(\omega) = \text{valor observado no lançamento}
\end{aligned}
$$

- Ambas são variáveis aleatórias?

Notação

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

$$\{\omega \in \Omega \mid X(\omega) = x\}$$
$$\{\omega \in \Omega \mid X(\omega) \leq x\}$$
$$\{\omega \in \Omega \mid X(\omega) \geq x\}$$
$$\{\omega \in \Omega \mid X(\omega) \in I\}$$

$\Rightarrow$ Jamais escreveremos $\mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq x\})$, mas sim $\mathbb{P}(X \leq x)$

Função de probabilidade acumulada

Definição: Seja X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$. 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, $\mathbb{P}(X \in B)$ pode ser obtida a partir de F_X, para todo $B \in \mathcal{B}(\mathbb{R})$.

Função de probabilidade acumulada

Teorema (Propriedades): Uma função de probabilidade acumulada de uma variável aleatória X satisfaz às propriedades abaixo:

1) $\lim_{x \to -\infty} F_X(x) = 0$ e $\lim_{x \to +\infty} F_X(x) = 1$
2) F_X é contínua à direita, ou seja, se $x_n \downarrow x$ então $F_X(x_n) \downarrow F_X(x)$
3) F_X é não decrescente, ou seja, se $x \leq y$ então $F_X(x) \leq F_X(y)$

Teorema: Se uma função $F : \mathbb{R} \to \mathbb{R}$ satisfaz às propriedades 1), 2) e 3) acima então existe um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e uma variável aleatória X nesse espaço tal que F é a função de probabilidade acumulada de X.

$\Rightarrow$ Para fins práticos, basta informar F_X!

Variáveis aleatórias discretas

Definição: Uma variável aleatória X definida sobre um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é dita *discreta* se a função X assume apenas uma quantidade enumerável de valores.

$$X(\Omega) = \{x_1, x_2, \dots\} \subset \mathbb{R}.$$

Definição: Definimos a *função de probabilidade* de X como sendo a probabilidade de X assumir algum valor em particular, ou seja,

$$p(x_i) = \mathbb{P}(X = x_i) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x_i\}).$$

Proposição: A função de probabilidade de X em $(\Omega, \mathcal{F}, \mathbb{P})$ satisfaz:

1) $0 \leq p(x_i) \leq 1, \forall i = 1, 2, \ldots$
2) $\sum_i p(x_i) = 1$

Podemos relacionar as funções de probabilidade acumulada e de probabilidade como:

$$F_X(x) = \sum_{\{i \mid x_i \leq x\}} p(x_i)$$

$$p(x_i) = F_X(x_i) - F_X(x_i^-),$$

onde $F_X(x_i^-) = \lim_{x \to x_i^-} F_X(x)$.

Exemplo: Considere o lançamento de uma moeda honesta duas vezes e estude a variável aleatória que conta o número de caras nos dois lançamentos.

Variáveis aleatórias contínuas

Definição: Uma variável aleatória X definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é dita *contínua* se existe uma função f_X não-negativa tal que a função de probabilidade acumulada de X, denotada por F_X, pode ser escrita como

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \forall x \in \mathbb{R}.$$

A função f_X é dita a *função densidade de probabilidade* de X.

Proposição: A função densidade de probabilidade f_X de uma variável aleatória X satisfaz:

1) $f_X(x) \geq 0, \forall x \in \mathbb{R}$
2) $\displaystyle\int_{-\infty}^{+\infty} f(t)\, dt = 1$
3) $F'_X(x) = f_X(x)$; em particular, F_X é uma função contínua

Conforme bem lembramos, vale que:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b)$$
$$= \int_a^b f_X(x)\, dx = F_X(b) - F_X(a).$$

Variáveis aleatórias mistas

Exemplo: Estude a variável aleatória cuja função de probabilidade acumulada é dada por

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0 \\ x/4, & \text{se } 0 \leq x < 1 \\ 1/2, & \text{se } 1 \leq x < 2 \\ 1, & \text{se } x \geq 2. \end{cases}$$

Variáveis aleatórias singulares

É possível construir uma função de probabilidade acumulada F que é contínua tal que $F'(x) = 0$, "para todo" $x \in \mathbb{R} \Rightarrow$ variável aleatória *singular*

Teorema: Toda função de probabilidade acumulada F pode ser decomposta como a ponderação de uma contínua, uma discreta e uma singular, ou seja,

$$F(x) = \alpha_d F^d(x) + \alpha_c F^c(x) + \alpha_s F^s(x),$$

com $\alpha_d + \alpha_c + \alpha_s = 1$ e $\alpha_d, \alpha_c, \alpha_s \geq 0$.

Exemplos – 1

Um experimento consiste em retirar cartas de um baralho de 52 cartas, bem embaralhado, até obter-se um ás. Seja X a variável aleatória que conta o número de cartas removidas. Qual é a função de probabilidade de X?

Exemplos – 2

Um graveto de comprimento unitário é partido em um ponto ao acaso, dando origem a dois pedaços. Qual é a probabilidade da razão do menor comprimento pro maior ser menor que ou igual a a, para qualquer $0 < a < 1$?

Distribuição de Bernoulli

Definição: Dizemos que $X \sim \text{Bern}(p)$ se

$$\begin{cases} \mathbb{P}(X = 0) & = 1 - p \\ \mathbb{P}(X = 1) & = p. \end{cases}$$

Sua função de probabilidade é dada por

$$\mathbb{P}(X = x) = p^x (1 - p)^{1-x}, \text{ para } x = 0, 1.$$

Distribuição Binomial

Definição: Dizemos que $X \sim \text{Binom}(n, p)$ se

$$\mathbb{P}(X = x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$$

Conta o número de sucessos (observações iguais a 1) em n repetições independentes de Bernoulli.

Exemplo: (O problema dos pontos) Dois jogadores A e B, igualmente qualificados, jogam um jogo no qual um total de seis pontos são necessários para se declarar um vencedor. A cada rodada, apenas um dos dois jogadores marca um ponto. Por algum incidente, o jogo precisou ser interrompido quanto o jogador A tem cinco pontos e o jogador B tem três pontos. Como deve ser uma divisão honesta do prêmio entre os dois jogadores?

Distribuição de Poisson

- Proposta pelo matemático francês Siméon Denis Poisson (1781–1840)
- Livro de 1837: *“Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile”* (Pesquisa sobre as Probabilidades em Vereditos Civis e Criminais)
- Introduzida indiretamente, e com pouca importância – ocupou somente uma página de seu livro
- Importância salientada em 1989, por L. von Bortkiewicz
- Útil para modelar número de eventos ocorridos em um intervalo fixo de tempo ou espaço, caso ocorram com uma taxa constante e conhecia e de modo independente do tempo desde a última ocorrência
- Iremos deduzir como limite de uma distribuição Binomial

Exemplo: Um comerciante acredita que consumidores chegam em sua loja a uma taxa média de 4,5 consumidores por hora. Seja X o número de consumidores que irão chegar durante um intervalo particular de uma hora mais tarde nesse dia. Estude a variável aleatória X, assumindo independência no número de consumidores chegando em intervalos de tempo disjuntos.

Definição: Dizemos que $X \sim \text{Poi}(\lambda)$ se

$$\mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0, 1, \ldots$$

Exemplo: Uma companhia de seguros introduziu uma nova apólice que cobre certas formas de lesão pessoal com uma indenização de \$ 100.000,00, ao custo de \$ 25,00 ao ano. Em média a seguradora paga 100 dessas indenizações anualmente. Sabendo-se que há mais de um milhão de segurados, qual a probabilidade que a seguradora precise pagar mais de 15 milhões de dólares em um ano?

Distribuição Geométrica

Definição: Dizemos que $X \sim \text{Geo}(p)$ se

$$\mathbb{P}(X = x) = p(1-p)^x, x = 0, 1, \ldots$$

Conta número de fracassos antes do primeiro sucesso em n repetições independentes de Bernoulli.

Observação: Também pode ser parametrizada contando o número de tentativas até o primeiro sucesso. Nesse caso, sua função de probabilidade é dada por:

$$p(1-p)^{x-1}, x = 1, 2, \ldots.$$

Sempre usaremos a parametrização com o número de fracassos, mas fique atento na hora de consultar a literatura!

Distribuição Geométrica

Exemplo: Uma linha de fabricação de um equipamento de precisão é interrompida na primeira ocorrência de um defeito. A partir da manutenção, o equipamento tem probabilidade de 0,01 de apresentar defeito em um dia qualquer. Deseja-se planejar o cronograma de manutenção preventiva e, para tal, decidiu-se avaliar probabilisticamente a espera até a produção ser interrompida. Qual seria o intervalo ideal para uma manutenção preventiva, se desejarmos uma probabilidade de, pelo menos 0,90 de que o defeito não ocorrerá?

Distribuição Geométrica

Teorema: (Perda de memória) A distribuição Geométrica é a única distribuição discreta satisfazendo a *propriedade de perda de memória*:

$$\mathbb{P}(X \geq m + n | X \geq m) = \mathbb{P}(X \geq n).$$

A variável "lembra" do presente, mas se "esquece" do passado.

Distribuição Exponencial

Exemplo: Uma fonte radioativa emite partículas com uma taxa λ. Se uma emissão acabou de ser observada, estude a distribuição de probabilidade que registra o tempo até a próxima emissão.

Distribuição exponencial

Definição: Dizemos que $X \sim \text{Exp}(\lambda)$ se

$$f_X(x) = \lambda e^{-\lambda x}, \text{ para } x \geq 0.$$

Observação: Uma outra forma de deduzir a distribuição exponencial é estudando o tempo até a ocorrência de um evento raro.

Observação: Analogamente à Poisson, λ é dito a taxa. Uma parametrização alternativa é em função da escala $\beta = 1/\lambda$.

Teorema: (Perda de memória) A distribuição Exponencial é a única distribuição contínua satisfazendo a *propriedade de perda de memória*:

$$\mathbb{P}(X \geq t + s | X \geq s) = \mathbb{P}(X \geq t).$$

Exemplo: O Chile encontra-se em uma das regiões do mundo que apresentam os maiores índices de ocorrência de terremotos. Suponha que o intervalo de tempo T entre a ocorrência de dois terremotos sucessivos possa ser modelado por uma variável aleatória exponencial com parâmetro λ (em 1/mes) desconhecido. Com base em dados históricos sabe-se que a probabilidade de que o intervalo de tempo entre dois terremotos exceda um ano é de 0,001%.

i) Usando a informação contida nos dados históricos, calcule o valor de λ.

ii) Calcule a probabilidade de que ocorram no máximo dois terremotos em um período de tempo de 6 meses.

Distribuição Normal

Definição: Dizemos que $X \sim N(\mu, \sigma^2)$ se

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

Observação: Primeira dedução: de Moivre, 1738: estudo dos coeficientes de $(a+b)^n$. Redescoberta por Gauss em 1809 (junto com o método de mínimos quadrados e o estimador de máxima verossimilhança), para medir erros de observações astronômicas.

Distribuição Normal

Teorema: Seja $X \sim N(\mu, \sigma^2)$. Então vale que:

- f_X é uma função densidade de probabilidade
- $\mathbb{E}[X] = \mu$ e $\mathbb{V}(X) = \sigma^2$
- $aX + b \sim N(a\mu + b, a^2\sigma^2)$, se $a \neq 0$
- Em particular, $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

Exemplo: Em um sistema monitorado de armazenamento, o estoque disponível de determinado item é elevado para S ao início de cada semana. A demanda pelo item em semanas sucessivas é modelado por variáveis aleatórias independentes, cada uma tendo distribuição normal com média $\mu = 250$ unidades e desvio padrão $\sigma = 45$ unidades. Qual é o valor de S tal que a proporção de semanas nas quais há falta de tal produto não supera $\alpha = 0{,}05$?

