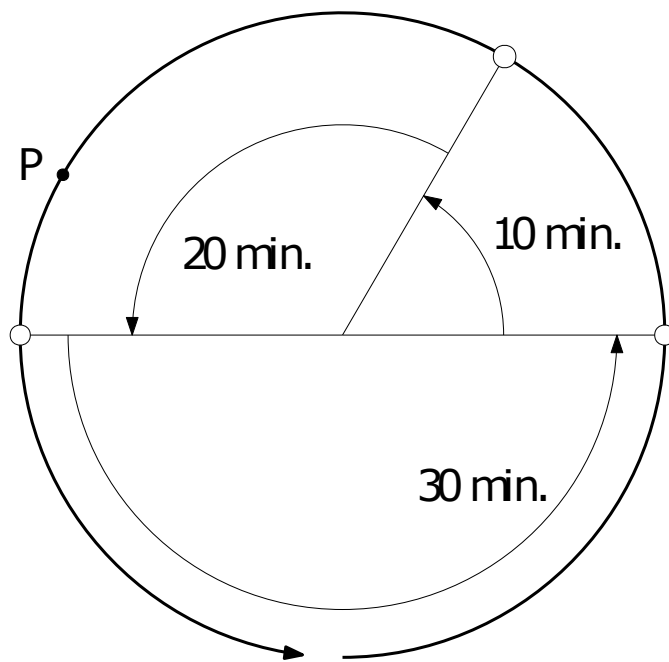


Probabilidades

Um curso básico de reoferecimento



○ Posição de um ônibus

Volta com 60 minutos

Alexandre Kawano & José Luis de Paiva

2023

Probabilidades

Um curso básico de reoferecimento

Alexandre Kawano

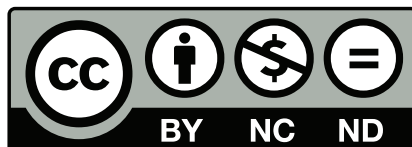
ORCID <https://orcid.org/0000-0002-2248-6422>

José Luis de Paiva

ORCID <https://orcid.org/0000-0001-9733-5744>

2023

“É permitida a reprodução parcial ou total desta obra, desde que citada a fonte e autoria, proibindo qualquer uso para fins comerciais.”



Este trabalho está licenciado com uma Licença Creative Commons - Atribuição-Não Comercial-SemDerivações 4.0 Internacional.

UNIVERSIDADE DE SÃO PAULO -USP

Reitor: Prof. Dr. Carlos Gilberto Carlotti Junior

Vice-reitora: Profa. Dra. Maria Arminda do Nascimento Arruda

ESCOLA POLITÉCNICA -EP

Diretor: Prof. Dr. Reinaldo Giudici

Vice-diretor: Prof. Dr. Sílvio Ikuyo Nabeta

Catálogo-na-publicação (CIP). Serviço de Biblioteca e Documentação
Escola Politécnica da Universidade de São Paulo. Maria Cristina Olaio
Villela CRB 1338/8^a

Kawano, Alexandre

Probabilidades: um curso básico de reoferecimento / A.Kawano,
J. L. Paiva. – São Paulo : Epusp, 2023.

1116 kb

ISBN 978-65-89190-25-7

DOI 10.11606/9786589190257

1. Probabilidades I. Paiva, José Luis de II.t.

Conteúdo

1	Introdução	3
2	Espaço Probabilístico	7
2.1	Introdução	7
2.2	Vídeo que acompanha este capítulo	8
2.3	Biografia de Kolmogorov	8
2.4	Conceitos preliminares	11
2.4.1	Conjuntos	11
2.4.2	Funções	13
2.4.3	Enumerabilidade	15
2.5	Espaços amostrais, eventos e probabilidade	15
2.5.1	Experimentos, eventos e espaço amostral	16
2.5.2	Exemplos de fixação	19
2.6	Problemas	21
3	Probabilidade condicionada	25
3.1	Introdução	25
3.2	Vídeo que acompanha este capítulo	25
3.3	Biografia de Thomas Bayes	26
3.4	Ideias preliminares	27
3.5	Probabilidade condicionada	30
3.5.1	Consequências da fórmula do condicionamento	31
3.5.2	Teorema da probabilidade total	34
3.6	Independência	34
3.7	Exemplos de fixação	35
3.8	Problemas	38

4	Variáveis aleatórias discretas	41
4.1	Introdução	41
4.2	Vídeo que acompanha este capítulo	42
4.3	Biografia de Pierre-Simon Laplace	42
4.4	Variáveis aleatórias discretas	45
4.4.1	A função variável aleatória	45
4.4.2	Experimentos e variáveis aleatórias	49
4.4.3	Independência de variáveis aleatórias. Independência sob condicionamento.	50
4.4.4	Função distribuição acumulada de uma variável aleatória discreta	53
4.4.5	Valor esperado de uma variável aleatória discreta	55
4.4.6	Variância de uma variável aleatória discreta	59
4.5	Modelos de variáveis aleatórias discretas	60
4.5.1	Binomial	60
4.5.2	Geométrica	61
4.5.3	Poisson	62
4.6	Condicionamento e modelos de V.A.s	63
4.7	Exemplos de fixação	66
4.8	Problemas	75
5	Variáveis aleatórias contínuas	79
5.1	Introdução	79
5.2	Vídeos que acompanham este capítulo	80
5.3	Atribuição de Probabilidades	80
5.3.1	Caso quando Ω é enumerável	81
5.3.2	Caso em que Ω não é enumerável. Necessidade de uma teoria específica para espaços amostrais não enumeráveis	85
5.4	Variáveis aleatórias contínuas	85
5.4.1	Construção do espaço de eventos E a partir de a imagem inversa de uma função definida em Ω . Abordagem para um curso de engenharia	87
5.4.2	Função probabilidade	89
5.5	Função de densidade de probabilidade multivariada	91

5.5.1	Marginalização	92
5.6	Variáveis aleatórias independentes	92
5.7	Função densidade de probabilidade condicionada	92
5.8	Soma de variáveis aleatórias	94
5.9	Transformação de variáveis aleatórias	94
5.10	Valor esperado e variância	96
5.11	Modelos de variáveis aleatórias contínuas	97
5.11.1	Variável aleatória uniforme	97
5.11.2	Variável aleatória exponencial	99
5.11.3	Variável aleatória normal	101
5.12	Exemplos de fixação	102
5.13	Problemas	113
6	Teorema do Limite Central	117
6.1	Introdução	117
6.2	Vídeo que acompanha este capítulo	117
6.3	Contexto histórico – Biografia de Patnuty Chebyshev	118
6.4	Noções de convergência	119
6.5	Lei fraca dos grandes números	124
6.6	Teorema do Limite Central	126
6.7	Exemplos de fixação	126
6.8	Problemas	130

Prefácio

Este livro é resultado da conjunção de dois fatores. O primeiro é a eclosão da pandemia de Covid19 em 2020, que levou alunos e professores aos cursos na modalidade *on-line*. O segundo é que este o público para o qual este pequeno texto se volta é o dos alunos que fazem o curso de probabilidades para engenharia, oferecido no ciclo básico, pela segunda vez, em no formato de re-oferecimento.

A consequência do primeiro fator é, por recomendação de especialistas em educação, que os conteúdos tradicionais dos cursos sejam reduzidos, devido às limitações impostas pelo ambiente social na pandemia. Dessa forma, este texto foca apenas temas que julgamos centrais e essenciais. Em contrapartida, os assuntos tratados são vistos com uma profundidade maior que a usual.

O segundo fator, o fato que este livro tem por público alvo alunos que tiveram alguma dificuldade durante o curso regular, leva a uma abordagem da teoria de probabilidades de uma forma que reflete a convicção pessoal dos autores. A nosso ver, a dificuldade da teoria tem origem na grande quantidade de subentendidos e na falta de clareza e explicitação das ideias. De fato, virtualmente todos os paradoxos da teoria básica de probabilidades têm origem ou na má formulação do problema ou em alguma falha teórica. Um exemplo, que é explorado no capítulo de variáveis aleatórias contínuas, o do paradoxo de Bertrand proposto em 1889, portanto antes da formalização da teoria feita por Kolmogorov nos anos 1930, só foi possível ser um paradoxo porque a teoria não era bem estabelecida na época de seu enunciado. Outros paradoxos como o dos dois filhos, o das três portas, o do ponto de ônibus, etc. são todos resolúveis a partir do momento que a teoria fica esclarecida.

O produto dos dois fatores mencionados nos levaram a escrever este livro com o objetivo de explicitar ao máximo os pontos básicos da teoria, nos limitando voluntariamente a temas dentro de um domínio estreito. Remetemos os leitores que buscam explorar a vastidão da área de Probabilidades

para Engenharia a textos voltados para o grande público.

Com o objetivo de fixar bem o conteúdo do curso, vemos todos os conceitos duas vezes, uma no contexto de variáveis aleatórias discretas e outra no das variáveis aleatórias contínuas. Assim, os estudantes tem oportunidade de serem expostos à teoria em dobro. Na segunda vez, é feito ainda o contraste com o que foi visto anteriormente.

Uma dificuldade que enfrentamos, mas que muitos autores evitam, é tratar a definição da função probabilidade quando as variáveis aleatórias envolvidas são contínuas. Em geral, isso é feito de forma bem feita quando se usa a teoria da medida, ou pelo menos quando se tem à disposição a integral de Lebesgue. Nos livros voltados para a Engenharia, não se toca no assunto e assume-se que tudo pode ser feito com alguma integral. No presente texto, explicitamente usamos a integral de Riemann, que é a única disponível aos alunos do ciclo básico de um curso de engenharia. Essa integral não permite a definição de medidas de conjuntos como por exemplo $\mathbb{Q} \cap [0, 1]$. Isso impede o uso de espaços de eventos ordinários, como os contendo intervalos do tipo $[a, b]$, $a, b \in \mathbb{R}$, a não ser que se coloque alguma restrição sobre a e b . Isso é feito aqui seguindo uma observação um tanto cínica de G. H. Hardy que os engenheiros ficariam muito felizes com um π com dez casas decimais.

A teoria que o livro contém é a padrão, com termos comuns, e por isso não há referências específicas, mas podemos dizer que algumas obras estiveram na base, como [Bar14] [Eva13] [Kop99] [GS01].

Agrademos aos alunos pelo retorno dado durante as aulas virtuais e em especial ao João Gabriel Brito pelos erros apontados no texto original.

Os autores.

Capítulo 1

Introdução

Uma disciplina de probabilidades tem um status especial entre os demais em um curso de engenharia. Sua posição pode ser compreendida quando contemplamos o desenvolvimento das ciências aplicadas ao longo da história, e principalmente no período entre o final do século 19 e o 20. Até meados do século 19, o paradigma da mecânica clássica prevalecia, e mais que uma verdade científica, havia a crença que tudo poderia ser explicado por leis determinísticas e equações diferenciais, uma vez as condições iniciais e de contorno fossem conhecidas. Laplace (1749 – 1827) [dL14] expressou bem essa convicção ao invocar o princípio de razão suficiente de Leibniz. O paradigma determinista continuou firme até e inclusive a teoria da relatividade, que fornece uma ligação entre a mecânica e o eletromagnetismo.

Dois tragédias com cientistas ilustram a transição do paradigma determinístico para o probabilístico.

Boltzmann (1844 – 1906) é o criador da teoria da mecânica estatística, e com ela ele propunha uma explicação para a segunda lei da termodinâmica para gases. Entretanto, sua teoria tinha por hipótese a existência de átomos e o comportamento coletivo estatístico para o conjunto que formava um gás. Boltzmann foi muito criticado e ridicularizado por suas ideias. Em vida, elas nunca foram aceitas. Suicidou-se em 1906.

Outra história menos trágica, mas nem por isso menos triste, é a de Gregor Mendel (1822 – 1884), o pai da Genética, cujas ideias foram completamente ignoradas [Gre18]. Ele publicou dois artigos em 1866, mas apenas em 1900, após sua morte, suas ideias foram redescobertas e aceitas. Uma das razões apontadas para a mudança de postura frente às suas propostas foi justamente a troca de paradigma do determinístico para o probabilístico [Zab89].

É desnecessário relembrar que o início da mecânica quântica se encontra no início do século 20, e que ela se afasta radicalmente do determinismo, colocando probabilidades no núcleo de sua teoria. O uso de probabilidades é necessário não porque há falta de informação, como havia dito Laplace [dL14] em 1814, mas porque a natureza é intrinsecamente não determinística.

Podemos interpretar [Zab89] as histórias de Boltzmann e Mendel como exemplos da mudança de paradigma, de uma visão de mundo determinística para uma probabilística. A interpretação probabilística do mundo físico requer muito mais abstração, pois os fenômenos, mesmo indiretamente, são observados e medidos, mas a explicação do que é observado deve ser dada por uma teoria subjacente para o acaso.

Até o final do século 19, os principais fenômenos probabilísticos em foco tinham origem basicamente em jogos de azar e em casos em que, na nomenclatura moderna, envolviam somente espaços amostrais finitos. E para isso, as teorias matemáticas disponíveis eram suficientes. Convencionou-se atribuir a origem da teoria das probabilidades à correspondência entre Pascal e Fermat ocorrida por volta de 1654 a respeito da solução matemática de um problema de apostas, mas certamente ideias desse naipe não florescem do nada. Certamente já havia estudos anteriores de, por exemplo Cardano, dos árabes, dos chineses e dos hindus feitos já no milênio anterior [Hac06]. Isaac Trodhunter é um autor talvez mais conhecido pelo monumental tratado sobre a história da resistência dos materiais [TP93], mas ele também escreveu em 1865 “A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace” [Tod65], que até hoje é a referência padrão para a história do desenvolvimento da teoria de probabilidades no período coberto por ele. Um pouco mais tarde, no momento em que entram em cena a mecânica estatística de Boltzmann e a mecânica quântica, a indisponibilidade de teorias matemáticas adequadas se tornou uma questão importante.

A teoria dos conjuntos foi fundada por Cantor em 1874 [Bel14], mas vários paradoxos vinham se acumulando, sendo, talvez, o mais conhecido do grande público o do Barbeiro de Sevilha (o barbeiro que barbeia todos que não se barbeiam se barbeia?). Os trabalhos de Zermelo e Frankel no início do século 20 resultaram na teoria axiomática que é usada atualmente na prática. Em outra frente, a teoria da medida de Lebesgue (1875 – 1941), que fundamenta a teoria da integral que leva seu nome, e está na base da teoria das probabilidades moderna, foi desenvolvida também no início do século 20 [Kop99], e colocada em bases matemáticas firmes por Kolmogorov (1903 – 1987) em 1933 [KM19].

Todas essas mudanças estão ligadas à mudança de paradigmas que ocorreu na transição entre os séculos 19 e 20, período rico também em convulsões sociais, entre elas a formação de vários estados modernos e várias revoluções. Basta mencionar o da Alemanha, Itália, a Revolução Meiji, a independência de nosso país e de outros, a revolução russa e a Primeira Guerra.

Um curso de probabilidades para engenharia marca essa transição histórica e oferece uma visão de mundo diferente, em que se coloca uma ordem no caos.

Capítulo 2

Espaço Probabilístico: Espaço amostral, eventos e Probabilidades

2.1 Introdução

Neste primeiro capítulo, vemos as bases da teoria de probabilidades, que consiste de um experimento probabilístico, um conjunto Ω que agrupa os resultados experimentais denominado espaço amostral, um conjunto E que agrupa subconjuntos do espaço amostral denominado espaço de eventos, e uma função $P : E \rightarrow [0, 1]$ definida em E que fornece as probabilidades dos elementos de E . Os eventos podem ser interpretadas como sentenças que envolvem os elementos de Ω . O terno (Ω, E, P) é o que se chama de *espaço probabilístico*.

Essa base de elementos probabilísticos é vista neste capítulo em um contexto muito particular, em que Ω , o espaço amostral, é enumerável. Esse fato crucial possibilita a simplificação da teoria e permite que o estudante focalize sua atenção sobre os elementos essenciais da teoria. Essencialmente essa simplificação se dá porque a função probabilidade $P : E \rightarrow [0, 1]$ pode ser definida para E consistindo do conjunto das partes de Ω , o que é em geral impossível se Ω não for enumerável.

As dificuldades técnicas que surgem quando Ω deixa de ser enumerável serão vistas em um capítulo posterior, quando analisarmos o caso das variáveis aleatórias contínuas.

Começamos fazendo uma rápida revisão do que entendemos por conjuntos e função e depois apresentamos os eventos, espaços amostrais e a

função probabilidade.

2.2 Vídeo que acompanha este capítulo

Um vídeo em que mostramos as bases de um espaço probabilístico - espaço amostral, eventos e função probabilidade para alunos de um curso de engenharia pode ser visto ao escanear-se o código QR abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Sobre as bases da teoria de probabilidades.

2.3 Biografia de Andrei Nikolaevich Kolmogorov – (Андре́й Николаевич Колмо́ров)

Mostramos aqui algumas passagens da biografia do matemático Russo Andrei Nikolaevich Kolmogorov [Enc10, Kol20, DL16], considerado o pai da teoria das probabilidades moderna, para ganharmos uma perspectiva mais humana da teoria.

Andrei Nikolaevich Kolmogorov nasceu em 1903, na cidade de Tambov -Rússia. Sua mãe, Marya, faleceu no parto e a criação de Andrei ficou aos cuidados da família materna, particularmente da sua tia Vera, que o adotou. O seu pai era um agrônomo de formação e foi exilado por participação em movimentos revolucionários anticzaristas e teria morrido em 1919, já na guerra civil.

Kolmogorov relata em um dos seus artigos biográficos: “Eu experimentei a alegria da “descoberta” da matemática cedo na vida, com 5 ou 6 anos, quando notei o padrão: $1 = 1^2$, $1+3 = 2^2$, $1+3+5 = 3^2$, $1+3+5+7 = 4^2$,” . Este e outros problemas aritméticos foram publicados em um jornal

escolar, editado na pequena escola infantil criada por suas tias, que atendia a 10 crianças de diferentes idades.

No ginásio, equivalente ao nosso ensino fundamental e médio, Kolmogorov usufruiu de uma atmosfera muito benéfica em uma escola particular de altíssima qualidade fundada por um grupo de intelectuais. As turmas eram pequenas (15 - 20 alunos) e uma parte considerável dos professores tinha forte atração por ciências. Apesar do alto desempenho em matemática, as paixões mais sérias de Kolmogorov eram biologia e história russa.

Em 1920, com a formação secundária completa, ingressou na Universidade de Moscou, no Departamento de Matemática e Física, não sem alguma hesitação para definir o caminho a seguir, pois tinha muitos interesses. Destacando-se os estudos em história russa antiga, artes e poesia, que permearam toda a sua formação acadêmica e de vida. Em paralelo, frequentou o Instituto Mendeleev de Tecnologia Química, com uma percepção de que a engenharia seria mais necessária que a ciência pura. Logo o seu interesse e brilhantismo em matemática suplantaram todas as suas dúvidas, quanto à grande relevância da matemática.

Participou de cursos e seminários ministrados pelos mais eminentes matemáticos russos. Os seminários, tratando da Teoria de Funções Analíticas, eram ministrados pelo brilhante professor Nikolay N. Luzin (Никола́й Николаевич Лузин), reconhecido mundialmente, que atentou ao talento de Kolmogorov e o convidou para um grupo de estudos com discussões semanais de tópicos matemáticos.

Ainda no período de graduação, Kolmogorov desenvolveu e publicou vários trabalhos em revistas no país e no exterior. Publicou, em 1922, um estudo na área de séries trigonométricas, que teve uma repercussão marcante na comunidade matemática, neste estudo apresentava um exemplo de uma função integrável cujas séries de Fourier divergiam em todos os pontos.

Graduou-se em 1925 na Universidade de Moscou e iniciou a pós - graduação sob orientação de Luzin. Dentro do amplo espectro de estudos, realizou trabalhos fundamentais nas condições de aplicabilidade da lei dos grandes números e da lei forte dos grandes números, consolidando assim os estudos anteriores de Bernoulli, Poisson, Chebychev e Markov. Na sequência publicou o estudo “Teoria geral da medida e o cálculo de probabilidades”, onde apresentou a primeira versão da sua concepção axiomática da teoria de probabilidades.

Em 1929, obtém seu título de doutorado e tornando-se pesquisador da Universidade de Moscou. Em 1931, assume como professor da Universidade de Moscou, onde permaneceu por toda a vida. Sempre manteve intensos

contatos científicos na Rússia e no exterior, especialmente na França e na Alemanha e, desde o início da sua carreira, grande parte dos seus artigos foi divulgada em publicações francesas e alemãs.

Quatro anos depois, em 1933, publicou uma monografia, em alemão, o hoje clássico “Grundbegriffe der Wahrscheinlichkeitsrechnung”, Fundamentos da Teoria da Probabilidade. Num texto sintético, combinou a noção de espaços amostrais e teoria da medida, consolidando a visão axiomática com base nos hoje conhecidos e consagrados axiomas de Kolmogorov, que, entre outros aspectos, consideravam conjuntos infinitos enumeráveis e não enumeráveis, ampliando a definição clássica de Laplace.

No texto, Kolmogorov frisou a necessidade da construção da teoria de probabilidade como uma teoria matemática pura e bem geral. A Teoria de Probabilidade efetivamente atingiu então o patamar de uma disciplina respeitável da matemática pura, como a álgebra e geometria. A repercussão foi tal que Kolmogorov é, por total mérito, reconhecido como o pai da probabilidade moderna.

Em 1936, irrompeu o caso Luzin, um simulacro dos cruéis e famosos expurgos stalinistas na prospecção e condenação dos “inimigos do povo”. O real propósito do processo era de perseguição ao grande matemático Luzin, de forma a expulsá-lo da Academia de Ciências Russa e de suas atividades acadêmicas. Tratava-se de um misto de disputa entre gerações de cientistas, pretensas acusações de plágio, interesses políticos e inclinações religiosas. Luzin foi efetivamente afastado de suas funções, porém não desterrado. Este evento insidioso e nebuloso envolveu Kolmogorov e outros matemáticos que, sob pressão-chantagem do regime stalinista, teriam se omitido, ou até mesmo corroborado parte das acusações. Tardamente, em 2012, a academia russa reverteu a decisão de expulsão.

Dentre outros trabalhos, no final dos anos 1930, Kolmogorov começou a estudar escoamentos turbulentos, empregando como suporte matemático de investigação a teoria de função de variáveis aleatórias de múltiplas variáveis. Suas publicações tiveram significativa influência na compreensão de escoamentos turbulentos e na sua modelagem matemática. Estes estudos foram desenvolvidos com seus estudantes A. M. Obukhov, A. S. Monin e A. M. Yaglom. Kolmogorov não participava da realização de experimentos, mas desenvolveu modelos com base em dados experimentais de outros pesquisadores.

Kolmogorov contribuiu de forma vigorosa não apenas como um matemático puro mas também em matemática aplicada, demonstrando magnífica habilidade para penetrar na essência dos problemas, como no caso do estudo de turbulência e movimento browniano. Um caso exemplar de seu

comprometimento foi o seu engajamento no esforço de guerra -Segunda grande guerra - quando publicou estudos sobre a teoria da artilharia com base em distribuições de probabilidades.

Em 1942, Kolmogorov casou-se com Anna Dmitrievna Egorova, que o acompanhou pelo resto da vida.

Kolmogorov foi influente membro da Academia Russa de Ciências e sempre teve amplo reconhecimento mundial e na Rússia, tendo sido agraciado com uma infinidade de prêmios e eleito para várias academias de ciências e artes em todo o mundo, algo que muito o envaidecia.

