

Introdução à Probabilidade

Notas de Aula

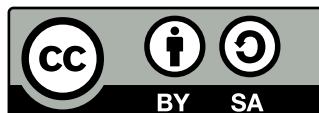
Leonardo T. Rolla

27 de agosto de 2024

Esse texto foi originalmente escrito em inglês e usado para lecionar uma disciplina em Warwick, no ano de 2022. Depois foi traduzido pelo autor usando Inteligência Artificial. Aqueles que quiserem melhorar o texto e enviar as modificações ao autor serão bem-vindos. O código-fonte pode ser obtido em:

<https://github.com/REA-Stat/apostila-probab> e <https://www.overleaf.com/read/pscdjdqjzbhc>

© 2023–2024 Departamento de Estatística, IME-USP.



Todos os direitos reservados. Permitido o uso nos termos licença Creative Commons Atribuição-CompartilhaIgual 4.0 Internacional.

Reutilização deste material

Você pode remixar, transformar, e criar a partir do material para qualquer fim, mesmo que comercial. Nesse caso, tem de distribuir as suas contribuições sob a mesma licença que o original. Você não pode aplicar termos jurídicos ou medidas de caráter tecnológico que restrinjam legalmente outros de fazerem algo que a licença permita.

Atribuição

Este material foi produzido originalmente por Leonardo T. Rolla (IME-USP), como parte do projeto *REA-Stat – Recursos Educacionais Abertos de Estatística* (<https://rea-stat.github.io/>), traduzido e adaptado de uma apostila em inglês produzida por Leonardo T. Rolla e Tessy Papavasileiou em 2022, com contribuições de Daniel Valesin e Giuseppe Cannizzaro.

Código-fonte

O código-fonte deste material está disponível em:

<https://github.com/REA-Stat/apostila-probab> e <https://www.overleaf.com/read/pscdjdqjzbhc>

Aviso legal

As pessoas e instituições aqui mencionadas não endossam a qualidade deste material e as opiniões nele contido, nem explícita nem implicitamente.

Qualquer erro contido neste material é responsabilidade de: Leonardo T. Rolla.

27 de agosto de 2024

Sumário

1	Espaços de probabilidade uniforme	6
1.1	Probabilidade em termos de conjuntos e funções	6
1.2	Espaço de Probabilidade Uniforme	8
2	Contagem	11
2.1	Fundamentos da Combinatória	11
2.2	Amostragem	18
3	Espaços de Probabilidade	25
3.1	Espaço amostral e espaço de eventos	25
3.2	Probabilidade	30
4	Probabilidade Condicional e Independência	37
4.1	Probabilidade Condicional	37
4.2	Lei da Probabilidade Total	42
4.3	Teorema de Bayes	44
4.4	Independência	45
5	Variáveis Aleatórias	48
5.1	Definição	48
5.2	Variáveis Aleatórias Discretas	50
5.3	As distribuições discretas mais comuns	53
6	Esperança	57
6.1	Definição e exemplos	57
6.2	Propriedades da esperança	59
6.3	Função de uma variável aleatória	61
6.4	Variância	63
7	Distribuições discretas multivariadas	66
7.1	Função de massa de probabilidade conjunta de duas variáveis	66
7.2	Esperança no caso bivariado discreto	67
7.3	Variáveis aleatórias discretas independentes	69
8	A Lei dos Grandes Números	72

9	Covariância	74
9.1	Definição	74
9.2	Propriedades	74
9.3	Somas de variáveis independentes duas a duas	77
10	Desigualdade de Chebyshev	78
10.1	Desigualdade de Markov	78
10.2	Desigualdade de Chebyshev	79
10.3	Prova da lei das médias	80
11	Coefficiente de Correlação	82
12	Teorema do Limite Central	85
13	Variáveis aleatórias contínuas	88
13.1	Função de densidade de probabilidade	88
13.2	Variáveis Uniformes	88
13.3	Distribuição Normal	89
13.4	Tempo de Vida Exponencial	90
13.5	Esperança	91
13.6	Variância	92
14	Uma única teoria para discretas e contínuas	95
14.1	Função de distribuição cumulativa	95
14.2	Casos discretos e contínuos	96
14.3	Esperança e variância	98
15	Distribuições Conjuntas e Independência	100
15.1	Função de Densidade Conjunta	100
15.2	Função de Distribuição Conjunta	100
15.3	Independência	101
15.4	Covariância e a lei das médias	103
16	Somas de variáveis aleatórias independentes	104
17	Momentos e funções geradoras de momentos	106
A	Somas úteis	111

B Exponenciais superam polinômios

113

1 Espaços de probabilidade uniforme

1.1 Probabilidade em termos de conjuntos e funções

Considere o seguinte experimento: uma caixa contém 4 bolas vermelhas, 6 bolas azuis e 10 bolas verdes. Escolhemos uma bola *aleatoriamente*. Qual é a *probabilidade* de essa bola ser vermelha? Aprendemos na escola que isso deveria ser o número de bolas vermelhas sobre o número total de bolas, ou seja, $\frac{4}{20} = 0.2 = 20\%$, e isso é de fato verdade sob certas suposições. Para entender as suposições que implicitamente fazemos ao realizar esse cálculo, fazemos as seguintes perguntas:

- O que é uma probabilidade como objeto matemático?
- Que outras perguntas poderíamos fazer sem alterar o experimento?
- A resposta seria a mesma se algumas bolas fossem mais difíceis de segurar (por exemplo, se tivessem tamanhos diferentes)?

Observe que quando perguntamos sobre uma probabilidade, precisamos determinar o evento cuja probabilidade nos interessa - enquanto a probabilidade de um evento específico (por exemplo, 'a bola é vermelha') é um número no intervalo de $[0, 1]$, a *probabilidade em si é um mapeamento que associa a cada evento um número*.

Há três possíveis resultados para este experimento: vermelho, azul e verde. Chamamos o conjunto de todos os resultados possíveis de *espaço amostral*, normalmente representado por Ω . Como objeto matemático, Ω é qualquer conjunto não vazio - neste caso, $\Omega = \text{vermelho}, \text{azul}, \text{verde}$.

No entanto, podemos fazer outras perguntas também. Por exemplo, podemos perguntar qual é a probabilidade de 'a bola ser azul ou verde' (o que seria equivalente a 'a bola não ser vermelha'). Em palavras, um evento é uma afirmação na qual você pode determinar se é verdadeira ou não após ver o resultado do experimento. Neste caso, todos os eventos possíveis são:

- 'A bola não é de nenhuma das três cores ou de qualquer outra cor' - matematicamente, isso será representado pelo conjunto vazio \emptyset , já que não contém nenhum dos resultados possíveis.
- 'A bola é vermelha' - representado por *vermelho*
- 'A bola é azul' - representado por *azul*

- 'A bola é verde' - representado por *verde*
- 'A bola é ou vermelha ou azul' - representado por *vermelho, azul*
- 'A bola é ou vermelha ou verde' - representado por *vermelho, verde*
- 'A bola é ou azul ou verde' - representado por *azul, verde*
- 'A bola é qualquer uma das cores vermelha, azul ou verde' - representado por *vermelho, azul, verde* = Ω .

Desta lista exaustiva, fica claro que todos os eventos são subconjuntos de Ω e, de fato, neste caso pelo menos, todos os subconjuntos de Ω são eventos. Chamamos a coleção de todos os eventos de *espaço de eventos* (ou, mais formalmente, σ -álgebra), normalmente representado por \mathcal{F} . Como objeto matemático, isso é uma coleção de subconjuntos do espaço amostral - veremos mais tarde que essa coleção deve satisfazer certas propriedades, mas por enquanto, supomos que ela inclui todos os subconjuntos, então $\mathcal{F} = \mathcal{P}(\Omega)$, onde \mathcal{P} denota o conjunto das partes (a coleção de todos os subconjuntos).

Já argumentamos que a probabilidade, normalmente representada por \mathbb{P} , é um mapeamento do espaço de eventos para números em $[0, 1]$ - o representamos como $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$. No entanto, com base em nossa intuição, esperamos que:

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(\textit{vermelho}) = 0,2$
- $\mathbb{P}(\textit{azul}) = 0,3$
- $\mathbb{P}(\textit{verde}) = 0,5$
- $\mathbb{P}(\textit{vermelho, azul}) = 0,5$
- $\mathbb{P}(\textit{vermelho, verde}) = 0,7$
- $\mathbb{P}(\textit{azul, verde}) = 0,8$
- $\mathbb{P}(\textit{vermelho, azul, verde}) = 1$.

Quais suposições implícitas estamos fazendo ao fazer esses cálculos, com base em nossa intuição?

- A probabilidade do evento \emptyset , que não inclui nenhum resultado, deve ser 0.
- A probabilidade do evento Ω , que inclui todos os resultados, deve ser 1.
- Quando um evento pode ser decomposto na união de dois eventos disjuntos, sua probabilidade deve ser a soma das duas probabilidades, por exemplo, $\mathbb{P}(\textit{vermelho, azul}) = \mathbb{P}(\textit{vermelho}) + \mathbb{P}(\textit{azul})$.

Essas são propriedades fundamentais que um mapeamento de probabilidade deve

ter.

In conclusion, the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ of sample space Ω , event space \mathcal{F} and probability \mathbb{P} form what is called a *probability space*.

Em conclusão, o trio $(\Omega, \mathcal{F}, \mathbb{P})$, composto pelo espaço amostral Ω , espaço de eventos \mathcal{F} e probabilidade \mathbb{P} , forma o que é chamado de um *espaço de probabilidade*.

1.2 Espaço de Probabilidade Uniforme

No exemplo anterior, a maneira como postulamos a probabilidade de cada cor tinha uma suposição implícita: que todas as bolas têm a mesma chance de serem escolhidas. Essa é uma suposição correta se as bolas têm o mesmo peso, tamanho, textura, etc.

E se, em vez de serem coloridas em vermelho, verde ou azul, as bolas fossem numeradas de 1 a 20? Sob a mesma suposição (que as bolas têm o mesmo peso, tamanho, textura, etc.), teríamos

$$\Omega = \{vermelho_1, \dots, vermelho_4, azul_1, \dots, azul_6, verde_1, \dots, verde_{10}\}$$

e \mathbb{P} seria tal que cada bola teria a mesma chance de ser escolhida, ou seja, cada elemento $\omega \in \Omega$ teria a mesma chance. Então, o evento 'a bola é vermelha' corresponderia ao evento $A = \{vermelho_1, \dots, vermelho_4\}$ e intuitivamente esperaríamos que a probabilidade de obter uma bola vermelha fosse igual a $\frac{4}{20}$, onde 4 é o número de elementos em A (e bolas vermelhas) e 20 é o número de todos os resultados possíveis.

Quando cada resultado é igualmente provável, temos um espaço de probabilidade uniforme.

Definição 1.1. Um *espaço de probabilidade uniforme* é definido como o trio $(\Omega, \mathcal{F}, \mathbb{P})$, onde

- Ω (o *espaço amostral*) é um conjunto finito não vazio de todos os resultados possíveis do experimento;
- \mathcal{F} (o *espaço de eventos*) é a coleção de todos os eventos, dada pelo conjunto de potências $\mathcal{P}(\Omega)$ de Ω ;

- $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ é um mapeamento do espaço de eventos para $[0, 1]$, satisfazendo

$$- \mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1, \quad (1.2a)$$

$$- \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \text{ para todo } A, B \in \mathcal{F} \text{ tal que } A \cap B = \emptyset$$

(aditividade finita) (1.2b)

$$- \mathbb{P}(\{\omega\}) = \mathbb{P}(\{\tilde{\omega}\}) \text{ para todo } \omega, \tilde{\omega} \in \Omega \quad (\text{uniforme}) \quad (1.2c)$$

Como resultado da suposição uniforme, calcular a probabilidade de qualquer evento se resume a calcular a cardinalidade do evento. Lembre-se de que um evento é um conjunto (um subconjunto do espaço amostral Ω) - a cardinalidade de um conjunto é o número de seus elementos. Para um conjunto $A \subseteq \Omega$, ele é denotado por $|A|$.

Para definir formalmente a cardinalidade, primeiro precisamos definir uma correspondência um a um entre dois conjuntos: dados dois conjuntos A e B , dizemos que eles estão em uma correspondência um a um se existir um mapa bijetivo entre eles, ou seja, uma função $f : A \rightarrow B$ que seja tanto injetiva quanto sobrejetiva.

Definição 1.3. Um conjunto A possui cardinalidade $n \in \mathbb{N}$ se estiver em uma correspondência um a um com $\{1, 2, \dots, n\}$ e A possui cardinalidade 0 se $A = \emptyset$.

Proposição 1.4. *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade uniforme. Então, para todo $\omega \in \Omega$*

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}, \quad (1.5)$$

e para todo $A \subseteq \Omega$ ($A \in \mathcal{F}$)

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}. \quad (1.6)$$

Demonstração. Uma vez que $\forall \omega_1, \omega_2 \in \Omega, \mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\})$, seja $p \in [0, 1]$ tal que

$$p := \mathbb{P}(\{\omega\}) \quad \forall \omega \in \Omega.$$

Uma vez que \mathbb{P} é uma medida de probabilidade

$$\begin{aligned} 1 = \mathbb{P}(\Omega) &= \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) \quad \text{por (1.2b)} \\ &= \sum_{\omega \in \Omega} p = p \sum_{\omega \in \Omega} 1 = p|\Omega|. \end{aligned}$$

Portanto

$$p = \frac{1}{|\Omega|}.$$

demonstrando (1.5). Vemos que (1.6) segue de (1.2b), pois

$$\begin{aligned} \mathbb{P}(A) &= \sum_{\omega \in A} \mathbb{P}(\{\omega\}), \quad \text{por (1.2b)} \\ &= \sum_{\omega \in A} p = p \sum_{\omega \in A} 1 = p|A| = \frac{|A|}{|\Omega|}. \end{aligned} \quad \square$$

Exercício 1.1. Considere uma urna com 50 bolas numeradas de 1 a 50. Suponha que elas sejam sorteadas uniformemente ao acaso. Após definir um espaço de probabilidade adequado, determine a probabilidade de que a primeira bola sorteada mostre um número divisível por 12.

Solução: Defina $(\Omega, \mathcal{F}, \mathbb{P})$ da seguinte forma: $\Omega = \{1, 2, 3, \dots, 50\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ e \mathbb{P} a medida de probabilidade uniforme. Então, o evento em questão é

$$E = \{12, 24, 36, 48\}.$$

Pela Proposição 1.4 (1.6),

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{4}{50} = \frac{2}{25}.$$

2 Contagem

2.1 Fundamentos da Combinatória

Em geral, a fórmula (1.6) afirma que, para calcular probabilidades, precisamos contar. Esta seção aprofunda o problema da contagem.

Exemplo 2.1. Se houver 30 pessoas em uma sala, qual é a probabilidade de pelo menos duas delas terem a mesma data de aniversário? (Assuma que ninguém nasceu em 29 de fevereiro e que cada dia tem a mesma chance de ser o aniversário de alguém).

Para responder à pergunta no Exemplo 2.1, precisamos calcular a cardinalidade do conjunto de todas as possíveis combinações de datas de aniversário, bem como a cardinalidade do conjunto de todas as possíveis combinações de datas de aniversário em que pelo menos duas são iguais. Como fazemos isso? Precisaremos usar o princípio fundamental da contagem, que nos permite calcular as cardinalidades de conjuntos grandes e complexos, onde a contagem explícita não é possível.

Primeiro, começamos identificando as regras fundamentais da contagem:

Regra da Correspondência Se A e B estão em correspondência um-para-um, então $|A| = |B|$.

Regra da Adição Se A_1, \dots, A_n são subconjuntos mutuamente disjuntos de algum conjunto, então

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{i=1}^n |A_i|.$$

Princípio Fundamental da Contagem Suponha que os elementos de um conjunto finito E podem ser determinados em k etapas sucessivas, com n_1 escolhas possíveis na etapa 1, n_2 escolhas possíveis na etapa 2, \dots , n_k escolhas possíveis na etapa k . Suponha também que escolhas diferentes levam a elementos diferentes. Então,

$$|E| = n_1 \cdot n_2 \cdot \dots \cdot n_k.$$

O conjunto de todas as combinações de datas de aniversário de 30 pessoas, como mostrado no Exemplo 2.1, é um exemplo de um conjunto de k -tuplas ordenadas

de elementos de um conjunto dado. Mais genericamente, isso é definido da seguinte forma:

Definição 2.2. Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$. Uma *sequência de comprimento* $k \in \mathbb{N}$ *de elementos de* A é uma k -tupla ordenada (a_1, \dots, a_k) tal que $a_i \in A$, $i = 1, 2, \dots, k$. Denotamos por $S_{n,k}(A)$ o conjunto de todas as sequências de comprimento k de elementos de A .

Por “ordenado”, queremos dizer que a ordem da sequência importa, por exemplo: $(a_1, a_2) \neq (a_2, a_1)$. Note também que repetições de elementos são permitidas, por exemplo, $(1, 1)$ é uma sequência de comprimento 2 de elementos de $\{1\}$.

Proposição 2.3. *Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$. O conjunto $S_{n,k}(A)$ de todas as sequências de comprimento $k \in \mathbb{N}$ de elementos de A tem cardinalidade n^k , ou seja,*

$$|S_{n,k}(A)| = n^k$$

Demonstração. Para construir um elemento arbitrário (a_1, a_2, \dots, a_k) de $S_{n,k}(A)$, realizamos as seguintes etapas:

- (a) escolher o primeiro valor a_1 . Existem $n_1 = |A| = n$ maneiras de fazer isso.
- (b) escolher o segundo valor a_2 . Existem $n_2 = |A| = n$ maneiras de fazer isso.

\vdots

- k . Escolher o valor a_k do k° elemento. Existem $n_k = |A| = n$ maneiras de fazer isso.

Portanto, encontramos

$$|S_{n,k}(A)| = n_1 \cdot n_2 \cdot \dots \cdot n_k = \underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ vezes}} = n^k.$$

□

Para completar o Exemplo 2.1, também precisamos calcular a cardinalidade do conjunto de todas as datas de aniversário possíveis em que pelo menos duas são iguais. É mais fácil e equivalente (por quê?) calcular a cardinalidade do conjunto em que nenhum dois aniversários são iguais. Como não pode haver repetições,

isso é um exemplo de uma "ordenação de comprimento 30 de elementos de $\{1, \dots, 365\}$ ". Mais genericamente, definimos ordenações de comprimento k da seguinte forma:

Definição 2.4. Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$ e seja $k \in \mathbb{N}$ tal que $k \leq n$. Uma *ordenação de comprimento k de elementos de A* é uma sequência de comprimento k de elementos de A sem repetições. Denotamos o conjunto de ordenações de comprimento k de elementos de A por $O_{n,k}(A)$. Assim, temos

$$O_{n,k}(A) = \{(a_1, \dots, a_k) : a_i \in A \forall i = 1, \dots, k, \quad a_i \neq a_j \forall i \neq j\}.$$

Proposição 2.5. *Seja A um conjunto finito de cardinalidades $n \in \mathbb{N}$ e $k \leq n$. Então*

$$|O_{n,k}(A)| = n(n-1) \dots (n-k+1).$$

Demonstração. Determinamos um elemento de $O_{n,k}(A)$ pelas seguintes etapas:

- (a) Escolhemos a_1 . Existem $n_1 = |A| = n$ opções para isso.
- (b) Escolhemos a_2 tal que $a_2 \neq a_1$. Existem $n_2 = n - 1$ opções para isso.
- (c) Escolhemos a_3 tal que $a_3 \neq a_2$ e $a_3 \neq a_1$. Existem $n_3 = n - 2$ opções para isso.
- \vdots
- k. Escolhemos a_k tal que $a_k \neq a_i \forall i = 1, 2, \dots, k-1$. Existem $n_k = n - (k-1)$ opções para isso.

Pelo princípio fundamental da contagem

$$|O_{n,k}(A)| = n_1 n_2 \dots n_k = n(n-1) \dots (n-k+1).$$

□

Exemplo 2.6 (continuação). Agora podemos responder à questão do Exemplo 2.1. Primeiro, calculamos a probabilidade de que duas pessoas não façam aniversário no mesmo dia, correspondendo ao evento B. Conforme discutido acima, B é uma ordenação de comprimento 30 ($k = 30$ de um conjunto de cardinalidade 365

($n = 365$) , então

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{365 \times \cdots \times (365 - 30 + 1)}{365^{30}} \approx 0,29$$

O evento de pelo menos duas pessoas fazerem aniversário no mesmo dia é o complemento do evento B, então a probabilidade de pelo menos duas entre 30 pessoas fazerem aniversário no mesmo dia será próxima de 71%.

Observações. Lemos $n!$ como n fatorial. A seguinte espera:

- $n!$ é o produto dos primeiros n números naturais e por suposição $0! = 1$.
- $n!$ é o número de maneiras pelas quais podemos ordenar os elementos de um conjunto de cardinalidade n ou equivalentemente o número de maneiras de colocar os elementos de um conjunto de cardinalidade n em uma linha.
- $n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$ é o número de maneiras de colocar k elementos de um conjunto de cardinalidade n consecutivos.

Exemplo 2.7. Agora consideramos uma questão ligeiramente diferente daquela do Exemplo 2.1: qual é a probabilidade de *exatamente* duas pessoas na sala fazerem aniversário no mesmo dia? Para construir tal exemplo, precisaríamos

1. Escolha as duas pessoas que fazem aniversário no mesmo dia.
2. Escolha um dia para o aniversário deles.
3. Escolha um dia para o aniversário de todos os outros, para que nenhum outro aniversário seja igual.

De quantas maneiras podemos escolher duas pessoas que fazem aniversário no mesmo dia? Precisamos escolher dois números de $C = \{1, \dots, 30\}$ – esta será uma sequência de comprimento 2 sem repetição, mas o que é diferente do que tínhamos antes é que a ordem não importa. Quer seja $(1, 2)$ ou $(2, 1)$, ainda é o mesmo par de pessoas com a mesma data de aniversário! Para corrigir isso, precisamos dividir por todas as maneiras possíveis pelas quais podemos ordenar os dois elementos - cada forma será uma sequência de comprimento 2 sem repetição, mas agora estamos escolhendo a partir de um conjunto de apenas dois pontos, então o conjunto de todas as possibilidades é $O_{2,2}$. Isto dá

$$\frac{|O_{30,2}|}{|O_{2,2}|} = \frac{30 \times 29}{2!} = \frac{30!}{28!2!}.$$

Denotamos isso por $\binom{30}{2}$. Agora, voltando à nossa questão: calculamos o número

de maneiras pelas quais podemos escolher as duas pessoas com o mesmo aniversário. Existem 365 maneiras de escolher o aniversário. Para as 28 pessoas restantes, haverá $364 \times \cdots \times (365 - 28)$ maneiras de escolher seus aniversários, já que todos precisam ser diferentes. Portanto, o número total de maneiras de selecionar um resultado no evento “exatamente duas pessoas fazem aniversário no mesmo dia” é

$$\frac{30!}{28!2!} 365 \times 364 \times \cdots \times (365 - 28).$$

Finalmente, para calcular a probabilidade de exatamente duas pessoas fazerem aniversário no mesmo dia, precisamos dividir pela cardinalidade de todas as combinações de aniversários possíveis dada por 365^{30} , o que dá aproximadamente 0,38 ou 38%.

Escolher duas pessoas de um conjunto de 30 é um exemplo de combinação de 2 elementos de $\{1, \dots, 30\}$. De forma mais geral, podemos perguntar o número de combinações de k elementos de um conjunto finito A .

Definição 2.8. Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$. Uma *combinação de k elementos de A* é um subconjunto de A com k elementos. Denotamos por $C_{n,k}(A)$ o conjunto de combinações de k elementos de A .

xxx xxx xxx xxx xxx xxx xxx xxx xxx

Proposição 2.9. *Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$ e $k \leq n$. Então*

$$|C_{n,k}(A)| = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Demonstração. Observe que uma ordenação de comprimento k dos elementos de A pode ser obtida *unicamente* pelos seguintes passos:

- (a) Escolher uma combinação de k elementos de A . Existem $n_1 = |C_{n,k}(A)|$ escolhas para isso.
- (b) Escolher uma permutação desses elementos. Pelo Corolário 2.11, existem $n_2 = k!$ escolhas para isso.

Pelo princípio fundamental da contagem,

$$|O_{n,k}(A)| = |C_{n,k}(A)| \cdot k!.$$

A equação acima pode ser resolvida para o único termo desconhecido, resultando em

$$|C_{n,k}(A)| = \frac{|O_{n,k}(A)|}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}.$$

Isso conclui a prova. \square

O número de ordenações de comprimento n quando a cardinalidade do conjunto também é n é um caso especial importante das ordenações:

Definição 2.10. Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$. Uma ordenação de comprimento n dos elementos de A é chamada de *permutação* de A .

Corolário 2.11. Seja A um conjunto finito de cardinalidade $n \in \mathbb{N}$. Então o número de permutações dos elementos de A é $n(n-1)(n-2)\dots 1 = n!$.

Demonstração. É suficiente tomar $k = n$ na Proposição 2.5. \square

Exercício 2.1. Um dado justo é lançado 8 vezes. Quantos resultados diferentes podemos obter que contenham o resultado 2 exatamente três vezes e o resultado 3 exatamente cinco vezes?