Definição: Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, um *vetor aleatório* é uma função $\mathbf{X} : \Omega \to \mathbb{R}^m$, representada por

$$\mathbf{X}(\omega) = (X_1(\omega), \ldots, X_m(\omega)),$$

onde cada componente X_i é uma variável aleatória.

Equivalente a dizer que

$$\bigcap_{i=1}^{m} \{\omega \in \Omega \mid X_i(\omega) \leq x_i\} = \bigcap_{i=1}^{m} X_i^{-1}(-\infty, x_i] \in \mathcal{F},$$

para todo $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$.

Função de probabilidade acumulada

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

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{x}}(x_1, \ldots, x_m) = \mathbb{P}(X_1 \leq x_1, \ldots, X_m \leq x_m), \forall \mathbf{x} \in \mathbb{R}^m.$$

Teorema: Valem as propriedades abaixo:

i) $F_{\mathbf{X}}(\mathbf{x})$ é não-decrescente em cada uma de suas coordenadas

ii) $F_{\mathbf{X}}(\mathbf{x})$ é contínua à direita em cada uma de suas coordenadas

iii) Se $x_j \to -\infty$ para algum j, então $F_{\mathbf{X}}(\mathbf{x}) \to 0$; se $x_j \to \infty$ para todo j, então $F_{\mathbf{X}}(\mathbf{x}) \to 1$

Função de probabilidade acumulada

Atenção! Ao contrário do caso univariado, essas três propriedades não implicam que existe um vetor aleatório cuja função de probabilidade acumulada seja tal.

Exemplo: A função

$$G(x,y) = \begin{cases} 1, & \text{se } x \geq 0, y \geq 0 \text{ e } x + y \geq 1 \\ 0, & \text{caso contrário} \end{cases}$$

satisfaz às três propriedades acima, porém não é função de probabilidade acumulada de nenhum vetor aleatório!

Função de probabilidade acumulada

Falta uma última propriedade adicional!

- Seja I o intervalo $(a, b]$
- Defina o operador

$$\begin{aligned}\Delta_{i,I} F_{\mathbf{X}}(\mathbf{x}) &= F_{\mathbf{X}}(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_m) - \\ &\qquad F_{\mathbf{X}}(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_m) \\ &= \mathbb{P}(X_1 \leq x_1, \ldots, X_{i-1} \leq x_{i-1}, a < X_i \leq b, \\ &\qquad X_{i+1} \leq x_{i+1}, \ldots, X_m \leq x_m),\end{aligned}$$

para $i = 1, \ldots, m$.

Função de probabilidade acumulada

Teorema: A função de probabilidade acumulada de um vetor aleatório $\mathbf{X}$ também satisfaz

iv) $0 \leq \Delta_{1,I_1} \ldots \Delta_{m,I_m} F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(a_1 < X_1 \leq b_1, \ldots, a_m < X_1 \leq b_m)$

Agora sim, pode-se provar que uma função $F : \mathbb{R}^m \to \mathbb{R}$ satisfazendo as propriedades i) – iv) é necessariamente a função de probabilidade acumulada de um vetor aleatório $\mathbf{X}$ de dimensão m.

Vetores aleatórios discretos

Se as componentes do vetor aleatório $\mathbf{X}$ são discretas, temos um *vetor aleatório discreto*. Sua *função de probabilidade conjunta* é definida por

$$p_{\mathbf{X}}(\mathbf{x}) = p(x_1, \ldots, x_m) = \mathbb{P}(X_1 = x_1, \ldots, X_m = x_m).$$

Teorema: A função de probabilidade conjunta de um vetor aleatório discreto $\mathbf{X}$ satisfaz às propriedades

i) $p_{\mathbf{X}}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^m$

ii) $\sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) = 1$

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
0	$1/16$	$1/16$	$1/8$
1	$1/8$	$1/8$	0
2	$1/16$	$1/8$	$1/8$
3	0	$1/8$	$1/16$

Através dessa tabela, podemos calcular facilmente probabilidades como:

- $\mathbb{P}(X = Y)$
- $\mathbb{P}(X > Y)$
- $\mathbb{P}(X > 1)$
- $\mathbb{P}(Y = 2)$

A distribuição multinomial

Considere um experimento com m possíveis resultados, cada um com probabilidade $p_i \geq 0$, para $i = 1, 2, \ldots, m$ e $\sum_{i=1}^{m} p_i = 1$. Tal experimento é repetido n vezes, de forma independente e observamos as variáveis $X_1, X_2, \ldots, X_m$, que correspondem ao número de ocorrências de cada um dos possíveis resultados dessas repetições. O vetor aleatório $\mathbf{X} = (X_1, \ldots, X_m)$ segue o modelo *multinomial* de parâmetros n e $\mathbf{p} = (p_1, \ldots, p_m)$. Sua função de probabilidade conjunta é dada por

$$p_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 = x_1, \ldots, X_m = x_m) = \frac{n!}{x_1! \ldots x_m!} p_1^{x_1} \ldots p_m^{x_m},$$

sendo que $\sum_{i=1}^{m} x_i = n$, para $0 \leq x_i \leq n$.

Exemplo: Considere 10 lançamentos independentes de um dado equilibrado, e seja X_i o número de ocorrências da face i, para $i \in \{1, 2, 3, 4, 5, 6\}$.

Vetores aleatórios contínuos

Definição: O vetor aleatório $\mathbf{X}$ é dito *contínuo* se existe uma função $f_{\mathbf{X}} : \mathbb{R}^m \to \mathbb{R}$, dita a *função densidade de probabilidade conjunta*, tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_{\mathbf{X}}(\mathbf{t})\ dt_m \ldots dt_1.$$

Atenção! NÃO é verdade que se as componentes de um vetor aleatório $\mathbf{X}$ são variáveis aleatórias contínuas, então $\mathbf{X}$ será um vetor aleatório contínuo!

Vetores aleatórios contínuos

Teorema: A função densidade de probabilidade conjunta de um vetor aleatório contínuo $\mathbf{X}$ satisfaz às propriedades:

i) $f_{\mathbf{X}}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^m$

ii) $\displaystyle\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(\mathbf{t})\, dt_m \ldots dt_1 = 1$

iii) $\displaystyle f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^m}{\partial x_1 \ldots \partial x_m} F_{\mathbf{X}}(x_1, \ldots, x_m)$

Vetores aleatórios contínuos

Seja $B \in \mathcal{B}(\mathbb{R}^m)$ um boreliano de $\mathbb{R}^m$. Então podemos calcular

$$\mathbb{P}(\mathbf{X} \in B) = \int \cdots \int_B f_{\mathbf{X}}(\mathbf{x})\ dx_m \ldots dx_1.$$

Ou seja, probabilidades são calculadas com integrais múltiplas.

Exemplo

Exemplo: Considere a função abaixo:

$$F(x,y) = \begin{cases} 0, & \text{se } x < 0 \text{ ou } y < 0; \\ \frac{x}{5}(1 - e^{-y}), & \text{se } 0 \leq x < 5 \text{ e } y \geq 0; \\ 1 - e^{-y}, & \text{se } x \geq 5 \text{ e } y \geq 0. \end{cases}$$

Verifique que F é uma função de probabilidade acumulada e encontre a função densidade de probabilidade associada.

Função de probabilidade acumulada marginal

- Seja $\mathbf{X} \in \mathbb{R}^m$ um vetor aleatório e o decomponha como $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, com $\mathbf{Y} \in \mathbb{R}^k$ e $\mathbf{Z} \in \mathbb{R}^{m-k}$
- Denote as componentes: $\mathbf{Y} = (y_1, \ldots, y_k)$ e $\mathbf{Z} = (z_1, \ldots, z_{m-k})$

Definição: A *função de probabilidade acumulada marginal* de $\mathbf{Y}$ é definida por

$$F_{\mathbf{Y}}(\mathbf{y}) = \lim_{\substack{z_i \to \infty \\ \forall i=1,\ldots,m-k}} F_{\mathbf{X}}(\mathbf{x}).$$

Exemplo

Exemplo: Dada a função de probabilidade acumulada abaixo

$$F(x,y) = \begin{cases} 0, & \text{se } x < 0 \text{ ou } y < 0; \\ \frac{x}{5}(1 - e^{-y}), & \text{se } 0 \le x < 5 \text{ e } y \ge 0; \\ 1 - e^{-y}, & \text{se } x \ge 5 \text{ e } y \ge 0, \end{cases}$$

encontre as acumuladas marginais de X e Y.

Distribuições marginais

Teorema: Sendo $\mathbf{X}$ um vetor aleatório discreto, a *função de probabilidade marginal* de $\mathbf{y}$ é dada por

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{z}} p_{\mathbf{X}}(\mathbf{y}, \mathbf{z}),$$

ou seja, somamos $p_{\mathbf{X}}$ somente nas coordenadas que não desejamos.

Teorema: Sendo $\mathbf{X}$ um vetor aleatório contínuo, a *função densidade de probabilidade marginal* de $\mathbf{y}$ é dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = \int \cdots \int_{\mathbb{R}^{m-k}} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) \; dz_{m-k} \dots dz_1,$$

ou seja, integramos $f_{\mathbf{X}}$ somente nas coordenadas que não desejamos.

Distribuições condicionais

Relembre a decomposição $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, com $\mathbf{Y} \in \mathbb{R}^k$ e $\mathbf{Z} \in \mathbb{R}^{m-k}$

- Se $\mathbf{X}$ é discreto, podemos calcular a *função de probabilidade condicional* de $\mathbf{Z}$, condicionado na observação que $\mathbf{Y} = \mathbf{y}$:

$$p_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}) = \frac{p_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{p_{\mathbf{Y}}(\mathbf{y})} = \frac{p_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{\sum_{\mathbf{z}} p_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}.$$

- Se $\mathbf{X}$ é contínuo, então definimos a *função densidade de probabilidade condicional* de $\mathbf{Z}$, condicionado na observação que $\mathbf{Y} = \mathbf{y}$:

$$f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{\int \cdots \int_{\mathbb{R}^{m-k}} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})\, dz_{m-k} \ldots dz_1}.$$

Um breve exemplo

Exemplo: As variáveis aleatórias X e Y tem função densidade de probabilidade conjunta dada por

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

Encontre a densidade condicional de $X|(Y=y)$.

Independência

Definição: Dizemos que as variáveis aleatórias $X_1, \dots, X_m$ são *independentes* se

$$\mathbb{P}(X_1 \in B_1, \dots, X_m \in B_m) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_m \in B_m), \forall B_i \in \mathcal{B}(\mathbb{R}).$$

Teorema: Denote por $F_{\mathbf{X}}$ a função de probabilidade acumulada do vetor aleatório $\mathbf{X} = (X_1, \dots, X_n)$, $f_{\mathbf{X}}$ a função densidade probabilidade ou função de probabilidade de $\mathbf{X}$, F_i a função de probabilidade acumulada de X_i e f_i a função densidade probabilidade marginal ou função de probabilidade marginal de X_i, para $i = 1, \dots, n$. Então temos que $X_1, \dots, X_n$ são independentes se e somente se vale alguma das condições abaixo:

- $F_{\mathbf{X}}(x_1, \dots, x_m) = F_1(x_1) \times \cdots \times F_m(x_m), \forall \mathbf{x} \in \mathbb{R}^m$
- $f_{\mathbf{X}}(x_1, \dots, x_m) = f_1(x_1) \times \cdots \times f_m(x_m), \forall \mathbf{x} \in \mathbb{R}^m$
- $f_{\mathbf{X}}(x_1, \dots, x_m) = h_1(x_1) \times \cdots \times h_m(x_m), \forall \mathbf{x} \in \mathbb{R}^m$, onde h_i é uma função não-negativa dependendo somente de x_i

Exemplos

A função densidade de probabilidade conjunta do vetor aleatório (X, Y) é dada por

$$f(x,y) = \begin{cases} ke^{-(x+2y)} & \text{para } x \geq 0 \text{ e } y \geq 0 \\ 0 & \text{caso contrário.} \end{cases}$$

Encontre o valor adequado de k, as distribuições marginais e diga se X e Y são independentes.