Diversas foram as áreas de atuação de Kolmogorov, sempre com contribuições de primeira linha, culminado em estudos que explicitamente foram remetidos ao seu nome, tais como: mecânica clássica, sistemas dinâmicos (Teorema KAM), teoria da informação (Complexidade de Kolmogorov), estatística (teste Kolmogorov-Smirnov), processos estocásticos (Chapman-Kolmogorov equations), análise funcional e topológica (Teorema de Riesz-Weyl-Kolmogorov). Além das publicações científicas, publicou centenas de artigos de divulgação de tópicos matemáticos, de divulgação científica, sobre teoria de poesias além de verbetes de matemática para enciclopédias.

Da sua convivência com antigas gerações de matemáticos, Kolmogorov comentava com o amigo V. I Arnold (Владимир Игоревич Арнольд) que depois dos 60 anos o matemático deveria deixar de fazer matemática. Ao longo de toda a vida e mais intensamente nos anos 1970 envolveu-se com pedagogia e no ensino de matemática em escolas secundárias, participando como membro de comitê executivo do governo, como professor em escola secundária, na orientação de trabalhos e na divulgação de artigos de ciências e artes, além de livros didáticos.

Em Moscou, em 1987, chegou ao fim a vida do grande A. N. Kolmogorov, um dos mais importantes e influentes matemáticos do século XX.

2.4 Conceitos preliminares

Apresentamos agora algumas ideias fundamentais para o entendimento das bases da teoria das probabilidades. Recomendamos não pular essa sessão.

2.4.1 Conjuntos

Um conjunto é uma coleção de objetos, que podem ser também de natureza matemática. Dois exemplos são o conjunto dos números naturais $\mathbb{N} =$

$\{1, 2, 3, \dots\}$ e o dos inteiros $\mathbb{Z} = \{0, -1, 1, -2, 2, -3, 3, \dots\}$. Os objetos que compõe os conjuntos são denominados *elementos*. Quando um elemento x faz parte de um conjunto A , escrevemos $x \in A$ e quando ele não faz parte, escrevemos $x \notin A$.

Tomemos dois conjuntos A e B . Suponha agora que todos os elementos de B são também elementos de A , como por exemplo $B = \{1, 2\}$ e $A = \{-1, 0, 1, 2, 4\}$. Quando isso ocorre, dizemos que B está contido em A e escrevemos $B \subset A$. Dizemos também que B é subconjunto de A . Note que sempre um conjunto está contido nele mesmo. Por exemplo, $\mathbb{N} \subset \mathbb{N}$, pois todo elemento $x \in \mathbb{N}$ pertence a \mathbb{N} . Quando um conjunto não está contido em um outro, escrevemos $B \not\subset A$. Por exemplo, $\{1, 3, 4\} \not\subset \{1, 3, 5, 6\}$.

Um caso muito importante ocorre com o conjunto dito *vazio*, que é denotado pelo símbolo \emptyset . Ele está contido em qualquer outro conjunto, isto é, dado qualquer conjunto A , sempre é verdade que $\emptyset \subset A$, mesmo que $A = \emptyset$, pois todo elemento de \emptyset (\emptyset não possui nenhum elemento!) pertence a A , qualquer que seja A .

É importante notar o uso das chaves $\{ \}$. Observe nos exemplos abaixo o uso correto dos símbolos \subset e \in .

$$\{1, 3, 5\} \subset \{0, 1, 2, 3, 4, 5, 6\}. \quad \{1, 3, 5\} \in \{\{1, 3, 5\}, \{0, 3\}, \{5, 6\}\}.$$

Dado um conjunto A , o conjunto formado por todos os subconjuntos de A é chamado de *conjunto das partes* de A , e é denotado por $\mathcal{P}(A)$. Por exemplo, se $A = \{a, b, c\}$, então

$$\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, A\}.$$

Note que como $\emptyset \subset A$, temos necessariamente $\emptyset \in \mathcal{P}(A)$. Além disso, como $A \subset A$, devemos ter $A \in \mathcal{P}(A)$ (Note o uso do símbolo \in e não de \subset).

Se A possui $n \geq 0$ elementos então $\mathcal{P}(A)$ possui 2^n elementos, como é fácil de ver. Só existem duas possibilidades para cada elemento de A , pertencer ou não pertencer a um subconjunto de A . No último exemplo acima, cada elemento do conjunto das partes $\mathcal{P}(A)$ pode ser descrito de uma forma ligeiramente diferente, assim: $\{a\} = (1, 0, 0)$, $\{b\} = (0, 1, 0)$, $\{b, c\} = (0, 1, 1)$, e assim por diante. Então é claro que a quantidade de elementos de $\mathcal{P}(A)$ é 2^3 .

Conjuntos podem ser descritos usando predicados, ou propriedades, como no exemplo

$$P = \{2(n - 1) : n \in \mathbb{N}\}.$$

Claramente, P é o conjunto dos números pares. P poderia ter sido descrito usando a notação usada acima, com uma lista e pontinhos: $P = \{0, 2, 4, \dots\}$.

Vamos assumir que o aluno já conhece os conjuntos dos números reais \mathbb{R} , dos inteiros \mathbb{Z} e dos racionais \mathbb{Q} , além do conjunto dos naturais \mathbb{N} .

Somente para deixar registrado e não deixar dúvidas quando formos resolver exercícios: O zero é par e $0 \notin \mathbb{N}$.

Operações com conjuntos

Dados dois conjuntos, definimos algumas operações que podem ser feitas com eles. A primeira a ser considerada é a *união*.

Definimos a união dos conjuntos A e B , denotada por $A \cup B$ por

$$A \cup B = \{x : x \in A \text{ ou } x \in B\}.$$

A interseção de A e B é definida por

$$A \cap B = \{x : x \in A \text{ e } x \in B\}.$$

Note que se A e B não tem elementos em comum, então $A \cap B = \emptyset$.

A diferença entre A e B é o conjunto

$$A \setminus B = \{x : x \in A \text{ e } x \notin B\}.$$

Um exemplo de diferença de conjuntos seria: $\{1, 3, 5\} \setminus \{3, 4, 5, 6, 7\} = \{1\}$.

2.4.2 Funções

Aqui abordamos o assunto da maneira intuitiva. Uma *função* é uma regra que associa a cada e todo elemento de um conjunto A denominado *domínio* a um e somente um elemento de um conjunto B denominado *contradomínio*. Não há a necessidade de todos os elementos de B participarem da regra. Uma função f que tem por domínio A e contradomínio B é denotada por $f : A \rightarrow B$.

O conjunto $\{y \in B : \text{existe } x \in A \text{ tal que } f(x) = y\}$ é a *imagem* da função f . A imagem de $f : A \rightarrow B$ é denotada por $\text{Im}(f)$. Se $\text{Im}(f) = B$, então f é dita ser *sobrejetora*.

Se para todo $y \in \text{Im}(f)$, existir apenas um elemento $x \in A$ tal que $f(x) = y$, então $f : A \rightarrow B$ é *injetora*. A função $g : \{-1, 0, 1\} \rightarrow \mathbb{Z}$ definida por $g(x) = x^2$ não é injetora, pois $g(-1) = g(1) = 1$. Essa função também não é sobrejetora, pois não há nenhum elemento x em $\{-1, 0, 1\}$ tal que $g(x) = -5 \in \mathbb{Z}$, por exemplo.

Quando uma função é injetora e sobrejetora, dizemos que ela é *bijetora*. Nesse caso, existe uma correspondência biunívoca entre A e B . Por exemplo, existe uma correspondência biunívoca entre o conjunto dos números naturais \mathbb{N} e o conjunto dos números pares $P = \{2(n-1) : n \in \mathbb{N}\}$. De fato, tomando qualquer número natural $n \in \mathbb{N}$, o número par correspondente é $2(n-1)$, e reciprocamente, dado um número par m , o número natural pertencente a \mathbb{N} correspondente é $1 + m/2$. A função bijetora que liga \mathbb{N} ao conjunto dos números pares é $h : \mathbb{N} \rightarrow P$ definida por $h(n) = 2(n-1)$.

Imagem inversa

Quando a função $f : A \rightarrow B$ é biunívoca, ela possui inversa. Isto é, se dado $y \in B$, $f(x) = y$, então existe uma (única) função $g : B \rightarrow A$ tal que $g(y) = x$. Essa função, denotada por $f^{-1} : B \rightarrow A$, é a *função inversa* de $f : A \rightarrow B$.

Existe um outro conceito correlacionado, porém diferente, mas com um nome semelhante, que é o da *imagem inversa* de uma função. Dada uma função $f : A \rightarrow B$ qualquer, não importando se ela é sobrejetora, injetora ou bijetora, definimos a imagem inversa de $C \subset B$ como sendo o conjunto

$$f^{-1}(C) = \{x \in A : f(x) \in C\}.$$

A definição da imagem inversa será muitíssimo importante quando estudarmos as **variáveis aleatórias** mais adiante. De fato, a conexão entre variáveis aleatórias e os **eventos**, que veremos neste capítulo é feita pelo conceito de imagem inversa.

Seja uma função $f : A \rightarrow B$ e dois subconjuntos C, D quaisquer de B . Valem as propriedades seguintes.

$$\begin{aligned} f^{-1}(C \cup D) &= f^{-1}(C) \cup f^{-1}(D), \\ f^{-1}(C \cap D) &= f^{-1}(C) \cap f^{-1}(D), \\ C \subset D &\Rightarrow f^{-1}(C) \subset f^{-1}(D), \\ f^{-1}(B) &= A, \\ f^{-1}(\emptyset) &= \emptyset. \end{aligned} \tag{2.1}$$

Observe que a afirmação $f(C \cap D) = f(C) \cap f(D)$ é falsa em geral. Por exemplo, quando consideramos $f : \{-1, 0, 1\} \rightarrow \mathbb{N}$ definida por $f(x) = x^2$, temos $f(\{-1\}) = \{1\}$, $f(\{1\}) = \{1\}$, então $f(\{-1\}) \cap f(\{1\}) = \{1\} \neq \emptyset$, mas $\{-1\} \cap \{1\} = \emptyset$. A extraordinariedade de (2.1) faz com que realmente a ideia da imagem inversa de uma função funcione bem quando falarmos de variáveis aleatórias.

2.4.3 Enumerabilidade

Dado um conjunto A , diremos que ele é um conjunto *finito*, se ele for vazio, ou se existir um número $N \in \mathbb{N}$ e função biunívoca $F : A \rightarrow \{n \in \mathbb{N} : n \leq N\}$ que relaciona \mathbb{N} e A . Deve-se tomar o cuidado para não empregar a palavra “limitado”, que será usada em outro contexto.

Diremos que um conjunto B é *enumerável* se ele for finito, ou se existir uma função bijetora $f : \mathbb{N} \rightarrow B$. Por exemplo, $\{a, b, c\}$, o conjunto dos números pares e o conjunto vazio são enumeráveis. O conjunto dos números racionais dentro do intervalo $]0, 1[$ também é enumerável, como mostramos abaixo, por um dispositivo que hoje é clássico. Basta considerar as tabelas

$$\begin{array}{cccccc}
 \frac{1}{1} & & & & & & \frac{1}{1} & & & & & & \\
 \frac{1}{1} & \frac{2}{2} & & & & & \frac{1}{1} & & & & & & \\
 \frac{2}{2} & \frac{2}{2} & & & & & \frac{2}{2} & & & & & & \\
 \frac{1}{2} & \frac{2}{2} & \frac{3}{3} & & & & \frac{1}{2} & \frac{2}{3} & & & & & \\
 \frac{3}{3} & \frac{3}{3} & \frac{3}{3} & & & \rightarrow & \frac{3}{3} & \frac{3}{3} & & & & & \\
 \frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & & & \frac{1}{4} & & \frac{3}{4} & & & & \\
 \frac{4}{4} & \frac{4}{4} & \frac{4}{4} & \frac{4}{4} & & & \frac{4}{4} & & \frac{4}{4} & & & & \\
 \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & \frac{5}{5} & & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & & & \\
 \frac{5}{5} & \frac{5}{5} & \frac{5}{5} & \frac{5}{5} & \frac{5}{5} & & \frac{5}{5} & \frac{5}{5} & \frac{5}{5} & \frac{5}{5} & & & \\
 \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & &
 \end{array}$$

em que na segunda foram eliminados os números repetidos. A correspondência biunívoca g entre os racionais entre $]0, 1[$ e \mathbb{N} é dada pela correspondência entre o número racional e sua posição na segunda tabela. Assim, $g(1) = 1$, $g(2) = \frac{1}{2}$, $g(3) = \frac{1}{3}$, $g(4) = \frac{2}{3}$ e assim por diante. Todos os números racionais em $]0, 1[$ são contemplados pela função g biunívoca.

Os números reais no intervalo $]0, 1[$ não são enumeráveis. E essa diferença é crucial para a teoria das probabilidades. Nos capítulos iniciais deste curso vamos trabalhar somente com conjuntos enumeráveis.

2.5 Espaços amostrais, eventos e probabilidade

Classicamente, um curso de probabilidades apresenta o que são espaços amostrais antes de eventos, mas aqui faremos uma inversão para uma compreensão intuitiva e que motive a introdução dos espaços amostrais.

2.5.1 Experimentos, eventos e espaço amostral

Tudo começa com um experimento, que chamamos de probabilístico. Trata-se de um procedimento cujo resultado é incerto, no sentido em que mesmo fixando-se todas os seus parâmetros, como temperatura, pressão e volume, o resultado do procedimento é diverso. Cada um de nós pode criar exemplos de experimentos probabilísticos.

Temos interesse em quantificar a chance que um determinado resultado tem de ocorrer. Por exemplo, lançamos um dado e desejamos saber qual é a chance de sair o número 6.

Na verdade, se quantificamos a chance de ocorrer o número 6 em um lançamento, começamos a nos interessar a perguntar sobre a probabilidade de ocorrer um número, digamos par, ou ainda, maior ou igual a 3. Ao quantificarmos, estamos apelando para conceitos matemáticos. A chance poderia ser expressa por uma função, por exemplo, $P(6)$ daria a probabilidade da ocorrência do número 6, no lançamento do dado. $P(3)$ seria a probabilidade de ocorrer o número 3, e assim por diante. Mas isso é completamente insatisfatório!

Precisamos refletir a linguagem natural nas manipulações matemáticas, e não conseguimos fazer isso dizendo apenas “6” ou “3”. Para expressar a ocorrência de 6 ou 3, seria melhor usarmos a linguagem dos conjuntos e escrever que temos interesse em quantificar a chance da ocorrência de $\{6, 3\} = \{6\} \cup \{3\}$.

Em um outro exemplo, poderíamos perguntar qual seria a chance que o lançamento de um dado fornecesse um número par e maior ou igual a 3. O resultado que desejamos pode ser escrito na linguagem dos conjuntos como $\{2, 4, 6\} \cap \{3, 4, 5, 6\}$.

Tendo estabelecido que é melhor usar a linguagem dos conjuntos para refletir a linguagem natural, podemos querer agora quantificar as chances. Elas são quantificadas por uma função chamada *Probabilidade*, cujo domínio é formado por conjuntos (como nos exemplos mostrados). Assim, a probabilidade de se obter um número par e maior ou igual a três em um lançamento de dado é expressa por $P(\{2, 4, 6\} \cap \{3, 4, 5, 6\})$.

O conjunto dos conjuntos que merecem ter probabilidades associadas é o conjunto dos *eventos*. Observe com atenção que podemos fazer operações do tipo “e”, “ou”, “exceto” com os eventos, bastando para isso usar respectivamente \cap , \cup , \setminus . Então, o resultado dessas operações deve ser também um elemento do espaço de eventos. Além disso, o conjunto vazio deve ser um evento também, pois poderíamos desejar saber qual é a chance de se obter um número par e ímpar. Isto é, desejamos saber quanto

vale $P(\{\text{número par} \cap \text{número ímpar}\}) = P(\emptyset)$. Claramente, devemos ter $P(\emptyset) = 0$.

Mas os eventos, que devem ser conjuntos, para que usemos a linguagem matemática, são formados a partir de que conjunto? Aqui entra a abstração matemática.

Formalmente, devemos partir de um conjunto base, denominado *espaço amostral*, de onde se extraem os eventos. *Os eventos são subconjuntos do espaço amostral*. Tipicamente o espaço amostral é abstratamente um conjunto que contém todos os resultados possíveis de um experimento. No caso do lançamento de um dado, o espaço amostral poderia ser $\{1, 2, 3, 4, 5, 6, 7, \pi\}$, mas não fazemos isso. Em geral, tomamos o menor conjunto possível, mas nem sempre isso é feito (para quem já estudou o tema antes, basta pensar na modelagem da altura das pessoas pela distribuição normal). Assim, no nosso exemplo acima, o espaço amostral é tipicamente tomado com sendo o conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$, a partir do qual é formado o conjunto dos eventos E , que é o conjunto das partes de Ω , $\mathcal{P}(\Omega)$. *Nesse exemplo*, o conjunto E dos eventos não pode ser menor que o das partes de Ω , pois seria fácil criar um evento no qual temos interesse que correspondesse exatamente a um elemento que não está em $\mathcal{P}(\Omega)$.

Um aspecto que deve-se ter em mente é que o espaço amostral é um ente abstrato *atemporal*. Ele existe na partida do experimento probabilístico. Ele só pode ser alterado formalmente através de um procedimento que veremos mais tarde, quando analisarmos o fenômeno do condicionamento.

Espaços amostrais enumeráveis são ditos *discretos*. Perceba com clareza que não são necessariamente finitos.

Por enquanto, nesta primeira parte do curso, vamos trabalhar somente com espaços amostrais Ω discretos, isto é, enumeráveis. Nesse caso, o espaço de eventos E *pode* ser tomado como sendo o conjunto das partes $\mathcal{P}(\Omega)$. Quando Ω não for enumerável, isso não é possível em geral, como veremos em no capítulo dedicado a variáveis aleatórias contínuas. Por outro lado, apesar de podermos tomarmos $E = \mathcal{P}(\Omega)$, quando Ω for enumerável, E também pode ser muito menor. O espaço de eventos poderia, por exemplo, ser simplesmente $E = \{\emptyset, \Omega\}$. Nesse caso, os únicos eventos que merecem ter probabilidade são “acontecer algo” e “não acontecer nada”. Formalmente, o espaço de eventos pode ser qualquer conjunto de subconjuntos de Ω que satisfaça os requisitos listados abaixo.

$$\left\{ \begin{array}{l} \Omega \in E. \\ C \in E \Rightarrow \Omega \setminus C \in E. \\ \{C_n\}_{n \in \mathbb{N}} \subset E \Rightarrow \bigcup_{n \in \mathbb{N}} C_n \subset E. \end{array} \right. \quad (2.2)$$

Essas propriedades definem o que é uma *sigma-álgebra*.

Reforçamos que nos primeiros capítulos, o espaço de eventos E , quando Ω é enumerável, será o conjunto das partes $\mathcal{P}(\Omega)$.

Dado o espaço amostral Ω e o espaço dos eventos E , podemos agora calcular as chances ou as probabilidades. A probabilidade é uma função definida sobre o espaço dos eventos E que tem como contradomínio o intervalo $[0, 1]$. Além disso, deve satisfazer mais algumas propriedades.

A função probabilidade $P : E \rightarrow [0, 1]$ deve satisfazer

1. $P(\Omega) = 1$,
2. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$, ou de uma maneira mais geral,
3. dado um conjunto enumerável de eventos $\{A_j\}_{j \in \mathbb{N}}$, mutuamente excludentes, devemos ter

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j),$$

em que por mutuamente excludentes, queremos dizer $i \neq j \Rightarrow A_i \cap A_j = \emptyset$.

Dessas propriedades, obtemos que $P(\emptyset) = 0$, como antecipado acima. De fato,

$$\Omega \cap \emptyset = \emptyset \Rightarrow 1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) \Rightarrow P(\emptyset) = 0.$$

O procedimento para a análise de qualquer problema que envolve probabilidades segue os seguintes passos:

1. Compreender qual é o *experimento*.
2. Abstrair o *espaço amostral* Ω .
3. Estabelecer o *espaço de eventos* E .
4. Determinar para qual elemento e de E desejamos calcular a probabilidade.
5. Explicitar a *função probabilidade* $P : E \rightarrow [0, 1]$.
6. Calcular $P(e)$.

2.5.2 Exemplos de fixação

Para fixarmos as ideias, elencamos agora alguns exemplos de fixação.

Exemplo do lançamento de uma moeda. Número finito de lançamentos Suponha que lançamos uma moeda duas vezes em ordem. Qual é a probabilidade que em um dos lançamentos tenhamos “Cara” (K), sabendo-se que os resultados (K, K) , (K, C) , (C, K) , (C, C) são equiprováveis?

A resolução segue o procedimento de análise prescrito acima.

1. Compreender qual é o *experimento*. Trata-se do lançamento de uma moeda duas vezes em ordem. Vamos denotar os resultados básicos por pares ordenados do tipo (a, b) , em que a e b são os resultados dos lançamentos respectivamente do primeiro e do segundo lançamentos. É importante notar que o resultado do experimento não é simplesmente “Cara” ou “Coroa”, mas sim um par ordenado.
2. Abstrair o *espaço amostral* Ω . O espaço amostral é o conjunto

$$\Omega = \{(K, K), (K, C), (C, K), (C, C)\}.$$

3. Estabelecer o *espaço de eventos* E . O espaço de eventos é o conjunto das partes de Ω . Nesse caso, E tem apenas $2^4 = 16$ elementos.
4. Determinar para qual elemento e de E desejamos calcular P . O nosso foco são os eventos que tem pelos menos uma Cara. Isto é, devemos calcular a probabilidade do evento

$$e = \{(K, K)\} \cup \{(K, C)\} \cup \{(C, K)\}.$$

5. Explicitar a *função probabilidade*. O enunciado diz que os resultados (K, K) , (K, C) , (C, K) , (C, C) do experimento são equiprováveis. Isso significa que cada um dos elementos $\{(K, K)\}$, $\{(K, C)\}$, $\{(C, K)\}$ e $\{(C, C)\} \in E$ tem a mesma probabilidade de ocorrer. Então, a função probabilidade é a função real $P : E \rightarrow [0, 1]$ atribui a cada elemento $\{(K, K)\}$, $\{(K, C)\}$, $\{(C, K)\}$, $\{(C, C)\}$ de E o valor $\frac{1}{4}$. Por exemplo, $P(\{(K, C)\}) = \frac{1}{4}$. As demais probabilidades dos outros elementos de E são calculadas usando as propriedades da função $P : E \rightarrow [0, 1]$.
6. Calcular $P(e)$. Devemos calcular $P(e) = P(\{(K, K)\} \cup \{(K, C)\} \cup \{(C, K)\})$. Notamos que os eventos $\{(K, K)\}$, $\{(K, C)\}$ e $\{(C, K)\}$

são mutuamente excludentes, e daí, devemos ter, pelas propriedades da função probabilidade,

$$P(e) = P(\{(K, K)\} \cup \{(K, C)\} \cup \{(C, K)\}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

Exemplo dos dois filhos Um casal tem dois filhos. Qual é a probabilidade que ambos sejam meninos sabendo-se que um deles é menino?

Novamente aplicamos o procedimento básico passo a passo.

1. Compreender qual é o *experimento*. A interpretação é importante. Imaginamos a situação hipotética em que do universo de casais com dois filhos, tomamos um casal e indagamos qual é a probabilidade que ambos os filhos sejam meninos, *já sabendo que um deles é menino*.
2. Abstrair o *espaço amostral* Ω . O espaço amostral é um conjunto que contém todos os resultados possíveis. Trata-se de uma arte, pois há liberdade criativa aqui. No nosso modelo, o espaço amostral é

$$\Omega = \{(O, O), (O, A), (A, O)\},$$

em que A e O são claros pelo contexto. Note que estamos usando pares ordenados. Então a ordem importa.