Solução. Cada resultado possível é completamente determinado especificando quais lançamentos resultam em um 2 - o restante será 3. Por exemplo, o resultado (2, 3, 2, 2, 3, 3, 3, 3) é determinado pelo subconjunto $\{1, 3, 4\}$ do conjunto $\{1, 2, 3, 4, 5, 6, 7, 8\}$, em que o primeiro corresponde às posições do 2 e o segundo é a numeração de todos os lançamentos. Portanto, para calcular o número de resultados possíveis, é suficiente calcular o número de subconjuntos de tamanho 3 de um conjunto com cardinalidade 8. Isso é exatamente $C_{8,3}$, que é igual a 56. \square

Exemplo 2.12. Vamos levar o exercício 2.1 um pouco mais adiante: como calcularíamos o número de resultados de 8 lançamentos de um dado que contenham exatamente três 2, três 4 e dois 5? Como antes, um resultado será completamente determinado especificando as posições de dois dos três possíveis resultados, por exemplo, 2 e 4. Por exemplo, o resultado (2, 4, 5, 2, 2, 5, 4, 4) corresponde aos conjuntos $A_2 = \{1, 4, 5\}$ e $A_4 = \{2, 7, 8\}$, onde A_2 é o conjunto

de posições do resultado 2 e da mesma forma para A_4 . Segue-se que $A_5 = \{3, 6\}$, pois essas são as únicas duas posições restantes.

Portanto, para responder à pergunta, precisamos contar quantas maneiras podemos escolher um subconjunto de $\{1, 2, 3, 4, 5, 6, 7, 8\}$ de tamanho 3 e, em seguida, outro subconjunto de tamanho 3 dos 5 elementos restantes. Temos $\binom{8}{3}$ escolhas para o primeiro conjunto e $\binom{5}{3}$ escolhas para o segundo. Aplicando o princípio fundamental da contagem, obtemos

$$\binom{8}{3} \binom{5}{3} = \frac{8!}{5!3!} \frac{5!}{3!2!} = \frac{8!}{3!3!2!}.$$

O exemplo acima ilustra a contagem das maneiras de dividir um conjunto em um número fixo de subconjuntos. Para definir isso formalmente, primeiro precisamos definir uma partição:

Definição 2.13. Seja A um conjunto de cardinalidade $n \in \mathbb{N}$ e $r \in \mathbb{N}$ tal que $r \leq n$. Uma *partição de A em r subconjuntos* é uma família $\{A_1, \dots, A_r\}$ de subconjuntos de A tal que

- (a) Cada subconjunto na família é não vazio: $A_i \neq \emptyset \forall i = 1, 2, \dots, r$.
- (b) Os subconjuntos na família são mutuamente disjuntos: $A_i \cap A_j = \emptyset \forall i \neq j$, $i, j = 1, 2, \dots, r$.
- (c) A união de todos os subconjuntos na família é igual a A : $\bigcup_{i=1}^r A_i = A$.

Proposição 2.14. *Seja A um conjunto finito com $|A| = n$. Seja $r \in \mathbb{N}$, $r \leq n$. Então, o número de partições de A em r subconjuntos $\{A_1, \dots, A_r\}$, de forma que $|A_1| = k_1$, $|A_2| = k_2 \dots$, $|A_r| = k_r$ ($k_1 + k_2 + \dots + k_r = n$, $1 \leq k_i \leq n$), é dado por*

$$\frac{n!}{k_1! k_2! \dots k_r!}.$$

Demonstração. Toda partição de A que satisfaça as suposições pode ser determinada de forma única pelos seguintes passos:

- (a) Escolher $A_1 \subseteq A$ de modo que $|A_1| = k_1$. Existem $\binom{n}{k_1}$ opções para este passo.
- (b) Escolher $A_2 \subseteq A$ de modo que $|A_2| = k_2$ e $A_1 \cap A_2 = \emptyset$ (o que implica que $A_2 \subseteq A \setminus A_1$). Existem $\binom{n-k_1}{k_2}$ opções para este passo.

- (c) Escolher $A_3 \subseteq A$ de modo que $|A_3| = k_3$, $A_3 \cap A_2 = \emptyset$ e $A_3 \cap A_1 = \emptyset$ (o que implica que $A_3 \subseteq A \setminus (A_1 \cup A_2)$). Existem $\binom{n-k_1-k_2}{k_3}$ opções para este passo.

\vdots

- r . Finalmente, escolher o conjunto restante $A_r \subseteq A$ de modo que $|A_r| = k_r$ e $A_i \cap A_j = \emptyset$ para todos $i = 1, 2, \dots, r-1$ (notamos que $A_r \subseteq A \setminus (A_1 \cup \dots \cup A_{r-1})$). Existem $\binom{n-(k_1+k_2+\dots+k_{r-1})}{k_r}$ opções para este passo.

Assim, pelo princípio fundamental da contagem, o número de partições de A em r subconjuntos $\{A_1, A_2, \dots, A_r\}$ de modo que $|A_1| = k_1, \dots, |A_r| = k_r$ com $k_1 + k_2 + \dots + k_r = n$ é

$$\begin{aligned} & \binom{n}{k_1} \binom{n-k_1}{k_2} \dots \binom{n-(k_1+\dots+k_{r-1})}{k_r} \\ &= \frac{n!}{k_1!(n-k_1)!} \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \dots \frac{(n-(k_1+\dots+k_{r-1}))!}{k_r!(n-(k_1+\dots+k_r))!} \\ &= \frac{n!}{k_1!k_2!\dots k_r!}, \end{aligned}$$

onde a última igualdade ocorre pela simplificação das frações e observando que $(n-(k_1+\dots+k_r))! = (n-n)! = 0! = 1$. \square

2.2 Amostragem

Selecionar um subconjunto de um conjunto maior também é chamado de amostragem. A amostragem nos permite usar informações sobre um pequeno grupo para fazer inferências sobre as propriedades ou preferências de um grupo maior. É uma ferramenta fundamental em estatísticas. Quando a população é "homogênea" (cada pessoa no grupo provavelmente tem as mesmas propriedades ou preferências), qualquer amostra escolhida aleatoriamente (com probabilidade uniforme) será representativa do grupo - ainda precisamos usar ferramentas matemáticas avançadas para quantificar a incerteza em nossas inferências sobre uma população, dada o tamanho da amostra, mas esse é o assunto de um módulo diferente. O que gostaríamos de entender agora é como a estrutura de probabilidade nas amostras varia quando a população é mista. Para entender a pergunta, considere o seguinte.

Exemplo 2.15. O professor de ST120 quer saber até que ponto os alunos entenderam o conceito de espaço de probabilidade. Dado o tamanho da turma, não é possível perguntar a cada aluno individualmente. Em vez disso, eles desejam amostrar um pequeno grupo de alunos e perguntar a eles. Como eles devem escolher os alunos?

Uma solução prática é conversar com alguns dos alunos na sala de aula. No entanto, existe um viés - os alunos que frequentam as aulas têm mais probabilidade de entender os conceitos! Portanto, o professor decide escolher aleatoriamente um número de números de identificação de alunos e enviá-los por e-mail, e vamos supor que todos os alunos respondam. Este seria um grupo representativo, se a turma fosse homogênea. No entanto, sabemos que a turma é composta por dois grupos de n_1 alunos de Matemática e n_2 alunos de Ciência da Computação. Para entender o viés, o professor precisa calcular a probabilidade de o grupo acabar com k_1 alunos de Matemática e k_2 alunos de Ciência da Computação. Qual seria essa probabilidade?

Existem $\binom{n_1}{k_1}$ maneiras de escolher k_1 alunos de Matemática e $\binom{n_2}{k_2}$ de escolher k_2 alunos de Ciência da Computação. Portanto, a cardinalidade do conjunto de todos os grupos de k_1 alunos de Matemática e k_2 alunos de Ciência da Computação será

$$\binom{n_1}{k_1} \binom{n_2}{k_2}$$

A cardinalidade do conjunto de todos os grupos de $k_1 + k_2$ alunos será $\binom{n_1+n_2}{k_1+k_2}$. Portanto, a probabilidade de escolher um grupo com k_1 alunos de Matemática e k_2 alunos de Ciência da Computação é

$$\frac{\binom{n_1}{k_1} \binom{n_2}{k_2}}{\binom{n_1+n_2}{k_1+k_2}}.$$

O próximo passo seria usar essa probabilidade para remover o viés, mas eles precisarão consultar os professores de módulos de estatísticas mais avançados sobre como fazer isso!

O exemplo acima é um caso de amostragem de uma população de tamanho $n \in \mathbb{N}$, que tem $n_1 \in \mathbb{N}$ ($n_1 \leq n$) indivíduos do tipo 1 e $n_2 = n - n_1$ indivíduos do tipo 2. Retiramos uma amostra de tamanho $k < N$ da população inteira 'aleatoriamente' *sem reposição* (ou seja, um indivíduo não pode ser escolhido

duas vezes). Então, a probabilidade de a amostra conter k_1 indivíduos do tipo 1 e k_2 indivíduos do tipo 2 é dada por

$$\frac{\binom{n_1}{k_1} \binom{n_2}{k_2}}{\binom{n}{k}}. \quad (2.16)$$

Exercício 2.2. Construa o espaço de probabilidade uniforme correspondente à amostragem sem reposição de uma população com dois tipos de indivíduos e prove (2.16). Em outras palavras, forneça o triplo $(\Omega, \mathcal{F}, \mathbb{P})$ de acordo com a proposição 1.4, de modo que, todas as realizações $\omega \in \Omega$ tenham a mesma probabilidade. O evento de interesse ω é o conjunto de todas as combinações que contêm k_1 números de 1 a n_1 e $k_2 = k - k_1$ números de $n_1 + 1$ a n .

Exemplo 2.17. Agora considere o seguinte problema: um biólogo ambiental está estudando o peso de uma espécie específica de peixe. Para fazer isso, eles retiram uma amostra dessa população de peixes, pesam os peixes e depois os liberam de volta à natureza. Suponha que não há maneira de saber se um peixe já foi amostrado. Além disso, suponha que a população consiste apenas em dois tipos de peixes - machos e fêmeas - que têm pesos médios ligeiramente diferentes e isso precisa ser levado em consideração ao remover o viés.

Se a população inteira de peixes for n , com n_1 machos e $n_2 = n - n_1$ fêmeas e o tamanho da amostra for k , qual é a probabilidade de ter k_1 machos e k_2 fêmeas na amostra?

A diferença entre este exemplo e o exemplo 2.15 é que agora os indivíduos podem ser escolhidos mais de uma vez. Como resultado, não é mais suficiente determinar apenas as posições dos indivíduos do tipo 1 na amostra, mas também precisamos especificar os indivíduos para que possamos acompanhar os escolhidos repetidamente. Portanto, passamos pelo seguinte processo:

- Escolha a posição dos machos (as posições restantes serão ocupadas pelas fêmeas) - há $\binom{k}{k_1}$ escolhas.
- Escolha os machos que são escolhidos - há $n_1^{k_1}$ escolhas.
- Escolha as fêmeas que são escolhidas - há $n_2^{k_2}$ escolhas.

Portanto, o número de amostras diferentes com k_1 machos e k_2 fêmeas será

$$\binom{k}{k_1} n_1^{k_1} n_2^{k_2}.$$

Dado que o número total de amostras possíveis é n^k , a probabilidade de escolher k_1 machos e k_2 fêmeas será dada por

$$\frac{\binom{k}{k_1} n_1^{k_1} n_2^{k_2}}{n^k} = \binom{k}{k_1} \left(\frac{n_1}{n}\right)^{k_1} \left(1 - \frac{n_1}{n}\right)^{k-k_1}. \quad (2.18)$$

O acima é um exemplo de amostragem de uma população de tamanho $n \in \mathbb{N}$, que possui $n_1 \in \mathbb{N}$ ($n_1 \leq n$) indivíduos do tipo 1 e $n_2 = n - n_1$ indivíduos do tipo 2. Nós tiramos uma amostra de tamanho $k < N$ da população inteira 'aleatoriamente' *com reposição* (ou seja, um indivíduo pode ser escolhido duas vezes). Então, a probabilidade de a amostra conter k_1 indivíduos do tipo 1 e $k_2 = k - k_1$ indivíduos do tipo 2 é dada por (2.18).

Exercício 2.3. Construa o espaço de probabilidade uniforme correspondente à amostragem com reposição de uma população com dois tipos de indivíduos e prove (2.18). Agora, o evento de interesse ω é o conjunto de todas as sequências que contêm k_1 números de 1 a n_1 e $k_2 = k - k_1$ números de $n_1 + 1$ a n .

Exercício 2.4. As probabilidades de amostragem não devem ser sensíveis a qual tipo foi considerado '1' e qual tipo foi considerado '2'. De fato,

$$\binom{k}{k_1} \left(\frac{n_1}{n}\right)^{k_1} \left(1 - \frac{n_1}{n}\right)^{k-k_1} = \binom{k}{k_2} \left(\frac{n_2}{n}\right)^{k_2} \left(1 - \frac{n_2}{n}\right)^{k-k_2}.$$

Supondo que $k_1 + k_2 = k$ e $n_1 + n_2 = n$, verifique a identidade acima.

Revisão de Combinatória

Conjunto finito A , $n = |A|$.

Sequências de comprimento k de elementos de A :

$$|S_{n,k}(A)| = n^k,$$

pois para especificar um elemento de $S_{n,k}(A)$, precisamos fazer k escolhas, e em cada escolha temos n opções.

Permutações (ou reorganizações) de elementos de A :

$$n!,$$

pois precisamos fazer n escolhas, e a cada escolha o número de opções diminui. (leia-se "fatorial de n ")

Ordenação de comprimento k de elementos de A :

$$|O_{n,k}(A)| = \frac{n!}{(n-k)!},$$

porque podemos obter qualquer elemento de $O_{n,k}(A)$ permutando os elementos de A , mantendo os primeiros k elementos e esquecendo a ordem dos $n-k$ restantes.

Subconjuntos de A com cardinalidade k :

$$|C_{n,k}(A)| = \binom{n}{k} = \frac{n!}{k!(n-k)!},$$

pois podemos obter qualquer elemento de $C_{n,k}(A)$ tomando os primeiros k elementos em uma permutação de A , mantendo os primeiros k elementos e depois esquecendo a ordem desses k elementos, assim como os $n-k$ restantes. (leia-se "combinação de n escolha k ")

Decomposição de A em conjuntos *rotulados* A_1, \dots, A_r de modo que $|A_j| = k_j$ para $j = 1, \dots, r$, onde $k_1 + \dots + k_r = n$.

$$\frac{n!}{k_1!k_2! \dots k_r!},$$

pois podemos obter a partição primeiro permutando todos os elementos de A , tomando A_1 como os primeiros k_1 elementos dessa permutação, A_2 como os próximos k_2 elementos e assim por diante, e depois esquecendo a ordem de cada bloco. Este é o número de maneiras pelas quais n bolas rotuladas podem ser distribuídas em r baldes rotulados sob a restrição de que, para cada $j = 1, \dots, r$, o balde número j recebe k_j bolas. As combinações $C_{n,k}(A)$ são apenas um caso particular de dois baldes rotulados como "dentro" e "fora".

Para fornecer uma justificação mais rigorosa para os termos no denominador (que chamamos de "esquecer a ordenação"), podemos obter o número raciocinando de trás para frente, como segue. Vamos produzir uma permutação de $A = \{1, \dots, n\}$ de duas maneiras. A primeira maneira tem três etapas:

- (a) Escolher k elementos de A para irem primeiro, e deixar os $n-k$ restantes

irem depois. Existem x possibilidades.

(b) Permutar os primeiros k elementos. Existem $k!$ possibilidades.

(c) Permutar os elementos restantes $n - k$. Existem $(n - k)!$ possibilidades.

No total, existem $x \cdot k! \cdot (n - k)!$ possibilidades. A segunda maneira é direta: apenas permutar os n elementos já existentes, existem $n!$ possibilidades. Portanto, $n! = x \cdot k! \cdot (n - k)!$, então acabamos de descobrir o que é x ! Portanto, $|C_{n,k}(A)| = \frac{n!}{k!(n-k)!}$.

Revisão de Amostragem

População com $n = n_1 + n_2$ indivíduos, onde n_1 são indivíduos do Tipo 1 e n_2 são indivíduos do Tipo 2.

Se tirarmos uma amostra de k indivíduos, *sem reposição*, então a chance de escolher k_1 elementos do Tipo 1 (e, portanto, $k_2 = k - k_1$ indivíduos do Tipo 2) é

$$\frac{\binom{n_1}{k_1} \cdot \binom{n_2}{k_2}}{\binom{n}{k}}.$$

Isso pode ser obtido assumindo que os indivíduos foram amostrados simultaneamente (então $|\Omega| = \binom{n}{k}$) ou um após o outro (então $|\Omega| = \frac{n!}{(n-k)!}$), e ambas as abordagens fornecem a mesma probabilidade (como esperado!). De fato, a segunda abordagem fornece

$$\frac{\binom{k}{k_1} \frac{n_1!}{(n_1-k_1)!} \frac{n_2!}{(n_2-k_2)!}}{\frac{n!}{(n-k)!}}$$

que se simplifica para a primeira fórmula se expandirmos o primeiro termo.

Se tirarmos uma amostra de k indivíduos, *com reposição*, então a chance de escolher k_1 elementos do Tipo 1 (e, portanto, $k_2 = k - k_1$ indivíduos do Tipo 2) é

$$\binom{k}{k_1} \left(\frac{n_1}{n}\right)^{k_1} \left(1 - \frac{n_1}{n}\right)^{k_2}.$$

Isso só pode ser alcançado assumindo que os indivíduos foram amostrados um após o outro, para que possamos devolvê-los à população (então $|\Omega| = n^k$). A fórmula acima é obtida após reescrever

$$\frac{\binom{k}{k_1} n_1^{k_1} n_2^{k_2}}{n^k}$$

de uma forma mais conveniente (ou significativa).

Observe que, se X denota o número de indivíduos do Tipo 1 na amostra com reposição, então

$$X \sim \text{Binom}(k, \frac{n_1}{n}),$$

o que pode ser verificado combinando a fórmula acima com a função de massa de probabilidade de uma variável aleatória binomial.

3 Espaços de Probabilidade

Na seção 1.2, definimos um espaço de probabilidade uniforme como o triplete $(\Omega, \mathcal{F}, \mathbb{P})$, onde Ω é um conjunto finito, $\mathcal{F} = \mathcal{P}(\Omega)$ e $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfaz as propriedades (1.2a) - (1.2c). Agora, vamos generalizar o conceito de espaço de probabilidade para permitir o seguinte:

- Espaços amostrais arbitrários Ω .
- Espaços de eventos que refletem informações parciais - não é necessário, e às vezes nem é possível, que $\mathcal{F} = \mathcal{P}(\Omega)$.
- Probabilidade definida em um espaço de eventos geral.

3.1 Espaço amostral e espaço de eventos

Definição 3.1. Um *espaço amostral* Ω é o conjunto de todos os possíveis resultados de um processo aleatório (ou experimento), ou seja, um processo cujo resultado não pode ser determinado antecipadamente. Pode ser qualquer conjunto.

Exemplo 3.2. Qual é o espaço amostral correspondente aos seguintes processos?

- O lançamento de uma moeda: $\Omega = \{Cara, Coroa\}$.
- O lançamento de um dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- O número de e-mails enviados por um endereço @google.com em um ano: $\Omega = \mathbb{N}$.
- O peso de uma maçã: $\Omega = [0, 1]$.
- A posição de um dardo lançado em um tabuleiro quadrado de tamanho 1: $\Omega = [0, 1] \times [0, 1]$.
- O preço das ações do Twitter em um ano: $\Omega = \mathbb{R}$.
- As flutuações de temperatura em Coventry em 2023: neste caso, o espaço amostral é uma função completa, mapeando o tempo t para um número em $[-50, 50]$.
- O estado do mundo em um ano! Neste caso, o espaço amostral não pode ser descrito, mas existe como conceito e pode ser parcialmente observado por meio de sua interação com processos que podem ser medidos (por exemplo, afetará as taxas de juros ou o número de internações hospitalares em um determinado dia no futuro, a frequência de eventos climáticos extremos,

etc).

Observação 1. Embora Ω possa ser qualquer conjunto em teoria, em ST120, consideraremos apenas os casos em que a cardinalidade de Ω é igual a n , para algum $n \in \mathbb{N}$ (espaço amostral finito), $|\Omega| = |\mathbb{N}|$ (espaço amostral contável) ou $|\Omega| = |\mathbb{R}|$ (espaço amostral não contável ou contínuo - para incluir intervalos ou produtos cartesianos de \mathbb{R} e seus intervalos).

Como já vimos no caso do espaço de probabilidade uniforme, eventos são subconjuntos de Ω . Mais geralmente, um evento é um subconjunto de Ω quando é possível dizer se algum resultado dado pertence ao conjunto (ou seja, 'o evento ocorreu') ou não, dadas as informações que temos sobre o resultado - observe que nem sempre temos informações completas sobre o resultado e o espaço de eventos reflete as informações que temos. O exemplo a seguir demonstra exatamente essa propriedade do espaço de eventos.

Exemplo 3.3. Suponha que eu lance um dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$) e relato as seguintes informações a dois alunos: eu digo a James se o resultado é um número par ou não, e digo a Lily o quociente de $(\omega - 1)$ dividido por dois (ou seja, relato 0 para $\{1, 2\}$, 1 para $\{3, 4\}$ e 2 para $\{5, 6\}$). Quais são os espaços de eventos correspondentes?

James só sabe se o resultado é ímpar ou par, então ele só pode dizer se pertence a $\{1, 3, 5\}$ ou $\{2, 4, 6\}$. Já que, pela construção de Ω , todos os resultados estão em Ω , ele também pode dizer que o resultado está em Ω e não em \emptyset - observe que tanto \emptyset quanto Ω são subconjuntos de Ω . Portanto, a coleção de eventos (ou seja, o espaço de eventos) correspondente às informações que James possui é

$$\mathcal{F}_J = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}.$$

(Pense sobre o motivo pelo qual James não pode afirmar com certeza se o resultado está em outro subconjunto).

Com base nas informações fornecidas a ela, Lily poderá dizer se o resultado está em $\{1, 2\}$, $\{3, 4\}$ ou $\{5, 6\}$. Ela também pode dizer se o resultado estará em $\{1, 2, 3, 4\}$, $\{1, 2, 5, 6\}$ ou $\{3, 4, 5, 6\}$ (por exemplo, $\{1, 2, 3, 4\}$ corresponde ao número informado a Lily sendo 0 ou 1) e ela pode dizer, por padrão, que o resultado estará em Ω e não em \emptyset (o \emptyset é definido como o complemento de Ω em relação a Ω , então todos os pontos em Ω que não estão em Ω , que é, é claro,

nenhum! É mais fácil pensar nisso como o evento em que o resultado não está no espaço amostral do que no evento em que nada acontece). Portanto, o espaço de eventos correspondente às informações fornecidas a Lily é

$$\mathcal{F}_L = \{\emptyset, \{1, 2\}, \{3, 4\}, \{5, 6\}, \{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{3, 4, 5, 6\}, \Omega\}.$$

Observe que há redundância na forma como as informações são codificadas no espaço de eventos - a informação corresponde a saber simultaneamente a resposta para 'o evento ocorreu ou não' para todos os eventos no espaço de eventos, mas saber, por exemplo, que tanto $\{1, 2\}$ quanto $\{1, 2, 3, 4\}$ aconteceram nos permite deduzir a resposta para tudo o mais. Que suposição estamos fazendo que nos permite dizer isso?

Se A e B são eventos, de acordo com nossa intuição, esperamos que o seguinte também sejam eventos:

- $A \cap B$ (Ambos A e B aconteceram).
- $A \cup B$ (Ou A ou B aconteceu).
- A^c (A não aconteceu).
- $A \setminus B = A \cap B^c$ (A aconteceu, mas B não aconteceu).

Notação. Quando escrevemos A^c , implicitamente estamos considerando o complemento em relação a um espaço amostral dado Ω . Uma notação mais explícita é escrever $\Omega \setminus A$.

Portanto, gostaríamos que o espaço de eventos fosse fechado sob as operações de união, interseção, complemento e diferença (quando dizemos que um conjunto é fechado sob uma operação, queremos dizer que, se aplicarmos a operação a quaisquer elementos do conjunto, o resultado ainda estará no conjunto). Será que isso é suficiente? Vamos considerar o seguinte

Exemplo 3.4. Considere o caso em que $\Omega = \mathbb{N}$ (ou seja, qualquer número natural pode ser o resultado do processo aleatório que consideramos) e suponha que temos informações suficientes para dizer se o evento $\{n\}$ aconteceu ou não, para qualquer $n \in \mathbb{N}$. De acordo com nossa intuição, se podemos dizer se qualquer resultado ω pertence a qualquer conjunto $\{n\}$ ou não (o que você pode dizer para ω quando $\omega \in \{n\}$?), então também deveríamos ser capazes de dizer se $\omega \in \{2n | n \in \mathbb{N}\}$ ou não, ou seja, deveríamos ser capazes de dizer se ω é par.

Portanto, esperamos

$$\bigcup_{n=0}^{\infty} \{2n\} = \{2n | n \in \mathbb{N}\}$$

estar no espaço de eventos. Agora estamos fazendo a suposição de que o espaço de eventos não é apenas fechado para uniões, mas também para uniões de conjuntos contáveis (ou seja, infinitos, mas com cardinalidade igual a $|\mathbb{N}|$).

Seguindo nossa intuição, definimos o espaço de eventos da seguinte forma

Definição 3.5. Seja Ω o espaço amostral e \mathcal{F} uma coleção de subconjuntos de Ω . \mathcal{F} é um *espaço de eventos* (também chamado de σ -álgebra) se satisfaz as seguintes condições:

- (i) $\Omega \in \mathcal{F}$.
- (ii) se $A \subseteq \Omega$, $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (\mathcal{F} é fechado sob complementos).
- (iii) se $\{A_n : n \in \mathbb{N}\}$ é tal que $A_n \in \mathcal{F} \forall n$, então

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

(\mathcal{F} é fechado sob uniões contáveis.)