Exemplos

A função densidade de probabilidade conjunta do vetor aleatório (X, Y) é dada por

$$f(x,y) = \begin{cases} 24xy & \text{para } x \geq 0, y \geq 0, \text{ e } x + y \leq 1 \\ 0 & \text{caso contrário.} \end{cases}$$

As variáveis aleatórias X e Y são independentes?

Exemplo

Suponha que n consumidores chegam em uma fila e estão aguardando atendimento. Seja X_i o tempo de atendimento ao consumidor i, para $i = 1, \ldots, n$. Assuma que o vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)$ tem função densidade de probabilidade conjunta dada por

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{c}{(2+\sum_{i=1}^{n} x_i)^{n+1}}, & \text{se } x_1, \ldots, x_n > 0; \\ 0, & \text{caso contrário.} \end{cases}$$

i) Encontre o valor de c.

ii) Assumindo $n = 5$, encontre a marginal de (X_1, X_4), e também de cada X_i. Calcule a probabilidade de $X_i > 3$.

iii) As componentes do vetor aleatório $\mathbf{X}$ são independentes?

iv) Encontre a distribuição condicional de $\mathbf{Y} = (X_3, X_4, X_5)$ dado $\mathbf{Z} = (X_1, X_2)$. Calcule a probabilidade de $X_3 > 3$ dado que $X_1 = 4$ e $X_2 = 6$.

Exemplo

Seja Z uma variável aleatória cuja FDP é dada por

$$f_Z(z) = \begin{cases} 2e^{-2z}, & \text{para } z > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Suponha que, dado $Z = z$, as duas variáveis aleatórias X_1 e X_2 são independentes com a mesma FDP dada por

$$f_{X|Z}(x|z) = \begin{cases} ze^{-zx}, & \text{para } x > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre a marginal de (X_1, X_2), e verifique se são independentes.
Observação: Z pode ser interpretada como a taxa de atendimento aos consumidores no exemplo anterior. Podemos também calcular a distribuição de $Z|(X_1, X_2)$, codificando o conhecimento sobre a taxa de atendimento após observar os tempos de atendimentos a dois consumidores.

Exemplo

O número de pessoas N que entra em uma loja em um determinado dia segue uma distribuição de Poisson de parâmetro λ. Sabe-se que uma proporção p dos clientes é homem e uma proporção $1-p$ é mulher. Seja X o número de clientes homens e Y o número de clientes mulheres que entram na loja em um determinado dia, de modo que $X+Y=N$. Encontre a função de probabilidade conjunta de X e Y.

Funções de variáveis e vetores aleatórios

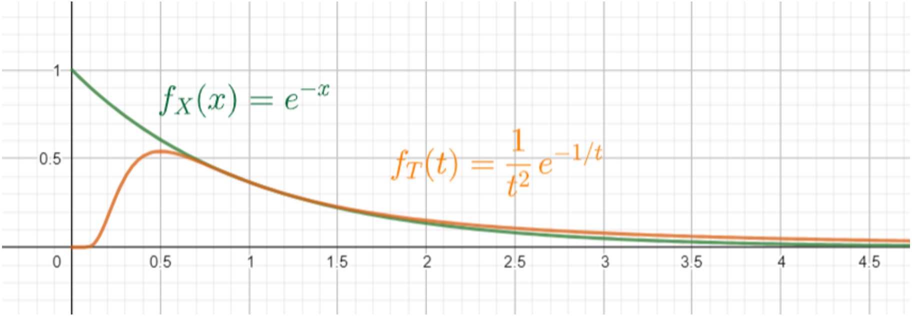
Se $\mathbf{X}$ é um vetor aleatório em $\mathbb{R}^m$ e $g : \mathbb{R}^m \to \mathbb{R}^k$ é uma função, então podemos estar interessados na distribuição de $\mathbf{Y} = g(\mathbf{X})$:

- Se D é o diâmetro de um rolamento esférico, então seu volume é dado por $\dfrac{\pi D^3}{6}$
- Se V é a voltagem em um circuito elétrico onde passa uma corrente i (valor fixo conhecido), então a potência é dada por $W = Vi$
- Se X e Y são peso e altura, respectivamente, de indivíduos em uma população, então o IMC é dado por X/Y^2

Exemplo: Seja X a taxa com a qual consumidores chegam em uma fila, e seja $T = 1/X$ o tempo médio de espera. Quem é a distribuição de T, a partir da distribuição de X?

Exemplo

No caso particular em que $X \sim \text{Exp}(1)$, temos que $f_T(t) = \dfrac{1}{t^2} e^{-1/t}$, para $t > 0$:



Exemplo

Se $X \sim \text{Unif}[-1, 1]$, qual é a distribuição de $Y = X^2$?

Funções (diferenciáveis e bijetivas) de va's

Teorema: Seja X uma variável aleatória contínua tomando valores no intervalo (a, b), e seja $Y = g(X)$, onde g é uma função bijetiva e diferenciável. Denote por (α, β) a imagem de (a, b) através de g, e por $h = g^{-1}$ a função inversa de g. Então Y é uma variável aleatória contínua cuja função densidade de probabilidade é dada por

$$f_Y(y) = \begin{cases} f_X(h(y)) \left| \frac{dh(y)}{dy} \right| & \text{para } \alpha < y < \beta, \\ 0 & \text{caso contrário.} \end{cases}$$

Exemplo

Um modelo para populações de organismos microscópicos em grandes ambientes é o de crescimento exponencial aqui descrito:

- No tempo 0, v organismos são introduzidos no meio, e seja X a taxa de crescimento;
- Após um tempo t, prevemos uma população de tamanho ve^{tX}.

Assuma que X tem a seguinte função densidade de probabilidade:

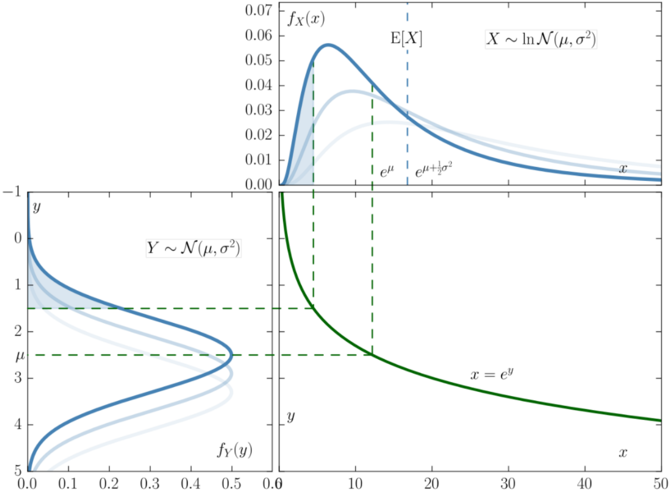
$$f_X(x) = \begin{cases} 3(1-x)^2, & \text{para } 0 < x < 1; \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre a distribuição de $Y = ve^{tX}$ para valores fixos e conhecidos de v e t, por exemplo, $v = 10$ e $t = 5$.

Exemplo

Dizemos que X tem distribuição lognormal de parâmetros μ e σ^2 se $X = e^Y$, onde $Y \sim N(\mu, \sigma^2)$. Encontre a função densidade de probabilidade de X.

Exemplo



Fonte: Wikipedia

Exemplo: Seja X_i o tempo que leva para o consumidor i ser atendido em uma fila, para $i = 1, 2$, independentes e identicamente distribuídos com distribuição Exp(2). Seja $Y = X_1 + X_2$ o tempo total de atendimento. Qual é a distribuição de Y?

Mais geralmente, temos...

Teorema: Seja $\mathbf{X}$ um vetor aleatório em $\mathbb{R}^m$ e $Y = g(\mathbf{X})$, onde $g : \mathbb{R}^m \to \mathbb{R}$. Para cada $y \in \mathbb{R}$ defina $A_y = \{\mathbf{x} \in \mathbb{R}^m \mid g(\mathbf{x}) \leq y\}$. Então temos que

$$F_Y(y) = \int \cdots \int_{A_y} f_{\mathbf{X}}(\mathbf{x})\ d\mathbf{x}.$$

Exemplo

O tempo X_i de atendimento do consumidor i em uma fila, medido em minutos, tem distribuição exponencial, para $i = 1, \ldots, n$. Cada consumidor leva em média $1/\lambda_i$ minutos para ser atendido. Assuma independência entre os tempos de atendimento de consumidores distintos. Encontre a distribuição das variáveis aleatórias abaixo:

a) O menor tempo de atendimento dentre os n consumidores
b) O maior tempo de atendimento dentre os n consumidores
c) O tempo de atendimento total dos n consumidores (assumindo $\lambda_1 = \cdots = \lambda_n = \lambda$)

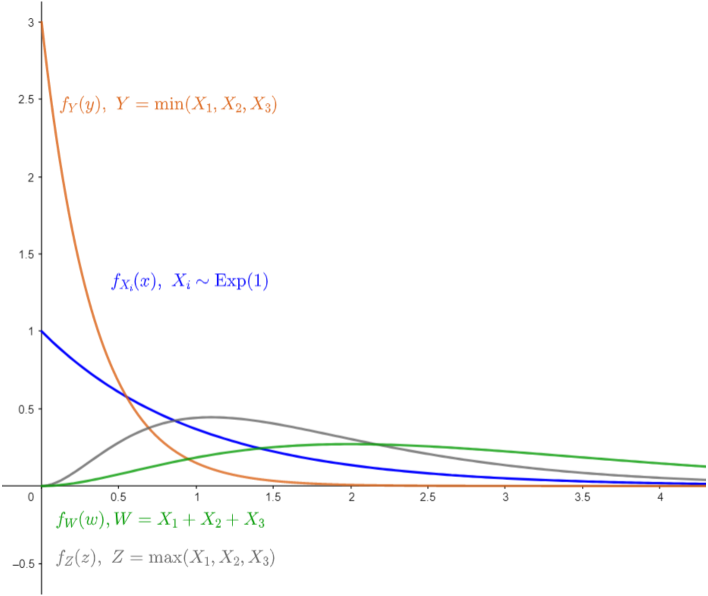
Portanto, temos que:

$$Y = \min_{1 \leq i \leq n} X_i \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$$

$$Z = \max_{1 \leq i \leq n} X_i \sim \text{distribuição sem nome}$$

$$W = X_1 + \cdots + X_n \sim \Gamma(n, \lambda) \text{ assumindo } \lambda_1 = \cdots = \lambda_n = \lambda$$

Exemplo



$f_Y(y),\ Y = \min(X_1, X_2, X_3)$
$f_{X_i}(x),\ X_i \sim \text{Exp}(1)$
$f_W(w), W = X_1 + X_2 + X_3$
$f_Z(z),\ Z = \max(X_1, X_2, X_3)$
3
2.5
2
1.5
1
0.5
0
−0.5
0.5
1
1.5
2
2.5
3
3.5
4

Exemplo: Seja S o paralelogramo limitado por $y = 2x$, $y = 2x - 2$, $y = x$ e $y = x + 1$. Calcule

$$\iint_S xy \; dxdy.$$

Funções vetoriais de vetores aleatórios ($g : \mathbb{R}^m \to \mathbb{R}^m$)