3. Estabelecer o *espaço de eventos* E . O espaço de eventos é o conjunto das partes de Ω . Nesse caso, E tem apenas $2^3 = 8$ elementos.
4. Determinar para qual elemento e de E desejamos calcular $P : E \rightarrow [0, 1]$. O nosso foco é o evento em que há dois meninos, isto é, $e = \{(O, O)\}$.
5. Explicitar a *função probabilidade*. Novamente, aqui entra a arte da modelagem. Vamos estabelecer que a probabilidade de se nascer um menino ou menina são equiprováveis. Então cada um dos eventos $\{(O, O)\}$, $\{(O, A)\}$ e $\{(A, O)\}$ do espaço de eventos E têm a mesma probabilidade.

Como

$$\{(O, O)\} \cup \{(O, A)\} \cup \{(A, O)\} = \Omega,$$

sendo $\{(O, O)\}$, $\{(O, A)\}$, $\{(A, O)\}$ mutuamente excludentes, igualmente prováveis, com $P(\{(O, O)\}) + P(\{(O, A)\}) + P(\{(A, O)\}) = 1$, devemos ter

$$P(e) = \frac{1}{3}.$$

Comentário sobre o exemplo dos dois filhos Muitos alunos perguntam se não poderíamos ter resolvido o problema *sem* considerar a ordem dos filhos. Isto é, se ao invés de usarmos pares ordenados, não poderíamos ter simplesmente usado conjuntos. Nesse caso, o espaço amostral seria

$$\Omega = \{\{O, O\}, \{O, A\}\},$$

pois $\{O, A\} = \{A, O\}$.

A resposta é sim. Esse seria um espaço amostral possível para a solução do problema. O espaço dos eventos seria o conjunto das partes de Ω como antes.

Desejamos calcular a probabilidade do evento $e = \{O, O\}$.

O que muda agora é a função probabilidade! $P(\{\{A, O\}\})$ deve ser o dobro de $P(\{\{O, O\}\})$, pois o conjunto $\{A, O\}$ compacta duas situações, uma em que o menino é o mais velho e outra em que a menina nasce antes. Como $P(\{\{A, O\}\}) + P(\{\{O, O\}\}) = 1$, novamente devemos ter necessariamente

$$P(e) = P(\{O, O\}) = \frac{1}{3},$$

como antes.

2.6 Problemas

1. Suponha que lançamos uma moeda duas vezes em ordem. Qual é a probabilidade de que no segundo lançamento tenhamos “Cara”, sabendo-se que em cada lançamento sair “Cara” ou “Coroa” são equiprováveis?
 - (a) Explícite o espaço amostral Ω .
 - (b) Explícite o espaço de eventos E .
 - (c) Explícite para qual elemento $e \in E$ deseja-se calcular a probabilidade.
 - (d) Explícite a função probabilidade.
 - (e) Calcule $P(e)$.

Comentário sobre o exercício: De forma intuitiva, o aluno poderia resolvê-lo imediatamente, mas é importante que todos os passos do procedimento de resolução de um exercício sejam seguidos. Em problemas mais complicados isso fará a diferença.

2. Lançamos uma moeda viciada até que uma cara apareça. A probabilidade de sair cara é $\frac{1}{4}$. O número de lançamentos N de coroas antes dessa primeira cara é anotado. Se sair cara no primeiro lançamento, então $N = 0$. Qual é a probabilidade que $N \geq 4$?
- Explicitite o espaço amostral Ω . Note que ele não varia em função do andamento do experimento. Ele é fixo.
 - O espaço amostral é finito?
 - O espaço amostral é enumerável?
 - Explicitite o espaço de eventos E .
 - Explicitite para qual elemento $e \in E$ deseja-se calcular a probabilidade.
 - Explicitite a função probabilidade.
 - Calcule $P(e)$.
3. Vamos repetir o problema anterior com uma pequena mudança. Lançamos uma moeda viciada sem parar, anotando se em cada vez sai cara ou coroa. A probabilidade sair cara é $\frac{1}{4}$. Denotamos o número de lançamentos N de coroas até a primeira cara. Se sair cara no primeiro lançamento, então $N = 0$. Qual é a probabilidade que $N \geq 4$?
- Explicitite o espaço amostral Ω . Note que ele não varia em função do andamento do experimento. Ele é fixo.
 - O espaço amostral é finito?
 - O espaço amostral é enumerável? A resposta agora, diferentemente do problema anterior é *não*.
 - Explique por quê o espaço amostral não pode ser enumerável.
4. Um lago é contaminado por partículas nocivas à saúde. Imagine que há trilhões delas no lago, mas são invisíveis a olho nú. Tomamos uma amostra da água do lago com 100ml. Desejamos calcular a probabilidade que na amostra tenhamos menos que 10 mil partículas.
- Explicitite o espaço amostral Ω .
 - Explicitite o espaço de eventos E .
 - Explicitite para qual elemento $e \in E$ deseja-se calcular a probabilidade.

5. Em uma caixa há uma caixa com 4 bolinhas numeradas de 1 a 4. Vamos fazer retiradas aleatórias da caixa. Qualquer bola pode ser sorteada com igual probabilidade. Retiram-se duas bolas, uma após a outra, sem reposição. Seja X_1 o número da primeira bolinha e X_2 o da segunda. Desejamos saber qual é a probabilidade que a segunda bola seja par.
- (a) Explícite o espaço amostral Ω . Note que ele não varia em função do andamento do experimento. Ele é fixo.
 - (b) Explícite o espaço de eventos E .
 - (c) Explícite para qual elemento $e \in E$ deseja-se calcular a probabilidade.
 - (d) Explícite a função probabilidade.
 - (e) Calcule $P(e)$.

Capítulo 3

Probabilidade condicionada

3.1 Introdução

Continuamos a ver as bases da teoria das probabilidades.

Esta aula é sobre o tema que caracteriza “Probabilidades” como um ramo de estudo da Matemática. Se olharmos os conceitos que apresentamos até aqui, eles são essencialmente aqueles que os alunos aprendem nos cursos médios de matemática ou nos cursos de cálculo. Até mesmo a definição da função probabilidades é feita por meio ou de somatórias ou de integrais. Entretanto, o ato de condicionar uma função (de probabilidade) a uma dada informação é algo novo, que tipifica o tema do nosso curso.

O efeito de se condicionar probabilidades a um evento pode ser conseguido alterando-se o espaço amostral, pois tudo passaria como se o novo espaço amostral fosse reduzido pela informação trazida pelo evento condicionante. Entretanto, não é esse o objetivo. O objetivo é obter-se o efeito da introdução de novas informações no cálculo de probabilidades *sem* se alterar o espaço amostral original Ω .

3.2 Vídeo que acompanha este capítulo

Um vídeo em acompanha este texto pode ser visto ao escanear-se o código QR abaixo. No vídeo resolvemos o exercício 1 que é proposto neste capítulo na página 38. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Exercício sobre probabilidades condicionadas.

3.3 Biografia de Thomas Bayes

Thomas Bayes [Enc10, Mcg15, Bel04] nasceu em Londres em 1702. Filho de Anne Carpenter e do pastor Joshua Bayes, religioso da linha não conformista – protestante – em relação às práticas da Igreja Anglicana. Os seus ancestrais fizeram feição fortuna na indústria de cutelaria e mantiveram esta riqueza por gerações.

A sua educação na infância e juventude foi feita em casa, com tutores. Como o ingresso nas prestigiosas universidades inglesas de Oxford e Cambridge não era possível para um protestante, Bayes ingressou na Universidade de Edimburgo – Escócia – em 1719. O seu objetivo principal era o estudo de teologia, visando o preparo para o ministério religioso. Registros indicam que no período que esteve em Edimburgo (1719-22) Bayes estudou também lógica, e supostamente matemática.

Bayes tornou-se ministro presbiteriano, tendo iniciado como ministro assistente do seu pai Joshua. Em 1733, tornou-se ministro em Tunbridge Wells, próxima a Londres, onde viveu até a sua morte em 1761. Bayes era considerado um pregador sem grande brilho, mas trabalhava e estudava intensamente, discutindo questões teológicas controversas, que eram divulgadas na forma de publicações.

Paralelamente, o reverendo Bayes desenvolvia estudos matemáticos, que registrava em seus cadernos de notas, tratando de tópicos como séries matemáticas e de análise numérica, e não necessariamente os divulgava. Era um newtoniano convicto e ativo defensor da teoria de cálculo diferencial (“fluxions”) de Newton, inclusive escreveu um artigo confrontando a intensa crítica ao trabalho de Newton publicada pelo eminente filósofo George Berkeley, que questionava os fundamentos lógicos da teoria newtoniana.

Bayes foi convidado e ingressou na Royal Society of London, em 1642, pelo reconhecimento dos seus trabalhos na área de matemática.

Há bastante especulação quanto à motivação de Bayes para o estudo da teoria de probabilidades. Poderia ter sido por influência de Moivre, autor

de “Doutrina do Acaso” e/ou por meio da revisão de um trabalho de Thomas Simpson que tratava da lei dos grandes números.

De fato, Bayes é reconhecido como quem estabeleceu originalmente as bases do entendimento do problema de cálculo da probabilidade inversa. Este trabalho fundamental, que propõe a solução do cálculo de probabilidades à posteriori, foi publicado postumamente, em 1763, após comunicação de seu amigo Richard Price na Royal Society e publicação nos anais desta. O artigo “An essay towards solving a problem in the doctrine of chances” –Um ensaio para solucionar um problema na doutrina das probabilidades– mostra uma clara discussão da distribuição binomial e resultados envolvendo a, hoje, denominada probabilidade condicionada.

Quando mencionamos teorema ou lei ou regra de Bayes não nos referimos exatamente ao trabalho de Bayes, e sim a uma forma generalizada proposta por Pierre Simon Laplace (1749-1827) em seu texto que consagra a teoria de probabilidade clássica “*Theorie Analytique des Probabilités*” – Teoria Analítica das Probabilidades – de 1821. Laplace expressou a ideia, sem utilizar qualquer equação, afirmando que a probabilidade de uma causa, para um dado evento, seria proporcional à probabilidade desse evento, dada essa causa. Por questão de mérito e justiça, alguns autores reivindicam que a conhecida lei de Bayes seria mais adequadamente nominada como lei de Bayes-Laplace.

O impacto no nosso atual cotidiano da Estatística e Inferência Bayesiana é impressionante em ampla e distintas aplicações. Para ficarmos em quatro exemplos: análise econômica, sistemas de comunicação, sistema de tradução automática e diagnóstico médico. No entanto, no século XIX e parte do século XX a lei de Bayes e os modelos baseados nesta geraram muita desconfiança e foram, em certa medida, desconsiderados. O questionamento das premissas e das metodologias da análise bayesiana pelos defensores da visão frequentista perdurou por todo o século passado e ainda suscita ricas e acaloradas discussões, apesar dos resultados incontestes.

3.4 Ideias preliminares

Em primeiro lugar, informamos que aquela conhecida fórmula para a probabilidade do evento A dado o evento B ,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.1)$$

não advém de nenhuma demonstração. Ela é *definida*.

Vamos tentar motivar a definição, mas antes atente ao fato que A e B são eventos, e portanto *conjuntos*, e por isso faz sentido a interseção $A \cap B$, que aparece no numerador de (3.1).

Exemplo 1. *Considere o seguinte experimento. Temos uma caixa com bolas numeradas de 1 a 10. Suponha que a probabilidade de se sortear a bola de número n é proporcional a n . Queremos saber qual é a probabilidade de se sortear um número par.*

Vamos aplicar o nosso procedimento de análise visto no capítulo anterior.

- a) *O experimento está entendido.*
- b) *O espaço amostral será o conjunto $\Omega_1 = \{1, 2, \dots, 10\}$.*
- c) *O espaço dos eventos E_1 é o conjunto formado por todos os subconjuntos de Ω_1 . Por exemplo, $\{2\}$, $\{6, 9\}$ são subconjuntos de Ω_1 , e portanto elementos de E_1 .*
- d) *O evento de interesse é $e = \{2, 4, 6, 8, 10\}$.*
- e) *O próximo passo é explicitar a função probabilidade $P_1 : E_1 \rightarrow [0, 1]$. Sabemos que $P_1(\{n\})$ é proporcional a n . Assim, existe $\alpha \in \mathbb{R}_+$ tal que $P_1(\{n\}) = \alpha n$. Como $\sum_{n=1}^{10} P_1(\{n\}) = 1$, devemos ter $\alpha \sum_{n=1}^{10} n = 1$. Logo, $\alpha = \frac{1}{55}$.*
- f) *O último passo é calcularmos $P_1(e) = \frac{1}{55}(2 + 4 + 6 + 8 + 10) = \frac{30}{55}$. Aproveitamos para dizermos que em particular, $P_1(\{2\}) = \frac{2}{55}$.*

Naturalmente, a probabilidade de se sortear um número ímpar é $1 - \frac{30}{55} = \frac{25}{55}$. Pensando em termos do nosso procedimento padrão, o evento “número ímpar” é o conjunto $\{1, 3, 5, 7, 9\}$. Ele é expresso por $\Omega_1 \setminus e$. Então, das propriedades da função probabilidade, temos necessariamente $P_1(\{1, 3, 5, 7, 9\}) = 1 - P(e) = 1 - \frac{30}{55} = \frac{25}{55}$.

Suponha agora que fizemos um sorteio, mas você não sabe qual foi o número sorteado. Qual é a probabilidade que o número tenha sido par? Note a atemporalidade dos problemas de probabilidades. O sorteio já ocorreu, como poderia também estar no futuro, e a resposta seria a mesma, isto é $P_1(e) = \frac{30}{55}$. \triangle

Agora, vamos alterar um pouco o enunciado do problema anterior.

Exemplo 2. Se dissermos que o número sorteado é par, qual é a probabilidade que ele seja o número 2? Essa probabilidade deve ser diferente de $\frac{2}{55}$, vista no exemplo 1, pois temos uma informação nova: “o número sorteado é par”. Para ver que essa probabilidade deve ser diferente de $\frac{2}{55}$, basta imaginar uma situação que dizemos que o número sorteado é par e menor que 3, e perguntamos a probabilidade de ele ser 2. Naturalmente, essa probabilidade vale 1, com certeza absoluta. Então, o fornecimento de informação altera probabilidades.

Aplicamos o nosso procedimento de resolução de problemas probabilísticos.

- a) O experimento está entendido. É o mesmo do exemplo 1.
- b) O espaço amostral não é o mesmo do exemplo 1, por causa da informação “o número sorteado é par”. O espaço amostral é o conjunto $\Omega_2 = \{2, 4, 6, 8, 10\}$.
- c) O espaço dos eventos E_2 é o conjunto de todos os subconjuntos de Ω_2 .
- d) O evento de interesse é o conjunto $e = \{2\}$.
- e) O próximo passo é explicitar a função probabilidade $P_2 : E_2 \rightarrow [0, 1]$. Note que a função probabilidade deve ser diferente de $P_1 : E_1 \rightarrow [0, 1]$, porque o domínio da função mudou! Era E_1 e agora é E_2 . Então devemos considerar obrigatoriamente uma nova função probabilidade $P_2 : E_2 \rightarrow [0, 1]$. É claro que, para começo de conversa, $P_2(\{2, 4, 6, 8, 10\}) = 1$, e não mais igual a $P_1(\{2, 4, 6, 8, 10\}) = \frac{30}{55}$, pois $\{2, 4, 6, 8, 10\} = \Omega_2$ é o todo agora. Temos a informação de que o número sorteado é par.

Então, como especificar uma função $P_2 : E_2 \rightarrow [0, 1]$ razoável? Uma é supor que a relação de proporcionalidade entre as chances de se obter 2, 4, ..., 10 se mantém. A probabilidade de se obter 10 deve continuar a ser 5 vezes maior do que a de se obter 2. Esse efeito é obtido multiplicando-se as probabilidades para os resultados 2, 4, ..., 10 obtidas no exemplo 1 por uma constante k de tal forma que a soma das probabilidades seja 1. Isto é:

$$k \underbrace{(P_1(\{2\}) + P_1(\{4\}) + P_1(\{6\}) + P_1(\{8\}) + P_1(\{10\}))}_{=P_1(\{2,4,6,8,10\})} = 1.$$

Essa constante de proporcionalidade não é nada mais que

$$k = \frac{1}{P_1(\{2, 4, 6, 8, 10\})} = \frac{1}{30/55} = \frac{1}{P_1(\text{Ser par})}.$$

Então, é razoável impormos que

$$P_2(\{n\}) = \frac{P_1(\{n\})}{P_1(\{2, 4, 6, 8, 10\})}.$$

f) O último passo é calcularmos

$$P_2(e) = \frac{\frac{2}{55}}{\frac{30}{55}} = \frac{2}{30},$$

que é maior do que $P_1(\{2\}) = \frac{2}{55}$, obtida no exemplo 1, como já era esperado, pois sabíamos que o número sorteado era par.

△

Observação 3.4.1. Note atentamente que até aqui *não* houve a introdução da probabilidade condicionada dada pela fórmula 3.1. O que fizemos foi alterar o espaço amostral, o espaço de eventos e a função probabilidades com o objetivo de incorporar a nova informação. A probabilidade condicionada será introduzida agora.

3.5 Probabilidade condicionada

Partimos de um problema probabilístico. Determinamos o seu espaço amostral Ω , o seu espaço de eventos E e sua função de probabilidade $P : E \rightarrow [0, 1]$. Calculamos a probabilidade de um evento particular $e \in E$.

Ocorre que uma nova informação é introduzida no problema. A questão que surge é saber se podemos manter os *mesmos* espaços amostral Ω e de eventos E para resolver o problema com a nova informação. A resposta é sim, desde que substituamos a função de probabilidade $P : E \rightarrow [0, 1]$. O domínio continua o mesmo, mas a definição da função de probabilidade muda. Ela muda exatamente de acordo com a fórmula 3.1 *por definição*.

Assim, suponha que $A, B \subset \Omega$ são elementos de E . A probabilidade do evento $A \in E$ é dada por $P(A)$. Quando é dado que o evento B ocorre, a probabilidade de A dado esse conhecimento passa a ser $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Perceba que essa definição nos leva exatamente ao resultado obtido no exemplo 2. A diferença é que lá, havíamos alterado o espaço amostral e o espaço de eventos. Aqui, com a definição de probabilidade condicionada, alteramos apenas a função probabilidade, que passa de $P : E \rightarrow [0, 1]$ para $P(\cdot|B) : E \rightarrow [0, 1]$. O domínio *não muda*. Bonito, não?

3.5.1 Consequências da fórmula do condicionamento

A primeira consequência da fórmula (3.1) é que podemos escrever imediatamente

$$P(A \cap B) = P(A|B)P(B). \quad (3.2)$$

Como $P(A \cap B) = P(B \cap A)$, devemos ter

$$P(A|B)P(B) = P(B|A)P(A), \quad (3.3)$$

que nos leva à famosa fórmula de Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (3.4)$$

que também é conhecida com o teorema da probabilidade inversa. Note que do lado esquerdo da equação (3.4), temos o conjunto B como condicionante e A é o evento sobre o qual desejamos calcular a probabilidade, ao passo que do lado direito, A passa a ser o evento condicionante. Os papéis de A e B se invertem.

Para ver a força dessa fórmula, vamos ver um exemplo marcante do período que vivemos.

Exemplo 3. *Suponha que 10% da população seja infectada por um vírus, cuja infecção em muitos casos passa despercebida, isto é, as pessoas são assintomáticas. Suponha que exista um teste que dê positivo se a pessoa realmente tiver o vírus com probabilidade de 60% (e portanto, mesmo se ela tiver o vírus, há 40% de chance do teste dar não positivo). Por outro lado, suponha que se a pessoa não tiver o vírus, o teste dê um falso positivo com 5% de chance. Uma pessoa é testada, com resultado positivo. Qual é a probabilidade que ela tenha o vírus?*

Vamos aplicar o nosso procedimento padrão para resolver esse problema.

- a) *No que consiste o experimento? Tomar uma pessoa qualquer da população. Então está implícito aqui que a probabilidade de se escolher qualquer pessoa é uniforme na população. Essa pessoa é testada e sabe-se que o teste deu positivo.*
- b) *Para determinarmos o espaço amostral, precisamos saber quais são os resultados do experimento. Eles são (V, P) , (\bar{V}, P) , (V, \bar{P}) , (\bar{V}, \bar{P}) , em que a barra sobre a letra significa a negação. V significa “pessoa com vírus”, P significa “teste positivo”. Então, o nosso espaço amostral que adotaremos será $\Omega = \{(V, P), (\bar{V}, P), (V, \bar{P}), (\bar{V}, \bar{P})\}$. Perceba que V*

não é evento, e nem sequer $\{V\}$. Um evento é um par ordenado do tipo $\{(V, P)\}$. Os eventos que temos interesse são

$$\begin{aligned} A &= \{(V, P), (V, \bar{P})\} \text{ (“Pessoa com vírus”),} \\ B &= \{(V, P), (\bar{V}, P)\} \text{ (“Testar positivo”).} \end{aligned}$$

Vamos precisar também do evento

$$C = \Omega \setminus A = \{(\bar{V}, P), (\bar{V}, \bar{P})\} \text{ (“Pessoa sem vírus”).}$$

Acima, observe o uso da notação $C = \Omega \setminus A$ (“Pessoa sem vírus”).

- c) O espaço de eventos E é o conjunto das partes de Ω .
- d) Antes de explicitarmos o evento sobre o qual temos interesse, vamos analisar a função probabilidade, pois essa é a novidade aqui. Ela é induzida pela informação “testar positivo”.

Concretamente, a novidade aqui é a introdução das especificações

$$\begin{aligned} P(\text{“testar +”} | \text{“Pessoa com vírus”}) &= P(B|A) = 0.60, \\ P(\text{“testar +”} | \text{“Pessoa sem vírus”}) &= P(B|C) = 0.05. \end{aligned} \quad (3.5)$$

Exploramos esse passo para explicitarmos vários elementos do problema.

- i) O evento $A = \{(V, P), (V, \bar{P})\}$ significa “pessoa com vírus”. Daí,

$$P(A) = P(\{(V, P), (V, \bar{P})\}) = 0.10, \quad (3.6)$$

de acordo com o enunciado.

- ii) O evento “pessoa sem vírus” tem probabilidade

$$P(C) = P(\Omega \setminus A) = 1 - 0.10 = 0.90. \quad (3.7)$$

- iii) O evento “testar +” é $B = \{(V, P), (\bar{V}, P)\}$. A probabilidade dele é

$$P(B) = P(\{(V, P), (\bar{V}, P)\}) = P(\{(V, P)\}) + P(\{(\bar{V}, P)\}),$$

pois $\{(V, P)\} \cap \{(\bar{V}, P)\} = \emptyset$.

O evento que temos interesse é $A = \{(V, P), (V, \bar{P})\}$, mas sabendo-se que a pessoa testou positivo, isto é, dado o evento $B = \{(V, P), (\bar{V}, P)\}$.

e) Agora especificamos a função probabilidade condicionada

$$P(\cdot | \text{“testar +”}) : E \rightarrow [0, 1].$$

Note que o nome da função não é P , mas sim $P(\cdot | \cdot)$. A barra não é um adjetivo para P . Trata-se de duas funções diferentes.