Exercício 3.1. Seja Ω um conjunto não vazio e $\mathcal{F} = \mathcal{P}(\Omega)$, ou seja, o conjunto de todos os subconjuntos de Ω . Então, \mathcal{F} é um espaço de eventos em Ω .

Solução. Precisamos mostrar que $\mathcal{F} = \mathcal{P}(\Omega)$ satisfaz as três propriedades da 3.5.

- (i) $\Omega \subseteq \Omega$ e, portanto, $\Omega \in \mathcal{P}(\Omega) = \mathcal{F}$.
- (ii) Suponha que $A \subseteq \Omega$, $A \in \mathcal{F}$. Então, $A^c = \Omega \setminus A \subseteq \Omega$. Assim, $A^c \in \mathcal{P}(\Omega) = \mathcal{F}$.
- (iii) Suponha que A_n são tais que $A_n \subseteq \Omega$, $\forall n \in \mathbb{N}$. Então

$$\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \Omega = \Omega \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{P}(\Omega) = \mathcal{F}.$$

Exercício 3.2. Seja $A \subseteq \Omega$ um subconjunto não vazio de Ω . Então

$$\{\emptyset, A, A^c, \Omega\}$$

é um espaço de eventos em Ω .

Solução. Precisamos mostrar que $\{\emptyset, A, A^c, \Omega\}$ satisfaz as três propriedades da

3.5.

- (i) $\Omega \in \mathcal{F}$ por definição.
- (ii) Verificamos que a propriedade é satisfeita para cada evento: $\emptyset^c = \Omega \in \mathcal{F}$, $A^c \in \mathcal{F}$, $(A^c)^c = A \in \mathcal{F}$, $\Omega^c = \emptyset \in \mathcal{F}$.
- (iii) Seja $\{B_n : n \in \mathbb{N}\}$ uma sequência de subconjuntos de Ω tal que $B_n \in \mathcal{F} \forall n$. Vamos considerar todas as possibilidades:
 - Se todos os conjuntos B_n forem idênticos e iguais a B (ou seja, $B \in \mathcal{F}$), ou existe uma subsequência B_{n_k} de conjuntos que são idênticos a B e o restante é todo \emptyset , então $\bigcup_{n=1}^{\infty} B_n = B \in \mathcal{F}$.
 - Se pelo menos um dos conjuntos for o espaço amostral (por exemplo, $B_i = \Omega$, para algum $i \in \mathbb{N}$), então

$$\bigcup_{n=1}^{\infty} B_n = \Omega$$

e, portanto, $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$.

- Se houver pelo menos um conjunto A e um conjunto A^c , então

$$\Omega = A \cup A^c \subseteq \bigcup_{n=1}^{\infty} B_n \subseteq \Omega,$$

onde o último relacionamento segue do fato de que todos os eventos são subconjuntos de Ω . Portanto, $\bigcup_{n=1}^{\infty} B_n = \Omega$ e, assim, $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$.

Proposição 3.6. *Seja \mathcal{F} um espaço de eventos em Ω . Então*

- (a) \mathcal{F} é fechado sob uniões finitas.
- (b) \mathcal{F} é fechado sob interseções finitas.
- (c) \mathcal{F} é fechado sob interseções contáveis.

Demonstração.

- (a) Sejam $A_1, \dots, A_n \in \mathcal{F}$. Defina $A_j = \emptyset \forall j > n$, portanto $A_n \in \mathcal{F} \forall n \geq 1$. Como os conjuntos vazios não contribuirão para a união, podemos mostrar que

$$\bigcup_{j=1}^n A_j = \bigcup_{j=1}^{\infty} A_j \in \mathcal{F},$$

pois \mathcal{F} é fechado sob uniões contáveis.

- (b) Sejam $A_1, \dots, A_n \in \mathcal{F}$. Queremos mostrar que $\bigcap_{j=1}^n A_j \in \mathcal{F}$. Pela lei de De Morgan (mostrando que cada elemento de um conjunto precisa pertencer ao outro conjunto também)

$$\bigcap_{j=1}^n A_j = \left(\bigcup_{j=1}^n A_j^c \right)^c.$$

Como o espaço de eventos \mathcal{F} é fechado sob complementos, $A_j^c \in \mathcal{F}$, para todo $j = 1, \dots, n$. Como é fechado sob uniões finitas (afirmação 1 da proposição, mostrada acima),

$$\bigcup_{j=1}^n A_j^c \in \mathcal{F}$$

Ao tomar o complemento mais uma vez, segue que $\bigcap_{j=1}^n A_j \in \mathcal{F}$.

- (c) A prova é semelhante à da afirmação 2 acima, observando que a lei de De Morgan também vale para uniões e interseções contáveis. Ou seja, podemos escrever

$$\bigcap_{j=1}^{\infty} A_j = \left(\bigcup_{j=1}^{\infty} A_j^c \right)^c.$$

□

3.2 Probabilidade

O último elemento da tripla do espaço de probabilidade é a medida de probabilidade \mathbb{P} . Na definição 1.1 do espaço de probabilidade uniforme, definimos a medida de probabilidade como uma função do espaço de eventos para $[0, 1]$, de modo que a probabilidade do evento Ω ('o resultado está no espaço amostral') é 1 e para dois eventos disjuntos A, B , $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ – a última propriedade pode ser generalizada por indução para a aditividade finita: se A_1, \dots, A_n são eventos disjuntos, então

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \mathbb{P}(A_k).$$

Isso é suficiente para espaços de probabilidade infinitos? Vamos considerar o seguinte

Exemplo 3.7. Seja $\Omega = \mathbb{N}^* = \{1, 2, \dots\}$ o conjunto dos números naturais positivos e $\mathcal{F} = \mathcal{P}(\Omega)$. Suponha que $\mathbb{P}(\{n\}) = \frac{1}{2^n}$, para todo $n \geq 1$. O que esperaríamos que o evento $\{2n | n \geq 1\}$ ('o resultado é um número par') seja?

Intuitivamente, somaríamos as probabilidades correspondentes ao resultado ser par, ou seja,

$$\mathbb{P}(\{2n | n \geq 1\}) = \sum_{n=1}^{\infty} \mathbb{P}(\{2n\}) = \sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{3}.$$

(Note que o evento $\{n\}$ corresponde a 'o resultado é n '). O cálculo acima não pode ser justificado, a menos que estendamos a propriedade da aditividade finita para valer também para uniões contáveis de eventos disjuntos. De fato, é isso que fazemos!

Definição 3.8 (Medida de Probabilidade). Dado um espaço amostral Ω e um espaço de eventos \mathcal{F} , uma função $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$ é chamada de *medida de probabilidade* se satisfaz

- (i) $\mathbb{P}(B) \in [0, 1]$ para todo $B \in \mathcal{F}$;
- (ii) $\mathbb{P}(\Omega) = 1$;
- (iii) (Aditividade Contável) Para todo $A_n \in \mathcal{F}$, $n \geq 1$ eventos disjuntos (ou seja, para todos $m, n \geq 1$ tais que $m \neq n$, $A_m \cap A_n = \emptyset$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Nós agora apresentamos a definição de um espaço de probabilidade abstrato.

Definição 3.9. Um *espaço de probabilidade* é definido como o triplo $(\Omega, \mathcal{F}, \mathbb{P})$, onde

- Ω (o *espaço amostral*) é o conjunto de todos os possíveis resultados do experimento (sempre assumimos que não é vazio);
- \mathcal{F} é um espaço de eventos de subconjuntos de Ω .
- \mathbb{P} é uma medida de probabilidade em \mathcal{F} .

Proposição 3.10. *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Então, \mathbb{P} possui as seguintes propriedades*

- (a) *Se $A, B \in \mathcal{F}$ tal que $A \subseteq B$, então*

$$\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A).$$

Observe que $B - A = B \cap A^c$ e deve ser interpretado como 'todos os elementos de B que não estão em A '.

- (b) *Para todo $A \in \mathcal{F}$,*

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

- (c) $\mathbb{P}(\emptyset) = 0$.

Demonstração.

- (a) Escrevemos B como a união do conjunto com todos os elementos em B que não estão em A e aqueles que estão, ou seja, $B = (B - A) \cup A$. Pela aditividade finita,

$$\mathbb{P}(B) = \mathbb{P}((B - A) \cup A) = \mathbb{P}(B - A) + \mathbb{P}(A),$$

o que prova a afirmação.

- (b) Usando a propriedade acima,

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

- (c) $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^c) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.

□

Podemos usar a aditividade contável para calcular a probabilidade de uma união de eventos disjuntos. Como podemos calcular a probabilidade de qualquer união de eventos? A seguinte proposição nos fornece uma maneira de fazer isso.

Proposição 3.11 (Fórmula de Inclusão-Exclusão). *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Então, para qualquer coleção finita A_1, \dots, A_n de eventos em \mathcal{F} ,*

temos

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

Observação 2. A fórmula 3.11 acima usa uma notação concisa e não é imediatamente fácil de interpretar. Para entendê-la melhor, vamos considerar alguns casos específicos.

- $n = 2$

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k=1}^2 A_k\right) &= \sum_{k=1}^2 (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq 2} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{1 \leq i \leq 2} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq 2} \mathbb{P}(A_{i_1} \cap A_{i_2}) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2). \end{aligned}$$

- $n = 3$

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k=1}^3 A_k\right) &= \sum_{k=1}^3 (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq 3} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{1 \leq i \leq 3} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq 3} \mathbb{P}(A_{i_1} \cap A_{i_2}) \\ &\quad + \sum_{1 \leq i_1 < i_2 < i_3 \leq 3} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) \\ &\quad - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Portanto, a soma $\sum_{1 \leq i_1 < \dots < i_k \leq 2}$ deve ser interpretada como a soma de todos os k -uplas (i_1, \dots, i_k) de números $\{1, \dots, n\}$ sem repetição (as desigualdades são estritas). Como vimos na seção 2, existem $\binom{n}{k}$ dessas k -uplas, portanto a soma terá $\binom{n}{k}$ parcelas.

Demonstração. Vamos provar o resultado apenas para $n = 2$ (a prova do passo de indução no caso geral é semelhante, mas mais confusa!). Escrevemos

$$A_1 \cup A_2 = (A_1 - B) \cup B \cup (A_2 - B),$$

onde $B = A_1 \cap A_2$. Os conjuntos $A_1 - B$, B e $A_2 - B$ são todos disjuntos, então podemos escrever

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}((A_1 - B) \cup B \cup (A_2 - B)) = \mathbb{P}(A_1 - B) + \mathbb{P}(B) + \mathbb{P}(A_2 - B)$$

usando a aditividade finita. Sabemos, pela proposição 3.10, que $\mathbb{P}(A_1 - B) = \mathbb{P}(A_1) - \mathbb{P}(B)$ e, da mesma forma, $\mathbb{P}(A_2 - B) = \mathbb{P}(A_2) - \mathbb{P}(B)$. Substituindo esses valores na fórmula acima, obtemos

$$\mathbb{P}(A_1 \cup A_2) = (\mathbb{P}(A_1) - \mathbb{P}(B)) + \mathbb{P}(B) + (\mathbb{P}(A_2) - \mathbb{P}(B)) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(B),$$

o que comprova a alegação. \square

Proposição 3.12. *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Se $A, B \in \mathcal{F}$ e $A \subseteq B$, então*

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

Demonstração. Uma vez que $A \subseteq B$, segue que $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A)$ ou, equivalentemente, $\mathbb{P}(A) = \mathbb{P}(B) - \mathbb{P}(B - A) \leq \mathbb{P}(B)$, onde a desigualdade decorre do fato de que as probabilidades são sempre não negativas. \square

Proposição 3.13 (Desigualdade de Boole). *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Se $A_1, \dots, A_n \in \mathcal{F}$, então*

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i) \quad (*)$$

Demonstração. Procedemos por indução. Para $n = 2$, observe que

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \underbrace{\mathbb{P}(A_1 \cap A_2)}_{\geq 0} \leq \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Portanto, (*) vale para $n = 2$. Suponha agora que (*) vale $\forall j \leq n$, então precisamos provar que vale para $n + 1$ eventos. Seja $A_1, \dots, A_{n+1} \in \mathcal{F}$, então, argumentando como acima, temos

$$\mathbb{P}\left(\bigcup_{j=1}^{n+1} A_j\right) = \mathbb{P}\left(\left(\bigcup_{j=1}^n A_j\right) \cup A_{n+1}\right)$$

$$\begin{aligned}
&= \mathbb{P} \left(\bigcup_{j=1}^n A_j \right) + \mathbb{P}(A_{n+1}) - \mathbb{P} \left(\left(\bigcup_{j=1}^n A_j \right) \cap A_{n+1} \right) \\
&\leq \mathbb{P} \left(\bigcup_{j=1}^n A_j \right) + \mathbb{P}(A_{n+1}) \\
&\leq \sum_{j=1}^n \mathbb{P}(A_j) + \mathbb{P}(A_{n+1}) = \sum_{j=1}^{n+1} \mathbb{P}(A_j). \quad \square
\end{aligned}$$

Revisão de espaços de probabilidade

Ω – *espaço amostral*: elementos $\omega \in \Omega$ são resultados, Ω é um conjunto não vazio.

\mathcal{F} – *espaço de eventos*: elementos $A \in \mathcal{F}$ são eventos ($A \subseteq \Omega$)

Deve satisfazer três condições:

- $\mathcal{F} \neq \emptyset$
- $A^c \in \mathcal{F}$ para todo $A \in \mathcal{F}$
- $(\cup_{n=1}^{\infty} A_n) \in \mathcal{F}$ para toda sequência de eventos A_1, A_2, A_3, \dots

Consequências dessas condições:

- $\Omega \in \mathcal{F}$ (de fato, tome $A \in \mathcal{F}$, $A \cup A^c = \Omega$)
- $\emptyset \in \mathcal{F}$ (de fato, $\Omega^c = \emptyset$)
- $(\cap_{n=1}^{\infty} A_n) \in \mathcal{F}$ para cada sequência de eventos A_1, A_2, A_3, \dots
(de fato, $\cap_{n=1}^{\infty} A_n = (\cup_{n=1}^{\infty} A_n^c)^c$)

\mathbb{P} – *medida de probabilidade*: $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$.

Deve satisfazer três condições:

- $\mathbb{P}(A) \geq 0$ para todo $A \in \mathcal{F}$
- $\mathbb{P}(\Omega) = 1$
- \mathbb{P} é aditiva contável: $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$
para cada sequência de eventos *disjuntos* A_1, A_2, A_3, \dots

O triplo $(\Omega, \mathcal{F}, \mathbb{P})$ é chamado de *espaço de probabilidade*.

Exemplo 3.14 (Espaços de probabilidade uniformes). Ω é um conjunto finito, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.

Exemplo 3.15. Não existe um espaço de probabilidade para modelar o experimento "escolher um número inteiro aleatoriamente". Por mais que

gostaríamos de dizer que um número inteiro X escolhido aleatoriamente será par com probabilidade $\frac{1}{2}$ e o último dígito em sua representação decimal será 7 com probabilidade $\frac{1}{10}$, não existe um espaço de probabilidade que possa modelar isso. Mais precisamente, para $\Omega = \mathbb{Z}$, $\mathcal{F} = \mathbb{P}(\mathbb{Z})$, não existe uma medida de probabilidade $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$ tal que $\mathbb{P}(\{j\}) = \mathbb{P}(\{k\})$ para todos $j, k \in \mathbb{Z}$. De fato, se $\mathbb{P}(\{j\}) > 0$, então $\mathbb{P}(\mathbb{Z}) = \sum_{x \in \mathbb{Z}} \mathbb{P}(\{x\}) = +\infty$, e se $\mathbb{P}(\{j\}) = 0$, então $\mathbb{P}(\mathbb{Z}) = \sum_{x \in \mathbb{Z}} \mathbb{P}(\{x\}) = 0$, e em ambos os casos \mathbb{P} viola os requisitos para ser uma medida de probabilidade, ou seja, $\mathbb{P}(\mathbb{Z}) = 1$.

4 Probabilidade Condicional e Independência

4.1 Probabilidade Condicional

Exemplo 4.1. Suponha que antes de lançar um dado justo, você faz uma aposta de uma libra que o resultado será 3. Seu amigo vê o resultado antes de você e lhe diz que o dado mostrou um número par. Você continuaria com a aposta ou desistiria dela? E se você fosse informado de que o resultado é ímpar? Como essas informações parciais sobre o resultado mudam a probabilidade?

Modelamos o espaço de probabilidade correspondente ao lançamento de um dado justo tomando $\Omega = \{1, 2, \dots, 6\}$, \mathcal{F} como o conjunto de todos os subconjuntos e \mathbb{P} como a probabilidade uniforme sobre ele. Então, o evento que nosso amigo nos diz que ocorreu é

$$B = \{2, 4, 6\}, \quad \text{e sua probabilidade é} \quad \mathbb{P}(B) = \frac{3}{6} = \frac{1}{2} > 0$$

O evento favorável para nós é

$$A = \{3\} \quad \text{e sua probabilidade é} \quad \mathbb{P}(A) = \frac{1}{6}.$$

Saber que o resultado é par pode ser interpretado como mudar o espaço amostral de Ω para B . Intuitivamente, assumiríamos que a probabilidade no novo espaço amostral permanece uniforme, mas a probabilidade de cada resultado muda de $\frac{1}{6}$ para $\frac{1}{3}$, pois agora existem apenas 3 resultados possíveis. Dado que nosso resultado preferido, o 3, não está no novo espaço amostral, esperaríamos que a probabilidade de obter 3 seja 0 e, portanto, faria sentido desistir da aposta. Se, por outro lado, nos dissessem que o resultado é ímpar, poderíamos reformular o espaço de probabilidade como um com espaço amostral $B^c = \{1, 3, 5\}$ e esperaríamos que a probabilidade de vencer a aposta fosse $\frac{1}{3}$, pois é um dos 3 resultados possíveis.

E se apostássemos em $\{2, 3\}$? Então estaríamos olhando para a quantidade de maneiras pelas quais ainda podemos vencer, dividida pela quantidade de resultados possíveis. Portanto, de acordo com nossa intuição, esperaríamos que

a probabilidade atualizada, dada que o evento B ocorreu, fosse

$$\mathbf{P}_B(A) = \frac{|A \cap B|}{|B|} = \frac{\left(\frac{|A \cap B|}{|\Omega|}\right)}{\left(\frac{|B|}{|\Omega|}\right)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Mas isso é uma probabilidade bem definida?

Proposição 4.2. *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $B \in \mathcal{F}$ tal que $\mathbb{P}(B) > 0$. Seja $\mathbf{P}_B : \mathcal{F} \rightarrow \mathbb{R}$ tal que*

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

Então, \mathbf{P}_B é uma medida de probabilidade.

Demonstração. Precisamos verificar se todas as propriedades das medidas de probabilidade são satisfeitas.

(i) Primeiro, precisamos mostrar que \mathbf{P}_B está definido para todo $A \in \mathcal{F}$ e assume valores em $[0, 1]$, ou seja, \mathbf{P}_B é um mapeamento de \mathcal{F} para $[0, 1]$, como deveria ser.

- Seja $A \in \mathcal{F}$. Assumimos que $B \in \mathcal{F}$ e, portanto, $A \cap B \in \mathcal{F}$, pois o espaço de eventos é fechado sob interseções. Portanto, $\mathbb{P}(A \cap B)$ está bem definido e, como $\mathbb{P}(B) > 0$, sua razão $\mathbf{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ está bem definida.
- $A \cap B \subseteq B$ e, portanto, $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ (proposição 3.12). Segue que $\mathbf{P}_B(A) \leq 1$. Da mesma forma, como $\mathbb{P}(A \cap B) \geq 0$ e $\mathbb{P}(B) > 0$, segue que $\mathbf{P}_B(A) \geq 0$. Portanto, $\mathbf{P}_B(A) \in [0, 1]$.

(ii)

$$\mathbf{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1,$$

conforme requerido.

(iii) (aditividade contável) Seja $A_n \in \mathcal{F}$ para todo $n \geq 1$, de modo que $A_n \cap A_m = \emptyset$ para todos $n \neq m$ (eventos disjuntos). Então

$$\begin{aligned} \mathbf{P}_B\left(\bigcup_{n=1}^{\infty} A_n\right) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right) \\ &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right) \quad (**) \end{aligned}$$

Agora, como $A_n \in \mathcal{F}$ e $B \in \mathcal{F}$ para todos $n \geq 1$, segue que $A_n \cap B \in \mathcal{F}$ para todos $n \geq 1$. Além disso, os eventos $A_n \cap B$ são disjuntos. De fato, para $n \neq m$

$$(A_n \cap B) \cap (A_m \cap B) \subseteq A_n \cap A_m = \emptyset.$$

Uma vez que \mathbb{P} é uma medida de probabilidade, ela é aditiva contável, o que implica que

$$(**) = \frac{1}{\mathbb{P}(B)} \sum_{n=1}^{\infty} \mathbb{P}(A_n \cap B) = \sum_{n=1}^{\infty} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n|B) = \sum_{n=1}^{\infty} \mathbf{P}_B(A_n).$$

□

Definição 4.3. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $B \in \mathcal{F}$ tal que $\mathbb{P}(B) > 0$. Para $A \in \mathcal{F}$, a *probabilidade condicional de A dado B* é denotada por $\mathbb{P}(A|B)$ e é definida como

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

Exercício 4.1. Um experimento consiste em lançar uma moeda justa 7 vezes.

- (a) Descreva o espaço de probabilidade associado a ele.
- (b) Seja E o evento correspondente a obter um número primo de caras. Qual é $\mathbb{P}(E)$?
- (c) Seja B o evento "Cara ocorre pelo menos 6 vezes". Qual é $\mathbb{P}(E|B)$?

Solução.

- (a) $\Omega = \{(a_1, \dots, a_7) : a_i \in \{H, T\}\} = S_{2,7}(\{H, T\})$, \mathcal{F} é o conjunto das partes de Ω e \mathbb{P} é a probabilidade uniforme, ou seja, \mathbb{P} é tal que

$$\forall A \in \mathcal{F} \quad \mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Lembrando que $|\Omega| = |S_{2,7}(\{H, T\})| = 2^7$.

- (b) Para $i = 1, \dots, 7$, seja A_i o evento "obtemos exatamente i caras". Os elementos de A_i podem ser caracterizados de forma única pela posição de H na sequência. Portanto, pelo princípio fundamental da contagem, $|A_i| = \binom{7}{i}$. Assim,

$$\mathbb{P}(A_i) = \frac{1}{2^7} \binom{7}{i}$$

Agora, observe que $A_i \cap A_j = \emptyset$ para $i \neq j$ (nenhum resultado tem tanto i quanto j caras) e

$$E = A_2 \cup A_3 \cup A_5 \cup A_7.$$

Então, pela aditividade finita

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(A_2) + \mathbb{P}(A_3) + \mathbb{P}(A_5) + \mathbb{P}(A_7) \\ &= \binom{7}{2} \frac{1}{2^7} + \binom{7}{3} \frac{1}{2^7} + \binom{7}{5} \frac{1}{2^7} + \binom{7}{7} \frac{1}{2^7} = \frac{78}{128}.\end{aligned}$$

- (c) B é o evento "H aparece pelo menos 6 vezes", então $B = A_6 \cup A_7$. Observe que,

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(A_6) + \mathbb{P}(A_7) = \binom{7}{6} \frac{1}{2^7} + \binom{7}{7} \frac{1}{2^7} \\ &= \frac{7!}{6!1!} \cdot \frac{1}{2^7} + \frac{7!}{6!0!} \cdot \frac{1}{2^7} = \frac{7+1}{2^7} = \frac{7+1}{2^7} = \frac{8}{2^7} = \frac{1}{2^4} > 0.\end{aligned}$$

Agora, podemos calcular $\mathbb{P}(E|B)$. Pela definição,

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

Como $E \cap B = (A_2 \cup A_3 \cup A_5 \cup A_7) \cap (A_6 \cup A_7) = A_7$, temos

$$\mathbb{P}(E|B) = \frac{\mathbb{P}(A_7)}{\mathbb{P}(B)} = \frac{1/2^7}{1/2^4} = \frac{2^4}{2^7} = \frac{1}{8}.$$

Exemplo 4.4. Um estudante compra 2 maçãs, 3 bananas e 5 cocos. Todos os dias o estudante escolhe uma fruta uniformemente ao acaso e a come.

O espaço amostral é o conjunto de todas as tríades que podem ser construídas com as frutas disponíveis, em que cada resultado corresponde à fruta comida em cada dia. Como no final dos três dias temos todas as informações, o espaço de eventos é o conjunto das partes do espaço amostral. Definimos os eventos $A_i = \{\text{o estudante come uma maçã no dia } i\}$, $B_i = \{\text{o estudante come uma banana no dia } i\}$ e $C_i = \{\text{o estudante come um coco no dia } i\}$.

- (a) Qual é a probabilidade de o estudante comer um coco no dia 1 e uma banana no dia 2? O evento 'o estudante come um coco no dia 1 e uma banana no dia 2' corresponde ao evento $C_1 \cap B_2$. Observe que a maneira como a informação sobre a probabilidade é codificada é por meio de probabilidades condicionais:

a afirmação '**todos os dias** o estudante escolhe uma fruta **uniformemente ao acaso** e a come' pode ser interpretada como a probabilidade condicional de escolher qualquer uma das frutas restantes uniformemente ao acaso, então sabemos que

$$\mathbb{P}(B_2|C_1) = \frac{3}{9}.$$

Segue da definição de probabilidade condicional que

$$\mathbb{P}(C_1 \cap B_2) = \mathbb{P}(B_2|C_1)\mathbb{P}(C_1) = \frac{3}{9} \frac{5}{10} = \frac{1}{6}.$$

Escrever a probabilidade de interseção de dois eventos como um produto de uma probabilidade condicional e uma probabilidade é chamado de 'regra da multiplicação' e pode ser estendido para interseções de mais de dois eventos. Por exemplo, consideremos a seguinte pergunta.