Teorema: Seja $\mathbf{X} = (X_1, \ldots, X_m)$ um vetor aleatório contínuo tomando valores em $S \subset \mathbb{R}^m$, e seja $\mathbf{Y} = g(\mathbf{X})$, onde $g : \mathbb{R}^m \to \mathbb{R}^m$ é uma função bijetiva e diferenciável, ou seja,
$Y_i = g_i\left(X_1, \ldots, X_m\right),\ i = 1, \ldots, m$. Denote por T a imagem de S através de g, e denote a inversa de g da seguinte forma:

$$x_i = h_i(y_1, \ldots, y_m),\ i = 1, \ldots, m.$$

Então o vetor aleatório $\mathbf{Y}$ é contínuo, com função densidade de probabilidade conjunta dada por

$$f_{\mathbf{Y}}\left(y_1, \ldots, y_m\right) = \begin{cases} f_{\mathbf{X}}\left(h_1(\mathbf{y}), \ldots, h_m(\mathbf{y})\right)|J| & \text{para } \mathbf{y} \in T \\ 0 & \text{caso contrário,} \end{cases}$$

onde J é o determinante

$$J = \det \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial y_1} & \cdots & \frac{\partial h_m}{\partial y_m} \end{bmatrix}.$$

Exemplo

Considere o vetor aleatório $\mathbf{X} = (X_1, X_2, X_3)$ com função densidade de probabilidade conjunta dada por

$$f_{\mathbf{X}}(\mathbf{x}) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)}, \text{ para } x_1, x_2, x_3 \geq 0.$$

Encontre a função densidade de probabilidade conjunta de $\mathbf{Y} = (Y_1, Y_2, Y_3)$, onde

$$\begin{aligned}
Y_1 &= X_1 \\
Y_2 &= X_1 + X_2 \\
Y_3 &= X_1 + X_2 + X_3.
\end{aligned}$$

Exemplo

Considere o vetor aleatório $\mathbf{X} = (X_1, X_2)$ com função densidade de probabilidade conjunta dada por

$$f_{\mathbf{X}}(\mathbf{x}) = 4x_1x_2, \text{ para } 0 < x_1, x_2 < 1.$$

Encontre a função densidade de probabilidade conjunta de $\mathbf{Y} = (Y_1, Y_2)$, onde

$$\begin{aligned} Y_1 &= X_1/X_2 \\ Y_2 &= X_1X_2. \end{aligned}$$

Exemplo

Sejam X_1 e X_2 variáveis aleatórias com distribuição normal padrão independentes. Encontre a distribuição de $X_1 + X_2$ e $X_1 - X_2$.

Exemplo

O surpreendente desse resultado, é que vale a recíproca:

Teorema: Sejam X_1 e X_2 variáveis aleatórias independentes e identicamente distribuídas com variância finita. Se $X_1 + X_2$ e $X_1 - X_2$ são independentes, então X_1 e X_2 têm distribuição normal.

Exemplo

Considere uma fila com dois consumidores, cujos tempos de atendimento X e Y, em minutos, são independentes e com distribuição exponencial de mesmo parâmetro λ. Considere as variáveis aleatórias $V = X + Y$ e $W = \dfrac{X}{X+Y}$. Note que V representa o tempo total de atendimento de ambos os consumidores e W representa a proporção do tempo total gasto no primeiro atendimento. Encontre a distribuição conjunta de V e W.

Exemplo

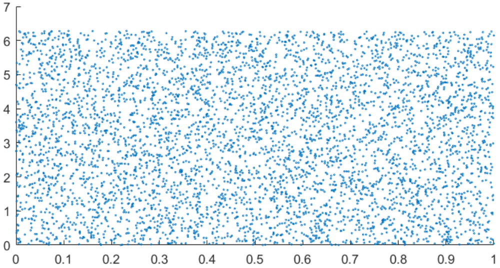
Um ponto (V, W) dentro do círculo de raio 1 centrado na origem é escolhido da seguinte forma:

- Primeiramente, um número R é escolhido ao acaso entre 0 e 1
- Em seguida um ângulo Θ é escolhido ao acaso entre 0 e 2π
- Finalmente, é feita a transformação

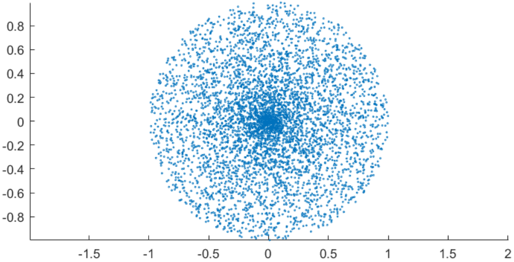
$$\begin{cases} V & = R\cos(\Theta) \\ W & = R\sin(\Theta) \end{cases}$$

Encontre a distribuição conjunta do vetor aleatório (V, W) e diga se tal distribuição é uniforme no círculo unitário.

Aplicações



7
6
5
4
3
2
1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1.5
-1
-0.5
0
0.5
1
1.5
2

Exemplo

Seja $\mathbf{X} \in \mathbb{R}^m$ um vetor aleatório com função densidade de probabilidade conjunta dada por $f_{\mathbf{X}}$. Defina $\mathbf{Y} = \mathbf{A}\mathbf{X}$, onde $\mathbf{A}$ é uma matriz $m \times m$ inversível. Mostre que a função densidade de probabilidade conjunta de $\mathbf{Y}$ é dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det(\mathbf{A})|} f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y}).$$

O que podemos falar sobre o suporte de $f_{\mathbf{Y}}$?

Exemplo

Dizemos que $\mathbf{X} \in \mathbb{R}^m$ tem distribuição normal multivariada de média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$ se

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

onde $\boldsymbol{\Sigma}$ é uma matriz simétrica e positiva-definida. Encontre a distribuição de $\mathbf{Y} = \mathbf{A}\mathbf{X}$, se $\mathbf{A}$ é uma matriz $m \times m$ inversível.

Média, valor esperado, esperança, etc...

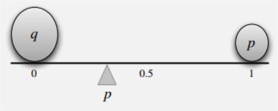
- Jogo no qual n resultados distintos podem ser obtidos $\Rightarrow$ cada com retorno x_i ao jogador, $i = 1, \ldots, n$
- Mesa cobra C para participar de uma rodada $\Rightarrow$ a mesa está levando vantagem?
- Jogar N vezes, resultado x_i tem probabilidade p_i de ocorrer $\Rightarrow$ x_i acontece 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

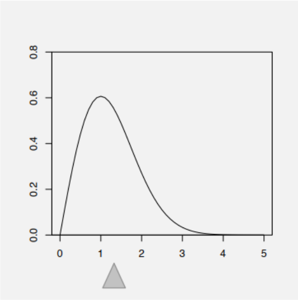
Média, valor esperado, esperança, etc...

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.





Coisas que já sabíamos mas é bom relembrar

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

O que pode acontecer com o valor esperado?

1) X discreta com função de probabilidade dada por

$$f_X(x) = \frac{1}{2|x|(|x|+1)}, \text{ para } x = \pm 1, \pm 2, \pm 3, \ldots.$$

2) X discreta com função de probabilidade dada por

$$f_X(x) = \frac{1}{x(x+1)}, \text{ para } x = 1, 2, 3, \ldots.$$

3) *(Distribuição de Cauchy)* X contínua com função densidade de probabilidade dada por

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \text{ para } x \in \mathbb{R}.$$

Quando o valor esperado existe?

Teorema: Seja X uma variável aleatória, e defina a suas *parte positiva* e *parte negativa*, respectivamente, por

$$X^+ = \max\{X, 0\} \quad \text{e} \quad X^- = -\min\{X, 0\}.$$

Então as seguintes afirmações são equivalentes:

i) $\mathbb{E}[X]$ existe e é finita

ii) Ambas $\mathbb{E}[X^+]$ e $\mathbb{E}[X^-]$ são finitas

iii) $\mathbb{E}[|X|]$ é finita

Motivação: 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, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$!

Lei do Estatístico Preguiçoso para vetores aleatórios

- Seja agora $\mathbf{X}$ um vetor aleatório em $\mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}$ uma função e $Y = g(\mathbf{X})$
- Se quisermos calcular $\mathbb{E}[Y] = \mathbb{E}[g(\mathbf{X})]$, também não precisamos encontrar a distribuição de Y primeiro!

Teorema: Seja $\mathbf{X}$ um vetor aleatório. Temos que

$$\mathbb{E}[g(\mathbf{X})] = \underbrace{\sum_{\mathbf{x}} g(\mathbf{x})\mathbb{P}(\mathbf{X} = \mathbf{x})}_{\text{se } \mathbf{X} \text{ é discreto}} = \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \; dx_1 \dots dx_m}_{\text{se } g(\mathbf{X}) \text{ é contínua}},$$

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

Exemplo

Um ponto (X, Y) é escolhido uniformemente no quadrado $[0, 1] \times [0, 1]$. Qual o valor esperado da coordenada x, da coordenada y, e do quadrado da distância até a origem?

Desigualdade de Markov

- Às vezes, por conta de termos pouca informação sobre X, não conseguimos calcular probabilidades a ela associadas, mas somente *estimar* alguns valores
- Podemos também estar interessados em estudar quanto que uma variável aleatória se desvia de um determinado valor (em particular a sua média)
- Chamaremos resultados nessa direção de *desigualdades de concentração*
- A desigualdade de concentração mais clássica é a *desigualdade de Markov*:

  Teorema: Seja X uma variável aleatória positiva e considere $t > 0$. Então

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

Exemplo: Numa empresa com 100 funcionários, o número médio de conexões simultâneas na Internet, em um certo período do dia, é de aproximadamente 30. Sabendo-se que atualmente a rede suporta no máximo 30 usuários simultaneamente, deseja-se avaliar a necessidade de aumentar esse número.

t	30	50	70	90
$\mathbb{E}[X]/t$	1	0,60	0,43	0,33

Exemplo: (Porque a desigualdade de Markov não é tão boa assim...) Considere $X \sim \text{Exp}(1/2)$. O que podemos dizer de $\mathbb{P}(X \geq t)$ para $t = 1, 2$ e 3?

Desigualdade de Markov

Note que tal desigualdade somente é interessante para valores de t suficientemente grandes, em particular, $t > \mathbb{E}[X]$.

Exemplo: Seja X uma variável aleatória com $\mathbb{E}[X] = 1$. Quanto é, no máximo, $\mathbb{P}(X \geq 100)$? Existe alguma variável aleatória X que atinja esse valor?

Propriedades

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 $\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i]$

Uma outra forma de calcular o valor esperado

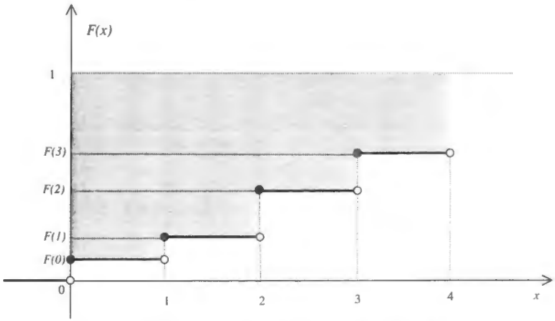
Teorema: Seja X uma variável aleatória. Se $\mathbb{E}[X]$ existe, então

$$\mathbb{E}[X] = \int_0^\infty 1 - F_X(x)\ dx - \int_{-\infty}^0 F_X(x)\ dx.$$

Interpretação gráfica – caso positivo

Se X é positiva e $\mathbb{E}[X]$ é finita, então

$$\mathbb{E}[X] = \int_0^\infty 1 - F_X(x)\ dx = \int_0^\infty \mathbb{P}(X > x)\ dx$$



Interpretação – caso discreto e positivo

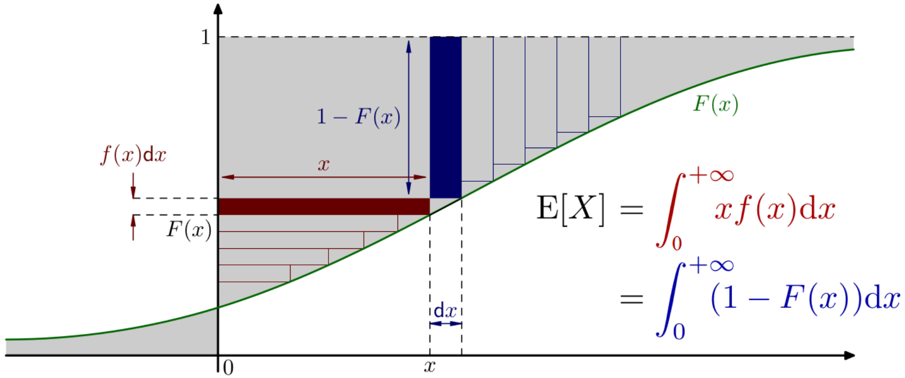
No caso discreto onde X assume os valores $0, 1, 2, \ldots$ tal expressão pode ser simplificada para

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\, dx = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$$

Intuição:

$$\begin{array}{cccccc}
\mathbb{P}(X=1) & \mathbb{P}(X=2) & \mathbb{P}(X=3) & \mathbb{P}(X=4) & \mathbb{P}(X=5) & \cdots \\
 & \mathbb{P}(X=2) & \mathbb{P}(X=3) & \mathbb{P}(X=4) & \mathbb{P}(X=5) & \cdots \\
 & & \mathbb{P}(X=3) & \mathbb{P}(X=4) & \mathbb{P}(X=5) & \cdots \\
 & & & \mathbb{P}(X=4) & \mathbb{P}(X=5) & \cdots \\
 & & & & \mathbb{P}(X=5) & \cdots
\end{array}$$

Intuição no caso contínuo:



Retornando ao caso geral...