Como vimos acima o evento “testar +” corresponde ao conjunto.

$$B = \{(V, P), (\bar{V}, P)\}.$$

Então,

$$P(\cdot | \text{“testar +”}) = \frac{P(\cdot \cap B)}{P(B)}.$$

Para a completa especificação dessa função, precisamos calcular $P(B)$, que é dada por

$$P(B) = P(\{(V, P)\}) + P(\{(\bar{V}, P)\}). \quad (3.8)$$

Relembrando os conjuntos $A = \{(V, P), (V, \bar{P})\}$ e $B = \{(V, P), (\bar{V}, P)\}$, vemos que $\{(V, P)\} = A \cap B$. Então, $P(\{(V, P)\}) = P(A \cap B)$. Ela é calculada usando os valores já dados em (3.5) e (3.6).

$$P(\{(V, P)\}) = P(A \cap B) = P(B|A)P(A) = 0.60 \times 0.10 = 0.06. \quad (3.9)$$

A probabilidade $P(\{(\bar{V}, P)\})$, que aparece em (3.8), é calculada da mesma forma.

$$P(\{(\bar{V}, P)\}) = P(B \cap C) = P(B|C)P(C) = 0.05 \times 0.90 = 0.045.$$

Dessa forma, ficamos com

$$P(B) = P(B|A)P(A) + P(B|C)P(C) = 0.06 + 0.045 = 0.105. \quad (3.10)$$

f) Para o evento $e = A = \{(V, P), (V, \bar{P})\}$ para o qual queremos calcular a probabilidade dado o evento condicionante B , temos

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

que é agora facilmente calculada, pois já calculamos acima o seu numerador (veja (3.9)). Assim, finalmente, temos a resposta pedida, que é

$$P(A|B) = \frac{0.06}{0.105} = 0.57,$$

Isto é, uma pessoa testada positiva tem 57% de probabilidade de ter o vírus, que não deixa de ser um resultado surpreendente.

△

3.5.2 Teorema da probabilidade total

Já usamos implicitamente o chamado Teorema da probabilidade total no exemplo oferecido acima sobre a testagem da presença de um vírus, na equação (3.10).

Explicitamente suponha que um evento $C \subset \Omega$ de um espaço amostral possa ser escrito como a união de eventos mutuamente excludentes, isto é

$$C = C_1 \cup C_2 \cup \dots \cup C_n,$$

com $C_i \cap C_j = \emptyset$ se $i \neq j$.

Então, para um evento $B \subset C$, temos a fórmula

$$P(B) = P(B \cap C) = P((B \cap C_1) \cup (B \cap C_2) \cup \dots \cup (B \cap C_n)) = \sum_{i=1}^n P(B|C_i)P(C_i).$$

3.6 Independência

A ideia de independência entre eventos de um experimento probabilístico é definida usando-se a probabilidade condicionada.

Diremos que dois eventos A e B de um espaço probabilístico (Ω, E, P) são independentes se

$$P(A|B) = P(A),$$

que na linguagem corrente, pode ser expressa dizendo-se que a probabilidade do evento A não é alterada pelo conhecimento do evento B . A consequência imediata dessa fórmula é que quando A e B são independentes, temos

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

Tratando de independência de eventos, a intuição funciona nos casos em que os eventos A e B são disjuntos, ou um está contido no outro. Por exemplo, se $A \cap B = \emptyset$, então se A ocorre, B não pode ocorrer e vice versa.

Suponha que $A \cap B = \emptyset$, com $P(A) \neq 0$, $P(B) \neq 0$. Nesse caso, $P(A|B) = 0$, mas $P(A) \neq 0$, isto é, $P(A|B) \neq P(A)$. Então de fato, quando A e B são mutuamente excludentes, A e B não são independentes, quando as probabilidades $P(A)$ e $P(B)$ não são nulas.

Entretanto, chamamos a atenção para o fato que a independência entre eventos é uma questão contábil, nem sempre coincidindo com a intuição. Por exemplo, considere a situação esquematizada na figura 3.1. Pode-se ver, fazendo as contas que os eventos A e B são independentes, mas A e C não são!

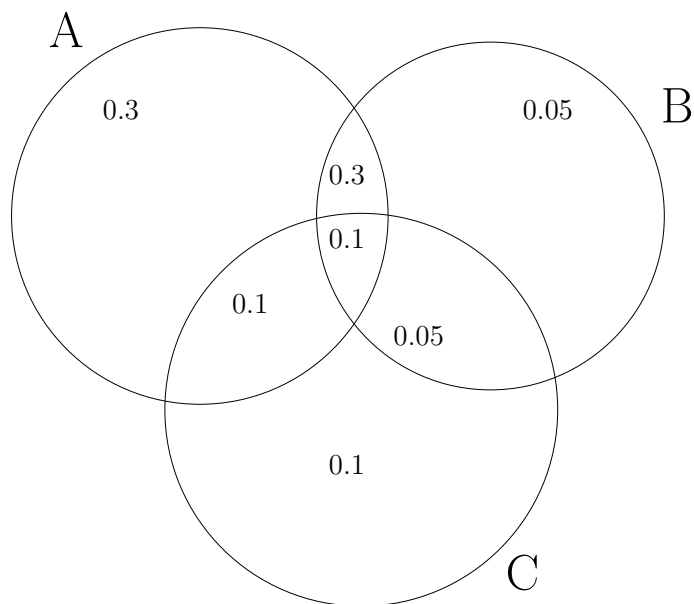


Figura 3.1: Eventos A , B e C . Os eventos A e B são independentes.

3.7 Exemplos de fixação

Para fixarmos as ideias, elencamos agora alguns exemplos de fixação.

Exemplo da casa de câmbio Considere uma cesta de 20 moedas diferentes que inclui o Dólar (EUA), Yuan (China), Rublo (Rússia), Peso (Argentina), Real (Brasil), Euro (ZE), Libra (GB) e o Iene (Japão). Um mercador aceita fazer uma transação A quando o pagamento for feito em Dólares, Yuans, Rublos, Pesos ou Reais. Ele aceita fazer uma transação B quando o pagamento somente for feito em Dólares, Euros, Libras ou Ienes. Não ocorre nenhuma transação (ou seja, não ocorre nem a transação A e nem a B) somente se a moeda não for aceita. A probabilidade de um cliente (que usa apenas uma moeda) usar qualquer uma das moedas da cesta é a mesma. Podemos afirmar que sabendo-se que o mercador e o cliente fizeram uma transação, a probabilidade que o Dólar tenha sido usado é maior do que a que o Real tenha sido usado?

Novamente aplicamos o procedimento básico passo a passo. Na resolução, vamos designar o evento “o mercador e o cliente fizeram uma transação” de C .

1. Compreender qual é o *experimento*. Imaginamos todos os encontros

entre clientes e mercadores do universo, passado presente e futuro. Um desses encontros é trazido a nós. Observamos a moeda que o cliente usa. Essa moeda é o resultado do experimento. Sabemos que houve uma transação.

2. Explicitamos o *espaço amostral* Ω . Ω é formado pelo conjunto das moedas da cesta com 20 elementos mencionada no enunciado.
3. Estabelecer o *espaço de eventos* E . O espaço de eventos é o conjunto das partes de Ω . Nesse caso, E tem 2^{20} elementos, ou seja mais de um milhão de elementos.
4. Determinar os eventos de interesse. Temos interesse no evento “o mercador e o cliente fizeram uma transação”, que concretamente é o conjunto

$$C = \{\text{Dólar, Yuan, Rublo, Peso, Real}\} \cup \{\text{Dólar, Euro, Libra, Iene}\},$$

que são as moedas aceitas para fazer uma transação.

$$C = \{\text{Dólar, Yuan, Rublo, Peso, Real, Euro, Libra, Iene}\}.$$

Temos interesse também nos eventos “Dólar” e “Real”. $D = \{\text{Dólar}\}$, $R = \{\text{Real}\}$.

5. Explicitar a *função probabilidade*. Pelo enunciado, $P(\{X\}) = 1/20$, para qualquer moeda X da cesta de 20 moedas mencionada no enunciado. Assim, $P(D) = P(R) = 1/20$, $P(C) = 8/20$.

Precisamos definir a função probabilidade condicionada. Ela é dada por

$$P(\cdot | C) = \frac{P(\cdot \cap C)}{P(C)}.$$

6. Agora estamos em condições de calcular as probabilidades de interesse $P(R|C)$ e $P(D|C)$. Concretamente,

$$P(R|C) = \frac{P(R \cap C)}{P(C)} = \frac{P(R)}{P(C)} = \frac{1/20}{8/20},$$

$$P(D|C) = \frac{P(D \cap C)}{P(C)} = \frac{P(D)}{P(C)} = \frac{1/20}{8/20}.$$

Então, a probabilidade que o Dólar tenha sido usado é exatamente a mesma que a do Real. Esse é um resultado contra-intuitivo, pois o Dólar é aceito nas duas transações, mas o Real somente em uma.

Exemplo do segundo resultado Seja uma caixa contendo bolas numeradas de 1 a 10. Fazemos um sorteio tirando uma bola de cada vez sem reposição. Qualquer bola tem a mesma chance de ter retirada como qualquer outra. Qual é a probabilidade que a primeira bola tenha sido ímpar, sabendo-se que a segunda é par?

Para responder a essa pergunta, vamos usar o nosso procedimento padrão.

- a) O experimento está entendido. O fato interessante é a inversão contida na pergunta. Teria sido natural perguntarmos sobre a probabilidade da segunda bola ser par, sabendo-se que a primeira é ímpar, mas aqui fizemos o contrário. A expressão chave aqui é “sem reposição”.
- b) Como dissemos, a modelagem é uma arte. O espaço amostral é uma abstração. Para nós, o espaço amostral será

$$\Omega = \{(\text{Par}, \text{Ímpar}), (\text{Ímpar}, \text{Par}), (\text{Par}, \text{Par}), (\text{Ímpar}, \text{Ímpar})\}.$$

O estudante poderia perguntar se não poderia ser composto pelos pares ordenados formados a partir dos números $1, 2, \dots, 10$. A resposta é: sim, poderia, mas o espaço de eventos e a função probabilidade seriam diferentes.

- c) O espaço dos eventos E é o conjunto de todos os subconjuntos de Ω . Então E tem somente 4 elementos.
- d) O evento de interesse é o conjunto $A = \{(\text{Ímpar}, \text{Par}), (\text{Ímpar}, \text{Ímpar})\}$ (primeira bola ímpar), sabendo-se que a segunda bola é par. Esse segundo evento é expresso pelo conjunto $B = \{(\text{Par}, \text{Par}), (\text{Ímpar}, \text{Par})\}$ (segunda bola par).
- e) A função probabilidade não condicionada $P : E \rightarrow [0, 1]$ especificada por

$$P((\text{Par}, \text{Par})) = P((\text{Ímpar}, \text{Ímpar})) = \frac{5}{10} \frac{4}{9},$$

$$P((\text{Par}, \text{Ímpar})) = P((\text{Ímpar}, \text{Par})) = \frac{5}{10} \frac{5}{9}.$$

A função probabilidade condicionada é dada por $P(\cdot|B) = \frac{P(\cdot \cap B)}{P(B)}$.

- f) Podemos calcular a probabilidade condicionada do evento de interesse por

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

em que $A \cap B = \{(\text{Impar}, \text{Par}), (\text{Impar}, \text{Impar})\} \cap \{(\text{Par}, \text{Par}), (\text{Impar}, \text{Par})\} = \{(\text{Impar}, \text{Par})\}$. Dai o numerador é dado por $P(A \cap B) = \frac{5}{10} \frac{5}{9}$.

O denominador é calculado por

$$\begin{aligned} P(B) &= P(\{(\text{Par}, \text{Par}), (\text{Impar}, \text{Par})\}) \\ &= P(\{(\text{Par}, \text{Par})\}) + P(\{(\text{Impar}, \text{Par})\}) \\ &= \frac{5}{10} \frac{4}{9} + \frac{5}{10} \frac{5}{9}, \end{aligned}$$

pois $\{(\text{Impar}, \text{Par})\}$ e $\{(\text{Impar}, \text{Impar})\}$ são eventos mutuamente excluídos.

A probabilidade condicionada pedida é

$$P(A|B) = \frac{\frac{5}{10} \frac{5}{9}}{\frac{5}{10} \frac{5}{9} + \frac{5}{10} \frac{4}{9}} = \frac{5}{9}.$$

Essa é a probabilidade que o primeiro número sorteado é ímpar, sabendo-se que segundo é par.

3.8 Problemas

1. Esse problema é objeto do vídeo que pode ser visto escaneando-se o QRcode abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Exercício sobre probabilidades condicionadas.

Suponha que o espaço amostral de um experimento é composto pelos números inteiros de 1 a 10. Cada um deles tem a mesma probabilidade de ser sorteado. É realizado um sorteio, e sabe-se que o resultado é solução da equação $(x - a)(x - b)^2 = 0$. $a \neq b$, $a, b \in \{1, \dots, 10\}$.

- (a) Explícite o espaço amostral Ω .
- (b) Explícite o espaço de eventos E .

- (c) Explícite quais são os elementos de E relevantes para o problema, que é resolvido usando-se probabilidade condicionada.
 - (d) Explícite a função probabilidade e a função probabilidade condicionada.
 - (e) Responda à pergunta: Dado que o número sorteado é solução da equação dada, a probabilidade que tenha sido o número a é menor do que a que tenha sido a raiz dupla b ?
2. Em uma caixa temos bolas idênticas numeradas de 1 a 10. Qualquer uma pode ser sorteada com igual probabilidade pois são todas iguais. Fazemos três retiradas sem reposição. Qual é a probabilidade que a segunda bola seja par sabendo-se que a terceira é par?
- (a) Explícite o espaço amostral Ω .
 - (b) Explícite o espaço de eventos E .
 - (c) Explícite quais são os elementos de E relevantes para o problema, que é resolvido usando-se probabilidade condicionada.
 - (d) Explícite a função probabilidade e a função probabilidade condicionada.
 - (e) Calcule a probabilidade que a terceira bola seja par sabendo-se que a segunda é par.
3. Em uma caixa temos bolas idênticas numeradas de 1 a 4. Qualquer uma pode ser sorteada com igual probabilidade pois são todas iguais. Retiramos uma a uma até sobrar uma única bola. Os eventos “a soma das primeiras duas bolas é ímpar” e “última bola sorteada é par” são independentes?
4. Determine x e y , que denotam probabilidades, para que o diagrama de Venn ilustrado na figura 3.2 corresponda a um espaço probabilístico com os eventos A e B independentes. O espaço amostral Ω é o conjunto $\Omega = A \cup B \cup C$.
5. Esse exercício já foi proposto na aula anterior. Agora o objetivo é resolvê-lo usando as ideias deste capítulo.

Em uma caixa há uma caixa com 4 bolinhas numeradas de 1 a 4. Vamos fazer retiradas aleatórias da caixa. Qualquer bola pode ser sorteada com igual probabilidade. Retiram-se duas bolas, uma após a outra, sem reposição. Seja X_1 o número da primeira bolinha e X_2

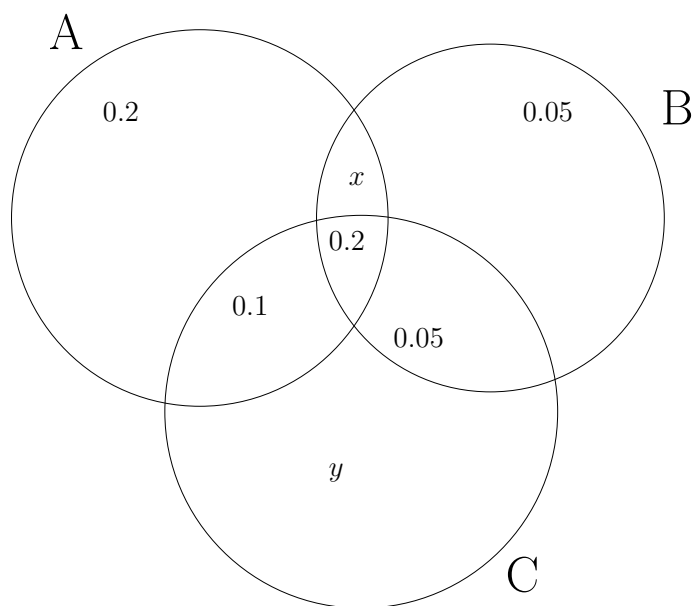


Figura 3.2: Diagrama de Venn.

o da segunda. Desejamos saber qual é a probabilidade de que a segunda bola seja par. Aplicando o procedimento padrão, começando pela especificação do espaço amostral Ω , resolva esse exercício usando probabilidade condicionada. A informação condicionante é o resultado da primeira bola.

Capítulo 4

Variáveis aleatórias discretas

4.1 Introdução

O tema deste capítulo é sobre variáveis aleatórias, que apesar do nome não são nem variáveis e nem aleatórias. Elas são funções reais definidas sobre o espaço amostral Ω , mas nem todas funções reais definidas sobre Ω podem ser chamadas legitimamente de variáveis aleatórias, pois é necessário que elas satisfaçam um requisito ligado ao espaço de eventos.

A ideia do uso de funções chamadas variáveis aleatórias simplifica muito a análise de problemas dentro da prática da engenharia, pois elas tendem a esconder o funcionamento de toda a teoria matemática de probabilidades detrás de um véu, em uma situação parecida com a de um motorista não precisar saber o que há debaixo do capô do carro, mas suas preocupações passam a ser outras. Isso significa que se por um lado muito do maquinário matemático fica funcionando ocultamente, outros conhecimentos específicos passam a ser necessários.

A utilidade das variáveis aleatórias fica acentuada quando se sabe que existem modelos prontos que seguem certas leis probabilísticas, que podemos dizer que são padronizadas, como a binomial, a geométrica e a Poisson, entre outras. Com a introdução delas, passamos a discutir distribuições de probabilidades, funções de distribuição acumulada, valores esperados e variâncias. O conceito de independência é retomado nesse novo contexto.

Neste capítulo, tudo ainda é feito no contexto em que o espaço amostral Ω é enumerável, o que facilita a apresentação dos conceitos. No próximo capítulo exploramos o caso de Ω não enumerável.

4.2 Vídeo que acompanha este capítulo

Um vídeo em que resolvemos o problema 7 proposto neste capítulo na página 76 pode ser visto ao escanear-se o código QR abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Exercício sobre variáveis aleatórias discretas.

4.3 Biografia de Pierre-Simon Laplace

Pierre-Simon Laplace [Enc10] [dL14] [OR21] [Bel14] nasceu em Beaumont-en-Auge – França – em 23 de março de 1749. Filho de camponeses, sabe-se que a sua infância e juventude foi marcada por dificuldades materiais, realidade humilde que Laplace sempre tentava ocultar. Com a amizade de vizinhos, teve o apoio para os estudos em um colégio e depois na Escola Militar de Beaumont.

Demonstrava interesse em estudos teológicos, incentivado pela família, e já mostrava grande habilidade matemática, sendo professor dessas disciplinas na escola de sua cidade natal. Com 18 anos, foi para Paris se apresentar a d’Alembert, grande autoridade científica à época, com recomendações de pessoas poderosas, mas sequer foi recebido pelo grande matemático. Numa nova e eficaz investida, enviou uma carta tratando de princípios gerais de mecânica a d’Alembert, que se impressionou com os escritos e imediatamente conseguiu a sua nomeação como professor de matemática para a Escola Militar em Paris.

Em 1773, deu o seu primeiro e enorme passo em mecânica celeste, aplicando as leis de gravitação de Newton ao sistema solar e estudando, principalmente, as perturbações mútuas das órbitas de planetas (Júpiter e Saturno), explicando assim um suposto comportamento instável destas. Demonstrou matematicamente a periodicidade das perturbações gravitacionais considerando as excentricidades e inclinações das órbitas, mostrando a estabilidade destas órbitas, com periodicidades extremamente longas. A

solução matemática dos problemas de interações gravitacionais mútuas no sistema solar revelava-se, até então, impossível. Newton concluíra que a intervenção divina era periodicamente necessária para ajustar as órbitas, mantendo-as estáveis. Como reconhecimento do seu trabalho, Laplace, com 24 anos, foi honrado com a sua nomeação para a Academia de Ciências. Estes estudos foram estendidos para a análise do sistema Lua-Terra-Sol, com uma descrição teórica rigorosa do movimento lunar.

Em 1796, publicou *Exposition du système du monde* – Exposição do sistema do mundo – um livro de divulgação dos seus estudos do movimento dos corpos do sistema solar com uma explicação para a origem do sistema solar a partir de uma nebulosa – nuvem gasosa. No período de 1798 a 1827, publicou a sua mais importante obra, os cinco volumes de *Traité de mécanique céleste* (Tratado de mecânica celeste), apresentando os métodos de cálculo e os resultados precisos da aplicação da lei de gravitação no estudo do movimento dos planetas e seus satélites no sistema solar. Com este livro, Laplace torna-se uma celebridade, o “Newton da França”.

Duas curiosidades relacionadas ao *Traité de mécanique céleste* podem ser lembradas. Uma advém do interesse de Laplace pelos resultados de seus cálculos, em detrimento de como obtê-los, condensando as deduções matemáticas e as justificando com a observação *Il est aisé à voir...* -É fácil ver que... Outra curiosidade é a relação de Laplace com Napoleão Bonaparte, que fora seu aluno na Escola Militar. Laplace presenteou o imperador com o Tratado de Mecânica, que, ciente do trabalho, provocou Laplace questionando por que razão não havia mencionado o autor do universo - Deus - em seu portentoso livro, e este teria respondido “Je n’avais pas besoin de cette hypothèse-là” (Eu não precisei desta hipótese).

Laplace vivenciou o Iluminismo em um período literalmente revolucionário da França - *Liberté, Égalité, Fraternité*. Teve sorte distinta de outros sábios, como Lavoisier - com quem desenvolveu estudos termoquímicos da respiração - e Condorcet. Atuou como professor e na organização da *École Polytechnique* e na *École Normale*. Exerceu diversas atividades públicas e tinha certa desenvoltura no meio político, adaptando-se às circunstâncias. No período napoleônico, teve brevíssima passagem como ministro do interior, e, posteriormente como senador e chanceler. Em 1817 recebeu o título de marquês.

A genialidade de Laplace perpassou o mundo da mecânica celeste, se estendendo para vários ramos da física-matemática, que constituem ferramentas extremamente atuais, tais como: função potencial para a teoria de campos – laplaciano – em eletromagnetismo e mecânica dos fluidos; transformada de Laplace para resolução de equações diferenciais para estudo

principalmente de sistemas dinâmicos; cálculos termoquímicos; propagação do som; teoria de probabilidade clássica.

O livro “The doctrine of Chances”, de autoria de Abraham De Moivre influenciou os trabalhos de Laplace em teoria de probabilidades, que publicou vários artigos e dois textos, o *Théorie analytique des probabilités* (Teoria analítica de probabilidades), em dois volumes, em 1812, e *Essai philosophique sur les probabilités* (Ensaio filosófico sobre as probabilidades), em 1814, obra de divulgação também em dois volumes, que teve notável sucesso. Ambos os textos apresentavam o que hoje se considera como os fundamentos da teoria de probabilidade clássica não axiomática.

O Teoria analítica estuda funções geradoras de momentos, aproximações assintóticas para a Lei dos grandes números (que hoje tratamos no Teorema do Limite Central), aplicações de cálculo de probabilidade em “filosofia natural” e “ciências morais” (questões ligadas às ações humanas). Além de tudo isso, dá uma interpretação mais abrangente para a probabilidade inversa do que a lei de Bayes.

No Ensaio filosófico, sendo um escrito de divulgação, nenhuma equação é apresentada. O texto discorre sobre os princípios gerais do cálculo das probabilidades, os métodos analíticos e aplicações do cálculo de probabilidades em diferentes situações, inclusive ciências morais, jogos, passando por tabelas de mortalidade e julgamentos. As reflexões de Laplace presentes no Ensaio Filosófico são bem abrangentes, ressaltando a importância do cálculo de probabilidades nas mais diversas aplicações nas ciências naturais.