- (b) Qual é a probabilidade de no terceiro dia o estudante comer a última maçã? Como existem exatamente duas maçãs, isso significa que o estudante comerá a primeira maçã no dia 1 ou no dia 2. Portanto, se A é o evento 'estudante come a última maçã no terceiro dia', podemos escrever

$$A = (A_1 \cap A_2^c \cap A_3) \cup (A_1^c \cap A_2 \cap A_3).$$

Observe que os eventos $A_1 \cap A_2^c \cap A_3$ e $A_1^c \cap A_2 \cap A_3$ são disjuntos, portanto

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A_1 \cap A_2^c \cap A_3) + \mathbb{P}(A_1^c \cap A_2 \cap A_3) \\ &= \mathbb{P}(A_1)\mathbb{P}(A_2^c|A_1)\mathbb{P}(A_3|A_1 \cap A_2^c) + \mathbb{P}(A_1^c)\mathbb{P}(A_2|A_1^c)\mathbb{P}(A_3|A_1^c \cap A_2) \\ &= \frac{2}{10} \cdot \frac{8}{9} \cdot \frac{1}{8} + \frac{8}{10} \cdot \frac{2}{9} \cdot \frac{1}{8} = \frac{1}{45} + \frac{1}{45} = \frac{2}{45}, \end{aligned}$$

usando a regra da multiplicação duas vezes.

Proposição 4.5 (Regra da Multiplicação). *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $A_1, \dots, A_n \in \mathcal{F}$ de modo que $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$. Então,*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Demonstração. Observe que para $k = 1, \dots, n-1$, $A_1 \cap \dots \cap A_k \supseteq A_1 \cap \dots \cap A_{n-1}$.

Portanto, pela Proposição 3.12 e pela hipótese

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_k) \geq \mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0.$$

que garante que todas as probabilidades condicionais no lado direito sejam bem definidas. O resultado segue de uma aplicação direta da definição de probabilidade condicional no lado direito:

$$\begin{aligned} & \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \frac{\mathbb{P}(A_3 \cap A_1 \cap A_2)}{\mathbb{P}(A_1 \cap A_2)} \cdots \frac{\mathbb{P}(A_1 \cap \cdots \cap A_n)}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})} \\ &= \mathbb{P}(A_1 \cap \cdots \cap A_n). \end{aligned} \quad \square$$

4.2 Lei da Probabilidade Total

Exemplo 4.6 (4.4 continuado). Suponha que agora nos seja perguntado para calcular a probabilidade de o estudante comer um coco no dia 2. Para calcular a probabilidade, precisamos condicionar o que aconteceu no dia 1, percorrendo todas as opções possíveis. Neste caso, existem duas opções que afetam o cálculo da probabilidade condicional: se o estudante também comeu um coco no dia 1 (evento C_1) ou não (evento C_1^c). Portanto,

$$\mathbb{P}(C_2) = \mathbb{P}(C_2|C_1) \cdot \mathbb{P}(C_1) + \mathbb{P}(C_2|C_1^c) \cdot \mathbb{P}(C_1^c) = \frac{4}{9} \cdot \frac{5}{10} + \frac{5}{9} \cdot \frac{5}{10} = \frac{1}{2}.$$

De onde vem esta fórmula? Nós escrevemos

$$C_2 = (C_2 \cap C_1) \cup (C_2 \cap C_1^c).$$

Assim, a partir da aditividade finita, segue que

$$\mathbb{P}(C_2) = \mathbb{P}(C_2 \cap C_1) + \mathbb{P}(C_2 \cap C_1^c).$$

Ao aplicar a regra da multiplicação às probabilidades condicionais acima, obtemos a fórmula que é um exemplo específico da *lei da probabilidade total*.

A lei da probabilidade total nos permite calcular a probabilidade de um evento, condicionando em todas as instâncias possíveis de um 'evento diferente', ou, de

forma mais formal, em todos os conjuntos de uma partição do espaço amostral.

Definição 4.7. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Seja $B_n \in \mathcal{F}$ para todos $n = 1, \dots, N$ (onde N é finito ou infinito). Então, a coleção de todos os B_n , $\{B_n : n = 1, \dots, N\}$, é chamada de *partição* de Ω se

- $B_n \neq \emptyset \quad \forall n = 1, \dots, N.$
- $B_n \cap B_m = \emptyset \quad \forall n \neq m$
- $\bigcup_{n=1}^N B_n = \Omega.$

Portanto, uma partição é uma coleção de eventos não vazios e disjuntos que abrange todo o espaço.

Proposição 4.8 (Lei da Probabilidade Total). *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\{B_n : n = 1, \dots, N\}$ com N finito ou infinito, seja uma partição de Ω tal que $\mathbb{P}(B_n) > 0, \quad \forall n = 1, \dots, N.$ Então, para todo $A \in \mathcal{F}$*

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

Demonstração. Observe que, uma vez que $\{B_n : n = 1, \dots, N\}$ forma uma partição de Ω , temos

$$A = A \cap \Omega = A \cap \bigcup_{n=1}^N B_n = \bigcup_{n=1}^N (A \cap B_n).$$

Além disso, uma vez que os B_n 's são disjuntos, os conjuntos $\{A \cap B_n : n = 1 \dots N\}$ também são disjuntos, portanto, pela aditividade finita/contável, temos

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n=1}^N A \cap B_n\right) = \sum_{n=1}^N \mathbb{P}(A \cap B_n) = \sum_{n=1}^N \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

Na última igualdade, usamos a definição de probabilidade condicional com a suposição de que $\mathbb{P}(B_n) > 0 \quad \forall n = 1, \dots, N.$ □

Exemplo 4.9. Um estudante enfrenta uma pergunta de múltipla escolha, com 4 opções. O estudante ou sabe a resposta ou escolhe uma das respostas de forma uniforme e aleatória. A probabilidade de o estudante saber a resposta é $\frac{2}{3}$.

(a) O estudante deseja calcular a probabilidade de responder corretamente.

Vamos começar definindo os eventos de interesse:

$$A = \{\text{o estudante responde corretamente}\}$$

$$B = \{\text{o estudante sabe a resposta}\}$$

As informações que temos sobre a probabilidade são que 'o estudante ou sabe a resposta (e, portanto, responde corretamente)' ou 'escolhe uma das respostas de forma uniforme e aleatória'. Isso pode ser expresso como $\mathbb{P}(A|B) = 1$ e $\mathbb{P}(A|B^c) = \frac{1}{4}$. Também nos foi dito que a probabilidade de o estudante saber a resposta é $\frac{2}{3}$. Portanto, $\mathbb{P}(B) = \frac{2}{3}$. Dadas essas informações, nos é pedido para encontrar $\mathbb{P}(A)$.

Uma vez que B e B^c formam uma partição do espaço amostral, aplicando a lei da probabilidade total obtemos

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) = 1 \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{3}{4}.$$

- (b) O professor gostaria de saber a probabilidade de o estudante saber a resposta se ele respondeu corretamente, ou seja, $\mathbb{P}(B|A)$. Como podemos usar as informações que temos para calcular isso? Escrevemos

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{2/3}{3/4} = \frac{8}{9}.$$

Isso é um exemplo específico do que é conhecido como a fórmula de Bayes.

4.3 Teorema de Bayes

Teorema 4.10 (Teorema de Bayes). *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\{B_n : n = 1, \dots, N\}$, com N finito ou infinito, seja uma partição de Ω tal que $\mathbb{P}(B_n) > 0 \quad \forall n = 1, \dots, N$. Então, para $A \in \mathcal{F}$ tal que $\mathbb{P}(A) > 0$*

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(A|B_n)\mathbb{P}(B_n)}{\sum_{j=1}^N \mathbb{P}(A|B_j)\mathbb{P}(B_j)} \quad \forall n = 1, \dots, N.$$

Demonstração. Pela definição de probabilidade condicional e como A é tal que $\mathbb{P}(A) > 0$, então pela definição de probabilidade condicional e pela lei da

probabilidade total:

$$\mathbb{P}(B_n|A) = \frac{\mathbb{P}(B_n \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_n)\mathbb{P}(B_n)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_n)\mathbb{P}(B_n)}{\sum_{j=1}^N \mathbb{P}(A|B_j)\mathbb{P}(B_j)}. \quad \square$$

Exemplo 4.11 (Falsos Positivos). Uma doença tem uma incidência de 1 em 100 na população. O teste diagnóstico disponível é tal que

- se você tem a doença, o teste é positivo com probabilidade $\frac{72}{100}$
- se você não tem a doença, o teste é positivo com probabilidade $\frac{5}{1000}$.

Uma pessoa recebe um resultado positivo. Qual é a probabilidade de ela realmente ter a doença?

Os dois eventos de interesse são $D = \{\text{a pessoa tem a doença}\}$ e $P = \{\text{a pessoa tem um teste positivo}\}$. Estamos interessados em $\mathbb{P}(D|P)$. As informações que temos são $\mathbb{P}(D) = \frac{1}{100}$, $\mathbb{P}(P|D) = \frac{72}{100}$ e $\mathbb{P}(P|D^c) = \frac{5}{1000}$. Pelo Teorema de Bayes

$$\mathbb{P}(D|P) = \frac{\mathbb{P}(P|D)\mathbb{P}(D)}{\mathbb{P}(P|D)\mathbb{P}(D) + \mathbb{P}(P|D^c)\mathbb{P}(D^c)} \approx 0.59.$$

O Teorema de Bayes nos permite calcular a probabilidade condicional de um evento, dado outro, em termos das probabilidades condicionais inversas. É particularmente útil em Estatística, levando a uma área inteira chamada Estatística Bayesiana: enquanto na probabilidade, estamos interessados em calcular probabilidades dadas um 'modelo' (ou seja, informações suficientes que determinam as probabilidades), na estatística, estamos interessados em escolher um modelo, dadas as observações que fazemos. O Teorema de Bayes nos permite conectar os dois.

4.4 Independência

Definição 4.12. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Dizemos que os eventos A e B são *independentes* se $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Uma maneira de pensar na independência é que o conhecimento sobre a ocorrência de um dos eventos não aumentará nem diminuirá a chance de o outro ocorrer. De fato, assumindo que $\mathbb{P}(B) > 0$, você pode verificar que A e B são independentes

se e somente se $\mathbb{P}(A|B) = \mathbb{P}(A)$ (exercício!). Em particular, se A e B são independentes, então A^c e B também são independentes.

Observação 3. As noções de eventos "independentes" e "disjuntos" são muito diferentes. Na verdade, essas noções são normalmente incompatíveis: dois eventos disjuntos são independentes se e somente se a probabilidade de um deles for 0 (exercício!).

Definição 4.13. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e A_1, A_2, \dots, A_n sejam eventos. Dizemos que os eventos A_1, \dots, A_n são *mutuamente independentes aos pares* se A_j e A_k forem independentes para todas as escolhas de j e k distintos. Dizemos que os eventos A_1, \dots, A_n são *mutuamente independentes*, se

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \mathbb{P}(A_{j_1})\mathbb{P}(A_{j_2}) \dots \mathbb{P}(A_{j_k})$$

para todo $k = 2, \dots, n$ e para todas as escolhas de $1 \leq j_1 < j_2 < \dots < j_k \leq n$.

No caso em que $n = 2$, independência aos pares é obviamente o mesmo que independência mútua. No caso em que $n = 3$, independência aos pares significa

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \\ \mathbb{P}(A_1 \cap A_3) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_3) \\ \mathbb{P}(A_2 \cap A_3) &= \mathbb{P}(A_2) \cdot \mathbb{P}(A_3).\end{aligned}$$

enquanto independência mútua significa

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \\ \mathbb{P}(A_1 \cap A_3) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_3) \\ \mathbb{P}(A_2 \cap A_3) &= \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \\ \mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3).\end{aligned}$$

Isso ilustra que a independência mútua é mais forte do que a independência aos pares. É difícil listar as condições para valores maiores de n . Por exemplo, se $n = 5$, a independência aos pares envolve $\binom{5}{2} = 10$ condições a serem verificadas, e a independência mútua envolve $2^5 - 5 - 1 = 26$ condições a serem verificadas.

Exemplo 4.14. Dois dados são lançados. Sejam

$$\begin{aligned}A_1 &= \{\text{o primeiro dado é par}\} \\A_2 &= \{\text{o segundo dado é ímpar}\} \\A_3 &= \{\text{soma dos dados é 7}\}.\end{aligned}$$

Esses eventos são independentes aos pares, pois

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= \frac{1}{4} = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \\ \mathbb{P}(A_1 \cap A_3) &= \frac{1}{12} = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3) \\ \mathbb{P}(A_2 \cap A_3) &= \frac{1}{12} = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3).\end{aligned}$$

Isso significa que, para cada par de eventos dessa família, o conhecimento sobre a ocorrência de um deles não afetará as chances de que os outros dois ocorram. Em particular, nem A_1 nem A_2 isoladamente afetarão as chances de A_3 . No entanto, saber que A_1 e A_2 ocorrem aumentará de fato a chance de que A_3 ocorra, como

$$\mathbb{P}(A_3|A_1 \cap A_2) = \frac{1}{3} \neq \frac{1}{6} = \mathbb{P}(A_3).$$

De maneira mais formal,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{12} \neq \frac{1}{24} = \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3).$$

Exemplo 4.15. Lance três moedas justas. Considere os eventos:

$$\begin{aligned}A_1 &= \text{Primeira moeda dá Cara} \\A_2 &= \text{Segunda moeda dá o mesmo que a primeira moeda} \\A_3 &= \text{Segunda moeda dá o mesmo que a terceira moeda} \\A_4 &= \text{Terceira dá Coroa}\end{aligned}$$

Então:

Esses eventos são independentes aos pares.

A_1 , A_2 e A_3 são mutuamente independentes.

A_1 , A_2 , A_3 e A_4 não são mutuamente independentes.

5 Variáveis Aleatórias

5.1 Definição

Com frequência, estamos interessados em uma quantidade que é determinada como resultado de um experimento dado.

Por exemplo, considere um jogo de azar em que dois dados são lançados e você recebe uma recompensa em dinheiro dada pelo valor máximo obtido entre os dois dados. Como você modela essa situação? Como de costume, cada resultado é um par $\omega = (\omega_1, \omega_2)$ onde ambos ω_1 e ω_2 estão em $\{1, 2, 3, 4, 5, 6\}$. Isso é, $\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. A recompensa é determinada pelos valores de ω_1 e ω_2 por meio da tabela a seguir:

	1	2	3	4	5	6
1	£1	£2	£3	£4	£5	£6
2	£2	£2	£3	£4	£5	£6
3	£3	£3	£3	£4	£5	£6
4	£4	£4	£4	£4	£5	£6
5	£5	£5	£5	£5	£5	£6
6	£6	£6	£6	£6	£6	£6

A palavra-chave aqui é 'determinada': mesmo que o resultado seja aleatório, ele é aleatório apenas porque o resultado ω é aleatório.

Matematicamente, isso significa que o prêmio X pode ser escrito como uma *função* do resultado ω . Em geral, uma variável aleatória é uma função

$$X : \Omega \rightarrow \mathbb{R}$$

do espaço amostral para o conjunto dos números reais. Mais formalmente, exigimos que as condições especificadas em termos de X sejam eventos aleatórios, ou seja, eventos que o observador pode determinar se ocorrem ou não.

Definição 5.1 (Variável Aleatória). Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Uma *variável aleatória* é uma função $X : \Omega \rightarrow \mathbb{R}$ tal que $\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$ para todo $a \in \mathbb{R}$.

No exemplo acima, podemos escrever X explicitamente como a função que atribui

a cada par (ω_1, ω_2) o valor máximo entre eles, ou seja, $X((x, y)) = \max\{x, y\}$.

Usando essa ferramenta, podemos fazer afirmações como

$$\mathbb{P}(X = 1) = \frac{1}{36}, \quad \mathbb{P}(X = 5) = \frac{1}{4}, \quad \dots$$

Notação. Por conveniência, usaremos a notação simplificada

$$\{X = 5\} = \{\omega \in \Omega : X(\omega) = 5\}, \quad \text{e } \mathbb{P}(X = 5) = \mathbb{P}(\{X = 5\}), \quad \text{etc.}$$

É muito útil considerar a medida de probabilidade no conjunto dos números reais induzida por uma variável aleatória. Se estamos interessados apenas no valor de X , podemos deixar $(\Omega, \mathcal{F}, \mathbb{P})$ de lado e tomar o espaço amostral como \mathbb{R} .

Definição 5.2 (Distribuição). A *distribuição* de uma variável aleatória X é a medida de probabilidade em \mathbb{R} denotada por \mathbb{P}_X e dada por

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

para subconjuntos B em algum espaço de eventos no conjunto dos números reais.

No exemplo anterior, $\mathbb{P}_X(B)$ é determinada por seus valores

$$\mathbb{P}_X(\{k\}) = \frac{2k-1}{36}, \quad k = 1, 2, 3, 4, 5, 6, \quad \mathbb{P}_X(\mathbb{R} \setminus \{1, 2, 3, 4, 5, 6\}) = 0.$$

Notação. O espaço de eventos em \mathbb{R} , denotado por $\mathcal{B}(\mathbb{R})$, é um espaço de eventos que contém conjuntos como $\{x\}$ e $(a, b]$. Não entraremos em detalhes sobre a descrição de $\mathcal{B}(\mathbb{R})$. Você pode pensar nisso como a coleção de todos os conjuntos que podem ser obtidos aplicando um número contável de operações de conjuntos (união, interseção, complemento) a intervalos.

Observação 4. O espaço Ω pode ser mais complicado do que \mathbb{R} . Em geral, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ pode ser mais simples do que $(\Omega, \mathcal{F}, \mathbb{P})$. Observe que \mathbb{P}_X foi construído a partir de \mathbb{P} e X . No entanto, não é possível reconstruir $(\Omega, \mathcal{F}, \mathbb{P})$ nem X a partir de \mathbb{P}_X .

Proposição 5.3. A função \mathbb{P}_X é uma medida de probabilidade em \mathbb{R} .

Demonstração. Lembrando da Definição 3.8. Verificamos as três condições:

- (i) $\mathbb{P}_X(B) = \mathbb{P}(X \in B) \in [0, 1]$.
- (ii) $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \mathbb{R}\}) = \mathbb{P}(\Omega) = 1$.
- (iii) Se $B_1, B_2, B_3, \dots \in \mathcal{B}(\mathbb{R})$ são disjuntos, então

$$\begin{aligned}
\mathbb{P}_X \left(\bigcup_{n=1}^{\infty} B_n \right) &= \mathbb{P} \left(\left\{ \omega : X(\omega) \in \bigcup_{n=1}^{\infty} B_n \right\} \right) \\
&= \mathbb{P} \left(\bigcup_{n=1}^{\infty} \{ \omega : X(\omega) \in B_n \} \right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(\{ \omega : X(\omega) \in B_n \}) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_X(B_n).
\end{aligned}$$

Nas desigualdades acima, usamos: definição de \mathbb{P}_X ; que a pré-imagem da união é a união da pré-imagem; que a pré-imagem de conjuntos disjuntos é disjunta, combinada com a aditividade contável de \mathbb{P} ; definição de \mathbb{P}_X .

Isso prova a proposição. \square

5.2 Variáveis Aleatórias Discretas

Definição 5.4 (Variável Aleatória Discreta). Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória. Dizemos que X e \mathbb{P}_X são *discretos* se existir um conjunto finito ou infinito contável $S \subseteq \mathbb{R}$ tal que $\mathbb{P}(X \in \mathbb{R} \setminus S) = 0$.

Definição 5.5 (Função de Massa de Probabilidade). Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória discreta. Definimos a *função de massa de probabilidade* de X como a função $p_X : \mathbb{R} \rightarrow [0, 1]$ dada por

$$p_X(x) = \mathbb{P}_X(\{x\}).$$

Observe que p_X é construído a partir de \mathbb{P}_X , e p_X é mais simples do que \mathbb{P}_X , porque p_X recebe como entrada um número e \mathbb{P}_X recebe como entrada um conjunto de números. Veremos abaixo que é possível reconstruir \mathbb{P}_X a partir de p_X no caso em que \mathbb{P}_X é discreto.

Definição 5.6 (Suporte discreto). Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória discreta. Definimos o *suporte discreto* de X , ou, mais

precisamente, o suporte discreto de sua distribuição \mathbb{P}_X , como o conjunto

$$\{x \in \mathbb{R} : p_X(x) > 0\}.$$

Como afirmado acima, para estudar \mathbb{P}_X , é suficiente conhecer p_X .

Proposição 5.7. *Seja X uma variável aleatória discreta. Então*

$$\mathbb{P}_X(B) = \sum_{x \in B \cap D_X} p_X(x)$$

para todo $B \in \mathcal{B}(\mathbb{R})$, onde D_X denota o suporte discreto de X .

Notação. Antes de escrever a prova, precisamos explicar o significado de $\sum_{x \in B \cap D}$. Uma vez que o conjunto $B \cap D$ é contável, podemos escrever $B \cap D = \{x_1, x_2, x_3, \dots\}$, e podemos considerar $\sum_{x \in B \cap D} p_X(x)$ como $\sum_{k=1}^{\infty} p_X(x_k)$. Precisamos ter cuidado aqui, pois usamos uma "enumeração" arbitrária de $B \cap D$. No entanto, uma vez que os termos na soma são não negativos, outra enumeração significaria reordenar os termos, o que não afeta o valor da soma.

Demonstração. Pela definição de \mathbb{P}_X ser discreta, existe um conjunto contável $S \subseteq \mathbb{R}$ tal que $\mathbb{P}_X(S^c) = 0$. Podemos decompor

$$\mathbb{P}_X(B) = \mathbb{P}_X(B \cap S) + \mathbb{P}_X(B \cap S^c).$$

O segundo termo é zero, porque, uma vez que $B \cap S^c \subseteq S^c$,

$$0 \leq \mathbb{P}_X(B \cap S^c) \leq \mathbb{P}_X(S^c) = 0.$$

Portanto,

$$\mathbb{P}_X(B) = \mathbb{P}_X(B \cap S) = \mathbb{P}_X(\cup_{x \in B \cap S} \{x\}) = \sum_{x \in B \cap S} \mathbb{P}_X(\{x\}).$$

Se substituirmos D^c em vez de B na fórmula acima, obtemos $\mathbb{P}_X(D^c) = \sum_{x \in D^c \cap S} \mathbb{P}_X(\{x\}) = 0$, porque $\mathbb{P}_X(\{x\}) = p_X(x) = 0$ para todo $x \in D^c$.

Pelo mesmo argumento,

$$\mathbb{P}_X(B) = \mathbb{P}_X(B \cap D) + \mathbb{P}_X(B \cap D^c) = \mathbb{P}_X(B \cap D) = \sum_{x \in B \cap D} p_X(x),$$

o que é o que queríamos provar. \square

É conveniente especificar a distribuição de uma variável aleatória dizendo o que é p_X . Quando dizemos "seja X uma variável aleatória discreta com função de massa de probabilidade tal e tal," o que queremos dizer? Isso realmente descreve uma variável aleatória? A próxima definição e proposição respondem a essa pergunta.

Definição 5.8 (Função de Massa de Probabilidade). Uma função $f : \mathbb{R} \rightarrow [0, 1]$ é uma *função de massa de probabilidade* se o conjunto D dado por $D = \{x : f(x) > 0\}$ é contável e $\sum_{x \in D} f(x) = 1$.

Proposição 5.9. *Seja $f : \mathbb{R} \rightarrow [0, 1]$ uma função de massa de probabilidade. Então existe um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e uma variável aleatória discreta X tal que $p_X(x) = f(x)$ para todo $x \in \mathbb{R}$.*

Demonstração. Tome $D = \{x : f(x) > 0\}$. Tome $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ e

$$\mathbb{P}(B) = \sum_{x \in B \cap D} f(x).$$

Finalmente, tome $X(x) = x$. Então X é uma variável aleatória. Além disso,

$$\mathbb{P}_X(D^c) = \sum_{x \in D \cap D^c} f(x) = 0,$$

porque a soma sobre um conjunto vazio sempre é igual a zero. Portanto, X é uma variável aleatória discreta. Vamos verificar que $p_X = p$.

Para $x \in D$, temos

$$p_X(x) = \mathbb{P}_X(\{x\}) = \mathbb{P}(\{x\}) = \sum_{z \in \{x\} \cap D} f(z) = \sum_{z \in \{x\}} f(z) = f(x)$$

porque a soma de um único fator é igual a esse fator.

Por outro lado, para $x \in D^c$, temos $f(x) = 0$ e

$$p_X(x) = \sum_{z \in \{x\} \cap D} f(z) = \sum_{z \in \emptyset} f(z) = 0 = f(x).$$

Portanto, $p_X = p$, como reivindicado, e isso completa a prova da proposição. \square

5.3 As distribuições discretas mais comuns

Definição 5.10 (Distribuição de Bernoulli). Dizemos que uma variável aleatória discreta X tem uma *distribuição de Bernoulli* com parâmetro $p \in [0, 1]$, denotado $X \sim \text{Bernoulli}(p)$, se sua função de massa de probabilidade for

$$p_X(x) = \begin{cases} p, & x = 1, \\ 1 - p, & x = 0, \\ 0, & \text{caso contrário.} \end{cases}$$

Exemplo 5.11. Seja $\Omega = \{C, T\}$ (cara ou coroa para um lançamento de moeda) com $\mathbb{P}(C) = p$ e $\mathbb{P}(T) = 1 - p$, e seja $X(C) = 1$ e $X(T) = 0$. Então, $X \sim \text{Bernoulli}(p)$.