Teorema: Seja X uma variável aleatória. Se $\mathbb{E}[X]$ existe, então

$$\mathbb{E}[X] = \int_0^\infty 1 - F_X(x)\ dx - \int_{-\infty}^0 F_X(x)\ dx.$$

Desigualdade de Jensen

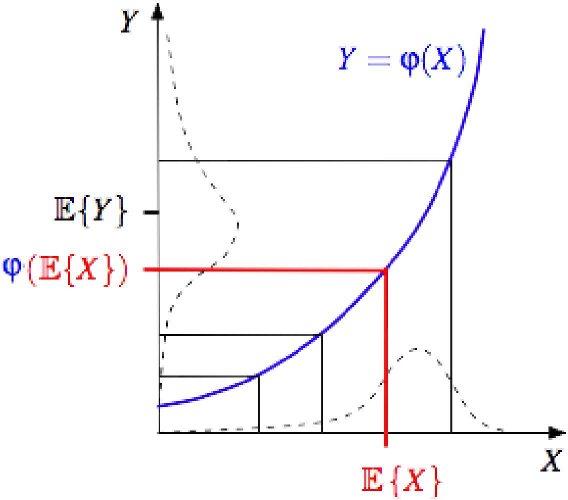
- Em geral, não há uma relação óbvia entre $\mathbb{E}[g(X)]$ e $g(\mathbb{E}[X])$
- Porém, sabemos relacionar tais quantidades em um caso particular

Definição: Dizemos que uma função $g : \mathbb{R} \to \mathbb{R}$ é *convexa* se para todo ponto $(x, g(x))$ existe uma reta que passa por tal ponto de modo que o gráfico de g está sempre acima de tal reta.

Teorema (Desigualdade de Jensen): Seja X uma variável aleatória com média finita e g uma função convexa. Então

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

Desigualdade de Jensen



Exemplos

1) A voltagem de uma corrente elétrica tem distribuição exponencial com média de 2V. Um voltímetro usado para medi-la está com problema, e qualquer medição acima de 3V é registrada como 3V. Em média, qual valor é registrado pelo voltímetro?

Exemplos

2) Vimos que se $X_1, \ldots, X_n$ são independentes, então $\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i]$. Porém, vejamos que a recíproca não é verdadeira.

Exemplo: Uma moeda equilibrada é lançada duas vezes, e denotamos por X_i o resultado observado, para $i = 1, 2$, onde $X_i = 0$ representa coroa e $X_i = 1$ representa cara. O par aleatório (U, V) é definido como

$$
\begin{aligned}
U &= X_1 + X_2 \\
V &= X_1 - X_2.
\end{aligned}
$$

Encontre a distribuição conjunta de U e V e calcule $\mathbb{E}[U]$, $\mathbb{E}[V]$ e $\mathbb{E}[UV]$. Podemos afirmar que U e V são independentes?

Exemplos

1) $X \sim \text{Bin}(n, p) \Rightarrow \mathbb{E}[X] = np$
2) Considere uma urna com b bolas brancas e v bolas vermelhas. Qual é o número médio de bolas brancas se fizermos n extrações com reposição? E se as extrações forem sem reposição? Considere $n < b + v$.

Exemplos

3) Se $X \leq Y$, então $\mathbb{E}[X] \leq \mathbb{E}[Y]$

Momentos

Motivação: Dada uma função $f : \mathbb{R} \to \mathbb{R}$, as suas derivadas $f(x_0)$, $f'(x_0)$, $f''(x_0)$, $\ldots$, $f^{(n)}(x_0)$, $\ldots$ nos dizem tendências de comportamento de f em torno de x_0. Conseguimos pensar em uma quantidade análoga para variáveis aleatórias?

Momentos

Definição: Seja X uma variável aleatória e considere $k \geq 1$ inteiro.

- O *momento de ordem* k de X é definido como $\mathbb{E}[X^k]$
- Dizemos que o k-ésimo momento de X existe se $\mathbb{E}[|X|^k]$ é finito
- Se $\mathbb{E}[X] = \mu < \infty$, definimos o *momento central de ordem* k de X como $\mathbb{E}[(X - \mu)^k]$, desde que essa quantidade exista.
- O *momento absoluto de ordem* k de X é definido por $\mathbb{E}[|X|^k]$.

Momentos

Teorema: Se $\mathbb{E}[|X|^k] < \infty$ para algum inteiro positivo k, então $\mathbb{E}[|X|^j] < \infty$, para todo inteiro $j < k$.

Um exemplo

Considere X uma variável aleatória contínua, com função densidade de probabilidade dada por

$$f(x) = ce^{-(x-3)^2/2}, \text{ para } x \in \mathbb{R}.$$

Calcule a média de X e todos seus momentos centrais.

Observação: Note que, nesse caso particular, não precisamos da constante normalizadora de X! Porém, isso não é regra geral.

Variância

Definição: Se $\mathbb{E}[X] = \mu$ existe e é finita, definimos a sua *variância* como o momento central de ordem 2, ou seja,

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

A raiz quadrada da variância é denominada o *desvio padrão* de X.

Teorema: Se X é uma variável aleatória tal que sua média e variância existem e são finitos, então:

1) $\mathbb{V}(aX + b) = a^2 \mathbb{V}(X)$
2) $\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
3) $\mathbb{V}(X) = 0$ se e somente se existe a tal que $\mathbb{P}(X = a) = 1$
4) Se $X_1, \ldots, X_n$ são independentes, então

$$\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n)$$

Coisas que já sabíamos mas é bom relembrar – parte 2

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

Desigualdade de Chebyshev

Intuitivamente, se a variável aleatória X tiver variância finita, podemos refinar a desigualdade de Markov, incorporando também tal informação. Temos então:

Teorema *(Desigualdade de Chebyshev)*: Seja X uma variável aleatória cuja variância existe e é finita. Então, para $t > 0$, vale que

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{V}(X)}{t^2}.$$

Exemplo

Qual é, no máximo, a probabilidade de uma variável aleatória com variância finita se desviar de mais de 3 desvios padrão de sua média?

Exemplo

Ao estimar a média de uma variável aleatória com variância finita através da média amostral, qual é, no máximo, a probabilidade de obtermos uma estimativa que diste mais de t unidades do valor verdadeiro?

Assimetria, curtose, etc...

Pergunta: Como interpretamos momentos centrais de ordem superior?

$$\mathbb{E}[(X-\mu)^k] = \mathbb{E}[(X-\mu)(X-\mu)^{k-1}]$$

Pela expressão acima, faz sentido interpretá-los como *desvios ponderados de X em relação à sua média.*

Definição: Os *coeficiente de assimetria* e *coeficiente de curtose* de X são definidos como

$$\alpha_3 = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3} \qquad \alpha_4 = \frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4}$$

supondo a existência do terceiro e quarto momento de X, respectivamente.

Pergunta: O que essas quantidades medem, especificamente?

Covariância e correlação

Definição: Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade, cujas respectivas médias μ_X e μ_Y existem e são finitas. Definimos a *covariância* entre X e Y como

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

Se os respectivos desvios-padrão σ_X e σ_Y existem e são finitos, definimos a *correlação* entre X e Y como

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

Covariância e correlação - propriedades

1) $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
2) X e Y independentes $\implies \text{Cov}(X, Y) = 0$ (porém a recíproca não é verdadeira!)
3) $\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$
4) $|\rho_{X,Y}| \leq 1$
5) $|\rho_{X,Y}| = 1$ se e somente se uma variável for função linear da outra

Esperança condicional: motivação

Dadas variáveis aleatórias X e Y, é possível calcular a distribuição condicional de X dado que $Y = y$:

$$f_{(X,Y)}(x, y) = 2, \text{ para } x > 0, \; y > 0 \text{ e } x + y \leq 1$$
$$\implies f_{X|Y}(x|y) = \frac{1}{1-y}, \text{ para } 0 < x < 1 - y < 1.$$

Esperança condicional: motivação

Sendo $X|Y = y$ uma variável aleatória com distribuição conhecida, podemos calcular seu valor esperado:

$$\mathbb{E}[X|Y = y] = \int_0^{1-y} x \frac{1}{1-y}\, dx = \frac{1-y}{2}.$$

Podemos calcular também, a sua variância:

$$\mathbb{V}(X|Y = y) = \mathbb{E}[X^2|Y = y] - \mathbb{E}[X|Y = y]^2 = \frac{(1-y)^2}{12}.$$

Esperança condicional

Definição: Dadas variáveis aleatórias X e Y, a *esperança condicional de* X *dado que* $Y = y$ é denotada por $\mathbb{E}[X|Y = y]$, e é calculada como o valor esperado usual da variável aleatória $X|Y = y$. Analogamente, a *variância condicional de* X *dado que* $Y = y$ é denotada por $\mathbb{V}(X|Y = y)$, e é calculada como a variância usual da variável aleatória $X|Y = y$.

Considere a função que a cada valor de $Y = y$ associa $\mathbb{E}[X|Y = y]$ Repare que isso é uma função da variável aleatória Y, que denotaremos por $\mathbb{E}[X|Y]$ Portanto, a esperança condicional $\mathbb{E}[X|Y]$ também é uma variável aleatória!

Exemplo

Um conjunto de n pacientes, em um mesmo hospital, é submetido a um tratamento que pode curar ou não uma doença, com certa probabilidade. Essa probabilidade, por sua vez, depende das condições hospitalares de aplicação do tratamento, e pode ter três valores possíveis, p_1, p_2 e p_3, com respectivas probabilidades $1/2$, $1/3$ e $1/6$. Dado um valor de probabilidade de cura, a cura de um paciente não interfere na de outros e vice-versa. Estamos interessados no valor esperado do número total de curados.

Lei da esperança iterada

Teorema *(Lei da esperança iterada)*: Sendo X e Y variáveis aleatórias no mesmo espaço de probabilidade, temos que $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$, desde que tais quantidades existam.

Exemplo

Um número Y é escolhido ao acaso no intervalo $[0, 1]$. A seguir, outro número X é escolhido ao acaso, no intervalo $[0, Y]$. Qual é a média de X?

Exemplo

Uma aplicação interessante da esperança condicional é calcular o valor esperado de uma soma de uma quantidade aleatória de variáveis aleatórias.

Exemplo: Suponha que o número de pessoas que entram em uma loja de departamentos em determinado dia seja uma variável aleatória com média 50. Suponha ainda que as quantias de dinheiro gastas por esses clientes sejam variáveis aleatórias independentes com média comum de R$ 80,00. Finalmente, suponha também que a quantia gasta por um cliente seja independente do número total de clientes que entram na loja. Qual é a quantidade esperada de dinheiro gasto na loja em um dado dia?