Laplace considerava a astronomia como a ciência mais perfeita e os seus estudos de mecânica celeste podem ser considerados como a quinta-essência do determinismo. A visão determinista da ciência e a probabilidade são harmonizadas, como expresso em trecho do Ensaio filosófico, nas palavras de Laplace: “A curva descrita por uma simples molécula de ar ou vapores é regida de um modo tão certo quanto as órbitas planetárias; não há diferença entre elas, a não ser a posta por nossa ignorância. A probabilidade se deve em parte a essa ignorância, em parte aos nossos conhecimentos”. No século seguinte, a Mecânica Quântica irrompe e abala as convicções da essência da interpretações da probabilidade clássica. Citando o próprio Laplace: “O que sabemos não é muito. O que não sabemos é imenso”.

Laplace morreu em 1827, em Paris.

4.4 Variáveis aleatórias discretas

Até aqui fixamos um procedimento para a análise de um problema probabilístico. Resumidamente, ele segue os passos:

1. Entender do problema e o experimento.
2. Explicitar o espaço amostral Ω .
3. Explicitar o espaço de eventos E .
4. Focalizar os eventos de interesse no problema.
5. Explicitar a função probabilidade $P : E \rightarrow [0, 1]$ e se for o caso, a função probabilidade condicionada a um evento dado $C \subset \Omega$, $P(\cdot | C) : E \rightarrow [0, 1]$.
6. Resolver o problema.

A ideia de variável aleatória simplifica esse procedimento na medida em que através dela são propostos *modelos prontos*.

De forma inicial e breve, dizemos que uma variável aleatória é uma função que tem por domínio o espaço amostral Ω e contradomínio numérico, tipicamente o conjunto dos números reais \mathbb{R} .

Observação 4.4.1. Note que apesar do nome, uma variável aleatória não é uma variável, mas sim uma função. Além disso, o domínio não é E , e sim Ω . O contradomínio não é restrito a $[0, 1]$.

Muitos afirmam que a grande vantagem das variáveis aleatórias é a transformação de um resultado experimental em um número. Que o contradomínio de uma variável aleatória seja um conjunto composto por números é certamente uma vantagem, mas a nosso ver não é a principal. A principal é que ela é uma *função*. Vamos esclarecer esse ponto ao longo do texto, mas em primeiro lugar, note que, por exemplo, o resultado do lançamento de um dado, ou o sorteio de uma bola numerada, já é naturalmente numérico. Que cara e coroa podem ser codificados em 0 e 1, e coisas do tipo, são elementares demais para serem a principal característica e vantagem da introdução da grande ideia das variáveis aleatórias.

4.4.1 A função variável aleatória

Seja um espaço probabilístico com o espaço amostral Ω , o espaço de eventos E e a função probabilidade $P : E \rightarrow [0, 1]$.

Como dissemos, uma variável aleatória é uma função que tem por contradomínio o conjunto dos números reais \mathbb{R} e domínio o espaço amostral Ω , mas ela deve ser uma função que satisfaz um requisito que explicitaremos abaixo. Antes, devemos fazer uma recordação da distinção entre função inversa e imagem inversa de uma função.

Dada uma função $f : W \rightarrow U$, ela tem inversa $f^{-1} : U \rightarrow W$ se f for bijetora. A imagem inversa de um subconjunto $A \subset U$ por f sempre existe, mesmo se f não for inversível. A imagem inversa de $A \subset U$ por f é o conjunto $\{w \in W \mid f(w) \in A\}$ dos elementos do domínio W . Infelizmente, a notação usada para a imagem inversa usa a simbologia f^{-1} :

$$f^{-1}(A) = \{w \in W \mid f(w) \in A\}.$$

Fique, portanto, atento ao uso do símbolo f^{-1} . Ele pode significar função inversa, que nem sempre existe, ou imagem inversa de um subconjunto do contradomínio, que sempre existe (mesmo sendo o conjunto vazio).

Exemplo 4. *Seja a função $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = 1 - x^2$. Essa função não tem inversa, pois não é bijetora. Mas, podemos afirmar que*

$$f^{-1}([0, 1]) = [-1, 1].$$

Podemos até dizer, por exemplo, que

$$f^{-1}([10, 12]) = \emptyset.$$

A imagem inversa de um subconjunto do contradomínio sempre existe.
 \triangle

Definição de variável aleatória. Uma variável aleatória é uma função do tipo

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

tal que a imagem inversa de conjuntos $] - \infty, a[$, $a > 0$ por X é elemento de E , isto é,

$$X^{-1}(] - \infty, a[) \in E, \quad \forall a > 0.$$

Isso significa que podemos calcular probabilidades do tipo $P(X^{-1}(] - \infty, a[))$.

Perceba que como a imagem inversa é um subconjunto do domínio, e o domínio é o espaço amostral Ω , faz sentido dizer que $X^{-1}(] - \infty, a[) \in E$, já que E é o conjunto dos subconjuntos de Ω , no nosso contexto discreto.

O uso de uma variável aleatória para se definir o espaço de eventos E . A essência da variável aleatória é que ela é uma função real definida em Ω cuja imagem inversa de $] - \infty, a[\subset \mathbb{R}$ seja elemento de E . Então em princípio, dada uma $X : \Omega \rightarrow \mathbb{R}$, podemos definir E . Isto é, não começamos com E e perguntamos se X é uma variável aleatória legítima, mas ao contrário. Dada uma X , criamos E para que X seja de fato uma variável aleatória. Devido às propriedades da imagem inversa de uma função (ver (2.1) do capítulo 2), podemos usar uma sigma-álgebra Σ de \mathbb{R} , isto é, Σ é uma coleção de subconjuntos de \mathbb{R} que satisfazem os requisitos (2.2) do capítulo 2, e fazer

$$E = \{X^{-1}(s) \mid s \in \Sigma\}.$$

Isso é possível porque a imagem inversa de uma sigma-álgebra por X é também uma sigma-álgebra. Além disso, X será automaticamente uma variável aleatória.

Abuso de linguagem

Destacamos um abuso de linguagem comum, que o estudante deve ficar atento. Fixamos um experimento com espaço amostral Ω , espaço de eventos E e função probabilidade $P : E \rightarrow [0, 1]$. Nele, consideramos a variável aleatória, que, nunca é demais enfatizar, é uma função $X : \Omega \rightarrow \mathbb{R}$.

A probabilidade do evento $X^{-1}(] - \infty, a]) \in E$ em geral, por um abuso de linguagem, é escrito da seguinte forma

$$P(X^{-1}(] - \infty, a]) = \underbrace{P(X < a)}_{\text{Abuso ling.}}$$

O motivo é a simplicidade da notação, mas o estudante deve ficar sempre atento.

Usando essa notação simplificada, podemos escrever a probabilidade de $X \in [a, b]$ por $P(a \leq X \leq b)$. Por extenso, essa probabilidade ficaria assim:

$$P(a \leq X \leq b) = P(\{u \in \Omega \mid X(u) \in [a, b]\}) = P(X^{-1}([a, b])).$$

Lembrando que valem as relações

$$\begin{aligned} X^{-1}(C \cup D) &= X^{-1}(C) \cup X^{-1}(D), \\ X^{-1}(C \cap D) &= X^{-1}(C) \cap X^{-1}(D), \\ C \subset D &\Rightarrow X^{-1}(C) \subset X^{-1}(D), \\ X^{-1}(B) &= A, \\ X^{-1}(\emptyset) &= \emptyset. \end{aligned} \tag{4.1}$$

vistas no capítulo 1 deste curso, temos então que

$$\begin{aligned}
 \underbrace{P(a \leq X \leq b)}_{\text{Abuso ling.}} &= P(X^{-1}([a, b])) = P(X^{-1}(]-\infty, b] \cap [a, +\infty[)) \\
 &= P(X^{-1}(]-\infty, b]) \cap X^{-1}([a, +\infty[)) \\
 &= \underbrace{P((X \geq a) e (X \leq b))}_{\text{Abuso ling.}}
 \end{aligned}$$

Analogamente, valem expressões do tipo

$$\begin{aligned}
 \underbrace{P(X \in [a, b] \cup [c, d])}_{\text{Abuso ling.}} &= P(X^{-1}([a, b] \cup [c, d])) \\
 &= P(X^{-1}([a, b]) \cup X^{-1}([c, d])) \\
 &= \underbrace{P((X \in [a, b]) \text{ ou } (X \in [c, d]))}_{\text{Abuso ling.}}.
 \end{aligned}$$

A conclusão é que podemos trabalhar com a probabilidade da interseção e união (e diferença) de eventos diretamente na notação simplificada $X \in A \cap B$, $X \in A \cup B$, $X \in A \setminus B$, etc., porque as propriedades (4.1) transformam-nas em operações sobre eventos. Tudo se passa por de trás de um véu que esconde o funcionamento da máquina.

Notação específica – Uso de letras maiúsculas e minúsculas. Depois que alertamos sobre o abuso de linguagem, esclarecemos que há um costume bastante difundido de denotar por uma letra maiúscula as variáveis aleatórias, enquanto denota-se por letras minúsculas o resultado da função variável aleatória. Por exemplo, Y seria uma variável aleatória e $Y(a) = y$ o resultado y da aplicação da variável aleatória Y no elemento do espaço amostral $a \in \Omega$.

Variáveis aleatórias *discretas*

Especializamos a discussão para o caso discreto. Muitos autores dizem que uma variável aleatória discreta é uma função que assume valores em um conjunto enumerável. Nós preferimos dizer que uma variável aleatória discreta é aquela cujo domínio, isto é, o espaço amostral Ω é enumerável. Dessa forma, o contradomínio também será, e evitamos várias questões técnicas quanto ao espaço de eventos E .

De qualquer forma, o estudante deve atentar ao fato que discreto não quer dizer finito e nem discreto no sentido mais coloquial, e sim *enumerável*.

Note que o conjunto dos racionais \mathbb{Q} é enumerável, então ele é perfeitamente elegível (mas está longe de ser o único) para ser o domínio (ou contradomínio) de uma variável aleatória discreta.

4.4.2 Experimentos e variáveis aleatórias

Nesta seção, mostramos como experimentos e variáveis aleatórias interagem, através de um exemplo.

Suponha que desejamos contar o número de caras em $N = 2$ lançamentos de uma moeda, que não sabemos se é viciada ou não. Para fixar as ideias, suponha que essa moeda tenha sido uma sorteada de uma caixa que continha apenas duas, uma viciada e outra honesta, através de um procedimento que dava qualquer uma das duas moedas com igual probabilidade. A moeda viciada dá cara com probabilidade $2/3$.

O experimento em si consiste dos passos:

1. Sortear uma moeda.
2. Lançá-la $N = 2$ vezes.

Denotamos cara por K , coroa por C , viciada por V e honesta por H . Um resultado experimental seria, por exemplo, a quadra (H, K, C) . O espaço amostral, que reúne todos os resultados experimentais é o conjunto

$$\Omega = \{(H, K, K), (H, K, C), \dots, (V, K, C), (V, C, K), (V, C, C)\},$$

que tem 2×2^2 elementos.

O espaço de eventos E é o conjunto das partes de Ω , que contém $2^{2 \times 2^2}$ elementos. Temos dois eventos em particular que têm grande importância. O primeiro é \mathcal{H} “moeda honesta” e o segundo é \mathcal{V} “moeda viciada”. Para deixarmos bem claro:

$$\begin{aligned} \mathcal{H} &= \{(H, K, K), (H, K, C), (H, C, K), (H, C, C)\}, \\ \mathcal{V} &= \{(V, K, K), (V, K, C), (V, C, K), (V, C, C)\}. \end{aligned}$$

Um evento do tipo (cara, cara) é o elemento

$$\{(H, K, K), (V, K, K)\} \in E.$$

Podemos facilmente definir a função probabilidade $P : E \rightarrow [0, 1]$, usando técnicas elementares. Por exemplo, $P(\{(H, K, K)\})$ é a probabilidade de se obter no sorteio uma moeda honesta e com ela obtermos em

seguida duas caras. Então, $P(\{(H, K, K)\}) = \frac{1}{2} \times (\frac{1}{2})^2$. A probabilidade de se obter no sorteio uma moeda viciada e com ela obtermos em seguida uma cara e uma coroa é $P(\{(V, K, C)\}) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{3}$. As outras probabilidades para os eventos elementares podem ser definidas de modo análogo. Todas as probabilidades para os elementos de E são calculadas facilmente.

Agora, definimos a variável aleatória $X : \Omega \rightarrow \mathbb{R}$ que fornece o número de caras, da seguinte forma. Dado o resultado experimental $r \in \Omega$:

1. Contar o número de caras em r .
2. Apagar a informação da ordem de caras e coroas.
3. Reportar o resultado da contagem $x = X(r)$.

O fato de apagarmos a informação sobre a ordem de caras e coroas não é desprezível, mas faz parte do modelo que usamos neste exemplo particular. O problema 7 da seção 4.8 na página 75, que é o tópico do vídeo que acompanha este texto, explora esse ponto.

Tendo em vista a discussão feita na seção 4.4.1 do que é uma variável aleatória, e sobre uso da linguagem informal (abuso de linguagem), podemos ver claramente que o evento $(X = 2)$ é o conjunto

$$(X = 2) = X^{-1}(\{2\}) = \{(H, K, K), (V, K, K)\}.$$

Sua probabilidade é calculada por

$$P(X = 2) = P(\{(H, K, K), (V, K, K)\}) = \frac{1}{2} \times \left(\frac{2}{3}\right)^2 + \frac{1}{2} \times \left(\frac{1}{2}\right)^2. \quad (4.2)$$

O estudante deve notar que a X , que dá o número de caras, correspondem eventos que contém o fato da moeda ser honesta ou viciada. Mais precisamente, a imagem inversa $X^{-1}(\{n\})$, $n = 0, 1, 2$, contém V ou H .

Outro ponto a ser notado é que apesar de X fornecer a quantidade de caras em um número fixo lançamentos de uma moeda, X não é distribuída de acordo com uma Binomial, que veremos abaixo.

4.4.3 Independência de variáveis aleatórias. Independência sob condicionamento.

Fixado o espaço probabilístico Ω , E , $P : E \rightarrow [0, 1]$, e dadas as variáveis aleatórias $X : \Omega \rightarrow \mathbb{R}$ e $Y : \Omega \rightarrow \mathbb{R}$, diremos que X e Y são independentes

se os eventos $X^{-1}(]-\infty, a])$ e $Y^{-1}(]-\infty, b])$ são independentes para todos $a, b \in \mathbb{R}$, isto é, que

$$P(X \leq a | Y \leq b) = P(X \leq a), \quad \forall a, b \in \mathbb{R}. \quad (4.3)$$

O estudante deve notar que se (4.3) vale, então como consequência, temos obrigatoriamente

$$P(Y \leq b | X \leq a) = P(Y \leq b), \quad \forall a, b \in \mathbb{R}, \quad (4.4)$$

pois

$$\begin{aligned} P(Y \leq b | X \leq a) &= \frac{P((Y \leq b) \cap (X \leq a))}{P(X \leq a)} \\ &= \frac{P(X \leq a | Y \leq b)P(Y \leq b)}{P(X \leq a)} = P(Y \leq b). \end{aligned}$$

Voltamos ao exemplo mostrado acima, em que de uma caixa sorteamos uma moeda que pode ser viciada ou honesta com igual probabilidade. Lembramos que a moeda viciada dá cara com probabilidade $2/3$. Após o sorteio da moeda, fazemos com ela dois lançamentos. No exemplo acima, definimos a variável aleatória X definida sobre o espaço amostral Ω , que fornece o número de caras nesses dois lançamentos. Vamos definir agora duas novas variáveis aleatórias. $X_1 : \Omega \rightarrow \mathbb{R}$ que dá o número de caras no primeiro lançamento e $X_2 : \Omega \rightarrow \mathbb{R}$ que faz o mesmo para o segundo lançamento. Para deixarmos bem claro: Realizamos o experimento uma vez e obtemos X_1 e X_2 . Devemos ter $X = X_1 + X_2$.

Perguntamos agora se X_1 e X_2 são independentes. A resposta é *não*.

A função probabilidade $P : E \rightarrow [0, 1]$ já foi dada na seção anterior. Precisamos agora da função probabilidade condicionada, que é obtida a partir de $P : E \rightarrow [0, 1]$.

Relembramos que estamos denotando o evento “moeda honesta” por \mathcal{H} e “moeda viciada” por \mathcal{V} . Observe a notação: $\mathcal{H} \in E, \mathcal{V} \in E$, pois são eventos, mas H e V apenas são designações para o fato da moeda ser honesta ou viciada nos ternos do tipo $(H, K, C), (V, K, C)$.

A probabilidade $P(X = 2 | \mathcal{H})$ de se obter duas caras quando a moeda é honesta pode ser obtida intuitivamente. É claro que $P(X = 2 | \mathcal{H}) = (1/2)^2$, mas ela pode ser também obtida a partir da função de probabilidade não

condicionada $P : E \rightarrow [0, 1]$ definida acima. De fato,

$$\begin{aligned} P(X = 2 | \mathcal{H}) &= P(X^{-1}(\{2\}) | \mathcal{H}) \\ &= P(\{(H, K, K), (V, K, K)\} | \{(H, K, K), (H, K, C), (H, C, K), (H, C, C)\}) \\ &= \frac{P(\{(H, K, K)\})}{P(\{(H, K, K), (H, K, C), (H, C, K), (H, C, C)\})} \\ &= \frac{\frac{1}{2} \times \left(\frac{1}{2}\right)^2}{\frac{1}{2} \times \left(\frac{1}{2}\right)^2 + \frac{1}{2} \times \left(\frac{1}{2}\right)^2 + \frac{1}{2} \times \left(\frac{1}{2}\right)^2 + \frac{1}{2} \times \left(\frac{1}{2}\right)^2} = \left(\frac{1}{2}\right)^2, \end{aligned}$$

como era esperado. O estudante deve ter em mente que o objetivo não era obter um resultado já conhecido, mas sim ver que o *método* funciona, e se ele funciona, ele pode ser aplicado em situações verdadeiramente complicadas.

De modo análogo, obtemos, por exemplo, $P(X = 2 | \mathcal{V}) = \left(\frac{2}{3}\right)^2$. As demais probabilidades que definem a função $P(\cdot | B) : E \rightarrow \mathbb{R}$, para os diversos eventos $B \in E$ são calculados de maneira similar.

É interessante notar que $P(X = 2)$, já calculada em (4.2), pode ser obtida também pela já conhecida fórmula

$$P(X = 2) = P(X = 2 | \mathcal{V})P(\mathcal{V}) + P(X = 2 | \mathcal{H})P(\mathcal{H}).$$

É só verificar.

Tendo as funções $P : E \rightarrow [0, 1]$ e $P(\cdot | \cdot) : E \times E \rightarrow \mathbb{R}$, podemos responder à pergunta se X_1 e X_2 são independentes. A resposta já sabemos: não são.

Calculemos $P(X_2 = 1)$ e $P(X_2 = 1 | X_1 = 1)$, respectivamente a probabilidade de obtermos uma cara no segundo lançamento e a probabilidade de se obter uma cara no segundo lançamento sabendo-se que no primeiro saiu cara.

$$\begin{aligned} P(X_2 = 1) &= P(X_2^{-1}(\{1\})) \\ &= P(\{(H, K, K), (H, C, K), (V, K, K), (V, C, K)\}) \\ &= \frac{1}{2} \times \left(\frac{1}{2}\right)^2 + \frac{1}{2} \times \left(\frac{1}{2}\right)^2 + \frac{1}{2} \times \left(\frac{2}{3}\right)^2 + \frac{1}{2} \times \frac{1}{3} \times \frac{2}{3} \\ &= \frac{7}{12}. \end{aligned}$$

De forma similar, obtemos (fazendo exatamente as mesmas contas!)

$$P(X_1 = 1) = \frac{7}{12}. \quad (4.5)$$

Agora, substituindo os valores obtidos em (4.2) e (4.5), obtemos

$$\begin{aligned} P(X_2 = 1|X_1 = 1) &= \frac{P(X_2 = 1) \cap (X_1 = 1)}{P(X_1 = 1)} \\ &= \frac{P(X = 2)}{P(X_1 = 1)} = \frac{25}{42}. \end{aligned}$$

Como $P(X_2 = 1|X_1 = 1) \neq P(X_2 = 1)$, as variáveis aleatórias X_1 e X_2 não são independentes. Isto é, sair cara no segundo lançamento *não* é independente do fato de sair cara no primeiro. Isso ocorre porque em X_1^{-1} e em X_2^{-1} estão contrabandeadas a informação sobre a moeda ser viciada ou não.

Note atentamente que se soubermos que a moeda sorteada é viciada (ou honesta), então é claro que sair cara no segundo lançamento é mesmo independente do fato de sair cara no primeiro. Esse fenômeno é conhecido como independência sob condicionamento.

4.4.4 Função distribuição acumulada de uma variável aleatória discreta

No contexto em que estamos, tratando de variáveis aleatórias discretas, definimos a *função distribuição acumulada* $F_X : \mathbb{R} \rightarrow [0, 1]$ de uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$ definida em um espaço probabilístico $(\Omega, E, P : E \rightarrow [0, 1])$ por

$$F_X(x) = P(X \leq x) = \sum_{x_j \leq x} P(X = x_j).$$

Na somatória acima, $\sum_{x_j \leq x} P(X = x_j)$ significa que devemos fazer a soma com todos os x_j tais que $x_j \leq x$.

Chamamos a atenção para a desigualdade. Não é $<$ e sim \leq *obrigatoriamente*.

Como as probabilidades são sempre menores ou iguais a 1, a função F_X nunca é decrescente. Além disso, ela apresenta saltos nos pontos $x_j \in \Omega$ em que $P(X = x_j) > 0$. Um gráfico típico é mostrado na figura 4.1. Observe onde estão as bolinhas fechadas nos extremos dos segmentos horizontais. É claro que para qualquer variável aleatória X , valem $\lim_{n \rightarrow -\infty} F_X(n) = 0$ e $\lim_{n \rightarrow +\infty} F_X(n) = 1$.

Quando tratamos de mais de uma variável aleatória, a função distribuição de probabilidade acumulada assume forma análoga, mas bidimensional.

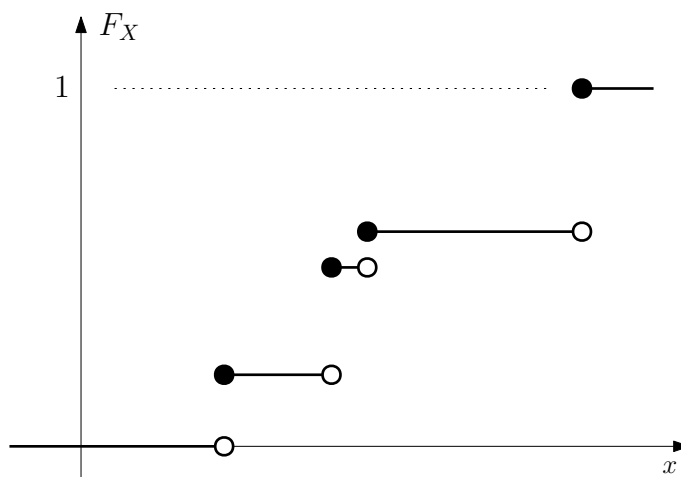


Figura 4.1: Um gráfico típico de função distribuição acumulada.

Por exemplo, a figura 4.2 uma função para duas variáveis aleatórias discretas.