Definição 5.12 (Distribuição Geométrica). Dizemos que uma variável aleatória discreta X tem uma *distribuição geométrica* com parâmetro $p \in (0, 1]$, denotado $X \sim \text{Geom}(p)$, se sua função de massa de probabilidade for

$$p_X(x) = \begin{cases} p \cdot (1 - p)^{x-1}, & x \in \mathbb{N}, \\ 0, & \text{caso contrário.} \end{cases}$$

Para ver que p_X é de fato uma função de massa de probabilidade, observe que

$$\sum_{k=0}^{\infty} p_X(k) = p \cdot \sum_{k=1}^{\infty} (1 - p)^{k-1} = p \cdot \sum_{\ell=0}^{\infty} (1 - p)^{\ell} = p \cdot \frac{1}{1 - (1 - p)} = p \cdot \frac{1}{p} = 1.$$

Variáveis aleatórias com distribuição geométrica surgem na seguinte situação. Suponha que realizamos repetidamente ensaios, cada um dos quais pode ser um sucesso ou um fracasso. Assuma que os ensaios são independentes e a probabilidade de sucesso é a mesma em cada um deles, igual a p . Então, o número de ensaios realizados até obtermos o primeiro sucesso segue uma distribuição geométrica com parâmetro p .

Observação 5. Algumas referências menos comuns usam uma definição diferente para a distribuição geométrica com parâmetro p : eles consideram a distribuição em \mathbb{N}_0 (em vez de \mathbb{N}) e a função de massa de probabilidade $\tilde{p}_X(k) = p \cdot (1 - p)^k$, para $k \in \mathbb{N}_0$. Uma variável aleatória com função de massa de probabilidade \tilde{p}_X

conta o número de ensaios *fracassados* realizados antes de obter o primeiro sucesso. Portanto, no caso de um sucesso já ocorrer no primeiro ensaio, o número de ensaios fracassados é zero.

Exemplo 5.13. Lançamos um dado repetidamente até obtermos um 6 pela primeira vez. Seja X o número total de vezes que lançamos o dado. Então, $X \sim \text{Geom}(\frac{1}{6})$.

Definição 5.14 (Distribuição binomial). Dizemos que uma variável aleatória discreta X tem uma *distribuição binomial* com parâmetros $n \in \mathbb{N}_0$ e $p \in [0, 1]$, denotado $X \sim \text{Binom}(n, p)$, se sua função de massa de probabilidade for

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \{0, 1, \dots, n\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Observe que p_X é de fato uma função de massa de probabilidade, pois

$$1 = (p + (1-p))^n = \sum_{k=0}^n \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Exemplo 5.15. Lance um dado dez vezes e seja X o número de vezes que sai um 5 ou um 6. Então, $X \sim \text{Binom}(10, \frac{1}{3})$.

Lembrando que $0^0 = 1$ e $0! = 1$.

Definição 5.16 (Distribuição de Poisson). Dizemos que uma variável aleatória discreta X tem uma *distribuição de Poisson* com parâmetro $\lambda \geq 0$, denotado por $X \sim \text{Poisson}(\lambda, p)$, se sua função de massa de probabilidade for

$$p_X(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!}, & x \in \mathbb{N}_0, \\ 0, & \text{caso contrário.} \end{cases}$$

Para mostrar que p_X é de fato uma função de massa de probabilidade, calculamos

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Variáveis aleatórias que contam ocorrências raras entre muitos ensaios (como: número de acidentes em uma estrada ao longo de um ano, número de erros

de digitação em uma página de livro) normalmente seguem a distribuição de Poisson. Mais precisamente, uma variável aleatória de Poisson pode ser usada para aproximar uma Binomial(n, p) quando n é grande, p é pequeno e $np = \lambda$ é fixo. De fato, se $X \sim \text{Binom}(n, p)$ para $p = \frac{\lambda}{n}$, então

$$\begin{aligned}\mathbb{P}(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{para } n \text{ grande}\end{aligned}$$

onde a aproximação segue das seguintes aproximações para n grande:

$$n(n-1)\cdots(n-k+1)/n^k \rightarrow 1, \quad \left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}, \quad \frac{\lambda}{n} \rightarrow 0.$$

Breve revisão de variáveis aleatórias

Variável aleatória:

$$X : \Omega \rightarrow \mathbb{R}$$

com a exigência de que $\{\omega : X(\omega) \leq a\} \in \mathcal{F}$ para todo $a \in \mathbb{R}$.

A *distribuição* de X é $\mathbb{P}_X : \mathcal{B} \rightarrow \mathbb{R}$, dada por

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega : X(\omega) \in B\})$$

é uma medida de probabilidade em \mathbb{R} , onde \mathcal{B} é o espaço de eventos em \mathbb{R} .

Uma variável aleatória X é *discreta* se existe um conjunto contável S tal que $\mathbb{P}_X(S^c) = 0$.

Para X discreto, definimos a *função de massa de probabilidade de X* como a função $p_X : \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$p_X(x) = \mathbb{P}_X(\{x\}).$$

e o *suporte discreto de X* como o conjunto $\{x \in \mathbb{R} : p_X(x) > 0\}$.

Se X é discreto, então

$$\mathbb{P}_X(B) = \sum_{x \in B \cap D_X} p_X(x)$$

para todo $B \in \mathcal{B}$, onde D_X denota o suporte discreto de X .

Dizemos que uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ é uma *função de massa de probabilidade* se $f(x) \geq 0$ para todo $x \in \mathbb{R}$, o conjunto $\{x \in \mathbb{R} : f(x) > 0\}$ é contável e

$$\sum_{x: f(x) > 0} f(x) = 1.$$

Dada uma função de massa de probabilidade f , é possível construir um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e uma variável aleatória X tal que f seja a função de massa de probabilidade de X .

6 Esperança

Se lançarmos um dado justo muitas vezes, esperamos que cada um dos seis resultados possíveis apareça cerca de um sexto do tempo, e assim a média dos números obtidos seria aproximadamente

$$\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{7}{2}.$$

Chamamos esse número de *esperança* de X .

6.1 Definição e exemplos

Definição 6.1 (Esperança). Seja X uma variável aleatória discreta. Definimos a *esperança* de X , denotada por $\mathbb{E}[X]$, como o número real dado por

$$\mathbb{E}[X] = \sum_{x: \mathbb{P}(X=x) > 0} x \cdot \mathbb{P}(X = x),$$

desde que essa soma convirja absolutamente, caso contrário $\mathbb{E}[X]$ não está definida.

Terminologia. Dizer que $\sum_n a_n$ *converge absolutamente* significa que $\sum_n |a_n|$ converge.

Definição 6.2 (Integrável). Dizemos que uma variável aleatória discreta X é *integrável* se a sua esperança estiver definida, ou seja, se a soma

$$\sum_{x: \mathbb{P}(X=x) > 0} |x| \cdot \mathbb{P}(X = x)$$

for convergente.

Notação. O fato de que $\mathbb{E}[X]$ depende de X fica evidente pelo fato de " X " aparecer em " $\mathbb{E}[X]$ " e o fato de que depende de \mathbb{P} fica em parte aparente pelo uso da mesma fonte em " \mathbb{E} " e " \mathbb{P} ".

Exemplo 6.3. Lance uma moeda justa 4 vezes e conte o número de caras.

$$\begin{aligned} \mathbb{E}[X] &= 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) + \\ &\quad + 3 \cdot \mathbb{P}(X = 3) + 4 \cdot \mathbb{P}(X = 4) \end{aligned}$$

$$\begin{aligned}
&= 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} \\
&= 2.
\end{aligned}$$

Exemplo 6.4 (Função indicadora). Seja $A \in \mathcal{F}$ e defina X como

$$X(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c. \end{cases}$$

Uma função assim é chamada de *função indicadora* do conjunto A e é denotada por $\mathbf{1}_A$. Nesse caso, $\mathbb{E}[X] = 0 \times \mathbb{P}(A^c) + 1 \times \mathbb{P}(A) = \mathbb{P}(A)$. Ou seja,

$$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A).$$

Exemplo 6.5. Lance um dado justo duas vezes e some os valores observados.

$$\begin{aligned}
\mathbb{E}[X] &= 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} + \\
&\quad + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7.
\end{aligned}$$

Exemplo 6.6. Pegue 3 cartas de um baralho de 52 cartas, uma após a outra e sem reposição, e conte quantas são damas de ouros.

$$\begin{aligned}
\mathbb{E}[X] &= 0 \times \frac{48 \cdot 47 \cdot 46}{52 \cdot 51 \cdot 50} + 1 \times \frac{3 \cdot 48 \cdot 47 \cdot 4}{52 \cdot 51 \cdot 50} + \\
&\quad + 2 \times \frac{3 \cdot 48 \cdot 4 \cdot 3}{52 \cdot 51 \cdot 50} + 3 \times \frac{4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50} = \frac{3}{13}.
\end{aligned}$$

Exemplo 6.7 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n=0}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} = \\
&= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.
\end{aligned}$$

Portanto, a esperança de uma variável aleatória distribuída como $\text{Poisson}(\lambda)$ é λ .

Exemplo 6.8 (Binomial). Se $X \sim \text{Binom}(n, p)$, então

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= n \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j+1} (1-p)^{n-j-1} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ &= np[p + (1-p)]^{n-1} = np.\end{aligned}$$

Exemplo 6.9 (Geométrica). Suponha que $X \sim \text{Geom}(p)$. Vamos calcular

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} n(1-p)^{n-1}p$$

através da diferenciação de uma série de potências. Escrevendo $x = 1 - p$, podemos desenvolver da seguinte forma:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=1}^{\infty} n \cdot p \cdot (1-p)^{n-1} = p \sum_{n=0}^{\infty} n \cdot x^{n-1} \\ &= p \sum_{n=0}^{\infty} \frac{d}{dx} [x^n] = p \cdot \frac{d}{dx} \left[\sum_{n=0}^{\infty} x^n \right] = p \cdot \frac{d}{dx} \left[\frac{1}{1-x} \right] \\ &= p \cdot \left(- (1-x)^{-2} \right) \cdot (-1) = \frac{1}{p}.\end{aligned}$$

Portanto, a esperança de uma variável aleatória distribuída como $\text{Geom}(p)$ é $\frac{1}{p}$. Sabemos que a série de potências $\sum_n x^n$ converge se $|x| < 1$, e estamos aceitando uma propriedade que diz que a série de potências pode ser diferenciada termo a termo dentro desse intervalo.

6.2 Propriedades da esperança

Nos exemplos acima, $2 = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}$, $7 = \frac{7}{2} + \frac{7}{2}$, $\frac{3}{13} = \frac{1}{13} + \frac{1}{13} + \frac{1}{13}$, e $np = p + \dots + p$. Isso não é uma coincidência. Vem do fato de que

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

para variáveis aleatórias discretas integráveis X e Y .

Teorema 6.10. *Sejam X e Y variáveis aleatórias discretas integráveis definidas*

no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Então:

- (1) $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ para todo $A \in \mathcal{F}$,
- (2) Se $0 \leq Z \leq X$ para todo $\omega \in \Omega$, então $0 \leq \mathbb{E}[Z] \leq \mathbb{E}[X]$,
- (3) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

Dizemos que a expectativa é *unitária*, *monótona* e *linear*.

Veremos uma prova mais adiante.

Exemplo 6.11. No Exemplo 6.3, podemos definir X_1, X_2, X_3 e X_4 como a função indicadora dos eventos em que o primeiro, segundo, terceiro e quarto lançamentos da moeda resultaram em Caras, respectivamente. Como $X = X_1 + X_2 + X_3 + X_4$, podemos obter a expectativa usando a linearidade, como em

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] + \mathbb{E}[X_4] = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 2,$$

em vez de calcular a função de massa de probabilidade de X .

Exemplo 6.12. No Exemplo 6.5, observe que $X = Y + Z$, onde Y e Z representam o resultado do primeiro e do segundo dados. Assim,

$$\mathbb{E}[X] = \mathbb{E}[Y] + \mathbb{E}[Z] = \frac{7}{2} + \frac{7}{2} = 7.$$

Exemplo 6.13. No Exemplo 6.6, observe que $X = X_1 + X_2 + X_3$, onde X_k é a indicadora de se a k -ésima carta é uma rainha. Diferentemente dos exemplos anteriores, observe que aqui X_1, X_2 e X_3 não são "independentes" (uma noção precisa de independência será introduzida mais adiante). No entanto, cada uma delas individualmente satisfaz $\mathbb{E}[X_k] = \frac{1}{13}$, e podemos calcular

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] = \frac{3}{13}.$$

Exemplo 6.14. No Exemplo 6.8, observe que X tem a mesma distribuição que $X_1 + \dots + X_n$, onde cada X_k é distribuída como Bernoulli(p), e, portanto,

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = (p + \dots + p) = np.$$

Nos exemplos anteriores, foi mais fácil calcular a expectativa usando a linearidade do que usando a distribuição de X . Em muitos outros casos, descrever a

distribuição de X de uma forma que nos permita calcular a expectativa pode ser muito difícil ou até mesmo impraticável, mas ainda assim pode ser possível calcular a expectativa usando a linearidade.

Exemplo 6.15. Uma gaveta contém 10 pares de meias, todos diferentes entre si. Alguém abre a gaveta no escuro e retira 6 meias dela. Qual é a expectativa de X , o número de pares formados pelas meias retiradas? É mais conveniente supor que as meias sejam retiradas em ordem, da 1-ª à 6-ª. Vamos contar quantas delas têm um par que também foi retirado. Isso dará um número N que é o dobro do número de pares, porque cada par será contado duas vezes, então $N = 2X$. Observe que $N = X_1 + \cdots + X_6$, onde $X_k = \mathbb{1}_{A_k}$ e A_k é o evento de que o par da k -ésima meia retirada também foi retirado. Então $\mathbb{P}(A_k) = \frac{5}{19}$ (exercício!) e, portanto, $\mathbb{E}[N] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_6] = 6 \cdot \frac{5}{19}$. Portanto, $\mathbb{E}[X] = \mathbb{E}[\frac{N}{2}] = \frac{15}{19}$. A combinatorial envolvida em mostrar que $\mathbb{P}(A_k) = \frac{5}{19}$ pode não ser muito fácil, mas é muito mais fácil do que descrever a distribuição de N .

6.3 Função de uma variável aleatória

Proposição 6.16. *Seja X uma variável aleatória discreta e $g : \mathbb{R} \rightarrow \mathbb{R}$ uma função qualquer. Então*

$$\mathbb{E}[g(X)] = \sum_{x \in D_X} g(x) \cdot \mathbb{P}(X = x),$$

se esta soma convergir absolutamente, e $\mathbb{E}[g(X)]$ é indefinida se não convergir. Na soma, D_X denota o suporte discreto de X .

Exemplo 6.17. Suponha que $p_X(x) = \frac{1}{3}$ para $x = 1, 2, 3$. Vamos calcular $\mathbb{E}[(X - 2)^2]$ de duas maneiras. Para a função $g(x) = (x - 2)^2$, queremos calcular $\mathbb{E}[g(X)]$.

A primeira maneira é a seguinte. Defina a variável aleatória $Z = g(X) = (X - 2)^2$ e calcule $\mathbb{E}[Z]$ a partir da definição. Começamos calculando $p_Z(0) = \mathbb{P}(X \in \{2\}) = \frac{1}{3}$ e $p_Z(1) = \mathbb{P}(X \in \{1, 3\}) = \frac{2}{3}$, obtendo a tabela

z	$p_Z(z)$
0	$\frac{1}{3}$
1	$\frac{2}{3}$

e, finalmente, $\mathbb{E}[Z] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$. Para a segunda maneira, basta escrever

x	$g(x)$	$p_X(x)$
1	1	$\frac{1}{3}$
2	0	$\frac{1}{3}$
3	1	$\frac{1}{3}$

e calcular $\mathbb{E}[g(X)] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$.

Exemplo 6.18. Se $X \sim \text{Poisson}(\lambda)$, então

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} n \frac{\lambda^n e^{-\lambda}}{(n-1)!} \\
 &= \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} + \sum_{n=1}^{\infty} (n-1) \frac{\lambda^n e^{-\lambda}}{(n-1)!} \\
 &= \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} + \sum_{n=2}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-2)!} \\
 &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} + \lambda^2 e^{-\lambda} \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} \\
 &= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \lambda^2 e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \\
 &= \lambda + \lambda^2.
 \end{aligned}$$

Embora tenha havido muita computação algébrica envolvida, a alternativa seria pior: definir $Z = X^2$, descrever o suporte discreto de Z , encontrar uma expressão para $p_Z(z)$, escrever $\mathbb{E}[Z] = \sum_z z \cdot p_Z(z)$ e, em seguida, tentar avaliar a soma.

Exemplo 6.19. Suponha que $X \sim \text{Geom}(p)$. Como antes, calcularemos

$$\mathbb{E}[X^2] = \sum_{n=1}^{\infty} n^2 (1-p)^{n-1} p$$

por meio da diferenciação de duas séries de potência. Para fazer isso, escrevemos $x = 1 - p$ e desenvolvemos da seguinte maneira:

$$\mathbb{E}[X^2] = \sum_{n=1}^{\infty} n^2 \cdot p \cdot x^{n-1}$$

$$\begin{aligned}
&= p \sum_{n=1}^{\infty} n \cdot x^{n-1} + p \sum_{n=1}^{\infty} n \cdot (n-1) \cdot x^{n-1} \\
&= p \sum_{n=0}^{\infty} n \cdot x^{n-1} + px \sum_{n=0}^{\infty} n \cdot (n-1) \cdot x^{n-2} \\
&= p \sum_{n=0}^{\infty} \frac{d}{dx} [x^n] + px \sum_{n=0}^{\infty} \frac{d^2}{dx^2} [x^n] \\
&= p \cdot \frac{d}{dx} \left[\sum_{n=0}^{\infty} x^n \right] + px \cdot \frac{d^2}{dx^2} \left[\sum_{n=0}^{\infty} x^n \right] \\
&= p \frac{1}{(1-x)^2} + px \frac{2}{(1-x)^3} \\
&= \frac{1}{p} + 2 \cdot \frac{1-p}{p^2} \\
&= \frac{2-p}{p^2}.
\end{aligned}$$

Como antes, sabemos que a série de potência $\sum_n x^n$ converge se $|x| < 1$, e estamos aceitando uma propriedade que diz que a série de potência pode ser diferenciada termo a termo dentro desse intervalo.

Exemplo 6.20. Suponha que $X \sim \text{Geom}(p)$. Para quais valores de t a função e^{tX} é integrável e qual é o valor de $\mathbb{E}[e^{tX}]$? Podemos escrever

$$\mathbb{E}[e^{tX}] = \sum_{n=1}^{\infty} e^{tn} \cdot p \cdot (1-p)^{n-1} = pe^t \sum_{k=0}^{\infty} [e^t \cdot (1-p)]^{n-1} = \frac{pe^t}{1 - [e^t \cdot (1-p)]}.$$

Isso pode ser reescrito como

$$\mathbb{E}[e^{tX}] = \frac{p}{e^{-t} + p - 1},$$

e é definido se $e^t \cdot (1-p) < 1$, ou alternativamente $t < \ln \frac{1}{1-p}$, e é indefinido caso contrário.

6.4 Variância

Aqui introduzimos outra quantidade fundamental que descreve a distribuição de uma variável aleatória. Enquanto $\mathbb{E}[X]$ fornece o número médio de X , agora definimos uma quantidade que quantifica o grau de dispersão de X em relação ao seu valor médio.

Definição 6.21 (Variáveis aleatórias quadrado-integráveis). Dizemos que uma variável aleatória discreta X é *quadrado-integrável* se X^2 for integrável, o que significa que a soma

$$\sum_{x: \mathbb{P}(X=x) > 0} x^2 \cdot \mathbb{P}(X = x)$$

é convergente. Observe que variáveis aleatórias quadrado-integráveis são automaticamente integráveis, pois $|x| \leq 1 + x^2$.

Definição 6.22 (Variância). Seja X uma variável aleatória discreta quadrado-integrável e denote $\mu = \mathbb{E}[X]$. Definimos a *variância* de X como

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

Embora esta fórmula seja a melhor definição para entender as propriedades da variância, muitas vezes existe uma maneira mais conveniente de calculá-la:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

que obtemos expandindo $\mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - \mu^2$.

Exemplo 6.23 (Poisson). Suponha que $X \sim \text{Poisson}(\lambda)$. Então

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda + \lambda^2 - \lambda = \lambda,$$

portanto, a variância de uma variável aleatória de Poisson é igual à sua expectativa.

Exemplo 6.24 (Geométrica). Suponha que $X \sim \text{Geom}(p)$. Então

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Exemplo 6.25 (Bernoulli). Suponha que $X \sim \text{Bernoulli}(p)$. Então

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p \cdot (1 - p).$$

Observe que

$$\text{Var}(aX) = a^2 \cdot \text{Var}(X)$$

o que significa que $\text{Var}(X)$ não está na mesma unidade de medida que X . Por

exemplo, se X é medido em metros, então $\mathbb{E}[X]$ também é medido em metros, mas $\text{Var}(X)$ é medido em metros quadrados.

Para quantificar a dispersão de X nas mesmas unidades de medida que X , precisamos calcular a raiz quadrada.

Definição 6.26 (Desvio padrão). Seja X uma variável aleatória discreta quadrado-integrável. Definimos o *desvio padrão* de X como

$$\sigma(X) = \sqrt{\text{Var}(X)}.$$

Ao contrário da variância, o desvio padrão satisfaz $\sigma(aX) = |a| \cdot \sigma(X)$.

O desvio padrão de uma variável aleatória de Poisson é $\sqrt{\lambda}$, de uma variável aleatória de Bernoulli é $\sqrt{p(1-p)}$, e de uma variável aleatória geométrica é $\sqrt{p^{-2} - p^{-1}}$.

7 Distribuições discretas multivariadas

7.1 Função de massa de probabilidade conjunta de duas variáveis

Definição 7.1 (Função de massa de probabilidade conjunta). Dadas duas variáveis aleatórias discretas X e Y , definimos a *função de massa de probabilidade conjunta de X e Y* , denotada por $p_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$, e dada por

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

ou, mais formalmente, $\mathbb{P}(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\})$.

Exemplo 7.2. Lance dois dados, e seja X o valor maior e Y o valor menor. Então, $p_{X,Y}(x, y)$ para $x = 1, \dots, 6$ e $y = 1, \dots, 6$ é dado pela tabela

	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
2	$\frac{0}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
3	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
4	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$
5	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{1}{36}$	$\frac{2}{36}$
6	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{0}{36}$	$\frac{1}{36}$

e $p(x, y) = 0$ se x ou y não pertencerem a $\{1, 2, 3, 4, 5, 6\}$.

Observe que

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x) \\ &= \sum_{y \in D_Y} \mathbb{P}(X = x, Y = y) + \mathbb{P}(X = x, Y \notin D_Y) \\ &= \sum_{y \in D_Y} \mathbb{P}(X = x, Y = y) \\ &= \sum_{y \in D_Y} p_{X,Y}(x, y) \end{aligned}$$

para cada $x \in \mathbb{R}$, onde D_Y denota o suporte discreto de Y .

Terminologia (Função de massa de probabilidade marginal). A fórmula acima

para calcular a função de massa de probabilidade de X a partir da função de massa de probabilidade conjunta de X e Y é chamada de *função de massa de probabilidade marginal*.

Exemplo 7.3. Uma sacola contém 1 bola vermelha, 2 verdes e 2 azuis. Retiramos 2 bolas da sacola, sem reposição. Seja X o número de bolas verdes retiradas, e Y o número de bolas vermelhas retiradas. Então, a função de massa de probabilidade conjunta de X e Y é dada pela célula central da tabela abaixo:

$y \backslash x$	0	1	2	total
0	0.1	0.4	0.1	0.6
1	0.2	0.2	0	0.4
total	0.3	0.6	0.1	1

Somando cada coluna, encontramos a função de massa de probabilidade marginal de X , que é dada por $p_X(0) = 0.3$, $p_X(1) = 0.6$, e $p_X(2) = 0.1$. Somando cada linha, encontramos a função de massa de probabilidade marginal de Y , que é dada por $p_Y(0) = 0.6$ e $p_Y(1) = 0.4$.

7.2 Esperança no caso bivariado discreto

Proposição 7.4 (Esperança no caso bivariado). *Sejam X e Y variáveis aleatórias discretas e $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ uma função qualquer. Então*

$$\mathbb{E}[g(X, Y)] = \sum_{x \in D_X} \sum_{y \in D_Y} g(x, y) \cdot \mathbb{P}(X = x, Y = y),$$

se essa soma convergir absolutamente, e $\mathbb{E}[g(X, Y)]$ é indefinida caso contrário. Os conjuntos D_X e D_Y na fórmula denotam os suportes discretos de X e Y .