Uma outra aplicação

Podemos também usar a lei da esperança iterada para calcular mais facilmente o valor esperado de produtos de variáveis aleatórias.

Teorema: Dado (X, Y) um vetor aleatório, temos que

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]].$$

Exemplo: Considere um círculo de raio unitário com centro na origem do plano cartesiano. Defina o par aleatório (X, Y) como sendo um ponto do primeiro quadrante escolhido ao acaso nesse círculo. Seja Z a área do retângulo formado pelos pontos $(\pm X, \pm Y)$. Assim, $Z = 4XY$. Qual é a área média desse retângulo?

Lei da variância iterada

- Lembremos que $\mathbb{V}(X|Y=y)$ é calculada como a variância usual da variável aleatória $X|Y=y$
- Porém, da mesma forma que $\mathbb{E}[X|Y]$ é variável aleatória, $\mathbb{V}(X|Y)$ também o é!
- A proposição abaixo nos dá uma fórmula análoga à lei da esperança iterada:

Teorema *(Lei da variância iterada)*:

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X|Y)] + \mathbb{V}(\mathbb{E}[X|Y]).$$

Exemplo

Suponha que o número de pessoas que chegam em uma estação de trem em qualquer instante t seja uma variável aleatória de Poisson com média λt. Se o primeiro trem chega na estação em um instante de tempo que é uniformemente distribuído ao longo de $(0, t_0)$ e independente do instante de chegada dos passageiros, quais são a média e a variância do número de passageiros que entram no trem?

Aplicação: predição e regressão linear

- Situação comum em Estatística: uma variável aleatória X é observada, e queremos prever o valor de outra variável aleatória Y
- Denote por $g(X)$ o preditor de Y. Ou seja, se $X = x$ é observado, então $g(x)$ é uma predição para o respectivo valor de Y
- Obviamente queremos escolher uma "boa" função g, de modo que $g(X)$ se aproxime de Y
- Mais precisamente, parece razoável buscarmos g que minimize $\mathbb{E}[(Y - g(X))^2]$, o *erro quadrático médio*
- Temos o seguinte resultado:
  Teorema: O preditor para Y que minimiza o erro quadrático médio é $\mathbb{E}[Y|X]$, ou seja,

$$\mathbb{E}[(Y - g(X))^2] \geq \mathbb{E}[(Y - \mathbb{E}[Y|X])^2]$$

Aplicação: predição e regressão linear

- Um modelo simples para relacionar duas variáveis aleatórias é

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

onde $\mathbb{E}[\varepsilon] = 0$ e ε é independente de X e Y.

- Portanto, se X é observado, a melhor previsão possível para Y é

$$\begin{aligned} \mathbb{E}[Y|X] &= \mathbb{E}[\beta_0 + \beta_1 X + \varepsilon|X] \\ &= \beta_0 + \beta_1 X + \mathbb{E}[\varepsilon|X] \\ &= \beta_0 + \beta_1 X, \end{aligned}$$

conforme nos é intuitivo!

- Isso justifica toda a modelagem *via* modelos lineares em Estatística
- A grande questão é: como estimar β_0 e β_1 a partir de dados observados?

Função geradora de momentos

Definição: A *função geradora de momentos* de uma variável aleatória X é definida como

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

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.

Usaremos esse resultado para encontrar distribuições de variáveis aleatórias de interesse.

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:

- $X \sim \text{Bern}(p) \implies \psi_X(t) = pe^t + 1 - p, \text{ para } t \in \mathbb{R}$
- $X \sim \text{Bin}(p) \implies \psi_X(t) = (pe^t + 1 - p)^n, \text{ para } t \in \mathbb{R}$
- $X \sim \text{Poi}(p) \implies \psi_X(t) = e^{-\lambda(1-e^t)}, \text{ para } t \in \mathbb{R}$
- $X \sim \mathcal{N}(\mu, \sigma^2) \implies \psi_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}, \text{ para } t \in \mathbb{R}$
- $X \sim \text{Exp}(\lambda) \implies \psi_X(t) = \dfrac{\lambda}{\lambda - t}, \text{ para } t < \lambda$
- $X \sim \text{Cauchy} \implies \psi_X(t)$ não existe!

Exemplos

1) Com esses resultados, podemos calcular muito mais facilmente certos momentos já conhecidos:
   - $X \sim \text{Bin}(p) \implies \mathbb{E}[X] = np, \mathbb{V}(X) = np(1-p)$
   - $X \sim \text{Poi}(p) \implies \mathbb{E}[X] = \mathbb{V}(X) = \lambda$
   - $X \sim \mathcal{N}(\mu, \sigma^2) \implies \mathbb{E}[X] = \mu, \mathbb{V}(X) = \sigma^2$

Exemplos

2) Nem sempre é mais fácil calcular momentos usando a função geradora de momentos:

   Seja $X \sim \mathcal{U}[a, b]$. Calcule a média e variância de X usando a sua função geradora de momentos.

Exemplos

3) Podemos calcular momentos até de distribuições cuja função densidade de probabilidade não nos é conhecida:

   Dizemos que X tem distribuição *log-normal de parâmetros* μ *e* σ^2 se $\ln(X)$ tem distribuição $N(\mu, \sigma^2)$. Encontre a média e variância de X.

Usaremos a função geradora de momentos também para encontrar mais facilmente distribuição da soma de variáveis aleatórias independentes. Mais precisamente:

Teorema: Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes com respectivas funções geradoras de momentos $\psi_1(t), \ldots, \psi_n(t)$, para $t \in (-\varepsilon, \varepsilon)$ e seja $Y = X_1 + \cdots + X_n$. Então a função geradora de momentos de Y existe e é dada por

$$\psi_Y(t) = \prod_{i=1}^{n} \psi_i(t), \text{ para } t \in (-\varepsilon, \varepsilon).$$

Soma de variáveis aleatórias independentes

Como a função geradora de momentos caracteriza a variável aleatória de modo único, "basta" fazemos o processo inverso:

- $X_1, \ldots, X_n \sim \text{Bern}(p) \implies X_1 + \cdots + X_n \sim \text{Bin}(n, p)$
- $X_1, \ldots, X_n \sim \text{Poisson}(\lambda) \implies X_1 + \cdots + X_n \sim \text{Poisson}(n\lambda)$
- $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2) \implies X_1 + \cdots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$

Função geradora de momentos multidimensional

Podemos estender a noção de função geradora de momentos para vetores aleatórios. Seja $\mathbf{X} = (X_1, \ldots, X_m)$ um vetor aleatório em $\mathbb{R}^m$. A *função geradora de momentos* do vetor aleatório $\mathbf{X}$ é definida por

$$\psi_{\mathbf{X}}(\mathbf{t}) = \psi_{\mathbf{X}}(t_1, \ldots, t_m) = \mathbb{E}[e^{\langle \mathbf{t}, \mathbf{X} \rangle}] = \mathbb{E}[e^{t_1 X_1 + \cdots + t_m X_m}],$$

para $\mathbf{t} \in (-\varepsilon, \varepsilon)^m$ onde tal quantidade seja finita.

Exemplo

Considere o vetor aleatório $\mathbf{X} = (X_1, X_2, X_3)$ com distribuição multinomial de parâmetros n e (p_1, p_2, p_3). Estude a função geradora de momentos de $\mathbf{X}$, identifique as suas distribuições marginais e calcule a covariância e correlação de pares de componentes de $\mathbf{X}$.

Uma caracterização da independência

Teorema: Seja $\mathbf{X} = (X_1, \dots, X_m)$ um vetor aleatório em $\mathbb{R}^m$ com função geradora de momentos $\psi_{\mathbf{X}}(\mathbf{t})$. Suponha que suas componentes são variáveis aleatórias com funções geradoras de momentos dadas por $\psi_1(t_1), \dots, \psi_m(t_m)$, respectivamente. Então as variáveis aleatórias $X_1, \dots, X_m$ são independentes se e somente se

$$\psi_{\mathbf{X}}(t_1, \dots, t_m) = \prod_{i=1}^{m} \psi_i(t_i).$$

Exemplo

Exemplo: Sejam X e Y variáveis aleatórias independentes normais padrão. Verifiquemos, de outra forma, que o par aleatório $U = X + Y$ e $V = X - Y$ também tem marginais normais e é independente.

Função característica

O fato da função geradora de momentos nem sempre existir (ou não existir para todo $t \in \mathbb{R}$) tem impactos positivos e negativos na teoria de probabilidades. Uma maneira de contornar os impactos negativos é definir uma quantidade análoga, que sempre existe:

Definição: A função característica de uma variável aleatória X é definida como

$$\phi_X(t) = \mathbb{E}[e^{itx}], \text{ para } t \in \mathbb{R}.$$

Teorema: A função característica satisfaz às seguintes propriedades:

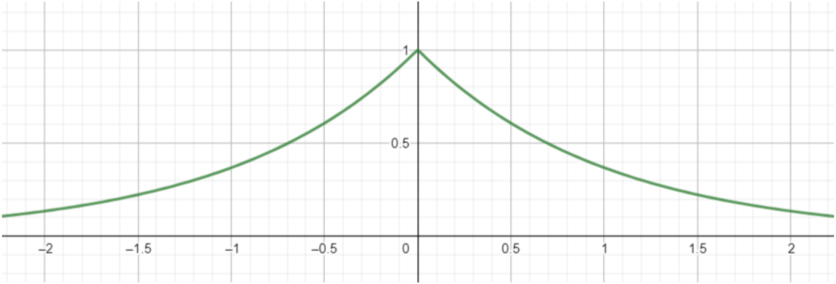
1) $|\phi_X(t)| \leq 1, \forall t \in \mathbb{R}$

2) $\phi_{X_1+\cdots+X_n}(t) = \prod_{j=1}^{n} \phi_j(t)$, se $X_1, \ldots, X_n$ são independentes

3) $\phi_X(t)$ também gera momentos:
$$\frac{d^n}{dt^n}\phi_X(t)\Big|_{t=0} = i^n \mathbb{E}[X^n], n = 1, 2, \ldots, \text{ se } \mathbb{E}[|X|^n] < \infty$$

Exemplos

- $X \sim \mathcal{N}(\mu, \sigma^2) \implies \phi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}, t \in \mathbb{R}$
- $X \sim \text{Exp}(\lambda) \implies \phi_X(t) = \dfrac{\lambda}{\lambda - it}, t \in \mathbb{R}$
- $X \sim \text{Cauchy} \implies \phi_X(t) = e^{-|t|}, t \in \mathbb{R}$



Lemas de Borel-Cantelli – preliminares

Sejam $A, A_1, A_2, \ldots$ eventos em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Vimos que:

- Se $A_n \uparrow A$, então $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$
- Se $A_n \downarrow A$, então $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$

Quando a sequência $A_1, A_2, \ldots$ é mais geral, definimos:

- $\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k =$ "ocorrência de infinitos dos eventos A_n"
- $\liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k =$ "ocorrência de todos os A_n, para n suficientemente grande"

O que podemos falar da probabilidade de algum desses eventos acontecer?

Lemas de Borel-Cantelli

Teorema: Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e sejam $A_1, A_2, \ldots$ eventos em $\mathcal{F}$. Temos então que

i) Se $\displaystyle\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, então $\mathbb{P}(\limsup A_n) = 0$

ii) Se $\displaystyle\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, e os eventos A_n são independentes, então $\mathbb{P}(\limsup A_n) = 1$

Exemplos

1) Um macaco imortal (e muito paciente!) bate em teclas ao acaso em um teclado, por tempo indefinido. Qual a probabilidade de que ele eventualmente digite a obra inteira de Shakespeare?

Exemplos

Apesar disso, tal evento é bastante improvável...