Cada valor no gráfico é calculado por

$$F_{X_1, X_2}(a, b) = P((X_1 \leq a) \cap (X_2 \leq b)) = \sum_{x_1 \leq a, x_2 \leq b} P((X_1 = x_1) \cap (X_2 = x_2)).$$

Marginalização. Quando temos duas variáveis aleatórias X e Y , com função distribuição de probabilidade acumulada $F_{X,Y} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$, perguntar por $F_X(x)$ é o mesmo que perguntar pela probabilidade de $X \leq x$, sem levar em conta o que ocorre com Y . Isto é, para Y , não há nenhuma restrição. Então, podemos escrever

$$F_X(x) = F_{X,Y}(x, +\infty) = \sum_{x_2 \in Y(\Omega)} \sum_{x_1 \leq x} P((X_1 = x_1) \cap (X_2 = x_2)).$$

Esse processo de obtenção de F_X a partir de $F_{X,Y}$ é chamado de *marginalização*.

A independência de variáveis aleatórias revisitada. Em vista da definição de função distribuição acumulada, duas variáveis aleatórias X e Y são independentes quando ocorrer

$$F_{X,Y}(x, y) = F_X(x) \times F_Y(y), \quad \forall (x, y) \in \mathbb{R}^2.$$

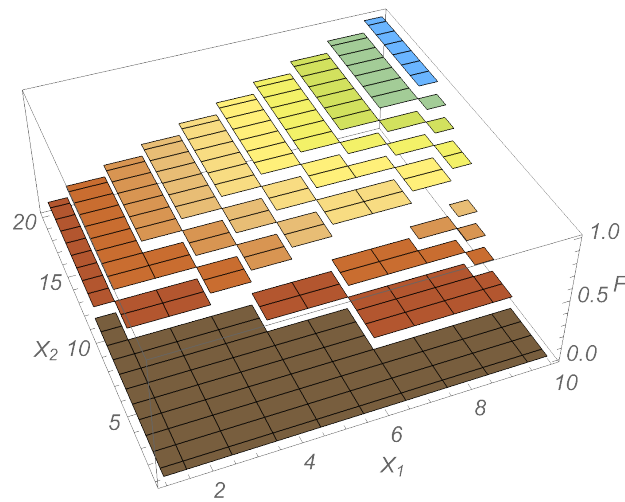


Figura 4.2: Função distribuição de probabilidade acumulada para duas variáveis aleatórias.

No caso de N variáveis aleatórias, X_1, X_2, \dots, X_N essa condição se torna

$$F_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_n) = F_{X_1}(x) \times F_{X_2}(x_2) \times \dots \times F_{X_N}(x_N),$$

Para todos os valores possíveis de x_1, x_2, \dots, x_N . Essa condição é equivalente a testar toda a independência de todas combinações as possíveis dos eventos no espaço de eventos E .

4.4.5 Valor esperado de uma variável aleatória discreta

A ideia de valor esperado é central na teoria de probabilidades. Infelizmente, ele é frequentemente confundido com o de média. Na teoria de probabilidades, a palavra média é usada às vezes como sinônimo de valor esperado, mas devemos ter bem claro quando os significados são distintos. Na teoria de probabilidades, a palavra média vai designar um parâmetro de uma distribuição de probabilidades.

De maneira geral, a palavra média está ligada a amostragem. Tomamos uma amostra e calculamos uma média, assim como fazemos no laboratório de física. A média é algo bem concreto. Por outro lado, o valor esperado é abstrato. Assim como a média surge naturalmente no contexto da amostragem, o valor esperado aparece naturalmente no contexto de jogos e na teoria das decisões.

Suponha que você queira saber o valor justo de uma aposta. Por exemplo, podemos convidá-lo a participar de um jogo em que lançamos uma moeda não viciada. Você escolhe cara ou coroa e jogamos a moeda. Se der o resultado que você havia escolhido, você ganha 1 real. Caso contrário, você não ganha nada. Pensando apenas no aspecto financeiro, qual é o máximo valor que você aceitaria pagar para entrar no jogo? Sendo uma pessoa racional, você responderia 0,5 real. E se você perguntasse a nós, que propusemos o jogo, caso você pagasse 0,4 real, nós aceitaríamos continuar pagando 1 real para ter o direito de participar do jogo? É claro que não, sendo nós professores de probabilidades! O valor justo para entrar no jogo, nas condições dadas acima é exatamente 0,5 real. E o que esse valor tem de especial? Ele é o valor hipotético, abstrato, de uma média global. Esse é o valor esperado, que é calculado usando-se a distribuição de probabilidades, que aliás é uma entidades abstrata, da variável aleatória.

O valor esperado $E[X]$ de uma variável aleatória discreta X que tem distribuição de probabilidades $P(X = x_i)$ é dado por

$$E[X] = \sum_{i \in \mathbb{N}} x_i P(X = x_i), \quad (4.6)$$

quando a soma $\sum_{i \in \mathbb{N}} |x_i| P(X = x_i)$ converge. Se essa soma não convergir, diremos que $E[X]$ não existe.

Exemplo 5. *Neste exemplo, tentamos deixar clara a distinção entre valor esperado e média.*

Tome uma classe de 100 alunos. Suponha que 50% tenha altura 160cm, 30% tenha altura 170cm e o restante tenha 180cm. Medimos a altura de todos os alunos e calculamos a média das alturas. Ela vale

$$m = 0,5 \times 160 + 0,30 \times 170 + 0,20 \times 180 = 167.$$

Essa é a média da altura dos alunos da classe. Se você tomar um grupo de 10 alunos, fazendo um sorteio em que qualquer aluno possa ser sorteado com igual probabilidade, a média da altura desses 10 alunos girará em torno de 167cm, mas muito provavelmente será diferente desse valor.

Suponha agora um experimento em que você sorteie um aluno e meça a altura dele. Seja a variável aleatória X que fornece essa altura. Perguntamos: Qual é o seu valor esperado?

Usando o nosso procedimento padrão de análise de problemas, sabemos que Ω é o conjunto dos alunos da classe que tem 100 pessoas, E é o conjunto das partes de Ω , que tem 2^{100} elementos e $P : E \rightarrow [0, 1]$ atribui o valor $1/100$ a cada elemento $\{a\}$ ($a \in \Omega$) de E .

Vamos calcular a distribuição de probabilidades da variável aleatória X .

$$\underbrace{P(X = 160)}_{\text{informal}} = P(X^{-1}(160)) = 0,5,$$

$$\underbrace{P(X = 170)}_{\text{informal}} = P(X^{-1}(170)) = 0,3,$$

$$\underbrace{P(X = 180)}_{\text{informal}} = P(X^{-1}(180)) = 0,2.$$

Qual é o evento experimental típico? Um conjunto do tipo $\{x\}$, em que x vale 160, 170 ou 180. Observe que se trata de um único valor. Dado esse valor x , qual é a sua média? Ora, x , pois só há um valor.

No entanto, podemos perguntar qual é o valor esperado de X , isto é, $E[X]$. Ele vale

$$E[X] = 160 \times E[X = 160] + 170 \times E[X = 170] + 180 \times E[X = 180] = 167,$$

com a interpretação que realizamos o experimento infinitas vezes, sempre partindo com a classe cheia com 100 alunos. Para cada vez que realizamos o experimento, tomamos um aluno e medimos a sua altura X .

Observe que a interpretação de média e valor esperado são completamente diferentes, apesar de seus valores coincidirem. No exemplo, o valor esperado se refere ao sorteio de um aluno. A média se refere ao conjunto de todos os alunos. \triangle

É importante notar que $E[X]$ é um número!

Neste texto introdutório, não vamos nos alongar sobre o tema, mas gostaríamos de deixar registrado que ele é tão interessante e cheio de nuances, que nem sempre faz sentido se tomar a soma que está na definição (4.6), e por isso há a restrição logo após a definição “quando a soma converge”. Essa restrição matemática é o reflexo da inexistência de sentido na prática. Basicamente, isso está ligado à ordem com que realizamos a soma.

Propriedades do valor esperado

Seja um experimento com espaço amostral Ω , espaço de eventos E , função probabilidade P .

Valem as seguintes propriedades

$$E[\alpha X_1 + \beta X_2] = \alpha E[X_1] + \beta E[X_2], \quad (4.7)$$

para quaisquer variáveis aleatórias X_1 e X_2 (tais que $E[X_1]$ e $E[X_2]$ existam) definidas nesse espaço probabilístico (Ω, E, P) com quaisquer $\alpha, \beta \in \mathbb{R}$. Essas variáveis aleatórias não precisam ser independentes.

Valor esperado de funções de variáveis aleatórias

Vamos ver como se mostra a relação (4.7), e aproveitamos para apresentar uma fórmula muito útil para o cálculo de funções de variáveis aleatórias.

Em primeiro lugar, suponha que no contexto de um espaço probabilístico (Ω, E, P) , tenhamos uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$. Podemos, sem dúvida criar outras variáveis aleatórias a partir dela. Exemplos: X^2 , \sqrt{X} , $\sin(X)$, e^X , ... a imaginação é o limite. Para deixar bem claro: Realizamos um experimento e obtemos um evento $e = \{a\} \in E$, com $a \in \Omega$. Com esse $a \in \Omega$, calculamos $x = X(a)$. Com esse mesmo $x = X(a)$, calculamos a nova variável aleatória, digamos $y = \sin(X)(a) = \sin(X(a)) = \sin(x)$.

Agora podemos perguntar qual é o valor esperado da variável aleatória $f(X)$, em que $f : \mathbb{R} \rightarrow \mathbb{R}$ é uma função real. Para simplificar a notação, vamos escrever $f(X)$ no lugar de $f \circ X$, como os estudantes devem estar acostumados nos cursos de Cálculo, e vamos denotar essa nova variável aleatória por $Y = f(X)$.

Sendo X uma variável aleatória discreta, a imagem da função composta $Y = f \circ X : \Omega \rightarrow \mathbb{R}$ é um conjunto enumerável $K = \{y_1, y_2, \dots\}$.

$$\begin{aligned} E[f(X)] &= \sum_{y_j \in K} y_j P(Y = y_j) = \sum_{y_j \in K} y_j \sum_{x_i: y_j = f(x_i)} P(X = x_i) \\ &= \sum_{x_i} f(x_i) P(X = x_i), \end{aligned}$$

isto é, só para deixarmos registrado, pela sua importância:

$$E[f(X)] = \sum_{x_i} f(x_i) P(X = x_i). \quad (4.8)$$

Valor esperado da soma de variáveis aleatórias. Agora podemos ver por quê (4.7) vale mesmo quando as variáveis não são independentes. De fato, dadas as variáveis aleatórias X e Y , $Z = X + Y$ é uma nova variável

aleatória. Então, de (4.8), obtemos

$$\begin{aligned} E[X + Y] &= \sum_{x_i, y_j} (x_i + y_j) P((X = x_i) \cap (Y = y_j)) \\ &= \sum_{x_i} \sum_{y_j} (x_i + y_j) P((X = x_i) \cap (Y = y_j)) \\ &= \sum_{x_i} x_i \sum_{y_j} P((X = x_i) \cap (Y = y_j)) \\ &\quad + \sum_{y_j} y_j \sum_{x_i} P((X = x_i) \cap (Y = y_j)). \end{aligned}$$

Agora, usando a marginalização, temos

$$\begin{aligned} E[X + Y] &= \sum_{x_i} x_i P(X = x_i) + \sum_{y_j} y_j P(Y = y_j) \\ &= E[X] + E[Y]. \end{aligned}$$

Na demonstração acima, em nenhum passo a independência de variáveis aleatórias foi evocada.

4.4.6 Variância de uma variável aleatória discreta

Poderíamos dar várias justificativas e interpretações para a variância, mas essas interpretações tem problemas inclusive aquela que a variância seria uma medida de dispersão. A interpretação correta é dada pela distribuição Normal, ou Gaussiana, via Teorema do Limite Central. Aqui nos contentaremos em dizer apenas que a variância $V[X]$ de uma variável aleatória X é definida por

$$V[X] = E[(X - E[X])^2],$$

quando os valores esperados existem.

Fazendo uma conta simples, vemos imediatamente que

$$V[X] = E[X^2] - (E[X])^2. \quad (4.9)$$

Além disso,

$$V[\alpha X] = \alpha^2 V[X].$$

e se X e Y forem variáveis aleatórias independentes, vale

$$V[\alpha X + \beta Y] = \alpha^2 V[X] + \beta^2 V[Y].$$

Note que agora, aparece a condição de independência.

O estudante deve ficar atento ao sinal na conta seguinte:

$$V[X - Y] = V[X] + V[Y].$$

As variâncias não se subtraem!

Desvio padrão. Nos cursos de Estatística apresenta-se o desvio padrão σ , que é a raiz quadrada da variância.

$$\sigma = \sqrt{V[X]}.$$

4.5 Modelos de variáveis aleatórias discretas

Depois que trabalhamos todos os detalhes com espaços amostrais, espaços amostrais, funções probabilidades (inclusive condicionada), imagem inversa de funções e funções compostas, apresentamos agora uma maneira de deixar tudo isso debaixo do capô. Fazendo uma analogia, modelos de variáveis aleatórias é como um carro: dirigimos sem conhecer os detalhes do motor.

Esse aspecto é ao mesmo tempo positivo e negativo. Positivo porque simplifica a solução de problemas probabilísticos e negativo porque o estudante não é motivado a se aprofundar na teoria.

Para as variáveis aleatórias abaixo, vamos usar a postura de esconder o espaço amostral, e o espaço de eventos, para passar a *ilusão* de simplicidade. O bom aluno deve estar atento.

4.5.1 Binomial

Uma variável aleatória binomial X com parâmetros $N \in \mathbb{N}$ e $p \in [0, 1]$, assume $X = n$ com probabilidade

$$P(X = n) = \binom{N}{n} p^n (1 - p)^{N-n}, \quad n \in \{0, \dots, N\}. \quad (4.10)$$

A variável aleatória binomial é usada quando o experimento probabilístico consiste em uma sucessão de N ensaios que podem dar somente dois resultados: sucesso ou não sucesso. X fornece o número de sucessos. Cada sucesso ocorre com probabilidade p .

Exemplo 6. Jogamos N vezes uma moeda viciada que dá cara com probabilidade p e coroa com probabilidade $1 - p$. A probabilidade que tenhamos exatamente n caras é dada por (4.10). \triangle

O valor esperado e a variância são dadas por

$$E[X] = Np, \quad V[X] = Np(1-p). \quad (4.11)$$

4.5.2 Geométrica

Suponha um experimento em que realizamos um ensaio em que pode dar somente dois resultados: sucesso ou não sucesso. A sucessão de ensaios continua até que ocorra o primeiro sucesso. A variável aleatória geométrica X fornece o número de *insucessos* até o primeiro sucesso. Para deixar claro: se o primeiro sucesso ocorre já no primeiro ensaio, então $X = 0$. A probabilidade que $X = n$, $n \geq 0$ é dada por

$$P(X = n) = (1-p)^n p. \quad (4.12)$$

Para $n < 0$, as probabilidades são nulas.

Há uma relação interessante para a função distribuição acumulada F_X que se vale das propriedades de uma progressão geométrica.

$$\begin{aligned} F_X(n) = P(X \leq n) &= 1 - P(X > n) = 1 - \sum_{j=n+1}^{+\infty} (1-p)^j p \\ &= 1 - p \sum_{j=n+1}^{+\infty} (1-p)^j = 1 - p \frac{(1-p)^{n+1}}{p} = 1 - (1-p)^{n+1}. \end{aligned}$$

Apenas para deixar registrado:

$$F_X(n) = 1 - (1-p)^{n+1}. \quad (4.13)$$

O valor esperado e a variância de X são dadas por

$$E[X] = \frac{1-p}{p}, \quad V[X] = \frac{1-p}{p^2}. \quad (4.14)$$

Falta de memória Como os resultados da sucessão de ensaios que compõe o experimento são independentes, é de se esperar que o fato de não ter ocorrido um sucesso até o n -ésimo lance, não altere nossas expectativas a

partir do $n + 1$ -ésimo lance em diante. Em outras palavras, o fato de não ter ocorrido uma cara até, digamos o milésimo lance de uma moeda não viciada, não implica que a probabilidade de sair cara no próximo lance tenha aumentado. Esse comportamento é o que chamamos de “falta de memória”.

Em termos matemáticos, dada a distribuição de probabilidades (4.12), podemos construir a partir dela a distribuição de probabilidade condicionada

$$P(X = (n + n_0) | X \geq n_0) = \frac{P(X = (n + n_0))}{P(X \geq n_0)}. \quad (4.15)$$

Mas,

$$P(X \geq n_0) = \sum_{j=n_0}^{+\infty} (1-p)^j p = p \frac{(1-p)^{n_0}}{p} = (1-p)^{n_0}.$$

De (4.15), temos então finalmente que

$$P(X = (n + n_0) | X \geq n_0) = \frac{(1-p)^{n+n_0} p}{(1-p)^{n_0}} = (1-p)^n p = P(X = n).$$

Isto é,

$$P(X = (n + n_0) | X \geq n_0) = P(X = n),$$

comprovando a propriedade de falta de memória.

4.5.3 Poisson

A variável aleatória de Poisson surge quando contamos o número de sucessos por “unidade de volume”, sendo que não há a priori um limite superior para esse número de sucessos. Por exemplo, sabemos que há λ partículas de um tipo qualquer por litro de água. Tomamos uma amostra de um litro e queremos saber a probabilidade que existam n dessas partículas na amostra.

Generaliza-se a “unidade de volume”, que pode ser “unidade de área”, “comprimento”, “tempo”, etc.

Se há n partículas por unidade de volume, digamos 10 partículas por litro, então há 20 partículas em dois litros, e assim por diante.

O (único) parâmetro da distribuição de probabilidades de uma variável aleatória de Poisson é seu valor esperado. Isto é, se X é uma variável aleatória de Poisson, então o único parâmetro é $\lambda = E[X]$.

A distribuição de probabilidades de X , Poisson com parâmetro λ é

$$P(X = n) = \frac{e^{-\lambda} \lambda^n}{n!}, \quad n \geq 0, \quad (4.16)$$

se $n < 0$, então $P(X = n) = 0$.

Por coincidência, a variância de X é igual a λ . Para deixarmos registrado:

$$E[X] = \lambda, \quad V[X] = \lambda. \quad (4.17)$$

Observe que a variável aleatória de Poisson não é limitada superiormente. Então, exemplos em que o limite superior é conhecido a priori não são de Poisson.

4.6 Probabilidades condicionadas e modelos de variáveis aleatórias

Dado um experimento com espaço probabilístico com espaço amostral Ω , espaço de eventos E , em muitas ocasiões quando trabalhamos com modelos de variáveis aleatórias, como por exemplo aquelas descritas na seção 4.5, é mais fácil determinarmos as probabilidades condicionadas do que a função probabilidade $P : E \rightarrow [0, 1]$ dos eventos básicos.

Se pensarmos bem, tudo em um modelo é condicionado a alguma informação, inclusive o próprio modelo. Em particular, o parâmetro de um dado modelo pode ser colocado como um evento condicionante. Por exemplo, podemos dar o parâmetro p da binomial, e assim, a distribuição de probabilidades (4.10) deve ser escrita da forma

$$P(X = n | p) = \binom{N}{n} p^n (1 - p)^{N-n}. \quad (4.18)$$

em que p indica o evento que p é o parâmetro da binomial.

É importante observar que a notação $P(X = n | p)$ é um grande abuso de notação. Em primeiro lugar, $X = n$ é a escrita informal do evento $X^{-1}(\{n\})$. Em segundo lugar, o “p” que aparece após a barra |, indica na verdade um evento $B \in E$ que sinaliza que dentre todas as possibilidades, o parâmetro da binomial é p . Finalmente, note atentamente que B e $X^{-1}(\{n\})$ devem estar e estão mesmo no mesmo espaço E . De fato, lembrando a definição de probabilidade condicionada, devemos fazer a interseção $B \cap X^{-1}(\{n\})$, e essa operação só faz sentido se ambos os subconjuntos participantes estão no mesmo conjunto E .

De qualquer forma, escrever (4.18) é muito prático, e calcular $P(X = n | p)$ mais fácil que calcular diretamente $P(X = n)$, em um contexto em que p segue uma distribuição de probabilidades. Já vimos isso no exemplo da seção 4.4.3, e o exemplo abaixo complementa o ponto.

Exemplo. Suponha que temos 2 moedas em uma caixa. Uma não é viciada, mas a outra dá cara com probabilidade $2/3$. Qualquer moeda da caixa pode ser sorteada com igual probabilidade. Com ela, fazemos 2 lançamentos e obtemos 2 caras. Conhecendo-se esse resultado, qual é a probabilidade que a moeda seja honesta?

Vamos resolver esse problema duas vezes. Na primeira usamos o nosso procedimento padrão, para que todas as engrenagens da teoria sejam mostradas, e na segunda usamos um método informal, que dá o mesmo resultado.

Resolução do exemplo com todos os detalhes. O espaço amostral Ω é formado por elementos do tipo (V, K, K) , (V, C, K) , (H, K, C) , etc., em que V indica que a moeda é viciada, H que é honesta, K que o resultado é cara e C que ele é coroa. O nosso espaço amostral tem 2^3 elementos. O espaço de eventos E é o conjunto das partes de Ω .

O próximo passo é determinarmos as funções de probabilidade que interveem no problema. Ao determinarmos as probabilidades dos eventos elementares $\{e\}$ correspondentes aos elementos $e \in \Omega$, isto é, $P(\{(V, K, K)\})$, $P(\{(V, C, K)\})$, etc., teremos bem definido a função $P : E \rightarrow [0, 1]$. Normalmente, a partir dela, definimos as funções de probabilidade condicionada $P(\cdot | B) : E \rightarrow [0, 1]$, para $B \in E$.

Ocorre que quando trabalhamos com modelos de variáveis aleatórias, como por exemplo, Binomial, Geométrica e a Poisson, às vezes é mais fácil determinarmos as probabilidades condicionadas e em seguida a probabilidade $P : E \rightarrow [0, 1]$. Esse é o caso no presente exemplo.

O evento “Moeda ser viciada” é o conjunto

$$\mathcal{V} = \{(V, K, K), (V, K, C), (V, C, K), (V, C, C)\}.$$

O evento “duas caras” é o conjunto

$$\mathcal{K} = \{(V, K, K), (H, K, K)\}.$$

Determinar $P(\{(V, K, K)\})$, ou seja a probabilidade de a moeda ser viciada e obtermos 2 caras com ela, por exemplo, é mais difícil que determinarmos $P(\mathcal{K} | \mathcal{V})$, pois essa última é dada por um modelo já pronto, que é o da Binomial. Sabendo-se que a moeda é viciada, a probabilidade de se obter duas caras em dois lançamentos é simplesmente

$$P(\mathcal{K} | \mathcal{V}) = \binom{2}{2} (p)^2 (1-p)^{2-2},$$

com $p = \frac{2}{3}$, então fazendo as substituições,

$$P(\mathcal{K} | \mathcal{V}) = \frac{4}{9}. \quad (4.19)$$

Agora, se quisermos calcular $P(\{(V, K, K)\})$, a probabilidade de a moeda ser viciada e obtermos 2 caras com ela, fazemos

$$P(\{(V, K, K)\}) = P(\mathcal{K} \cap \mathcal{V}) = P(\mathcal{K} | \mathcal{V})P(\mathcal{V}).$$

Note que pelo enunciado, $P(\mathcal{V}) = \frac{1}{2}$. Então, de (4.19), temos

$$P(\{(V, K, K)\}) = \frac{4}{9} \times \frac{1}{2} = \frac{2}{9}.$$

De forma análoga, denotando por

$$\mathcal{H} = \{(H, K, K), (H, K, C), (H, C, K), (H, C, C)\}$$

o evento “moeda ser honesta”, podemos calcular $P(\{(H, K, K)\})$ por

$$P(\{(H, K, K)\}) = P(\mathcal{K} | \mathcal{H}) = \left(\frac{1}{2}\right)^2 \frac{1}{2} = \frac{1}{8},$$

pois a moeda é honesta, e as duas moedas da caixa podem ser sorteadas com igual probabilidade (1/2).