Demonstração. Seja $Z = g(X, Y)$. Primeiro, observamos que, para cada $z \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(Z = z) &= \mathbb{P}(g(X, Y) = z) \\ &= \mathbb{P}((X, Y) \in g^{-1}(z)) \\ &= \sum_{(x, y) \in g^{-1}(z) \cap D} \mathbb{P}(X = x, Y = y), \end{aligned}$$

onde $D = \{(x, y) : \mathbb{P}(X = x, Y = y) > 0\}$. O suporte de Z é dado pela imagem $D_Z = g(D)$, que é contável. Finalmente,

$$\begin{aligned}
\text{Esperança}[Z] &= \sum_{z \in D_Z} z \cdot \mathbb{P}(Z = z) \\
&= \sum_{z \in D_Z} \sum_{(x,y) \in \text{ing}^{-1}(z) \cap D} z \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_{z \in D_Z} \sum_{(x,y) \in \text{ing}^{-1}(z) \cap D} g(x, y) \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y) \in D} g(x, y) \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in D_X} \sum_{y \in D_Y} g(x, y) \cdot \mathbb{P}(X = x, Y = y),
\end{aligned}$$

e a soma converge absolutamente se e somente se Z for integrável. Isso conclui a prova. \square

Prova da Proposição 6.16. Se tomarmos $Y = 0$ e aplicarmos a Proposição 7.4 com $\tilde{g}(x, y) = g(x)$, obtemos

$$\begin{aligned}
\text{Esperança}[X] &= \text{Esperança}[\tilde{g}(X, Y)] \\
&= \sum_{x \in D_X} \sum_{y \in \{0\}} \tilde{g}(x, y) \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in D_X} g(x) \cdot \mathbb{P}(X = x),
\end{aligned}$$

e as somas convergem absolutamente se e somente se $\text{Esperança}[X]$ estiver definida. Isso conclui a prova da Proposição 6.16. \square

Corolário 7.5. *A esperança é linear.*

Demonstração. Suponhamos que X e Y são integráveis, e que $a, b \in \mathbb{R}$. Usando a Proposição 7.4 com $g(x, y) = ax + by$, $g(x, y) = x$ e $g(x, y) = y$, obtemos

$$\begin{aligned}
\mathbb{E}[aX + bY] &= \sum_{x \in D_X} \sum_{y \in D_Y} (ax + by) \cdot \mathbb{P}(X = x, Y = y) \\
&= a \sum_{x \in D_X} \sum_{y \in D_Y} x \cdot \mathbb{P}(X = x, Y = y) + b \sum_{x \in D_X} \sum_{y \in D_Y} y \cdot \mathbb{P}(X = x, Y = y) \\
&= a \mathbb{E}[X] + b \mathbb{E}[Y],
\end{aligned}$$

o que é o que queríamos provar. \square

7.3 Variáveis aleatórias discretas independentes

Definição 7.6 (Independência). Duas variáveis aleatórias discretas X e Y são *independentes* se

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

para todo $x, y \in \mathbb{R}$.

Exemplo 7.7. Lance uma moeda justa 5 vezes. Seja X o número de Caras nos três primeiros lances e Y o número de Caras nos dois últimos lances. Então, a função de massa de probabilidade conjunta é mostrada na tabela

$y \backslash x$	0	1	2	3	total
0	1/32	3/32	3/32	1/32	1/4
1	2/32	6/32	6/32	2/32	1/2
2	1/32	3/32	3/32	1/32	1/4
total	1/8	3/8	3/8	1/8	1

Observe como cada entrada no meio da tabela é dada pelo produto de sua soma de coluna e soma de linha, o que significa exatamente que $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$.

Definição 7.8 (Independência par a par). Dizemos que uma coleção de variáveis aleatórias discretas X_1, X_2, X_3, \dots é *independente par a par* se X_j e X_k forem independentes para todo $j \neq k$.

Definição 7.9 (Independência mútua). Dizemos que uma coleção de variáveis aleatórias discretas X_1, X_2, X_3, \dots é *mutuamente independentes* se, para todo k e todo x_1, x_2, \dots, x_k , tivermos

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdots \mathbb{P}(X_k = x_k).$$

Teorema 7.10 (Esperança de variáveis aleatórias independentes). Se X e Y são variáveis aleatórias discretas independentes integráveis, então XY é integrável e

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

Demonstração. Usando a Proposição 7.4, temos:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x \in D_X} \sum_{y \in D_Y} xy \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in D_X} x \cdot \left(\sum_{y \in D_Y} y \cdot \mathbb{P}(X = x) \mathbb{P}(Y = y) \right) \\
&= \sum_{x \in D_X} x \cdot \mathbb{P}(X = x) \cdot \left(\sum_{y \in D_Y} y \cdot \mathbb{P}(Y = y) \right) \\
&= \left(\sum_{y \in D_Y} y \cdot \mathbb{P}(Y = y) \right) \cdot \left(\sum_{x \in D_X} x \cdot \mathbb{P}(X = x) \right) \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y].
\end{aligned}$$

Para usar a Proposição 7.4, deveríamos ter sabido que XY era integrável desde o início. Isso pode ser verificado usando exatamente o mesmo desenvolvimento com $|x|$ em vez de x e $|y|$ em vez de y . Isso prova o teorema. \square

Exemplo 7.11. Lance um dado justo duas vezes e multiplique os valores observados.

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{36} (1 \cdot 1 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 4 + 8 \cdot 2 + 9 \cdot 1 + 10 \cdot 2 + \\
&\quad + 12 \cdot 4 + 15 \cdot 2 + 16 \cdot 1 + 18 \cdot 2 + 20 \cdot 2 + 24 \cdot 2 + 25 \cdot 1 + 30 \cdot 2 + 36 \cdot 1) \\
&= \frac{49}{4}.
\end{aligned}$$

Uma solução mais simples é observar que $X = YZ$, onde Y e Z representam o primeiro e o segundo lançamento do dado. Usando o teorema acima,

$$\mathbb{E}[X] = \mathbb{E}[Y] \cdot \mathbb{E}[Z] = \frac{7}{2} \cdot \frac{7}{2} = \frac{49}{4}.$$

Observe como o cálculo foi simplificado.

Proposição 7.12. Se X_1, \dots, X_n são variáveis aleatórias discretas integráveis independentes duas a duas, então

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i).$$

Demonstração. Seja $\mu_i := \mathbb{E}[X_i]$, $i = 1, 2, \dots, n$ e defina $\mu := \sum_{i=1}^n \mu_i = \mathbb{E}[\sum_{i=1}^n X_i]$ pela linearidade da esperança. Então,

$$\begin{aligned}
 \text{Var}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \mu\right)^2\right] \\
 &= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2\right] \\
 &= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n (X_i - \mu_i)(X_j - \mu_j)\right] \\
 &= \mathbb{E}\left[\sum_{i=1}^n (X_i - \mu_i)^2\right] + \mathbb{E}\left[\sum_{i \neq j} (X_i - \mu_i)(X_j - \mu_j)\right] \\
 &= \sum_{i=1}^n \mathbb{E}[(X_i - \mu_i)^2] + \sum_{i \neq j} \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\
 &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \left(\mathbb{E}[X_i X_j] - \mu_i \mu_j\right) \\
 &= \sum_{i=1}^n \text{Var}(X_i). \quad \square
 \end{aligned}$$

8 A Lei dos Grandes Números

Um dos principais tópicos da Probabilidade é a "lei dos grandes números". Ela afirma que a soma de um grande número de variáveis independentes tende a se aproximar da esperança. Mais precisamente, a média observada $\frac{X_1 + \dots + X_n}{n}$, que é aleatória, se aproximará da média teórica $\mathbb{E}[\frac{X_1 + \dots + X_n}{n}]$, que é determinística!

A manifestação mais simples desse fenômeno diz respeito às frequências relativas. Imagine que executamos um experimento muitas vezes, sob as mesmas condições, e contamos quantas vezes resultou em sucesso e quantas vezes resultou em fracasso. Tomando X_n como o indicador de que o n -ésimo teste resultou em sucesso, a frequência relativa de sucesso é exatamente dada por $\frac{X_1 + \dots + X_n}{n}$. Enquanto escrevia esse preâmbulo, os palestrantes simularam o lançamento de uma moeda justa um milhão de vezes e obtiveram "Cara" 499.947 vezes. Repetindo o mesmo procedimento, eles obtiveram "Cara" 499.508 vezes, depois 500.318 vezes e, em seguida, 500.512 vezes. Obviamente, algo está acontecendo aqui. Conforme previsto pela lei dos grandes números, a frequência relativa sempre foi muito próxima da probabilidade de obter "Cara" em cada lançamento da moeda, que é exatamente $\frac{1}{2}$. Intuitivamente, tendemos a associar a probabilidade de sucesso à frequência relativa de sucessos, e essa associação está quase arraigada em nosso pensamento.

Mais genericamente, o resultado relevante de cada experimento não precisa ser 0 ou 1 para representar fracasso ou sucesso. Pode ser qualquer variável aleatória. Antes de escrever este parágrafo, os palestrantes lançaram um dado 10 vezes, e a soma dos valores obtidos foi 38. Repetindo o mesmo procedimento, eles obtiveram 33, 28, 37 e, finalmente, 43. Não parece muito com essa soma estar bem concentrada perto de algum valor determinístico. No entanto, as palestras prosseguiram com a simulação de 10.000 lançamentos do dado, e a soma dos resultados foi 35.082. Repetindo esse procedimento, eles obtiveram uma soma de 34.769, depois 35.419 e, finalmente, 34.691. Conforme previsto pela lei dos grandes números, quando o número de lançamentos dos dados era grande, a média observada estava sempre próxima da média teórica, dada por $\mathbb{E}[X_1] = \frac{7}{2}$.

Na teoria da Probabilidade, a lei dos grandes números não é apenas uma história ou um fenômeno misterioso, é um teorema que pode ter muitas formulações diferentes. Aqui consideraremos a versão mais simples possível.

Teorema 8.1 (Lei dos Grandes Números). *Se X_1, X_2, X_3, \dots forem variáveis aleatórias discretas independentes duas a duas e com variância quadrática integrável, com a mesma média μ e mesma variância σ^2 , então, para todo $a > 0$ e $n \in \mathbb{N}$,*

$$\mathbb{P}\left(\mu - a \leq \frac{X_1 + \dots + X_n}{n} \leq \mu + a\right) \geq 1 - \frac{\sigma^2}{a^2 n}.$$

A lei dos grandes números é comumente conhecida como a Lei dos Grandes Números.

Chamamos a atenção para o fato de que, não importa quão pequeno seja a , essa probabilidade pode ser aproximada por 1, desde que escolhamos um valor grande o suficiente para n (é claro, se a for muito pequeno, precisaremos de um valor realmente muito grande para n).

Nosso próximo objetivo é entender como tal estimativa de probabilidades surgiu. Até agora, usamos probabilidades para calcular esperanças e desvios padrão, e agora de repente estamos usando o desvio padrão para fazer estimativas extremamente interessantes sobre probabilidades!

Essa empreitada terá duas partes: entender qual é a variância da soma de muitas variáveis aleatórias e entender como a variância de uma variável aleatória fornece estimativas sobre a probabilidade de ela se desviar de sua média.

9 Covariância

9.1 Definição

O que acontece com a variância quando adicionamos variáveis?

Exemplo 9.1. Lance cinco moedas justas. Seja X o número de "Cara" nos primeiros dois lançamentos de moedas, e Y o número de "Cara" nos últimos três lançamentos de moedas. Seja $Z = X + Y$ o número total de "Cara" em todos os cinco lançamentos. Após cálculos (que omitimos), obtemos $\text{Var}(X) = \frac{1}{2}$, $\text{Var}(Y) = \frac{3}{4}$ e $\text{Var}(Z) = \frac{5}{4}$. Neste caso, a relação $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ foi verificada. Isso é uma coincidência? Em quais condições essa relação é verificada?

Exemplo 9.2. Suponha que $X \sim \text{Bernoulli}(\frac{1}{2})$ e seja $Y = X$. Neste caso, temos $\text{Var}(X) = \frac{1}{4}$, $\text{Var}(Y) = \frac{1}{4}$, $\text{Var}(X + Y) = 1$, então $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$.

Para entender melhor o que acontece com $\text{Var}(X + Y)$, expandimos:

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}[(X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y])]^2 \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \cdot \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].\end{aligned}$$

Definição 9.3 (Covariância). Suponha que X e Y são variáveis aleatórias discretas de variância quadrática integrável. Definimos a *covariância de X e Y* como

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

A partir dos cálculos anteriores, vemos que $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ se e somente se $\text{Cov}(X, Y) = 0$. Quando essa condição é satisfeita, dizemos que X e Y são *não correlacionados*.

9.2 Propriedades

Vejamos as principais propriedades da covariância. Trocando X e Y , temos

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

Substituindo X no lugar de Y , obtemos

$$\text{Cov}(X, X) = \text{Var}(X)$$

Proposição 9.4. *Suponha que X , Y e Z são variáveis aleatórias discretas de variância quadrática integrável. Então*

$$\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$$

para todo $a, b \in \mathbb{R}$.

Demonstração. Basta expandir e agrupar:

$$\begin{aligned} \text{Cov}(aX + bY, Z) &= \mathbb{E}[(aX + bY) - \mathbb{E}[aX + bY])(Z - \mathbb{E}[Z])] \\ &= \mathbb{E}[(aX - \mathbb{E}[aX] + bY - \mathbb{E}[bY])(Z - \mathbb{E}[Z])] \\ &= \mathbb{E}[(aX - \mathbb{E}[aX])(Z - \mathbb{E}[Z]) + (bY - \mathbb{E}[bY])(Z - \mathbb{E}[Z])] \\ &= a \mathbb{E}[(X - \mathbb{E}[X])(Z - \mathbb{E}[Z])] + b \mathbb{E}[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])] \\ &= a \text{Cov}(X, Z) + b \text{Cov}(Y, Z), \end{aligned}$$

provando a identidade. □

Corolário 9.5. *Suponha que $X_1, \dots, X_n, Y_1, \dots, Y_m$ são variáveis aleatórias discretas de variância quadrática integrável. Então*

$$\text{Cov}\left(\sum_{j=1}^n a_j X_j, \sum_{k=1}^m b_k Y_k\right) = \sum_{j=1}^n \sum_{k=1}^m a_j b_k \text{Cov}(X_j, Y_k)$$

para todo $a_1, \dots, a_n, b_1, \dots, b_m \in \mathbb{R}$.

Demonstração. Usando a proposição anterior repetidamente e a simetria,

$$\begin{aligned} \text{Cov}\left(\sum_{j=1}^n a_j X_j, \sum_{k=1}^m b_k Y_k\right) &= \sum_{j=1}^n a_j \text{Cov}\left(X_j, \sum_{k=1}^m b_k Y_k\right) \\ &= \sum_{j=1}^n a_j \text{Cov}\left(\sum_{k=1}^m b_k Y_k, X_j\right) \\ &= \sum_{j=1}^n \sum_{k=1}^m a_j b_k \text{Cov}(Y_k, X_j) \end{aligned}$$

$$= \sum_{j=1}^n \sum_{k=1}^m a_j b_k \text{Cov}(X_j, Y_j),$$

o que prova a identidade declarada. \square

Corolário 9.6. *Sejam X_1, \dots, X_n variáveis aleatórias discretas de variância quadrática integrável. Então*

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k)$$

Demonstração. Usando as propriedades anteriores,

$$\begin{aligned} \text{Var}\left(\sum_{k=1}^n X_k\right) &= \text{Cov}\left(\sum_{j=1}^n X_j, \sum_{k=1}^n X_k\right) \\ &= \sum_{j=1}^n \sum_{k=1}^n \text{Cov}(X_j, X_k) \\ &= \sum_{k=1}^n \text{Cov}(X_k, X_k) + \sum_{j \neq k} \text{Cov}(X_j, X_k) \\ &= \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k). \end{aligned} \quad \square$$

Definição 9.7. Dizemos que uma coleção de variáveis aleatórias discretas de variância quadrática integrável X_1, X_2, X_3, \dots é *não correlacionada* se $\text{Cov}(X_j, X_k) = 0$ para cada $j \neq k$.

Corolário 9.8. *Se X_1, \dots, X_n são variáveis aleatórias discretas não correlacionadas, então*

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k).$$

Demonstração. Aplique a fórmula anterior e observe que a covariância entre termos diferentes é zero. \square

Isso nos dá a "lei da raiz quadrada": se X_1, \dots, X_n são variáveis aleatórias discretas não correlacionadas com a mesma média μ e variância σ^2 , então

$$\mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \mu \quad \text{e} \quad \sigma\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma}{\sqrt{n}}.$$

Isso começa a explicar por que a lei das médias emerge quando somamos muitas variáveis aleatórias.

9.3 Somas de variáveis independentes duas a duas

Talvez uma expressão mais conveniente para a covariância:

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

Se X e Y são independentes, então, pelo Teorema 7.10, $\text{Cov}(X, Y) = 0$.

Corolário 9.9. *Se X_1, \dots, X_n são variáveis aleatórias discretas de variância quadrática integrável independentes duas a duas, então*

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k).$$

Pela observação anterior, uma família de variáveis aleatórias discretas independentes duas a duas também é não correlacionada.

10 Desigualdade de Chebyshev

Agora veremos como a média e o desvio padrão de uma variável aleatória nos permitem fazer algumas estimativas sobre probabilidades envolvendo a variável aleatória.

10.1 Desigualdade de Markov

Para chegarmos lá, começamos com algo mais modesto: a Desigualdade de Markov. Ela nos permite dizer algo sobre a distribuição de uma variável aleatória com base no conhecimento de sua expectativa.

Teorema 10.1 (Desigualdade de Markov). *Seja X uma variável aleatória discreta integrável não negativa. Então,*

$$\mathbb{P}(X > x) \leq \frac{\mathbb{E}[X]}{x}$$

para todo $x > 0$.

Demonstração. Fixe $x > 0$. Defina a variável aleatória

$$Y := \begin{cases} x & \text{se } X \geq x; \\ 0 & \text{caso contrário.} \end{cases}$$

Temos que $X \geq Y$, porque:

- se $X \geq x$, então $Y = x$, então $X \geq Y$;
- se $X \in [0, x)$, então $Y = 0$, então $X \geq Y$.

Isso também nos dá $\mathbb{E}[X] \geq \mathbb{E}[Y]$. Em seguida, observe que Y é uma variável aleatória discreta (ela assume apenas os valores 0 e x) com

$$p_Y(x) = \mathbb{P}(X \geq x), \quad p_Y(0) = \mathbb{P}(X < x).$$

Portanto,

$$\mathbb{E}[X] \geq \mathbb{E}[Y] = 0 \cdot p_Y(0) + x \cdot p_Y(x) = x \cdot p_Y(x) = x \cdot \mathbb{P}(X \geq x).$$

Rearranjando isso, obtemos a desigualdade desejada. \square

Exemplo 10.2. Suponha que uma empresa produza em média 50 itens por semana. Você pode estimar a probabilidade de que a produção de uma semana específica exceda 75 itens? Seja X o número de itens que a empresa produz a cada semana. Pela afirmação, sabemos que $X \geq 0$ e $\mathbb{E}[X] = 50$. Portanto, as premissas da Desigualdade de Markov são satisfeitas e podemos deduzir uma estimativa sobre a probabilidade solicitada, que é $\mathbb{P}(X \geq 75)$. Portanto,

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}[X]}{75} = \frac{50}{75} = \frac{2}{3}.$$

É interessante notar que a Desigualdade de Markov nem sempre fornece uma limitação útil. De fato, se X for uma variável aleatória não negativa com expectativa igual a μ , e $x \in (0, \mu]$, então na desigualdade

$$\mathbb{P}(X \geq x) \leq \frac{\mu}{x},$$

o lado direito é maior que 1, então a limitação apenas nos diz que a probabilidade é menor ou igual a 1, mas já sabíamos disso!

10.2 Desigualdade de Chebyshev

Enquanto a Desigualdade de Markov fornece uma limitação sobre a probabilidade de uma variável aleatória ser grande, a Desigualdade de Chebyshev fornece uma limitação sobre a probabilidade de uma variável aleatória estar longe de sua expectativa.

Teorema 10.3 (Desigualdade de Chebyshev). *Seja X uma variável aleatória discreta de variância quadrática integrável. Então,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

para todo $a > 0$.

É importante destacar que aqui não fazemos suposições quanto ao sinal de X .

Demonstração. Seja $a > 0$. Defina $Y := (X - \mathbb{E}[X])^2$. Portanto, Y é não negativa e

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

Em particular, Y é integrável. A seguir, observe que os seguintes eventos são os mesmos:

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

Portanto, pela Desigualdade de Markov, temos

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}(Y \geq a^2) \leq \frac{\mathbb{E}[Y]}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

Isso conclui a prova do teorema. \square

Exemplo 10.4. Suponha que $\mathbb{E}[X] = 10$ e $\sigma(X) = 2$. Vamos encontrar uma estimativa da probabilidade de que $6 \leq X \leq 14$. Tomando $x = 4$ na Desigualdade de Chebyshev,

$$\mathbb{P}(6 \leq X \leq 14) \geq \mathbb{P}(6 < X < 14) = 1 - \mathbb{P}(|X - \mathbb{E}[X]| \geq 4) \geq 1 - \frac{2^2}{4^2} = \frac{3}{4}.$$

10.3 Prova da lei das médias

Lembrando que X_1, X_2, X_3, \dots são variáveis aleatórias discretas de variância quadrática integrável independentes duas a duas com a mesma média μ e mesma variância σ^2 .

Queremos mostrar que, para todo $a > 0$ e $n \in \mathbb{N}$,

$$\mathbb{P}\left(\mu - a \leq \frac{X_1 + \dots + X_n}{n} \leq \mu + a\right) \geq 1 - \frac{\sigma^2}{a^2 n}.$$

Usando a linearidade da expectativa e o fato de que todos os X_k têm a mesma média μ ,

$$\mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n} \mathbb{E}[X_1 + \dots + X_n] = \frac{1}{n} \cdot (\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]) = \mu.$$

Usando o Corolário 9.9 e o fato de que todos os X_k têm a mesma variância σ^2 ,

$$\begin{aligned} \text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) &= \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) \\ &= \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_n)) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2} \cdot n\sigma^2 \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Usando a Desigualdade de Chebyshev, obtemos:

$$\begin{aligned}
\mathbb{P}\left(\mu - a \leq \frac{X_1 + \dots + X_n}{n} \leq \mu + a\right) &= 1 - \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > a\right) \\
&\geq 1 - \frac{\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right)}{a^2} \\
&= 1 - \frac{\sigma^2}{a^2 \cdot n},
\end{aligned}$$

o que conclui a prova.

11 Coeficiente de Correlação

A covariância é uma quantidade útil que descreve como duas variáveis aleatórias variam juntas. No entanto, ela tem uma desvantagem: não é invariante à escala. Para explicar o que isso significa, suponha que X e Y sejam duas variáveis aleatórias, ambas medindo comprimentos em metros. Vamos assumir que U e V fornecem as mesmas medidas que X e Y , respectivamente, mas em centímetros, ou seja, $U = 100X$ e $V = 100Y$. Então,

$$\text{Cov}(U, V) = \text{Cov}(100X, 100Y) = 100 \cdot 100 \cdot \text{Cov}(X, Y) = 10^4 \cdot \text{Cov}(X, Y).$$

Isso significa que alterar a escala também altera a covariância. Para obter uma quantidade invariante à escala, fazemos a seguinte definição.

Definição (Coeficiente de Correlação). Sejam X e Y variáveis aleatórias discretas quadrado-integráveis com variância positiva. O *coeficiente de correlação* entre X e Y é definido como

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Conforme prometido, o coeficiente de correlação não muda quando escalamos ou deslocamos as variáveis aleatórias.

Proposição. Sejam X e Y variáveis aleatórias quadrado-integráveis com variância positiva. Para qualquer $a, b, c, d \in \mathbb{R}$ com $a, c > 0$, temos

$$\rho(aX + b, cY + d) = \rho(X, Y).$$

Demonstração. Primeiro, observamos que a covariância entre uma constante e qualquer outra variável aleatória é igual a zero. De fato,

$$\text{Cov}(b, Y) = \mathbb{E}[bY] - \mathbb{E}[b] \cdot \mathbb{E}[Y] = b \cdot \mathbb{E}[Y] - b \cdot \mathbb{E}[Y] = 0.$$

Portanto,

$$\text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y).$$

Substituindo isso na fórmula do coeficiente de correlação,

$$\begin{aligned}\rho(aX + b, cY + d) &= \frac{\text{Cov}(aX + b, cY + d)}{\sigma(aX + b) \cdot \sigma(cY + d)} \\ &= \frac{ac \cdot \text{Cov}(X, Y)}{a \cdot \sigma(X) \cdot c \cdot \sigma(Y)} \\ &= \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} = \rho(X, Y). \quad \square\end{aligned}$$

Além disso, como $\sigma(-Y) = \sigma(Y)$ e $\text{Cov}(X, Y) = \text{Cov}(Y, X)$, o coeficiente de correlação também satisfaz.

$$\rho(X, Y) = \rho(Y, X) \text{ e } \rho(X, -Y) = -\rho(X, Y).$$

A próxima proposição descreve ainda mais em que sentido o coeficiente de correlação $\rho(X, Y)$ é um índice adimensional que quantifica o quão bem X e Y estão alinhados, veja a Figura 11.1 para uma descrição visual.

Proposição. *Sejam X e Y variáveis aleatórias quadrado-integráveis com variância positiva. Então*

$$-1 \leq \rho(X, Y) \leq 1.$$

Nos casos extremos, $\rho(X, X) = 1$ e $\rho(X, -X) = -1$.