- Hamlet tem 130.000 caracteres
- A probabilidade do macaco digitar tal obra é de aproximadamente $4{,}4 \times 10^{-360.783}$, considerando pontuação e maiúsculas
- Se cada próton do universo observável for um macaco digitando, incessantemente, 2.000 caracteres por minuto, desde o Big Bang até o fim (estimado) do Universo, para termos uma chance em um trilhão de observar Hamlet digitada por inteiro (1 em 10^{12}), deveríamos ter $10^{360.641}$ universos formados por macacos atômicos!
- Em um desses universos, a chance de um documento de meros 79 caracteres ser digitado corretamente é de menos de 1 em um trilhão!

Exemplos

2) Considere um jogo infinito, onde na rodada n o jogador perde 2^n reais com probabilidade $\dfrac{1}{2^n+1}$ ou ganha 1 real com probabilidade $\dfrac{2^n}{2^n+1}$. O que podemos falar sobre o patrimônio do jogador ao "final" do jogo?

Exemplos

3) Suponha que lançamos uma moeda independentemente infinitas vezes, de modo que a probabilidade de observar cara no n-ésimo lançamento é $1/n$. Qual a probabilidade de observarmos infinitas caras? E se essa probabilidade fosse $1/n^2$?

Observação: Note que a independência não pode ser removida na segunda afirmação!

Convergência de variáveis aleatórias: motivação

Pergunta central: Dada uma sequência de variáveis aleatórias $X_1, X_2, \ldots$ (também denotada por $(X_n)_{n\in\mathbb{N}}$) no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, o que acontece no limite quando $n \to \infty$?

Razões para considerar essa pergunta:

- Estimadores em Estatística: Estimar parâmetro de distribuição $f(x|\theta)$ através de um estimador $\hat{\theta}_n = T(X_1, \ldots, X_n)$. Gostaríamos de garantir que $\hat{\theta}_n \approx \theta$, se n é "grande"
- Aproximações de distribuições: Se n é "grande" e p é "pequeno", então $\mathrm{Bin}(n,p) \approx \mathrm{Poi}(\lambda = np) \approx N(np, np(1-p))$
- Teorema Central do Limite: Seja $S = X_1 + \cdots + X_n$, onde as variáveis aleatórias X_i são independentes entre si, com médias finitas e variâncias finitas e positivas. Então, se n é "grande"

$$\frac{S - \mathbb{E}[S]}{\sqrt{\mathbb{V}(S)}} \approx N(0,1).$$

$\Longrightarrow$ Mas o que significa duas variáveis aleatórias serem "parecidas"?

Tipos de convergência de variáveis aleatórias: motivação

- Dois números x e y estão próximos se
$$|x - y| < \varepsilon,$$
para uma tolerância ε escolhida
- Duas funções $f, g : \mathbb{R} \to \mathbb{R}$ estão próximas se
$$|f(x) - g(x)| < \varepsilon, \forall x \in \mathbb{R},$$
para uma tolerância ε escolhida
- Dadas duas variáveis aleatórias X e Y, definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, como dizemos que uma está "próxima" da outra?
  - $\mathbb{P}(X \in A) \approx \mathbb{P}(Y \in A), \forall A \in \mathcal{B}(\mathbb{R})$?
  - $X(\omega) \approx Y(\omega), \forall \omega \in \Omega$?
  - X e Y têm momentos "parecidos"?
  - $\mathbb{E}[(X - Y)^2] \approx 0$?

$\Longrightarrow$ Veremos que podemos definir esse conceito de várias formas diferentes, e também como se relacionam entre si!

Convergência quase certa

Motivação: Como variáveis aleatórias são funções de Ω em $\mathbb{R}$, podemos estudar a convergência pontual de tais funções.

Definição: Dizemos que uma sequência de variáveis aleatórias $(X_n)_{n\in\mathbb{N}}$ *converge quase certamente* para uma variável aleatória X, todas definidas no mesmo espaço de probabilidade, se

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = \mathbb{P}\left(\left\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Denotamos tal fato por $X_n \stackrel{qc}{\to} X$.

Intuição: É o tipo mais restrito de convergência. Diz que o conjunto de eventos $\omega \in \Omega$ para os quais $X_n(\omega)$ não se aproxima de $X(\omega)$ tem probabilidade zero.

Convergência quase certa: exemplos

- Considere uma sequência $X_1, X_2, \ldots$ de variáveis aleatórias Bernoulli, cada uma com sua própria probabilidade de sucesso. O que significa $X_n \stackrel{qc}{\to} 0$?
- **Lei forte dos grandes números**: Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com média finita. Então $\overline{X}_n \stackrel{qc}{\to} \mu$.
- Suponha que n números são escolhidos uniformemente no intervalo [0, 1]. Seja X_n a variável aleatória representando a menor observação. Mostre que $X_n \stackrel{qc}{\to} 0$.

Convergência em probabilidade

Motivação: Em vez de olhar para a convergência pontual das variáveis aleatórias, tentar estudar o conjunto onde elas estão "longe" do seu limite.

Definição: Dizemos que uma sequência de variáveis aleatórias $(X_n)_{n\in\mathbb{N}}$ *converge em probabilidade* para uma variável aleatória X, todas definidas no mesmo espaço de probabilidade, se

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0, \forall \varepsilon > 0.$$

Denotamos esse fato por $X_n \xrightarrow{p} X$.

Intuição: Fixada uma tolerância ε, a probabilidade dos eventos $\omega \in \Omega$ tais que $|X_n(\omega) - X(\omega)| > \varepsilon$ torna-se cada vez menor, a medida que n cresce. Ou seja, eventos "não usuais", com respeito a X, ocorrem cada vez "menos".

Convergência em probabilidade: exemplos

- **Consistência da média amostral, ou lei fraca dos grandes números**: Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com média e variância finitas. Então $\overline{X}_n \xrightarrow{p} \mu$, pela desigualdade de Chebyshev.
- Considere $X_n \sim \mathcal{U}[0, 1/n]$. Mostre que $X_n \xrightarrow{p} 0$.
- **Histogramas**: Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, e sejam $a < b$ constantes. Defina $Y_i = 1$ se $a \leq X_i < b$, e $Y_i = 0$, caso contrário. Então $\overline{Y}_n \xrightarrow{p} \mathbb{P}(a \leq X_1 < b)$.

Convergência em distribuição

Definição: Dizemos que uma sequência de variáveis aleatórias $(X_n)_{n\in\mathbb{N}}$ *converge em distribuição* para uma variável aleatória X se

$$\lim_{n\to\infty} F_n(x) = F_X(x), \forall x \in \mathbb{R} \text{ onde } F_X \text{ é contínua,}$$

onde F_n é a função de probabilidade acumulada de X_n e F_X é a função de probabilidade acumulada de X. Denotamos esse fato por $X_n \xrightarrow{d} X$.

Intuição: As funções de probabilidade acumulada de X_n se aproximam cada vez mais da função de probabilidade acumulada de X.

Porque é razoável? Como a função de probabilidade acumulada nos diz toda a informação sobre a distribuição de uma variável aleatória, tal convergência parece razoável, para propósitos de cálculos de probabilidades.

Convergência em distribuição: exemplos

- Considere a sequência de variáveis aleatórias cujas funções de probabilidade são dadas por

$$F_n(x) = \begin{cases} 1 - \left(1 - \frac{1}{n}\right)^{nx}, & \text{se } x > 0 \\ 0, & \text{se } x \leq 0. \end{cases}$$

  Mostre que $X_n \xrightarrow{d} X$, onde $X \sim \text{Exp}(1)$.

- Considere $X_n \sim \mathcal{U}[0, 1/n]$. Mostre que $X_n \xrightarrow{d} 0$.
- **Teorema Central do Limite**: Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância σ^2, e seja $\overline{X}$ a média amostral. Então
  $\dfrac{\sqrt{n}}{\sigma}(\overline{X} - \mu) \xrightarrow{d} X$, onde $X \sim N(0, 1)$.

Convergência em média r

Motivação: Dado $r > 0$, a quantidade $\mathbb{E}[|X - Y|^r]$ pode ser uma medida de distância entre as variáveis aleatórias X e Y.

Definição: Dizemos que uma sequência de variáveis aleatórias $(X_n)_{n\in\mathbb{N}}$ *converge em média* r para uma variável aleatória X, todas definidas no mesmo espaço de probabilidade, se

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0.$$

Denotamos tal fato por $X_n \xrightarrow{r} X$.

Intuição: Quanto maior r, mais peso damos aos desvios de X_n de X, de modo que r controla a "velocidade" com a qual X_n se aproxima de X.

Exemplo: Considere $X_n \sim \mathcal{U}[0, 1/n]$. Provamos que $X_n \xrightarrow{d} 0$ e $X_n \xrightarrow{p} 0$. Provemos agora que $X_n \xrightarrow{r} 0, \forall r \geq 1$.

Relação entre os tipos de convergência

Os quatro tipos de convergência que já vimos relacionam-se da seguinte forma:

$$\begin{array}{ccccc} \xrightarrow{s} & \underset{s\geq r\geq 1}{\Longrightarrow} & \xrightarrow{r} & & \\ & & \Downarrow & & \\ \xrightarrow{qc} & \Longrightarrow & \xrightarrow{p} & \Longrightarrow & \xrightarrow{d} \end{array}$$

Porém, em geral, as recíprocas não são verdadeiras, ou seja,

$$\xrightarrow{d} \quad \not\Longrightarrow \quad \xrightarrow{p} \quad \not\Longrightarrow \quad \xrightarrow{qc} \text{ e ainda } \xrightarrow{p} \quad \not\Longrightarrow \quad \xrightarrow{r}$$

Vejamos alguns exemplos e condições para valerem as recíprocas.

Contra-exemplos

1) **Exemplo que** $\xrightarrow{d} \quad \not\Longrightarrow \quad \xrightarrow{p}$: Seja $\Omega = \{a, b\}, \mathcal{F} = 2^{\Omega}$ e $\mathbb{P}$ uniforme. Defina as variáveis aleatórias

$$X_n(\omega) = \begin{cases} 1, & \text{se } \omega = a \\ 0, & \text{se } \omega = b \end{cases}, \ \forall n \in \mathbb{N} \quad \text{e} \quad X(\omega) = \begin{cases} 0, & \text{se } \omega = a \\ 1, & \text{se } \omega = b \end{cases}$$

Temos que $X_n \xrightarrow{d} X$ mas não vale a convergência em probabilidade.

Contra-exemplos

2) **Exemplo que** $\xrightarrow{p} \quad \not\Longrightarrow \quad \xrightarrow{qc}$: Seja $X_n \sim \text{Bern}(1/n)$, para $n \in \mathbb{N}$. Temos que $X_n \xrightarrow{p} 0$ mas não vale a convergência quase certa.

Contra-exemplos

3) **Exemplo que** $\xrightarrow{p} \;\not\Longrightarrow\; \xrightarrow{r}$: Seja X_n tal que $\mathbb{P}(X_n = n^2) = 1/n$ e $\mathbb{P}(X_n = 0) = 1 - 1/n$. Temos que $X_n \xrightarrow{p} 0$ mas não vale a convergência em média r, para qualquer $r \geq 1$.

Quando valem algumas recíprocas

Porém, em algumas situações especiais, valem as recíprocas:

Quando $\xrightarrow{d}$ **implica** $\xrightarrow{p}$: Se a sequência $(X_n)_{n\in\mathbb{N}}$ converge em distribuição para uma constante c, então $X_n \xrightarrow{p} c$.

$\xrightarrow{p}$ **"quase" implica** $\xrightarrow{qc}$: Se a sequência $(X_n)_{n\in\mathbb{N}}$ converge em probabilidade para X, então existe uma sub-sequência $(X_{n_k})_{k\in\mathbb{N}}$ que converge quase certamente para X.