A probabilidade de obtermos duas caras, é dada por $P(\mathcal{K})$, que é calculada por

$$P(\mathcal{K}) = P(\{(V, K, K)\}) + P(\{(H, K, K)\}) = \frac{2}{9} + \frac{1}{8}.$$

Finalmente, podemos calcular a probabilidade pedida, que é a de a moeda ser honesta dado que em dois lances obtivemos duas caras.

$$P(\mathcal{H} | \mathcal{K}) = \frac{P(\{(H, K, K)\})}{P(\mathcal{K})} = \frac{\frac{1}{8}}{\frac{2}{9} + \frac{1}{8}} = \frac{9}{25} = 0,36.$$

A probabilidade que seja viciada é $1 - 0,36 = 0,64$.

Resolução do exemplo pelo método informal. Pela definição de probabilidade condicionada,

$$P(\text{“moeda honesta”} | \text{“sair duas caras”}) = \frac{P(\text{“moeda honesta” e “sair duas caras”})}{P(\text{“sair duas caras”})}.$$

O numerador é calculado por

$$P(\text{“moeda honesta” e “sair duas caras”}) = P(\text{“sair duas caras”} \mid \text{“moeda honesta”}) \times \frac{1}{2},$$

pois a moeda honesta é sorteada da caixa com probabilidade $1/2$.

O enunciado nos diz que

$$P(\text{“sair duas caras”} \mid \text{“moeda honesta”}) = \left(\frac{1}{2}\right)^2.$$

O denominador é calculado pelo teorema da probabilidade total, visto no capítulo da aula anterior.

$$\begin{aligned} P(\text{“sair duas caras”}) &= P(\text{“sair duas caras”} \mid \text{“moeda honesta”}) \times \frac{1}{2} \\ &\quad + P(\text{“sair duas caras”} \mid \text{“moeda viciada”}) \times \frac{1}{2}. \end{aligned}$$

$$P(\text{“sair duas caras”} \mid \text{“moeda viciada”}) = \left(\frac{2}{3}\right)^2.$$

Juntando tudo, temos

$$P(\text{“moeda honesta”} \mid \text{“sair duas caras”}) = \frac{\left(\frac{1}{2}\right)^2 \frac{1}{2}}{\left(\frac{1}{2}\right)^2 \frac{1}{2} + \left(\frac{2}{3}\right)^2 \frac{1}{2}},$$

que é o mesmo resultado obtido acima. Entender o funcionamento da teoria é mais importante que o resultado em si, pois afinal de contas, esse é um curso universitário.

4.7 Exemplos de fixação

Para fixarmos as ideias, elencamos agora alguns exemplos de fixação.

Marginalização. Sejam duas variáveis aleatórias X e Y que tem distribuição de probabilidades dada pela tabela 4.1.

Calculemos $P(X = 1)$, $P(X = 2)$, $P(X = 3)$ e $P(Y = 2)$.

As probabilidades pedidas são calculadas por marginalização.

$$P(X = 1) = \frac{1}{36} + \frac{2}{36} + \frac{4}{36} = \frac{7}{36},$$

Tabela 4.1: Distribuição de probabilidades

	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
$Y = 2$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{6}{36}$
$Y = 3$	$\frac{4}{36}$	$\frac{7}{36}$	$\frac{9}{36}$

que simplesmente reflete o fato que

$$P(X = 1) = P((X = 1) \cap (Y = 1)) + P((X = 1) \cap (Y = 2)) + P((X = 1) \cap (Y = 3)).$$

Analogamente,

$$P(X = 2) = \frac{1}{36} + \frac{5}{36} + \frac{7}{36} = \frac{13}{36}.$$

$$P(X = 3) = \frac{1}{36} + \frac{6}{36} + \frac{9}{36} = \frac{16}{36}.$$

$$P(Y = 2) = \frac{2}{36} + \frac{5}{36} + \frac{6}{36} = \frac{13}{36}.$$

Independência de variáveis aleatórias Sejam duas variáveis aleatórias X e Y que tem distribuição de probabilidades dada pela tabela 4.1. X e Y são variáveis aleatórias independentes?

Para responder a essa pergunta, basta verificar se *sempre* vale $P((X = i) \cap (Y = j)) = P(X = i) \times P(Y = j)$. Claramente isso é falso, pois

$$P((X = 1) \cap (Y = 2)) = \frac{2}{36} \neq \frac{7}{36} \times \frac{13}{36}.$$

Função distribuição acumulada. Usando os dados da tabela 4.1, obter o gráfico de F_X .

O estudante deve prestar bastante atenção nos detalhes da figura 4.3:

- Onde estão as bolas abertas e fechadas.
- O menor valor de $F(x)$ é zero e ocorre à esquerda do gráfico (mesmo que com $x \rightarrow -\infty$).

- O maior valor de $F(x)$ ocorre à direita é sempre 1 (mesmo que com $x \rightarrow +\infty$).
- A função tem um número enumerável de descontinuidades.
- Onde a função não é descontínua, a função é constante.
- Os pontos onde as funções são descontínuas correspondem aos valores que a variável aleatória assume.
- A função nunca é decrescente.

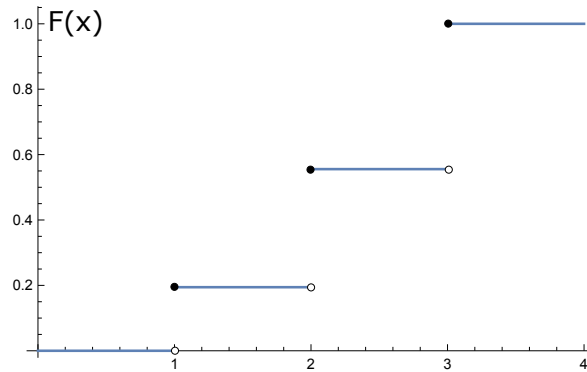


Figura 4.3: Função distribuição acumulada.

Valor esperado. Usando os dados da tabela 4.1, calculemos $E[X + Y^2]$.

Como o valor esperado da soma de variáveis aleatórias é igual à soma dos respectivos valores esperados, temos

$$E[X + Y^2] = E[X] + E[Y^2].$$

Já havíamos calculado os valores de $P(X = 1)$, $P(X = 2)$ e $P(X = 3)$.

$$P(X = 1) = \frac{7}{36}, \quad P(X = 2) = \frac{13}{36}, \quad P(X = 3) = \frac{16}{36}.$$

Desses valores obtemos

$$E[X] = 1 \times \frac{7}{36} + 2 \times \frac{13}{36} + 3 \times \frac{16}{36} = \frac{9}{4}.$$

Precisamos calcular $P(Y = 1)$, $P(Y = 2)$ e $P(Y = 3)$. Temos

$$\begin{aligned} P(Y = 1) &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}, \\ P(Y = 2) &= \frac{2}{36} + \frac{5}{36} + \frac{6}{36} = \frac{13}{36}, \\ P(Y = 3) &= \frac{4}{36} + \frac{7}{36} + \frac{9}{36} = \frac{20}{36}. \end{aligned}$$

A partir desses valores, temos

$$P(Y^2 = 1) = \frac{3}{36}, \quad P(Y^2 = 4) = \frac{13}{36}, \quad P(Y^2 = 9) = \frac{20}{36}.$$

Daí,

$$E[Y^2] = 1 \times \frac{3}{36} + 4 \times \frac{13}{36} + 9 \times \frac{20}{36} = \frac{235}{36}.$$

Só por curiosidade, $E[Y] = \frac{89}{36}$. Note então que $E[Y^2] \neq E[Y]^2 = \frac{7921}{1296}$.

Finalmente,

$$E[X + Y^2] = E[X] + E[Y^2] = \frac{9}{4} + \frac{235}{36} = \frac{79}{9}.$$

Variância. Usando os dados da tabela 4.1, calculemos $V[XY]$.

Começamos com a fórmula básica

$$V[XY] = E[(XY)^2] - E[XY]^2,$$

obtida de (4.9).

$$E[(XY)^2] = \sum_{i=1}^3 \sum_{j=1}^3 x_i^2 y_j^2 P((X = x_i) \cap (Y = y_j)) = \frac{445}{12}.$$

$$E[XY] = \sum_{i=1}^3 \sum_{j=1}^3 x_i y_j P((X = x_i) \cap (Y = y_j)) = \frac{445}{12} = \frac{67}{12}.$$

Então,

$$V[XY] = \frac{445}{12} - \left(\frac{67}{12}\right)^2 = \frac{851}{144}.$$

Acima já havíamos calculado $E[X] = \frac{7}{36}$ e $E[Y] = \frac{7}{36}$. Note que

$$E[XY] = \frac{67}{12} \neq E[X]E[Y] = \frac{623}{1296}.$$

Variável aleatória Binomial. Considere os inteiros na reta real. A distância entre dois inteiros vizinhos é 1. Suponha que coloquemos uma partícula na origem. Jogamos uma moeda não viciada $6N$ vezes. Calcule a probabilidade de a partícula estar em uma posição cuja distância à origem é maior ou igual a $2N$. Calcule também o valor esperado para essa distância e a sua variância.

Há pelo menos três maneiras de se resolver esse problema. O primeiro é pela interpretação direta (e correta) do problema. Na segunda, mudamos o experimento dado por um outro que dá o mesmo resultado, e na terceira o problema é resolvido no estilo informal, como dirigindo um carro sem saber o que tem no tanque de combustível, que também dá a resposta correta. Ao fim tecemos um comentário sobre as posturas que um estudante pode tomar.

Analisemos agora a primeira solução. O experimento fala explicitamente em $6N$ lançamentos de uma moeda não viciada. Então o espaço amostral Ω é composto por resultados experimentais do tipo

$$\underbrace{(K, K, C, C, C, C, K, C, K, K, \dots)}_{6N \text{ termos}} \in \Omega.$$

Observe novamente que não há passado, presente ou futuro. O que existe são pontos, ou elementos de um conjunto.

Se considerarmos apenas os resultados possíveis, Ω tem 2^{6N} elementos. O espaço de eventos E tem $2^{(2^{6N})}$ elementos. A função probabilidade $P : E \rightarrow [0, 1]$ que atribui valores a cada um dos elementos de E é facilmente descrita. Basta informar que cada evento $e \in E$ com um elemento, isto é, e é do tipo $e = \{(K, C, C, C, K, \dots, K)\}$ tem probabilidade $(1/2)^{6N}$, pois a moeda não é viciada.

Definimos a variável aleatória $X : \Omega \rightarrow \mathbb{R}$ que conta o número de caras em um resultado experimental.

Por exemplo, se $6N = 6$, $X((C, K, C, K, C, C)) = 2$, pois o resultado experimental (C, K, C, K, C, C) tem apenas duas caras.

Lembramos que a notação $P(X = 1)$ é um abuso de linguagem, como foi detalhado ao longo do texto. Na verdade, $P(X = 1)$ significa $P(X^{-1}(\{1\}))$, que faz sentido pois $X^{-1}(\{1\}) \in E$. Perceba bem como tudo se encaixa! A questão agora é saber quanto vale $P(X = n)$, para $n \in \{0, \dots, 6N\}$. Muito simples, pois se dissermos que um sucesso é “sair cara”, então $X = n$ é o número de sucessos em $6N$ tentativas, sendo que em uma tentativa a

probabilidade de sucesso é $p = 1/2$. Então,

$$P(X = n) = \binom{6N}{n} \left(\frac{1}{2}\right)^n \left(1 - \frac{1}{2}\right)^{6N-n} = \binom{6N}{n} \left(\frac{1}{2}\right)^{6N},$$

isto é, X é uma variável aleatória com distribuição binomial de parâmetros $6N$ tentativas e probabilidade de sucesso $p = 1/2$. Aproveitamos agora para deixar registrado que

$$E[X] = 3N, \quad V[X] = \frac{3N}{2},$$

conforme as fórmulas dadas em (4.11).

Agora contamos os passos para a direita e esquerda na reta real. Ao todo, andamos para a direita X posições, e para a esquerda $6N - X$ posições. Então, a posição após $6N$ lances da moeda será $Y = X - (6N - X) = 2X - 6N$. Note que criamos uma nova variável aleatória $Y : \Omega \rightarrow \mathbb{R}$ definida por $Y(a) = 2X(a) - 6N$, para todo $a \in \Omega$.

A probabilidade pedida, de que a partícula esteja a uma distância igual ou maior a $2N$ da origem é calculada por

$$P(|Y| \geq 2N) = P((Y \geq 2N) \cup Y \leq -2N) = P(Y \geq 2N) + P(Y \leq -2N).$$

Finalmente,

$$\begin{aligned} P(Y \geq 2N) &= P(2X - 6N \geq 2N) = P(X \geq 4N), \\ P(Y \leq -2N) &= P(2X - 6N \leq -2N) = P(X \leq 2N). \end{aligned}$$

Daí, a probabilidade pedida é

$$P(|Y| \geq 2N) = \sum_{n=4N}^{6N} \binom{6N}{n} \frac{1}{2^{6N}} + \sum_{n=0}^{2N} \binom{6N}{n} \frac{1}{2^{6N}}.$$

Na tabela 4.2 são mostrados alguns resultados concretos.

Tabela 4.2: Alguns resultados para diferentes valores de N

	$N = 5$	$N = 10$	$N = 20$	$N = 50$
$P(Y \geq 2N)$	0.0987371	0.0134893	0.0003304	8.01488×10^{-9}

Falta agora calcularmos o valor esperado e a variância da distância até a origem, isto é, devemos calcular $E[Y]$ e $V[X]$.

$$E[Y] = 2E[X] - 6N = 2 \times 3N - 6N = 0,$$

como já era de se esperar, e

$$V[Y] = 2^2V[X] = 4 \times \frac{3N}{2} = 6N.$$

Segunda solução. Na segunda solução, mudamos ligeiramente o experimento dado. Não consideramos um experimento com $6N$ lances de moeda, mas sim consideramos um experimento com um lance de moeda apenas. Depois, construímos uma variável aleatória que resolve o problema, que envolve os $6N$ lançamentos.

Seja a variável aleatória X_i que assume $X_i = 1$ se o experimento der cara e $X_i = -1$ se der coroa na i -ésima jogada. Isto é, estamos assumindo que $\Omega_i = \{K, C\}$ em cada lance da moeda. Nesse caso, o espaço de eventos E_i tem somente 4 elementos e a função probabilidade é muito simples. $P_i : E \rightarrow [0, 1]$ é descrita conhecendo-se $P_i(\{K\}) = P_i(\{C\}) = 1/2$. Nesse caso, as probabilidades $P_i(X = 1)$ e $P_i(X = -1)$ são muito fáceis de serem calculadas

$$P_i(X = 1) = P_i(X^{-1}(\{1\})) = P_i(K) = \frac{1}{2},$$

e analogamente, $P_i(X = -1) = \frac{1}{2}$.

O valor esperado e a variância são dadas por

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

$$V[X_i] = E[X_i^2] - E[X_i]^2 = 1 - 0^2 = 1.$$

A posição da partícula é dada pela variável aleatória Y dada pela soma

$$Y = \sum_{i=1}^{6N} X_i,$$

que tem valor esperado e variância

$$E[Y] = 0, \quad V[Y] = \sum_{i=1}^{6N} 1 = 6N,$$

como antes. Note que usamos a independência de variáveis aleatórias para calcularmos a variância da soma.

A probabilidade pedida é dada por $P(|Y| \geq 2N)$, exatamente como antes, então deste ponto em diante, a análise é exatamente igual ao que fizemos na primeira solução.

A diferença aqui está no experimento, que modificamos ligeiramente.

Terceira solução, no estilo informal. Vamos ver um outro modo de se resolver o problema, muito usado na prática, que dá o resultado correto. Agora, olhamos apenas para a variável aleatória, que esconde todas as engrenagens debaixo do capô.

Temos o lançamento de uma moeda por $6N$ vezes. Se der cara, avançamos uma posição na reta real. Caso contrário, recuamos uma posição. Logo, com $X = n$ sucessos nesses $6N$ lances teremos avançado n casas e recuado $6N - n$. Então a posição final será $Y = n - (6N - n) = 2n - 6N$. O número de sucessos em $6N$ lances é dada pela distribuição binomial. Dai

$$P(X = n) = \binom{6N}{n} \left(\frac{1}{2}\right)^n \left(1 - \frac{1}{2}\right)^{6N-n}.$$

Para que a partícula fique a uma distância igual ou superior a $2N$ da origem à direita, basta que o número de sucessos n satisfaça $n - (6N - n) \geq 2N$, isto é que $n \geq 4N$.

Para que a partícula fique a uma distância igual ou superior a $2N$ da origem à esquerda, basta que o número de não sucessos n satisfaça $(6N - n) - n \geq 2N$, isto é que $n \leq 2N$.

Isto é, a probabilidade pedida é

$$P(X \geq 4N) + P(X \leq 2N),$$

que é o mesmo resultado obtido pela primeira resolução, só que de uma maneira mais intuitiva e rápida. Aqui não mencionamos o espaço amostral e nem o espaço de eventos.

Comentário sobre as resoluções. Muitas vezes é possível se resolver problemas de forma intuitiva, sem entrarmos no mérito da teoria. Entretanto, esse tipo de resolução fica fora do escopo de um curso universitário. A universidade deve preparar cidadãos capazes de desenvolver novas teorias, ou *pelo menos* ter raciocínio crítico, e para isso, é mesmo necessário uma formação que privilegie a base do conhecimento.

Não se deve procurar apenas o resultado, mas sim buscar as origens, de ver que todas as peças de uma grande teoria se encaixam. Essas peças depois serão usadas na construção de outras teorias mais intrincadas.

Variável aleatória Geométrica. Vamos analisar quanto tempo devemos jogar para ganhar na Mega-Sena. Seja X a variável aleatória que dá o número de insucessos (apostas) até o primeiro sucesso, que é ganhar o

prêmio. Um elemento típico do espaço amostral Ω é $(F, F, F, F, \dots, F, S)$, em que F é falha e S é sucesso. $X : \Omega \rightarrow \mathbb{R}$ fornece o número de F 's antes do primeiro S . A probabilidade que $X = n$ é dada por

$$P(X = n) = P(X^{-1}(\{n\})) = P(\{(F, F, F, F, \dots, F, S)\}),$$

que é simplesmente

$$P(X = n) = (1 - p)^n p,$$

em que p é a probabilidade de ganhar o prêmio em uma aposta.

Suponhamos que façamos uma aposta por semana. A aposta é feita apenas com 6 números. A probabilidade que se sorteiem exatamente os 6 números da nossa aposta é dada pela razão entre os casos favoráveis e os casos possíveis. Daí

$$p = \frac{1}{\binom{60}{6}} = \frac{1}{50063860} = 1.99745 \times 10^{-8}.$$

A fórmula (4.14) nos revela que o valor esperado para o número de semanas antes do primeiro sucesso é

$$E[X] = \frac{1 - p}{p} = 5.00639 \times 10^7 \text{ semanas,}$$

ou seja, um pouco mais que um milhão de anos!

Variável aleatória de Poisson. Assumindo que o número de óbitos por dia por Covid19 (dentro de um certo período relativamente longo) seja aproximadamente constante, e que o número médio de óbitos por dia em Osasco seja 5,2, qual é a probabilidade que não ocorra nenhum óbito em Osasco em um dia fixo?

Para respondermos a essa pergunta, devemos notar que há um “volume” considerado, que é “um dia”, e que dentro desse volume temos um número médio de “partículas”, que é o número médio de óbitos. Estamos dentro do contexto para o uso de uma variável aleatória de Poisson.

Chamemos essa variável aleatória de X . Sabemos que $E[X] = 5,2$. A distribuição de probabilidades é dada por

$$P(X = n) = \frac{e^{-5,2} 5,2^n}{n!}, \quad n \geq 0.$$

como aplicação direta da fórmula (4.16), com $\lambda = E[X] = 5,2$.

Então, para $n = 0$, teremos

$$P(X = 0) = e^{-5,2} = 5,51656 \times 10^{-3}.$$

Mostramos também um gráfico com as probabilidades $P(X_n)$ para $n = 0, \dots, 8$ na figura 4.4.

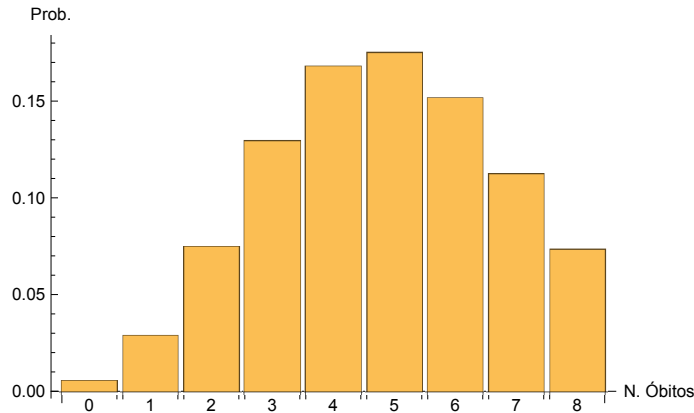


Figura 4.4: Gráfico do exemplo da variável aleatória de Poisson - Covid19.

4.8 Problemas

1. Faça um gráfico da função de distribuição acumulada F_X da variável aleatória geométrica de parâmetro $p = 1/3$.
2. O objetivo deste problema é explorar a diferença entre média de uma amostra e valor esperado de uma variável aleatória. Exploramos também o cálculo da variância de uma variável aleatória.

Lance uma moeda não viciada 10 vezes, atribuindo o valor 1 se sair cara, e 2 se sair coroa. Calcule a média obtida. Defina agora a variável aleatória X que descreve os resultados obtidos, sendo cada lançamento um experimento individual. Calcule o valor esperado $E[X]$. Defina a variável aleatória Y sendo a média amostral dos lançamentos calculada por $Y = \frac{1}{N} \sum_{j=1}^N X_j$. Calcule $V[Y]$.

3. Esse problema explora questão da validação científica de um problema.

Um uma caixa há duas moedas. Uma não viciada, e outra com duas caras. Faz-se o sorteio de uma delas, sendo que ambas podem sair

com igual probabilidade. Com essa moeda, fizemos dois lances e o resultado foi um par de caras. Qual é a probabilidade que a moeda seja a viciada? E se o resultado fosse uma cara e uma coroa?

4. Suponha que temos 10 moedas em uma caixa. A metade é não viciada, mas temos 2 moedas que dão cara com probabilidade $1/3$, e o restante dá cara com probabilidade $2/3$. Qualquer moeda da caixa pode ser sorteada com igual probabilidade. Com ela, fazemos 5 lançamentos e obtemos 2 caras. Conhecendo-se esse resultado, qual é a probabilidade que a moeda seja honesta?
5. O objetivo deste problema é investigarmos a questão da independência sob condicionamento.

Suponha que temos 10 moedas em uma caixa. Seis moedas são honestas e quatro são viciadas. A moeda viciada dá cara com probabilidade $2/3$. Qualquer moeda da caixa pode ser sorteada com igual probabilidade. Sorteamos uma e com ela, fazemos 2 lançamentos. Calcule:

- a) A probabilidade que no segundo lançamento tenhamos uma cara.
 - b) A probabilidade que no segundo lançamento tenhamos uma coroa sabendo-se que no primeiro lançamento obtivemos uma cara.
6. Sabemos que uma variável aleatória geométrica X tem a propriedade da falta de memória. Isso significa que a probabilidade que X seja muito maior que $aE[X]$, com $a > 1$, isto é $P(X > aE[X])$, independe de a ? Então o que significa X ter a propriedade da falta de memória?
 7. Este problema é resolvido no vídeo que acompanha este capítulo. Ele pode ser visto ao escanear-se o QR code abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Exercício sobre variáveis aleatórias discretas.