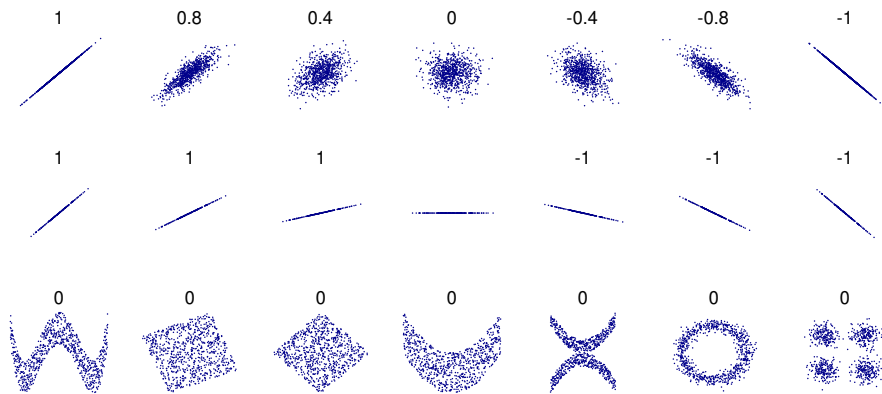


Figura 11.1: Ilustração de $\rho(X, Y)$ assumindo que o par (X, Y) tem a mesma probabilidade de estar em cada ponto na nuvem representada. (retirado da Wikipedia)

Demonstração. Observe que

$$\left(\frac{X - \mathbb{E}[X]}{\sigma(X)} - \frac{Y - \mathbb{E}[Y]}{\sigma(Y)} \right)^2 \geq 0$$

Tomando a expectativa e expandindo, obtemos

$$\frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\sigma^2(X)} + \frac{\mathbb{E}[(Y - \mathbb{E}[Y])^2]}{\sigma^2(Y)} - 2 \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sigma(X)\sigma(Y)} \geq 0,$$

o que significa

$$2\rho(X, Y) \leq 2,$$

logo $\rho(X, Y) \leq 1$. O mesmo argumento com $-Y$ no lugar de Y nos dá $\rho(X, Y) = -\rho(X, -Y) \leq 1$, portanto $\rho(X, Y) \geq -1$. Finalmente,

$$\rho(X, X) = \frac{\text{Cov}(X, X)}{\sigma(X) \cdot \sigma(X)} = 1$$

e $\rho(X, -X) = -\rho(X, X) = -1$. □

12 Teorema do Limite Central

Teorema 12.1 (Teorema do Limite Central). *Sejam X_1, X_2, X_3, \dots variáveis aleatórias independentes entre si, quadrado-integráveis, com a mesma distribuição. Denote sua média por μ e variância por $\sigma^2 > 0$. Então, para todo $a < b$*

$$\mathbb{P}\left(a \leq \frac{X_1 + \dots + X_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b\right) \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

A aproximação “ \approx ” significa que a probabilidade se aproxima o quanto desejarmos da integral (como mostrado na Figura 12.1) se escolhermos n suficientemente grande.

Esse fenômeno notável está no cerne da estatística e da maioria das ciências naturais. Ele afirma que, *independentemente da distribuição de X* , se adicionarmos um número suficiente de amostras de X , só veremos sua média μ e variância σ^2 .

Não demonstraremos o Teorema do Limite Central neste módulo, pois são necessárias ferramentas mais avançadas.

No caso de $X \sim \text{Bernoulli}(\frac{1}{2})$, o que corresponde a lançar uma moeda justa, podemos visualizar como a distribuição de $X_1 + \dots + X_n$ se aproxima da função $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, como ilustrado na Figura 12.2.

Exemplo 12.2. Ao contar os votos em uma eleição muito disputada, 25.301 votos já foram contados: 12.636 para o Candidato A e 12.665 para o Candidato B. Ainda faltam 400 votos para serem contados. Qual é a probabilidade de que o Candidato B vença a eleição?

Supondo que cada voto seja como uma moeda justa, a pergunta que estamos

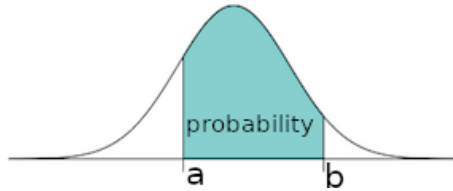


Figura 12.1: Gráfico de $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ e a probabilidade de que $Z \in [a, b]$ representada pela área esverdeada entre os pontos a e b .

fazendo é:

$$\mathbb{P}(X_1 + \cdots + X_{400} \geq 215)$$

onde X_1, \dots, X_n são independentes e têm distribuição Bernoulli($\frac{1}{2}$). Pela simetria, isso é o mesmo que:

$$\mathbb{P}(X_1 + \cdots + X_{400} \leq 185)$$

e, portanto, isso é igual a:

$$\frac{1}{2} \cdot \left[1 - \mathbb{P}(185 < X_1 + \cdots + X_{400} < 215) \right]$$

Dado que $\mu = \frac{1}{2}$ e $\sigma = \frac{1}{2}$, reescrevemos convenientemente o evento como:

$$\frac{1}{2} - \frac{1}{2} \cdot \mathbb{P}\left(-1.5 < \frac{X_1 + \cdots + X_{400} - 400 \cdot \mu}{\sigma\sqrt{400}} < 1.5\right)$$

e, usando o Teorema do Limite Central, aproximamos por:

$$\frac{1}{2} - \frac{1}{2} \int_{-1.5}^{1.5} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0.07$$

Você não deve tentar calcular esta integral em casa, a única maneira de obter esse valor é consultando uma tabela, veremos mais sobre isso posteriormente. Portanto, a resposta é 0.07, ou 7

Observe que fornecemos apenas uma resposta aproximada com uma figura significativa. Para obter mais precisão do que isso, seriam necessárias considerações mais cuidadosas e seriam o tema de módulos mais avançados.

Também podemos usar as "versões de um lado" do Teorema do Limite Central. Dessa forma, o exemplo anterior fica simplificado:

$$\begin{aligned} \mathbb{P}(X_1 + \cdots + X_{400} \geq 215) &= \mathbb{P}\left(\frac{X_1 + \cdots + X_{400} - 400 \cdot \mu}{\sigma\sqrt{400}} \geq 1.5\right) \\ &\approx \int_{1.5}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \end{aligned}$$

≈ 0.07 .

Se reescrevermos o Teorema do Limite Central como

$$\mathbb{P}\left(\mu + \frac{\sigma}{\sqrt{n}}a \leq \frac{X_1 + \cdots + X_n}{n} \leq \mu + \frac{\sigma}{\sqrt{n}}b\right) \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

obtemos uma boa descrição do comportamento estatístico da média observada $\frac{X_1 + \cdots + X_n}{n}$. Conforme previsto pela lei das médias, a média observada está concentrada em torno de μ , mas agora podemos dizer algo mais preciso. A média observada flutua como $\mu + \sigma \frac{1}{\sqrt{n}}Z$, onde Z é essa "coisa" descrita por

$$\mathbb{P}(a \leq Z \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Variáveis aleatórias descritas em termos de integrais são chamadas de *variáveis aleatórias contínuas*, que é o tema da próxima seção.

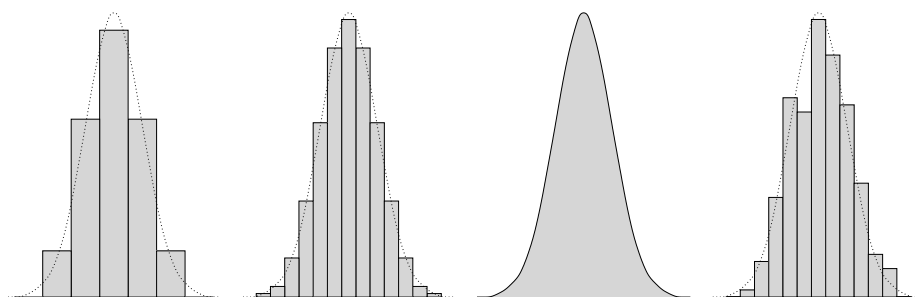


Figura 12.2: Função de probabilidade de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ para S_n com distribuições $\text{Binom}(4, \frac{1}{2})$ e $\text{Binom}(16, \frac{1}{2})$ para valores entre -3 e 3 . A área de cada retângulo é dada pela função de probabilidade. O terceiro gráfico é a função de densidade de uma normal padrão, assim como as linhas pontilhadas. O quarto gráfico representa as frequências relativas de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ para S_n com distribuição $\text{Binom}(16, \frac{1}{2})$, em um experimento real com 200 amostras.

13 Variáveis aleatórias contínuas

13.1 Função de densidade de probabilidade

Considere a variável aleatória X definida informalmente da seguinte maneira: "Seja X um número escolhido no intervalo $(0, 1)$ de forma uniforme ao acaso." Ao pensarmos com cuidado, percebemos que essa descrição é um pouco desconcertante. A palavra "uniforme" deveria significar que qualquer número $x \in (0, 1)$ tem igual probabilidade de ser escolhido, ou seja, $\mathbb{P}(X = x)$ deveria ser o mesmo para cada x . No entanto, existem infinitos $x \in (0, 1)$, o que forçaria $\mathbb{P}(X = x)$ a ser zero. Isso de fato acontecerá para todas as variáveis aleatórias com distribuições contínuas, que definimos agora.

Definição 13.1. Dizemos que uma variável aleatória X é *contínua*, com *função de densidade de probabilidade* $f_X : \mathbb{R} \rightarrow \mathbb{R}$, se

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

para todo $a < b \in \mathbb{R}$.

Uma densidade f_X especifica a "probabilidade por unidade de comprimento." É algo análogo à função de probabilidade de massa, mas não exatamente. A probabilidade de X estar em um pequeno intervalo de comprimento Δx é dada por $f_X(x)\Delta x$, em vez de $f_X(x)$, e f_X em si pode assumir valores muito grandes em pequenos intervalos. Portanto, é " $f_X(x)\Delta x$ " que é análogo a $p_X(x)$.

Uma função de densidade deve satisfazer necessariamente

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

13.2 Variáveis Uniformes

Sejam a e b em \mathbb{R} com $a < b$. Uma variável aleatória X tem a distribuição uniforme (contínua) em (a, b) se

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

Escrevemos $X \sim \mathcal{U}(a, b)$.

Observe que, para qualquer intervalo $[c, d] \subseteq [a, b]$, temos

$$\mathbb{P}(X \in [c, d]) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a},$$

ou seja, o intervalo $[c, d]$ tem uma probabilidade dada pela *proporção* do seu comprimento dentro do intervalo $[a, b]$. Por outro lado, se $c < a$ e $d \in [a, b]$, então

$$\mathbb{P}(X \in [c, d]) = \frac{d-a}{b-a}$$

porque a parte do intervalo $[c, d]$ que não se sobrepõe com $[a, b]$ não conta.

13.3 Distribuição Normal

Seja $\mu \in \mathbb{R}$ e $\sigma > 0$. Uma variável aleatória X tem distribuição normal (ou gaussiana) com parâmetros μ e σ^2 se tiver uma função de densidade de probabilidade dada por

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

para todo $x \in \mathbb{R}$. Escrevemos $X \sim \mathcal{N}(\mu, \sigma^2)$.

O parâmetro μ dá o centro da função de densidade, e o parâmetro σ^2 especifica a escala com que essa densidade está sendo alongada. Através de uma mudança de variáveis na integral, podemos ver que $X \sim \mathcal{N}(\mu, \sigma^2)$ é equivalente a $X = \mu + \sigma \cdot Z$ com $Z \sim \mathcal{N}(0, 1)$. De fato, definindo $Z = \frac{X-\mu}{\sigma}$,

$$\begin{aligned} \mathbb{P}(a \leq Z \leq b) &= \mathbb{P}(\mu + a\sigma \leq X \leq \mu + b\sigma) \\ &= \int_{\mu+a\sigma}^{\mu+b\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz. \end{aligned}$$

Observação 6. Não é fácil ver que $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. Ao substituir, obtemos

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \cdot e^{-u^2} du.$$

Portanto, é suficiente verificar que $\int_{-\infty}^{+\infty} e^{-x^2} = \sqrt{\pi}$. Na maioria dos livros didáticos, isso é feito usando coordenadas polares para integrais em \mathbb{R}^2 , mas não queremos usar esse método. Em vez disso, usamos um truque diferente: trocamos as integrais iteradas em

$$\int_0^{+\infty} \left(\int_0^{+\infty} ye^{-(1+x^2)y^2} dy \right) dx = \int_0^{+\infty} \left(\int_0^{+\infty} ye^{-x^2y^2} e^{-y^2} dx \right) dy,$$

o que podemos fazer porque o integrando é não-negativo. A primeira integral pode ser calculada como

$$\lim_{z \rightarrow +\infty} \int_0^z ye^{-(1+x^2)y^2} dy = \lim_{z \rightarrow +\infty} \frac{-1}{2(1+x^2)} \left[e^{-(1+x^2)y^2} \right]_0^z = \frac{1}{2(1+x^2)}$$

e

$$\int_0^{+\infty} \left(\int_0^{+\infty} ye^{-(1+x^2)y^2} dy \right) dx = \lim_{z \rightarrow +\infty} \int_0^z \frac{1}{2(1+x^2)} dx = \lim_{z \rightarrow +\infty} \frac{\arctan z}{2} = \frac{\pi}{4}.$$

A segunda integral pode ser reescrita como

$$\int_0^{+\infty} \left(\int_0^{+\infty} e^{-u^2} du \right) e^{-y^2} dy.$$

Desta forma, concluímos que $\int_0^{+\infty} e^{-x^2} = \frac{\sqrt{\pi}}{2}$. Por simetria, $\int_{-\infty}^{+\infty} e^{-x^2} = \sqrt{\pi}$, que era o que queríamos.

13.4 Tempo de Vida Exponencial

Seja $\lambda > 0$. Uma variável aleatória X tem distribuição exponencial com parâmetro λ se

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

Escrevemos $X \sim \text{Exp}(\lambda)$.

Observe que

$$\mathbb{P}(X > t) = e^{-\lambda t}.$$

A distribuição exponencial é comumente usada para modelar a vida útil de entidades que têm a propriedade de *falta de memória*, normalmente objetos inanimados que não sofrem efeitos de envelhecimento. Para explicar o que isso significa, vamos pensar nas lâmpadas. Suponhamos que a vida útil das lâmpadas de uma determinada marca tem uma distribuição Exponencial(λ) (vamos assumir que ligamos a luz e não a desligamos até a lâmpada queimar). Então, a propriedade de falta de memória significa que, independentemente de a lâmpada ter sido ativada recentemente ou ter estado ativa por um certo período de tempo, a distribuição do tempo restante de vida é a mesma. Matematicamente, isso é expresso pela seguinte identidade, que vale para todos $s, t \geq 0$:

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s).$$

13.5 Esperança

Definição 13.2 (Esperança). Seja X uma variável aleatória contínua com densidade f_X . Definimos a *esperança* de X , denotada por $\mathbb{E}[X]$, como o número real dado por

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx,$$

desde que esta integral convirja absolutamente. Se a integral convergir absolutamente, dizemos que X é *integrável*; caso contrário, $\mathbb{E}[X]$ não está definida.

Exemplo 13.3 (Uniforme). Se $X \sim \mathcal{U}[a, b]$, então

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

Isso significa que a esperança de uma variável aleatória com distribuição uniforme no intervalo $[a, b]$ é o ponto médio do intervalo.

Exemplo 13.4 (Exponencial). Se $X \sim \text{Exp}(\lambda)$, então, integrando por partes,

$$\mathbb{E}[X] = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \lim_{u \rightarrow +\infty} \left[-x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right]_0^u = \frac{1}{\lambda}.$$

Exemplo 13.5 (Normal). Suponha que $X \sim \mathcal{N}(0, 1)$, Então, substituindo $u = x^2/2$,

$$\int_0^{+\infty} x \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \lim_{z \rightarrow +\infty} \left[\frac{-e^{-x^2/2}}{\sqrt{2\pi}} \right]_0^z = \frac{1}{\sqrt{2\pi}}.$$

Por simetria,

$$\int_{-\infty}^0 x \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -\frac{1}{\sqrt{2\pi}}$$

e, portanto, $\mathbb{E}[X] = 0$.

Exemplo 13.6 (Cauchy). Suponha que X é uma variável aleatória com densidade

$$f_X(x) = \frac{1}{\pi \cdot (1 + x^2)}.$$

Então

$$\int_0^{+\infty} x f_X(x) dx = \int_0^{+\infty} \frac{x}{\pi \cdot (1 + x^2)} dx \geq \int_1^{+\infty} \frac{x}{\pi \cdot (1 + x^2)} dx,$$

e assim

$$\int_0^{+\infty} x f_X(x) dx \geq \int_1^{+\infty} \frac{1}{2\pi x} dx = \frac{1}{2\pi} \lim_{z \rightarrow \infty} \ln z = +\infty.$$

Neste caso, $\mathbb{E}[X]$ não é definida, apesar da simetria. Finalmente, temos um exemplo de uma variável aleatória que não é integrável!

13.6 Variância

Proposição 13.7. *Seja X uma variável aleatória contínua com densidade f_X . Seja $g : \mathbb{R} \rightarrow \mathbb{R}$ uma função piecewise contínua. Então*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

se a integral convergir absolutamente, e $\mathbb{E}[g(X)]$ é indefinida se não convergir absolutamente.

A analogia com a função de massa de probabilidade está resumida na Tabela 1.

Exemplo 13.8 (Uniforme). Se $X \sim \mathcal{U}[a, b]$, então

$$\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}.$$

Exemplo 13.9 (Exponencial). Se $X \sim \text{Exp}(\lambda)$, então, integrando por partes duas vezes,

$$\mathbb{E}[X^2] = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \lim_{z \rightarrow +\infty} \left[-x^2 e^{-\lambda x} - \frac{2x}{\lambda} e^{-\lambda x} - \frac{2}{\lambda^2} e^{-\lambda x} \right]_0^z = \frac{2}{\lambda^2}.$$

Exemplo 13.10 (Normal). Se $X \sim \mathcal{N}(0, 1)$, então, integrando por partes,

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ &= 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x \cdot (x e^{-x^2/2}) dx \\ &= \frac{2}{\sqrt{2\pi}} \lim_{u \rightarrow +\infty} \left[-x e^{-x^2/2} + \int_0^u e^{-x^2/2} dx \right]_0^u \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-x^2/2} dx = 1. \end{aligned}$$

Definição 13.11 (Quadrado-integrável). Como no caso de variáveis aleatórias discretas, dizemos que uma variável aleatória contínua é *quadrado-integrável* se X^2 for integrável.

Assim como no caso discreto, se uma variável aleatória for quadrado-integrável, ela será automaticamente integrável (porque $|x| \leq 1 + x^2$).

p_X	f_X
$p_X : \mathbb{R} \rightarrow \mathbb{R}$	$f_X : \mathbb{R} \rightarrow \mathbb{R}$
$p_X(x) \geq 0 \ \forall x \in \mathbb{R}$	$f_X(x) \geq 0 \ \forall x \in \mathbb{R}$
$\mathbb{P}(X = x) = p_X(x)$	$\mathbb{P}(X = x) = 0$
$\mathbb{P}(a \leq X \leq b) = \sum_{a \leq x \leq b} p_X(x)$	$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$
$p_X(x) = \mathbb{P}(X = x)$	definido implicitamente acima
$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x) dx = 1$
$p_X(x) \leq 1 \ \forall x$	$f_X(x)$ pode ser > 1
$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$

Tabela 1: Função de massa de probabilidade e função de densidade de probabilidade.

Definição 13.12 (Variância). Seja X uma variável aleatória contínua quadrado-integrável com densidade f_X e média $\mu = \mathbb{E}[X]$. Definimos a *variância* de X como

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

Como fizemos no caso de variáveis aleatórias discretas, podemos expandir a definição de variância para obter uma fórmula alternativa:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

que usamos nos exemplos a seguir.

Exemplo 13.13 (Uniforme). Se $X \sim \mathcal{U}[a, b]$, então

$$\text{Var}(X) = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b - a)^2}{12}.$$

Exemplo 13.14 (Exponencial). Se $X \sim \text{Exp}(\lambda)$, então

$$\text{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Exemplo 13.15 (Normal). Se $X \sim \mathcal{N}(0, 1)$, então

$$\text{Var}(X) = 1 - 0^2 = 1.$$

14 Uma única teoria para discretas e contínuas

14.1 Função de distribuição cumulativa

Uma maneira de especificar uma distribuição de probabilidade em \mathbb{R} é dizer o quanto de probabilidade está à esquerda de cada ponto x . Em termos de uma variável aleatória X com a distribuição dada, essa probabilidade é uma função de x .

Definição 14.1. Seja X uma variável aleatória. A *função de distribuição cumulativa* de X é a função $F_X : \mathbb{R} \rightarrow [0, \infty)$ definida por

$$F_X(x) = \mathbb{P}(X \leq x)$$

para todo $x \in \mathbb{R}$.

Outras probabilidades a partir de F_X

Agora, observamos que, embora $F_X(x)$ seja definido como $\mathbb{P}(X \leq x)$, é possível usar F_X para obter outras probabilidades envolvendo X . Fórmulas importantes são

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F_X(x)$$

e, para $x < y$,

$$\mathbb{P}(x < X \leq y) = \mathbb{P}(X \leq y) - \mathbb{P}(X \leq x) = F_X(y) - F_X(x).$$

Observe também que

$$\mathbb{P}(X = x) > 0 \quad \text{se e somente se} \quad F_X \text{ tem um salto em } x$$

e, nesse caso, o tamanho do salto é a probabilidade de $X = x$.

F_X **determina** \mathbb{P}_X

Proposição 14.2. Se X e Y são duas variáveis aleatórias com $F_X = F_Y$, então X e Y têm a mesma distribuição.

Essa proposição nos diz que a função de distribuição cumulativa realmente codifica a distribuição de uma variável aleatória (no sentido de que, dada a função de distribuição cumulativa, há apenas uma distribuição correspondente a ela).

Já vimos que F_X determina $\mathbb{P}_X(\{x\})$ para cada $x \in \mathbb{R}$ e $\mathbb{P}_X((a, b])$ para todo $a < b \in \mathbb{R}$.

Faltam-nos as ferramentas necessárias para provar que ela determina $\mathbb{P}_X(B)$ para todo $B \in \mathcal{B}$.

14.2 Casos discretos e contínuos

Para obter uma primeira ideia de como se parece uma função de distribuição cumulativa, consideremos o caso em que X é discreto e tem suporte discreto contido em \mathbb{N}_0 , de modo que

$$\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1.$$

Em seguida, observe primeiro que $F_X(x) = \mathbb{P}(X \leq x) = 0$ para todos os $x < 0$. A seguir, e para todo $x \in [0, 1)$,

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X < 0) + \mathbb{P}(X = 0) + \mathbb{P}(0 < X \leq x) = 0 + p_X(0) + 0 = p_X(0).$$

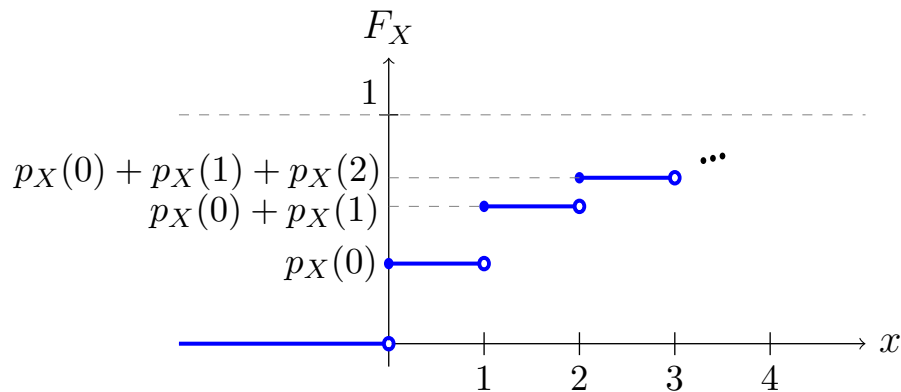


Figura 14.1: Função de distribuição cumulativa de uma variável aleatória discreta.

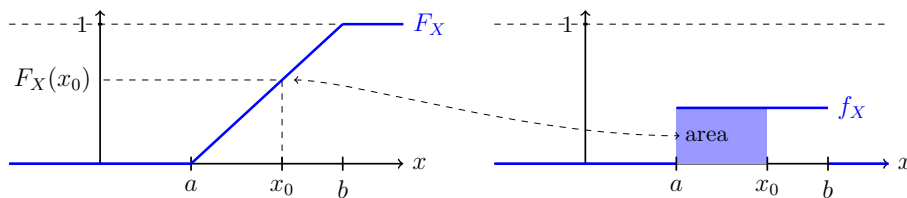


Figura 14.2: Função de distribuição cumulativa de uma variável aleatória uniforme.

Ao argumentar de maneira semelhante, concluímos que

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0, \\ p_X(0) & \text{se } x \in [0, 1), \\ p_X(0) + p_X(1) & \text{se } x \in [1, 2), \\ p_X(0) + p_X(1) + p_X(2) & \text{se } x \in [2, 3) \\ \dots & \end{cases}$$

O gráfico de F_X se parece com o da Figura 14.1.

Se X for contínuo, então

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x]) = \int_{-\infty}^x f_X(y) dy.$$

O teorema fundamental do Cálculo implica então que, nos pontos onde f_X é

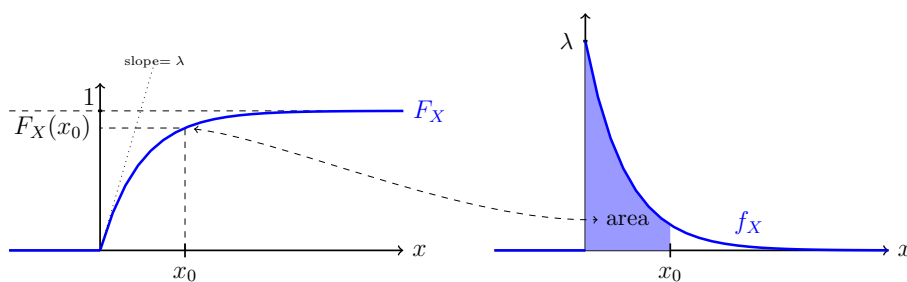


Figura 14.3: Função de distribuição cumulativa de uma variável aleatória exponencial.

contínuo,

$$f_X(x) = \frac{d}{dx}F'_X(x),$$

ou seja, a função de distribuição cumulativa é diferenciável e sua derivada é a função de densidade de probabilidade. O gráfico de F_X se parece com o das Figuras 14.2 e 14.3.