Resultados importantes

Um importante resultado que facilita aferir a convergência em distribuição é o seguinte:

Teorema: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias e X uma variável aleatória, onde ϕ_n e ϕ_X denotam as respectivas funções características e ψ_n e ψ_X denotam as respectivas funções geradoras de momentos, supondo suas existências em torno de 0. Então:

- $X_n \stackrel{d}{\to} X$ se e somente se $\lim_{n\to\infty} \phi_n(t) = \phi(t), \forall t \in \mathbb{R}$.
- $X_n \stackrel{d}{\to} X$ se $\lim_{n\to\infty} \psi_n(t) = \psi(t)$ para todos t onde tais funções existam.

Resultados importantes

Finalmente, os dois resultados abaixo serão úteis ao trabalharmos com as leis dos grandes números:

Teorema: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias e X uma variável aleatória, todas definidas no mesmo espaço de probabilidade, e seja também $g : \mathbb{R} \to \mathbb{R}$ uma função contínua. Então se X_n converge para X quase certamente, em probabilidade ou em distribuição então $g(X_n)$ converge da mesma forma para $g(X)$.

Teorema *(Teorema de Slutsky)*: Sejam $(X_n)_{n\in\mathbb{N}}$ e $(Y_n)_{n\in\mathbb{N}}$ sequências de variáveis aleatórias e X uma variável aleatória, todas definidas no mesmo espaço de probabilidade tais que valem as convergências $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{p} c$, com c constante. Então:

i) $X_n + Y_n \xrightarrow{d} X + c$

ii) $X_nY_n \xrightarrow{d} cX$

iii) Se $c \neq 0$, então $\dfrac{X_n}{Y_n} \xrightarrow{d} \dfrac{X}{c}$, desde que $\mathbb{P}(Y_n = 0) = 0, \forall n$.

Um exemplo

A fim de ilustrar esses últimos resultados, vejamos um exemplo.

Exemplo: Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com distribuição $N(\mu, \sigma^2)$. Estude o comportamento assintótico dos estimadores média e variância amostral, dados respectivamente por

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

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Leis dos Grandes Números: introdução histórica

- Importantes para o desenvolvimento da teoria das Probabilidades. Estão intimamente ligadas à sua interpretação frequentista
- Século XVI Cardano afirmou, sem provas, que a acurácia de estatísticas empíricas tende a melhorar com o aumento no número de amostras
- Primeira versão de uma Lei dos Grandes Números: provada no século XVIII por Jacob Bernoulli em seu trabalho "*Ars Conjecturandi*"
- Primeira utilização do nome "Lei dos Grandes Números": através dos trabalhos de Poisson (século XIX)
- Transição entre os séculos XIX e XX: trabalhos de Chebyshev, Markov, Borel, Cantelli, Kolmogorov e Khinchin

Leis dos grandes números: intuição

- $(\Omega, \mathcal{F}, \mathbb{P})$ espaço de probabilidade, $\Omega =$ possíveis resultados de um experimento
- Dado $A \in \mathcal{F}$ seja n_A o número de ocorrências de A dentre n realizações do experimento
- Intuitivamente, $\lim_{n\to\infty} \frac{n_A}{n} = \mathbb{P}(A)$, porém como provamos a validade desse limite?
- Após a axiomatização de Kolmogorov, no início do século XX, temos as ferramentas necessárias para torná-lo rigoroso!
- Sejam $X_1, X_2, \dots \sim \text{Bern}(p = \mathbb{P}(A))$ variáveis aleatórias independentes e identicamente distribuídas $\implies \frac{n_A}{n} = \overline{X}_n$.
- É verdade que $\overline{X}_n \xrightarrow{p} p$? Ou ainda que $\overline{X}_n \xrightarrow{qc} p$?

Leis dos grandes números: formulação

Sejam $X_1, X_2, \ldots$ variáveis aleatórias definidas no mesmo espaço de probabilidade, cujas esperanças existem e são finitas.

- Dizemos que a sequência $(X_n)_{n\in\mathbb{N}}$ satisfaz a *lei fraca dos grandes números* se
$$\overline{X}_n - \mathbb{E}[\overline{X}_n] \xrightarrow{p} 0$$
- Dizemos que a sequência $(X_n)_{n\in\mathbb{N}}$ satisfaz a *lei forte dos grandes números* se
$$\overline{X}_n - \mathbb{E}[\overline{X}_n] \xrightarrow{qc} 0$$

Portanto, queremos responder a seguinte pergunta:

Dada uma particular sequência de variáveis aleatórias $(X_n)_{n\in\mathbb{N}}$, sob quais condições ela satisfaz a lei forte ou fraca dos grandes números?

Desde o século XVIII até o início do século XX temos várias respostas para essa pergunta!

Algumas leis dos grandes números

- Lei Fraca de Bernoulli: $(X_n)_{n\in\mathbb{N}}$ iid com distribuição Bern(p)
- Lei Fraca de Chebyshev: $(X_n)_{n\in\mathbb{N}}$ independentes dois-a-dois, com variância finita e uniformemente limitadas
- Lei Fraca de Khintchine: $(X_n)_{n\in\mathbb{N}}$ iid com média finita
- Lei Forte de Borel: $(X_n)_{n\in\mathbb{N}}$ iid com distribuição Bern(p)
- 1a. Lei Forte de Kolmogorov: $(X_n)_{n\in\mathbb{N}}$ independente, com média finita e satisfazendo $\displaystyle\sum_{n=1}^{+\infty}\frac{\mathbb{V}(X_n)}{n^2} < +\infty$
- Lei Forte de Kolmogorov: $(X_n)_{n\in\mathbb{N}}$ iid com média finita

Uma importante aplicação

- Seja $X_1, X_2, \ldots$ uma amostra aleatória de uma variável aleatória X, cuja função de probabilidade acumulada denotamos por F_X.

  Como estimamos F_X a partir das amostras? Quão boa é essa estimativa?

- Definimos a *função de probabilidade acumulada empírica* como

$$F_n(x) = \frac{1}{n}[\text{quantidade de observações abaixo de } x].$$

- Provemos que, fixado $x_0 \in \mathbb{R}$, então $F_n(x_0) \stackrel{qc}{\to} F(x_0)$.

Uma importante aplicação

- Porém, tal resultado pode ser muito melhorado! O *Teorema de Glivenko-Cantelli* ou *Teorema Fundamental da Estatística* afirma que tal convergência é uniforme em todo $\mathbb{R}$, ou seja,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \stackrel{qc}{\to} 0.$$

- Tal resultado é fundamental em testes como o teste de Kolmogorov-Smirnov e o teste de Shapiro-Wilk
- Porém, é um resultado assintótico, ou seja, só vale no limite quando $n \to \infty$! Felizmente temos também o *Teorema de Dvoretzky–Kiefer–Wolfowitz–Massart*, que nos dá uma estimativa para a velocidade de convergência:

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}, \quad \forall\varepsilon > 0$$

Exemplos

1) Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas, com distribuição $[0, 1]$. Defina $Y_n = (X_1 \dots X_n)^{1/n}$. O que se pode falar da convergência de Y_n?

Exemplos

2) A quantidade de partículas radioativas emitidas por certo elemento durante unidades de tempo consecutivas são modeladas por distribuições de Poisson independentes X_n de parâmetro comum $\lambda > 0$. Estude o limite

$$\lim_{n\to\infty} \frac{X_1^2 + \cdots + X_n^2}{n}.$$

Exemplos

3) Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com distribuição $N(0,1)$. Estude a convergência de

$$\frac{X_1^2 + \cdots + X_n^2}{(X_1 - 1)^2 + \cdots + (X_n - 1)^2}.$$

Exemplos

4) Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade comum dada por

$$f(x) = \begin{cases} e^{-(x+1/2)}, & \text{para } x \geq -1/2 \\ 0, & \text{para } x < -1/2. \end{cases}$$

Mostre que $X_1 + \cdots + X_n \xrightarrow{qc} \infty$.

Teorema Central do Limite: motivação

- Considere $(X_n)_{n\in\mathbb{N}}$ uma sequência independente e identicamente distribuída de variáveis aleatórias de média zero e variância 1
- As leis fraca e forte dos grandes números nos dizem, respectivamente, que

$$\overline{X}_n \xrightarrow{p} 0 \quad \text{e} \quad \overline{X}_n \xrightarrow{qc} 0$$

- Porém, para fazer um teste de hipótese ou construir um intervalo de confiança, precisamos saber como se dá essa convergência!
- O *Teorema Central do Limite* nos diz uma resposta para essa pergunta:

$$\sqrt{n}\,\overline{X}_n \xrightarrow{d} \mathcal{N}(0,1)$$

- Analogamente às Leis dos Grandes Números, há toda uma *classe* de Teoremas Centrais do Limite, ou seja, condições na sequência de interesse tais que se dá a convergência em distribuição para uma variável aleatória normal

Teorema Central do Limite: história

- As primeiras versões datam do século XVIII, onde DeMoivre o postulou para o caso de variáveis com distribuição de Bernoulli
- Trabalho quase 100 anos esquecido, resgatado no início do século XIX por Laplace, ainda não recebendo a devida atenção
- Sua importância foi discernida somente na transição entre os séculos XIX e XX, com os trabalhos de Lyapunov
- O nome "Teorema Central do Limite" aparece pela primeira vez em 1920, em um trabalho de G. Pólya, onde o termo "central" deriva da sua importância em teoria das Probabilidades
- Porém, a escola francesa de Probabilidade interpreta o termo "central" no sentido de *descrever o comportamento do centro da distribuição, em oposição à sua cauda*, chamando tal resultado de *"Teorema do Limite Central"*
- Outros personagens importantes: Cauchy, Bernstein, Lindberg, Lévy, Kolmogorov, Feller, etc.

Dois Teoremas Centrais do Limite

Teorema *(TCL para variáveis aleatórias iid)*: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância $0 < \sigma^2 < \infty$. Então temos que

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, 1).$$

Provemos uma versão sob hipóteses mais fortes, onde assumimos que as variáveis aleatórias têm função geradora de momentos definidas em torno de zero.

Dois Teoremas Centrais do Limite

Podemos enfraquecer a hipótese das variáveis aleatórias serem identicamente distribuídas, a um pequeno custo de uma hipótese adicional

Teorema *(TCL de Lyapunov)*: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes, com respectivas médias μ_n e variâncias $0 < \sigma_n^2 < \infty$. Seja $s_n^2 = \sum_{k=1}^n \sigma_k^2$. Assuma que a *condição de Lyapunov* é satisfeita:

$$\exists\ \delta > 0 \ \Bigg| \ \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}[|X_k - \mu_k|^{2+\delta}] = 0.$$

Então vale que

$$\frac{n}{s_n}(\overline{X}_n - \mathbb{E}[\overline{X}_n]) \xrightarrow{d} N(0,1).$$

Teorema de Berry-Esseen

Assim como as Leis dos Grandes Números, o Teorema Central do Limite é um resultado assintótico! Felizmente temos como saber algo sobre a "distância" para o limite.

Teorema: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência variáveis aleatórias independentes e identicamente distribuídas, de média zero e variância $0 < \sigma^2 < \infty$. Assuma que $\mathbb{E}[|X_i|^3] < \infty$ e seja $Z \sim N(0,1)$. Então temos que

$$\left|\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\overline{Y}_n \leq x\right) - \mathbb{P}(Z \leq x)\right| \leq \frac{C\ \mathbb{E}[|X_i|^3]}{\sigma^3\sqrt{n}},\ \forall x \in \mathbb{R},$$

onde C é uma constante positiva.

Atualmente, sabe-se que $0{,}4097 \leq C \leq 0{,}4748$.

Exemplos

1) *(Teorema Central do Limite de deMoivre-Laplace)* Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com distribuição de Bernoulli de parâmetro p, e defina $S_n = X_1 + \cdots + X_n$. Então

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

Exemplos

2) Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância 2. Encontre o limite em distribuição de:

   a) $$\frac{\sqrt{n}(X_1+\cdots+X_n)}{X_1^2+\cdots+X_n^2}$$

   b) $$\frac{X_1+\cdots+X_n}{\sqrt{X_1^2+\cdots+X_n^2}}$$