14.3 Esperança e variância

É possível fornecer uma definição unificada de esperança de uma variável aleatória, sem assumir que ela seja discreta ou que tenha uma densidade. Há uma fórmula mágica usando F_X que funciona simultaneamente para qualquer tipo de variável aleatória. Não vamos nos preocupar em fornecer tal fórmula, mas é importante ter em mente que a esperança é algo que pode ser definido para qualquer variável aleatória limitada (e, desde que algumas somas ou integrais sejam convergentes, também pode ser definida para variáveis aleatórias ilimitadas). Novamente, dizemos que X é *integrável* se $\mathbb{E}[X]$ for definido e finito.

Essa definição geral de esperança ainda satisfaz as três propriedades:

- Unitária: $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$,
- Monótona: Se $0 \leq Z \leq X$ para todo $\omega \in \Omega$, então $0 \leq \mathbb{E}[Z] \leq \mathbb{E}[X]$,
- Linear: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

desde que X e Y sejam integráveis.

Não provaremos essas propriedades. Claro, não poderíamos possivelmente prová-las, pois nem mesmo fornecemos a definição geral de esperança. Mas mesmo que tivéssemos escrito a fórmula, com as ferramentas atuais não seríamos capazes de provar que a esperança é linear em geral. A ideia da prova é a seguinte: quaisquer variáveis aleatórias X e Y podem ser aproximadas por variáveis aleatórias discretas X' e Y' e, uma vez que $\mathbb{E}[X' + Y'] = \mathbb{E}[X'] + \mathbb{E}[Y']$, concluímos que $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Mais uma vez, e X é *quadraticamente integrável* se $\mathbb{E}[X^2]$ for finito, e observe que se X é quadraticamente integrável, então ele é automaticamente integrável (porque $|x| \leq 1 + x^2$).

Definição 14.3 (Variância). A *variância* de uma variável aleatória quadrática-

mente integrável X é definida como

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

Observamos que a desigualdade de Chebyshev é válida para qualquer variável aleatória quadraticamente integrável. De fato, na prova dada na Seção 10.2, usamos apenas as propriedades de esperança mencionadas acima e nada mais.

Definição 14.4 (Covariância). Também definimos a *covariância* de duas variáveis aleatórias quadraticamente integráveis como

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

e dizemos que elas são *não correlacionadas* se sua covariância for zero.

Observe que a covariância possui todas as propriedades mencionadas na Seção 9.2. De fato, a prova dessas propriedades usou apenas as três propriedades de esperança mencionadas acima e nada mais. Em particular, o Corolário 9.8 vale em geral, ou seja,

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

desde que X_1, \dots, X_n sejam não correlacionados.

A independência é discutida na próxima seção.

15 Distribuições Conjuntas e Independência

15.1 Função de Densidade Conjunta

Agora explicaremos o que significa para duas variáveis aleatórias seguirem uma distribuição conjunta contínua.

Definição 15.1 (Densidade Conjunta). Duas variáveis aleatórias X e Y definidas no mesmo espaço de probabilidade são chamadas de *conjuntamente contínuas* com uma *função de densidade conjunta* $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ se

$$\mathbb{P}(a_1 \leq X \leq a_2, b_1 \leq Y \leq b_2) = \int_{a_1}^{a_2} \left(\int_{b_1}^{b_2} f_{X,Y}(x, y) dy \right) dx$$

para todo $a_1 < a_2$ e $b_1 < b_2$.

As *funções de densidade marginal* são dadas por:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

Em resumo, densidades marginais são obtidas da densidade conjunta "integrando" a outra variável.

Exemplo 15.2. Considere o quadrado $Q = \{(x, y) \in \mathbb{R}^2 : |x| + |y| < 1\}$, e suponha que a função de densidade conjunta de X e Y seja $f_{X,Y}(x, y) = \frac{1}{2} \mathbb{1}_Q(x, y)$. Podemos determinar a função de densidade marginal de X integrando:

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{2} \mathbb{1}_Q(x, y) dy = (1 - |x|) \mathbb{1}_{[-1,1]}(x).$$

Se integrarmos em relação a x , obtemos $f_Y(y) = (1 - |y|) \mathbb{1}_{[-1,1]}(y)$.

15.2 Função de Distribuição Conjunta

Definição 15.3. Sejam X, Y variáveis aleatórias. A *função de distribuição conjunta* de (X, Y) é a função $F_{X,Y} : \mathbb{R}^2 \rightarrow [0, 1]$ dada por

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

para todo $x, y \in \mathbb{R}$.

Para variáveis discretas, essa definição se reduz a

$$F_{X,Y}(x,y) = \sum_{s \leq x} \sum_{t \leq y} p_{X,Y}(s,t)$$

para todo $x, y \in \mathbb{R}$. Para variáveis aleatórias conjuntamente contínuas, a função de distribuição conjunta é dada por

$$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s,t) \, dt \, ds.$$

De qualquer forma, é possível recuperar a função de massa de probabilidade conjunta e a função de densidade de probabilidade conjunta a partir da função de distribuição conjunta, mas não entraremos em detalhes sobre isso.

15.3 Independência

Definição 15.4 (Independência de Duas Variáveis Aleatórias). Dadas duas variáveis aleatórias X e Y , dizemos que X e Y são *independentes* se

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

para todos os conjuntos $A, B \in \mathcal{B}$.

Para variáveis aleatórias discretas, essa definição é equivalente a

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

para todos os valores $x, y \in \mathbb{R}$, conforme explicado na Seção 7.3.

Surpreendentemente, a função de distribuição conjunta é capaz de capturar se duas variáveis aleatórias são independentes ou não.

Proposição 15.5. *Duas variáveis aleatórias X e Y são independentes se e somente se*

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

para todo $x, y \in \mathbb{R}$.

Fragmento da prova. Para $A = (a, b]$ e $B = (c, d]$, expandimos para obter

$$\begin{aligned}
\mathbb{P}(a < X \leq b, c < Y \leq d) &= \mathbb{P}(a < X \leq b, Y \leq d) - \mathbb{P}(a < X \leq b, Y \leq c) \\
&= \mathbb{P}(X \leq b, Y \leq d) - \mathbb{P}(X \leq a, Y \leq d) - \mathbb{P}(a < X \leq b, Y \leq c) \\
&= F_{X,Y}(b, d) - F_{X,Y}(a, d) - [F_{X,Y}(b, c) - F_{X,Y}(a, c)] \\
&= F_X(b)F_Y(d) - F_X(a)F_Y(d) - [F_X(b)F_Y(c) - F_X(a)F_Y(c)] \\
&= [F_X(b) - F_X(a)] \cdot [F_Y(d) - F_Y(c)] \\
&= \mathbb{P}(a < X \leq b)\mathbb{P}(c < Y \leq d).
\end{aligned}$$

Portanto, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ quando tanto A quanto B são compostos por uma união de intervalos finitos desse tipo (abertos à esquerda e fechados à direita). O caso de conjuntos mais gerais A e B requer ferramentas que atualmente não possuímos. \square

Para variáveis aleatórias conjuntamente contínuas, a situação é um pouco mais complicada. Suponha que X e Y tenham f_X e f_Y como funções de densidade. Então, elas são independentes se e somente se forem conjuntamente contínuas com uma função de densidade conjunta dada por

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

para todo $x, y \in \mathbb{R}$. No entanto, se nos for fornecida uma função de densidade conjunta $f_{X,Y}$ e desejarmos mostrar que X e Y *não* são independentes, não basta encontrar um único ponto (x, y) tal que $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$. Precisamos verificar que $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$ para todos os $x \in [a, b]$ e todos os $y \in [c, d]$ para alguns intervalos não degenerados $[a, b]$ e $[c, d]$. Isso ocorre porque, ao contrário da função de massa de probabilidade, as funções de densidade de probabilidade não são únicas, e podemos modificá-las em um único ponto, fazendo com que a identidade mostrada acima deixe de ser válida.

Teorema 15.6. *Se X e Y são variáveis aleatórias independentes integráveis, então XY é integrável e*

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

Em particular, variáveis aleatórias independentes quadrado-integráveis são não correlacionadas.

Nós fornecemos uma prova assumindo que X e Y são discretas. Uma prova que funcione no caso geral requer ferramentas que atualmente não temos. Quanto à linearidade da esperança, ela se baseia no fato de que qualquer variável aleatória pode ser aproximada por variáveis aleatórias discretas, reduzindo o problema ao caso que já conhecemos.

Definição 15.7 (Independência par a par). Dizemos que uma coleção de variáveis aleatórias X_1, X_2, X_3, \dots é *independente par a par* se X_j e X_k forem independentes para todo $j \neq k$.

Definição 15.8 (Independência mútua). Dizemos que uma coleção de variáveis aleatórias discretas X_1, X_2, X_3, \dots é *mutuamente independentes* se, para todo k e todo $A_1, \dots, A_k \in \mathcal{B}$, tivermos

$$\mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_k \in A_k).$$

Existem condições análogas para independência mútua quando as variáveis aleatórias são contínuas ou discretas, e também uma condição equivalente em termos de funções de distribuição cumulativa conjunta. Mas não entraremos em detalhes sobre isso.

15.4 Covariância e a lei das médias

Como mencionado na seção anterior, se X_1, \dots, X_n são variáveis aleatórias não correlacionadas, então

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n).$$

Usando isso e a desigualdade de Chebyshev, podemos novamente provar a lei das médias para qualquer sequência de variáveis aleatórias não correlacionadas com a mesma média μ e a mesma variância σ^2 , sem assumir que sejam discretas. A prova é idêntica à vista na Seção 10.3.

O Teorema do Limite Central também se aplica a qualquer sequência de variáveis aleatórias quadrado-integráveis mutuamente independentes com a mesma distribuição, sem assumir que sejam discretas.

16 Somas de variáveis aleatórias independentes

Somas de variáveis aleatórias independentes surgem em muitos contextos diferentes. Dadas duas variáveis aleatórias independentes X e Y , qual é a distribuição de $X + Y$?

Se X e Y são ambas discretas, $X + Y$ é discreta e sua função de massa de probabilidade pode ser calculada usando a Lei da Probabilidade Total:

$$\begin{aligned} p_{X+Y}(z) &= \sum_x \mathbb{P}(X = x, Y = z - x) = \sum_x \mathbb{P}(X = x) \mathbb{P}(Y = z - x) \\ &= \sum_x p_X(x) p_Y(z - x). \end{aligned}$$

Exemplo 16.1. Suponha que $X \sim \text{Binom}(n, p)$ e $Y \sim \text{Binom}(m, p)$. A função de massa de probabilidade de $X + Y$ pode ser obtida da seguinte forma:

$$\begin{aligned} p_{X+Y}(k) &= \sum_{j=0}^{\infty} \mathbb{P}(X = j) \mathbb{P}(Y = k - j) \\ &= \sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-k+j} \\ &= p^k (1-p)^{m+n-k} \sum_{j=0}^k \binom{n}{j} \binom{m}{k-j} \\ &= \binom{n+m}{k} p^k (1-p)^{m+n-k}. \end{aligned}$$

Quando as variáveis X e Y são independentes e têm densidades f_X e f_Y , temos a relação análoga a seguir

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx.$$

Exemplo 16.2 (Exponenciais e Gama). Sejam X e Y independentes, ambas com a distribuição Exponencial com parâmetro $\lambda > 0$, ou seja,

$$f_X(x) = f_Y(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0; \\ 0 & \text{caso contrário.} \end{cases}$$

Seja $Z := X + Y$. Queremos calcular

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) \, dx.$$

Agora, observe que o produto dentro da integral é igual a zero quando $x < 0$ (pois $f_X(x) = 0$ nesse caso) e quando $x > z$ (pois $f_Y(z-x) = 0$ nesse caso). A integral é então igual a

$$\int_0^z \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)} \, dx = \lambda^2 \cdot \int_0^z e^{-\lambda z} \, dx = \lambda^2 \cdot z \cdot e^{-\lambda z}.$$

A distribuição acima corresponde a uma distribuição Gama com parâmetros 2 e λ . Em geral, Z tem distribuição Gama com parâmetros n e λ se a sua densidade for dada por $f_Z(z) = \frac{\lambda^n}{(n-1)!} \cdot z^{n-1} \cdot e^{-\lambda z}$ para $z \geq 0$.

O caso em que X e Y são normais é tão importante que o apresentamos como uma proposição.

Proposição 16.3. *Se $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ são independentes. Então $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Demonstração. Como $X_1 - \mu_1 \sim \mathcal{N}(0, \sigma_1^2)$ e $X_2 - \mu_2 \sim \mathcal{N}(0, \sigma_2^2)$, podemos supor que $\mu_1 = \mu_2 = 0$. Após manipulações algébricas longas e trabalhosas, é possível obter

$$f_{X+Y}(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} e^{-\frac{(z-x)^2}{2\sigma_2^2}} e^{-\frac{x^2}{2\sigma_1^2}} \, dx = \dots = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \cdot e^{-\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)}}.$$

Portanto, f_{X+Y} é a densidade correspondente à distribuição $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$, que era o que queríamos mostrar. \square

17 Momentos e funções geradoras de momentos

Definição 17.1. Dada uma variável aleatória X , definimos a *função geradora de momentos de X* como a função M_X dada por

$$M_X(t) = \mathbb{E}[e^{tX}]$$

para os valores de t para os quais e^{tX} é integrável.

Exemplo 17.2 (Geométrica). Se $X \sim \text{Geom}(p)$, então

$$M_X(t) = \sum_{n=1}^{\infty} e^{tn} p(1-p)^{n-1} = \begin{cases} \frac{p}{e^{-t} + p - 1}, & t < \ln \frac{1}{1-p}, \\ +\infty, & t \geq \ln \frac{1}{1-p}. \end{cases}$$

Exemplo 17.3 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então

$$M_X(t) = \sum_{n=0}^{\infty} e^{tn} \frac{e^{-\lambda} \lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

Exemplo 17.4 (Normal). Seja X uma variável aleatória normal com parâmetros μ e σ^2 , ou seja,

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

A função geradora de momentos de X pode ser calculada da seguinte forma:

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x) \, dx \\ &= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp \left\{ -\frac{1}{2\sigma^2} [(x-\mu)^2 - 2\sigma^2 tx] \right\} \, dx \quad (17.5) \end{aligned}$$

Agora, completamos o quadrado:

$$\begin{aligned} (x-\mu)^2 - 2\sigma^2 tx &= x^2 - 2(\mu + \sigma^2 t) \cdot x + \mu^2 \\ &= x^2 - 2(\mu + \sigma^2 t) \cdot x + \mu^2 \pm (2\mu\sigma^2 t + \sigma^4 t^2) \\ &= (x - \mu - \sigma^2 t)^2 - 2\mu\sigma^2 t - \sigma^4 t^2. \end{aligned}$$

Isso dá

$$-\frac{1}{2\sigma^2}[(x-\mu)^2 - 2\sigma^2 tx] = -\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} + t\mu + \frac{\sigma^2 t^2}{2}.$$

A integral em (17.5) então se torna

$$\exp\left\{t\mu + \frac{\sigma^2 t^2}{2}\right\} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right\} dx.$$

Agora, observe que a função sendo integrada é a função de densidade de probabilidade de $\mathcal{N}(\mu + \sigma^2 t, \sigma^2)$, então a integral é igual a 1. Em conclusão,

$$M_X(t) = \exp\left\{t\mu + \frac{\sigma^2 t^2}{2}\right\}, \quad t \in \mathbb{R}.$$

Definição 17.6 (Momentos). Definimos o k -ésimo momento de uma variável aleatória X como $\mathbb{E}[X^k]$ se X^k for integrável.

O nome "função geradora de momentos" vem do seguinte fato.

Proposição 17.7. Se $M_X(t)$ é definido em $(-a, a)$ para algum $a > 0$, então X tem todos os momentos e eles são dados por

$$\mathbb{E}[X^k] = M_X^{(k)}(0),$$

onde $M_X^{(k)}$ denota a k -ésima derivada da função M_X .

Não temos as ferramentas para provar esta proposição, mas se estivermos dispostos a ser atrevidos, podemos fazer:

$$\frac{d^k}{dt^k} M_X(t) = \frac{d^k}{dt^k} \mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d^k}{dt^k} e^{tX}\right] = \mathbb{E}[X^k e^{tX}]$$

e, avaliando em $t = 0$, obtemos a proposição.

Exemplo 17.8 (Geométrica). Se $X \sim \text{Geom}(p)$, então

$$\mathbb{E}[X] = M_X'(0) = \frac{1}{p}, \quad \mathbb{E}[X^2] = M_X''(0) = \frac{2}{p^2} - \frac{1}{p}, \quad \text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1-p}{p^2}.$$

Exemplo 17.9 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então

$$\mathbb{E}[X] = M_X'(0) = \lambda, \quad \mathbb{E}[X^2] = M_X''(0) = \lambda^2 + \lambda, \quad \text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda.$$

Exemplo 17.10 (Normal). Seja $X \sim \mathcal{N}(\mu, \sigma^2)$. Vamos calcular sua média (ou seja, sua esperança) e variância. No exemplo anterior, provamos que $M_X(t)$ está definido para todos $t \in \mathbb{R}$, com

$$M_X(t) = \exp \left\{ \frac{t^2 \sigma^2}{2} + \mu t \right\}.$$

Em seguida, usamos o teorema acima para calcular

$$\mathbb{E}[X] = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = \left[(t\sigma^2 + \mu) \cdot \exp \left\{ \frac{t^2 \sigma^2}{2} + \mu t \right\} \right]_{t=0} = \mu$$

e

$$\mathbb{E}[X^2] = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} = \dots = \sigma^2 + \mu^2.$$

Portanto,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Esta é a razão pela qual X é dito ser uma variável aleatória normal com média μ e variância σ^2 .

Proposição 17.11. *Para todos $a, b \in \mathbb{R}$,*

$$M_{aX+b}(t) = e^{tb} \cdot M_X(at)$$

para todo t no qual as funções geradoras de momentos estão definidas.

Demonstração. Nós calculamos

$$M_{aX+b}(t) = \mathbb{E}[e^{t(aX+b)}] = e^{tb} \cdot \mathbb{E}[e^{t(aX)}] = e^{tb} \cdot M_X(at). \quad \square$$

Proposição 17.12. *Quando X e Y são independentes,*

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

para todo t no qual as funções geradoras de momentos estão definidas.

Demonstração. Se X e Y são independentes, então e^{tX} e e^{tY} também são.

Portanto,

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} \cdot e^{tY}] = \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] = M_X(t) \cdot M_Y(t). \quad \square$$

Exemplo 17.13 (Soma de variáveis de Poisson independentes). Sejam $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ independentes. Então

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)} = M_Z(t),$$

onde $Z \sim \text{Poisson}(\lambda + \mu)$. Isso implica que $X + Y \sim \text{Poisson}(\lambda + \mu)$?

O exemplo acima nos faz pensar se conhecer a função geradora de momentos de uma variável aleatória nos diz qual é a distribuição da variável aleatória. E, de fato, este é o caso.

Teorema 17.14 (A função geradora de momentos determina a distribuição). *Dadas duas variáveis aleatórias X e Y , se existe $a > 0$ tal que $M_X(t)$ e $M_Y(t)$ são finitas e coincidem para todo $t \in [-a, a]$, então X e Y têm a mesma distribuição.*

Também omitimos a prova deste teorema, observando que ela requer ferramentas ainda mais difíceis de construir do que outras provas omitidas nestas notas.

Exemplo 17.15 (Soma de variáveis de Poisson independentes). Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ são independentes, então $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Exemplo 17.16 (Soma de variáveis normais independentes). Sejam X e Y duas variáveis normais independentes, com médias μ_X e μ_Y e variâncias σ_X^2 e σ_Y^2 , respectivamente. Também sejam a e $b \in \mathbb{R}$ com $a \neq 0$. Vamos determinar as distribuições de $aX + b$ e de $X + Y$. Na última aula, mostramos que

$$M_X(t) = \exp \left\{ \frac{t^2 \sigma_X^2}{2} + t \mu_X \right\}, \quad M_Y(t) = \exp \left\{ \frac{t^2 \sigma_Y^2}{2} + t \mu_Y \right\}.$$

Temos que

$$M_{aX+b}(t) = e^{tb} \cdot M_X(at) = e^{tb} \cdot \exp \left\{ a \mu_X t + \frac{a^2 \sigma_X^2 t^2}{2} \right\} = \exp \left\{ (a \mu_X + b) t + \frac{a^2 \sigma_X^2 t^2}{2} \right\}.$$

Portanto, $aX + b$ tem a mesma função geradora de momentos que uma variável aleatória $\mathcal{N}(a\mu + b, a^2\sigma^2)$. Já que esta função geradora de momentos é definida em uma vizinhança da origem (na verdade, em toda a reta real), concluímos

que $aX + b \sim \mathcal{N}(a\mu_X + b, a^2\sigma_X^2)$.

Em seguida, como X e Y são independentes, temos

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) = \exp \left\{ \mu_X t + \frac{\sigma_X^2 t^2}{2} \right\} \cdot \exp \left\{ \mu_Y t + \frac{\sigma_Y^2 t^2}{2} \right\} \\ &= \exp \left\{ (\mu_X + \mu_Y)t + \frac{(\sigma_X^2 + \sigma_Y^2)t^2}{2} \right\}. \end{aligned}$$

Isso mostra que $X + Y$ tem a mesma função geradora de momentos que uma variável aleatória $\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Já que esta função geradora de momentos é definida em um intervalo aberto que contém zero, concluímos que $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

A Somas úteis

$$\begin{aligned}
 (a+b)^n &= \sum_{j=0}^n \binom{n}{j} a^j b^{n-j} & a, b \in \mathbb{R}, n \in \mathbb{N}_0 \\
 \sum_{n=0}^{\infty} x^n &= \frac{1}{1-x} & 0 < x < 1 \\
 \sum_{n=0}^{\infty} nx^{n-1} &= \frac{1}{(1-x)^2} & 0 < x < 1 \\
 \sum_{n=0}^{\infty} n(n-1)x^{n-2} &= \frac{2}{(1-x)^3} & 0 < x < 1 \\
 \sum_{n=0}^{\infty} n(n-1)(n-2)x^{n-3} &= \frac{3!}{(1-x)^4} & 0 < x < 1 \\
 \sum_{k=0}^{\infty} \frac{x^k}{k!} &= e^x & x \in \mathbb{R} \\
 \sum_{k=1}^n k &= \frac{n(n+1)}{2} & n \in \mathbb{N} \\
 \sum_{k=1}^n k^2 &= \frac{n(n+1)(2n+1)}{6} & n \in \mathbb{N}
 \end{aligned}$$

As cinco primeiras são: teorema binomial, série geométrica, derivada da série geométrica, segunda derivada da série geométrica, terceira derivada da série geométrica. Acontece que é legítimo diferenciar uma série da forma $\sum_n a_n x^n$ termo a termo, mas não estamos preocupados com os detalhes de por que isso é verdade.

A quinta é a chamada "série de Taylor" da função exponencial. Para verificar se a fórmula faz sentido, observe que ambos os lados resultam em 1 para $x = 0$ e cada lado é igual à sua própria derivada. Esses dois fatos implicam que ambos os lados são iguais para todos os valores de x , mas não estamos preocupados com os detalhes disso.

As duas últimas fórmulas, uma vez escritas, podem ser provadas por indução (suponha que sejam corretas para um certo valor de n , mostre que são corretas para $n + 1$). Se você está curioso sobre como essas fórmulas surgiram, elas podem ser derivadas fazendo inicialmente um palpite educado de que devem

ser representadas por polinômios um grau maior do que o termo somado e, em seguida, usando os dois ou três primeiros termos para escrever um sistema de equações para os coeficientes.

B Exponenciais superam polinômios

Para todo $x \geq 0$ e $n \in \mathbb{N}$,

$$e^x \geq 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

Demonstração. Para $n = 0$, já sabemos que $e^x \geq 1$. Para $n = 1$,

$$e^x = 1 + \int_0^x e^x \, dx \geq 1 + \int_0^x 1 \, dx = 1 + x$$

Para $n = 2$,

$$e^x = 1 + \int_0^x e^x \, dx \geq 1 + \int_0^x (1 + x) \, dx = 1 + x + \frac{x^2}{2}$$

Para $n = 3$,

$$e^x = 1 + \int_0^x e^x \, dx \geq 1 + \int_0^x (1 + x + \frac{x^2}{2}) \, dx = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!}.$$

O padrão é claro. □

Isso implica que

$$\frac{a_0 + a_1x + \cdots + a_nx^n}{e^{ax}}$$

tende a 0 à medida que $x \rightarrow +\infty$, para todo $a > 0$.

Demonstração. De fato, como $e^{ax} \geq \frac{a^{n+1}}{(n+1)!}x^{n+1}$, cada termo em

$$\frac{a_0}{e^{ax}} + \frac{a_1x}{e^{ax}} + \cdots + \frac{a_nx^n}{e^{ax}}$$

está se aproximando de zero à medida que x aumenta. □

Isso é útil ao calcular integrais impróprias que incluem polinômios e funções exponenciais.