Ele explora a influência dos modelos de variáveis aleatórias sobre os resultados obtidos, e chama a atenção sobre tipo de informação que há nas variáveis aleatórias binomial e geométrica.

Em uma caixa há duas moedas, que podem ser sorteadas com igual probabilidade. Uma delas não é viciada e a outra dá cara com probabilidade $2/3$. Realizamos três experimentos. No primeiro, sorteamos uma moeda da caixa, lançamo-la até que saia uma cara. O resultado foram duas coroas antes da primeira cara. No segundo, lançamos a moeda três vezes e o resultado foi que houve duas coroas. No terceiro experimento, lançamos a moeda também três vezes, e o resultado foi a sequência coroa, coroa, cara. Em cada um dos experimentos, qual é a probabilidade que a moeda sorteada da caixa seja viciada? Suponha que seu objetivo nesses experimentos tenha sido obter evidências que a moeda seja, na verdade, honesta.

8. Uma pessoa carregava três folhas com dados sobre o número de casos de Covid19 e o número de óbitos. Cada folha correspondia a uma cidade diferente do Estado de São Paulo sem identificação, sendo os dados relativos a um único dia. Os dados estão resumidos na tabela 4.3.

Tabela 4.3: Dados sobre o número de óbitos em 3 cidades em um único dia

	Cidade A	Cidade B	Cidade C
N. casos	43	141	42
N. óbitos	3	1	5

Sabe-se que A, B e C são certamente Osasco, Piracicaba e Jundiaí, não necessariamente nesta ordem.

Assumindo que o número de óbitos por dia seja aproximadamente constante (dentro de um certo período relativamente longo), e que o número médio de óbitos por dia para Osasco, Piracicaba e Jundiaí sejam respectivamente 5, 2, 2, 1 e 2, 7, qual é a probabilidade que a cidade A seja Osasco? Piracicaba? Jundiaí?

Capítulo 5

Variáveis aleatórias contínuas

5.1 Introdução

Neste capítulo temos dois objetivos. O primeiro e mais importante é mostrar que existem muito casos em que uma teoria de probabilidades que só contemplem espaços amostrais enumeráveis não é satisfatória. Veremos que quando os espaços amostrais tem infinitos elementos, mesmo em uma quantidade enumerável, a definição de uma função de probabilidade em um espaço de eventos pode ser problemática. Como exemplo, mostramos que a probabilidade de se escolher “completamente ao acaso” um número par entre os naturais não é $\frac{1}{2}$. Por causa das dificuldades na definição da função de probabilidades, a expressão “completamente ao acaso” é em muitos casos desprovida de sentido, como será mostrado no texto e discutido no vídeo que o acompanha.

Para expandirmos o alcance da teoria de probabilidades, lançamos mão de funções densidade de probabilidade, através das quais definimos funções de probabilidades para os casos em que o espaço amostral não é enumerável. Devido às restrições ligadas à integral de Riemann, que é a que temos à disposição no Ciclo Básico de um curso de Engenharia, o espaço de eventos deve ser necessariamente restrito a uma sigma-álgebra particular. Aqui fazemos uma homenagem a partir de uma observação cínica de G. H. Hardy feita em seu livro *Em defesa de um matemático* [Har00], que um engenheiro ficaria feliz com um π com dez casas decimais.

Vamos trabalhar com conceitos já vistos para o caso discreto, como valor esperado, variância, e independência, mas agora em um contexto mais geral.

O segundo objetivo, com importância relativa menor que a primeira, é apresentar três modelos de variáveis aleatórias contínuas, que são a uni-

forme, a exponencial e a normal. Devido ao escopo deste texto, que é esclarecer os aspectos mais delicados da teoria, não nos retemos muito no desenvolvimento desses modelos, e remetemos os leitores à obras mais genéricas como por exemplo, [Dan13].

5.2 Vídeos que acompanham este capítulo

Este texto acompanha a aula sobre variáveis aleatórias contínuas do curso de reoferecimento Probabilidades para a Escola Politécnica da Universidade de São Paulo. Acompanham o texto dois vídeos, que podem ser vistos escaneando-se os códigos QR abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Sobre a vacuidade da expressão “completamente ao acaso”.



Soma de variáveis aleatórias

5.3 A problemática atribuição de probabilidades quando o espaço amostral tem infinitos elementos

É fácil construirmos experimentos aleatórios que dão resultados contra-intuitivos quando trabalhamos com espaços amostrais Ω que não são finitos, mesmo enumeráveis.

5.3.1 Caso quando Ω é enumerável

Usando uma linguagem bem informal, considere o experimento de se sortear “completamente ao acaso” um número natural, isto é, do conjunto \mathbb{N} . Qual é a probabilidade que ele seja par? Já adiantamos que a priori a resposta $1/2$ é fruto de uma ilusão. Ela é uma possível resposta, mas não a única. De fato, honestamente a resposta pode ser qualquer número entre zero e um.

A expressão “completamente ao acaso” em geral é muito pouco informativa, como o exemplo abaixo indica. Retornaremos a essa questão no exemplo de fixação da página 102.

A probabilidade de sortearmos um número par “completamente ao acaso” dentre os naturais não é $1/2$. Para entendermos o que ocorre, vamos aplicar o nosso método de análise, explicitando o espaço probabilístico $(\Omega, E, P : E \rightarrow [0, 1])$. O experimento é claro. O espaço amostral é o conjunto $\Omega = \{1, 2, 3, \dots\}$. Sendo Ω enumerável, podemos tomar E , o espaço de eventos, como sendo o conjunto das partes de Ω , mas como já sabemos, podemos também tomar para E qualquer uma outra sigma-álgebra.

O próximo passo seria a análise da função probabilidade $P : E \rightarrow [0, 1]$, mas neste ponto devemos chamar a atenção para a ilusão mencionada acima, que induziria à resposta $P(\text{“par”}) = 1/2$.

A origem dessa ilusão está em se esquecer que Ω é apenas um conjunto, e um conjunto é apenas uma coleção sem imposição de ordem sobre seus elementos. Por causa da ordem $1, 2, 3, \dots$, somos induzidos pela aparente simetria, responder $P(\text{“par”}) = 1/2$. Entretanto, podemos também apresentar Ω de outra forma. Colocamos os ímpares e os pares em linhas separadas, rearranjando-os tomando apenas um ímpar antes de tomar dois pares, conforme ilustrado na figura 5.1.

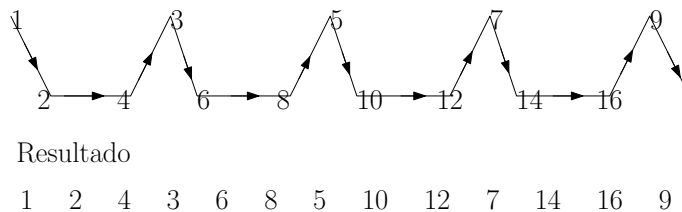


Figura 5.1: Reordenamento dos naturais.

Agora a densidade dos pares é de dois em três números. Logo, agora

tenderíamos a responder que probabilidade $P(\text{“par”})$ seria $2/3$ e não mais $1/2$.

A manipulação acima é possível porque o conjunto dos números naturais contém infinitos elementos. A tentativa de se responder que a probabilidade de se sortear um número par dentre os naturais seria $1/2$ porque metade deles é par, também não funciona, pois a metade de infinito continua sendo infinito e a razão ∞/∞ não está definida.

O estudante poderia ainda tentar atribuir exatamente a mesma probabilidade $\epsilon > 0$ a cada número natural, e somar os ϵ 's correspondentes aos pares para dar uma resposta, mas essa estratégia também não funciona, pois para qualquer $\epsilon > 0$ teremos

$$\sum_{n \in \mathbb{N}} \epsilon = +\infty.$$

Aproveitamos a ocasião para acabar de uma vez por todas com a *fake news* de que dx é um número muito pequeno. Essa fantasia simplesmente é falsa, assim como tomar um $\epsilon > 0$ tendendo a zero. Para deixar claro: também é uma ilusão, pois $\lim_{\epsilon \rightarrow 0} \epsilon = 0$, e nada mais.

Para salvar a ilusão de que $P(\text{“par”}) = 1/2$, podemos usar várias estratégias, todas arbitrárias. Aqui mostramos apenas duas.

Primeira estratégia. Usamos uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$ definida por $X(n) = 0$ se n for par e $X(n) = 1$ se n for ímpar. Usamos X para definir o espaço de eventos, que é uma sigma-álgebra, ao invés de simplesmente dizer que E é o conjunto das partes de Ω , como já foi mencionado na seção 4.4.1 do capítulo 4, mais precisamente na página 47. A imagem de X é apenas o conjunto $\{0, 1\}$. Tome o conjunto das partes de $\{0, 1\}$ e use-o para definir o espaço de eventos. Ele será o conjunto $E = \{X^{-1}(\{0, 1\}), X^{-1}(\{0\}), X^{-1}(\{1\}), X^{-1}(\emptyset)\}$. Note que as uniões enumeráveis, interseções e complementos de elementos de E ainda resultam em outros elementos de E , e daí E é uma sigma-álgebra.

Agora, definimos a função probabilidade $P : E \rightarrow [0, 1]$ por $P(\{0\}) = \frac{1}{2}$ e $P(\{1\}) = \frac{1}{2}$. Com esses valores, temos $P(X^{-1}(\{0, 1\})) = 1$ e naturalmente $P(X^{-1}(\emptyset)) = 0$.

Dessa forma, a probabilidade de se sortear um número par é $1/2$. Note que não atribuímos probabilidades a cada evento do tipo $\{n\}$ com $n \in \mathbb{N}$, mas sim coletivamente. Aqui é impossível se responder a questões como saber a probabilidade de se sortear um número, digamos menor

que 10, pois esse fato não é um elemento de E , em outras palavras, não é um evento.

Segunda estratégia. Atribuímos probabilidades aos eventos $\{n\} \in E$, com $n \in \mathbb{N}$, de tal forma que

$$\sum_{n=1}^{+\infty} P(\{2n\}) = \sum_{n=1}^{+\infty} P(\{2n-1\}) = \frac{1}{2}.$$

Nesse caso, $P(\text{“par”}) = P(\text{“ímpar”}) = 1/2$.

Há várias formas de se fazer isso. Sabendo-se que

$$\sum_{n=1}^{+\infty} \frac{1}{2^n} = 1, \quad (5.1)$$

uma dessas formas seria atribuir a probabilidade $\frac{1}{2^n}$ de se sortear um número no conjunto finito $C_{10}(n) \doteq \{10(n-1)+1, 10(n-1)+2, 10(n-1)+3, \dots, 10(n-1)+10\} \subset E$ com dez elementos e impor que sabendo-se que um número foi sorteado em $C_{10}(n)$, então a probabilidade de se sortear um par é $1/2$. Isso é, usando o teorema da probabilidade total, teríamos

$$\begin{aligned} P(\text{“par”}) &= \sum_{n \in \mathbb{N}} P(\text{“par”} | C_{10}(n)) P(C_{10}(n)) \\ &= \sum_{n \in \mathbb{N}} \frac{1}{2} \frac{1}{2^n} = \frac{1}{2} \sum_{n \in \mathbb{N}} \frac{1}{2^n} = \frac{1}{2}. \end{aligned}$$

Naturalmente, essa é apenas uma das maneiras arbitrárias de se fazer com que $P(\text{“par”})$ seja, por um capricho, igual a $1/2$.

A probabilidade de se sortear um racional dentro do intervalo $[0, 1]$ é zero. A discussão acima já sinaliza que quando trabalhamos com conjuntos infinitos, mesmo quando são enumeráveis, tudo precisa ser feito com muito cuidado. Inquirimos agora qual é a probabilidade de se sortear um número racional do intervalo $[0, 1]$, quando dizemos que a probabilidade de se sortear qualquer um número pertencente a um intervalo $[a, b] \subset [0, 1]$ é simplesmente $(b - a)$. Então é claro que a probabilidade de se sortear um número no intervalo $[0, 1]$ é $(1 - 0) = 1$. O que estamos perguntando é qual é a probabilidade de se sortear um número no conjunto $\mathbb{Q} \cap [0, 1]$, ou em símbolos, $P(\mathbb{Q} \cap [0, 1])$. Note que aqui não estamos usando a integração

de Riemann, porque esse cálculo é simplesmente impossível quando se usa essa integral.

Antes de iniciarmos a discussão, relembremos que o conjunto dos racionais entre 0 e 1 é denso em $[0, 1]$ no sentido que na vizinhança de qualquer elemento não racional de $[0, 1]$ existe pelo menos um racional. Há uma quantidade infinita, mas enumerável de racionais em $[0, 1]$, como já mostrado no capítulo 1 deste curso.

Apesar dos racionais em $[0, 1]$ serem muitos (infinitos), a probabilidade $P(\mathbb{Q} \cap [0, 1])$ é exatamente zero. Não muito pequena, mas realmente zero. Para ver a razão, nos valem novamente de (5.1).

Como $\mathbb{Q} \cap [0, 1]$ é um conjunto enumerável (reveja o capítulo 1 deste curso), podemos fazer uma ordenação dos seus elementos dizendo qual deles é o primeiro, o segundo, e assim por diante. Escolhemos $\epsilon > 0$ arbitrariamente pequeno e cobrimos o primeiro elemento, isto é, colocamo-lo dentro de um intervalo de comprimento $\epsilon \frac{1}{2}$. Cobrimos o segundo elemento com um intervalo de comprimento $\epsilon \frac{1}{2^2}$, o terceiro com um intervalo de comprimento $\epsilon \frac{1}{2^3}$ e assim por diante. Logo, a probabilidade de sortearmos um racional em $[0, 1]$ fica estritamente majorada pela soma

$$\epsilon \sum_{n=1}^{+\infty} \frac{1}{2^n} = \epsilon,$$

isto é $P(\mathbb{Q} \cap [0, 1]) < \epsilon$.

Como $\epsilon > 0$ era arbitrário, a única possibilidade é termos

$$P(\mathbb{Q} \cap [0, 1]) = 0.$$

Então, apesar de não intuitivo, quando o espaço amostral Ω é enumerável, e portanto quanto estamos no caso de variáveis aleatórias discretas, é possível atribuir probabilidades não nulas a todos os eventos de E do tipo $\{x\}$, com $x \in \Omega$.

No exemplo acima, o estudante poderia imaginar se não seria possível atribuir uma probabilidade não nula a cada evento do tipo $\{x\}$, $x \in [0, 1]$ com probabilidade “rapidamente decrescente” e somarmos as probabilidades quando $x \in [0, 1]$, para obtermos diretamente $P([0, 1])$. A resposta é não.

5.3.2 Caso em que Ω não é enumerável. Necessidade de uma teoria específica para espaços amostrais não enumeráveis

Suponha que fosse possível atribuir probabilidades não nulas a todos os eventos do tipo $\{x\}$, com $x \in \Omega$, quando Ω não é enumerável. Então, definindo

$$\Omega_n \doteq \{x \in \Omega \mid P(\{x\}) > \frac{1}{n}\},$$

teremos necessariamente que

$$\Omega = \bigcup_{n \in \mathbb{N}} \Omega_n. \quad (5.2)$$

Se o requisito $P(\Omega) = 1$ deve ser satisfeito, então, cada um dos conjuntos Ω_n tem um número finito de elementos, aliás, no máximo n , pois senão $P(\Omega) = 1$ jamais poderia ser verdade.

Logo, por (5.2), como a união enumerável de enumeráveis é também enumerável, temos que Ω deve ter no máximo um número enumerável de elementos x tais que a probabilidade $P(\{x\})$ é não nula.

Isso significa, por exemplo, que atribuir probabilidades não nulas a todos os elementos de um intervalo do tipo $[a, b] \subset \mathbb{R}$ não é possível. Em suma, o procedimento que havíamos adotado no caso de variáveis aleatórias discretas, de especificarmos o conjunto das partes de Ω para o espaço de eventos E , é simplesmente impossível. Então é necessária uma nova abordagem para tratar de variáveis aleatórias não discretas.

5.4 Variáveis aleatórias contínuas

Vamos tratar de um caso muito específico de um tipo de variáveis aleatórias que não são discretas, as chamadas variáveis aleatórias contínuas. O estudante deve atentar ao fato que uma variável aleatória não ser discreta não significa que ela seja contínua. É também falso que se uma variável aleatória não é contínua então ela é discreta. As variáveis aleatórias contínuas são um caso especialíssimo de não discretas!

Suponha um experimento cujo espaço amostral seja Ω com uma quantidade não enumerável de elementos. O próximo passo é a especificação do espaço de eventos. Para isso, consideramos preliminarmente uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$, que, como já sabemos, é uma função, e que assume

valores não nulos em um subconjunto não enumerável de \mathbb{R} . Desejamos especificar $P(X \leq a)$ para todo $a \in A \subset \mathbb{R}$, em que $A \subset \mathbb{R}$ é um subconjunto de números reais conveniente¹. Esse desejo nos leva ao espaço de eventos.

Da teoria que desenvolvemos no capítulo 4, sabemos que $(X \leq a)$ é um abuso de linguagem que se refere ao conjunto $\{\omega \in \Omega \mid X(\omega) \leq a\} = X^{-1}(]-\infty, a])$, com $a \in A$, como explicitado na página 47. Como desejamos calcular a sua probabilidade, ele deve ser um elemento do espaço de eventos. Para deixarmos claro: $X^{-1}(]-\infty, a]) \in E$ é um requisito para que $X : \Omega \rightarrow \mathbb{R}$ seja uma variável aleatória.

O caso de variáveis discretas é simples o suficiente para que no caso padrão possamos especificar o espaço de eventos como sendo o conjunto das partes de Ω . Dessa forma, a imagem inversa de qualquer subconjunto de \mathbb{R} por qualquer função definida em Ω será elemento do espaço de eventos E . Entretanto, quando Ω não é enumerável, seu conjunto das partes contém tantos elementos que lá comparecem subconjuntos de Ω para os quais não é possível atribuir um sentido para suas probabilidades. Como a função probabilidade tem como domínio o espaço de eventos E , E deve conter somente conjuntos para os quais a probabilidade faz sentido. Dada uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$, vamos usá-la para construir E na seção 5.4.1, sempre lembrando que nosso desejo está focado no cálculo de coisas a partir de $P(X \leq a)$ para todo $a \in A \subset \mathbb{R}$.

Neste ponto, intervém o que dissemos anteriormente sobre um curso universitário. O estudante de ter postura crítica, especialmente em momentos distópicos como aquele que vivemos nesses anos 2020. A razão é que desejamos ter o poder de calcular probabilidades do maior número possível de eventos. Entretanto, esse poder está limitado às armas que um curso de cálculo para engenharia fornece. No caso presente, estamos limitados pelo escopo da integral de Riemann. Por essa razão, não podemos colocar no espaço de eventos, por exemplo, o conjunto $X^{-1}(\mathbb{R} \setminus \mathbb{Q})$, que é subconjunto de Ω que fornecem X irracional, pois mais tarde não será possível calcular a sua probabilidade, como veremos. Então, não procederemos como em outros cursos de probabilidades, em que se especifica uma sigma-álgebra de eventos bastante geral, ou ainda em outros que não explicitam E e nem se preocupam com a possibilidade de se calcular probabilidades quando temos à disposição somente a integral de Riemann. Faremos algo mais modesto, com o compromisso de termos um curso universitário coerente.

Por exemplo, a integral de Riemann da função de Dirichlet $f : [0, 1] \rightarrow \mathbb{R}$

¹A necessidade da introdução desse conjunto $A \subset \mathbb{R}$ vem da limitação de termos à disposição somente a integral de Riemann em um curso de Engenharia.

definida por

$$f(x) = \begin{cases} 0, & x \in [0, 1] \cap \mathbb{Q}, \\ 1, & x \in [0, 1] \setminus \mathbb{Q}, \end{cases} \quad (5.3)$$

não existe, como o estudante que já fez um curso de Cálculo I pode rapidamente verificar. Como esse é um ponto importante, faremos uma rápida recordação para mostrar que f definida acima não é Riemann-integrável.

Rápida recordação da integral de Riemann. Considere o intervalo $[a, b] \subset \mathbb{R}$ e um conjunto de pontos em $[a, b]$ com um número finito de elementos $\{x_0 = a, x_1, \dots, x_n = b\} \subset [a, b]$ qualquer que satisfaça $x_0 \leq x_1 \leq \dots \leq x_n$. Qualquer conjunto P com essas propriedades é uma partição de $[a, b]$. Se Q for uma partição de $[a, b]$ tal que $P \subset Q$, então dizemos que Q é um refinamento de P .

Considere uma função não negativa e limitada $g : [a, b] \rightarrow [0, +\infty[$, $g \leq M$. Uma soma de Riemann $S_{P,g}$ correspondente à partição P e à função g é qualquer número no conjunto

$$\left\{ \sum_{i=1}^n g(\xi_i)(x_i - x_{i-1}) : \xi_i \in [x_{i-1}, x_i] \right\}.$$

Dizemos que g é Riemann integrável em $[a, b]$ se dado qualquer $\epsilon > 0$, existe um número I e uma partição P_ϵ tal que para toda partição Q que refina P_ϵ , vale

$$|I - S_{Q,g}| < \epsilon,$$

para toda soma de Riemann correspondente a Q e g . Quando I existe, I é a integral de g no intervalo $[a, b]$, denotada por $\int_{[a,b]} g(x) dx$.

Agora podemos ver claramente porque a função de Dirichlet (5.3) não é Riemann integrável. Dada qualquer partição de $[0, 1]$, podemos sempre exibir somas de Riemann que valem 1 e 0, e daí o tal número I , que daria a integral de f não existe.

5.4.1 Construção do espaço de eventos E a partir de a imagem inversa de uma função definida em Ω . Abordagem para um curso de engenharia

G. H. Hardy em seu livro *Em defesa de um matemático* [Har00], quando comenta sobre a beleza e a seriedade de Teoremas, ele comenta, um pouco

cinicamente, qual seria o grau de precisão que os engenheiros precisariam, e disse “sejam generosos se o fixarmos em dez algarismos”. Seguindo em seu texto ele diz que para um engenheiro, o número π poderia ser muito bem 3.14159265.

Devido à limitação imposta pelo uso da integral de Riemann, que é a que temos à disposição, vamos levar a sugestão de Hardy a sério, e consideremos $A \subset \mathbb{Q} \subset \mathbb{R}$, A contendo somente números reais cuja representação decimal tem apenas $N_{prec} = 10$ casas decimais.

Para construir o espaço de eventos E , a partir de uma função $X : \Omega \rightarrow \mathbb{R}$, considere intervalos reais do tipo $] - \infty, a]$ e $] - \infty, a[$, com $a \in A$.

Para simplificar a discussão, a partir desse ponto assumimos que $\Omega = \mathbb{R}$ a função $X : \Omega \rightarrow \mathbb{R}$ é a identidade.

Lembramos que uma sigma-álgebra \mathcal{F} de subconjuntos de \mathbb{R} é uma coleção de conjuntos que satisfaz três condições.

1. $\mathbb{R} \in \mathcal{F}$.
2. $C \in \mathcal{F} \Rightarrow \mathbb{R} \setminus C \in \mathcal{F}$.
3. $\{C_n\}_{n \in \mathbb{N}} \subset \mathcal{F} \Rightarrow \bigcup_{n \in \mathbb{N}} C_n \in \mathcal{F}$.

Tome agora a intersecção Σ de todas as sigma-álgebras que contém os intervalos $] - \infty, a]$ e $] - \infty, a[$ com $a \in A$. Agora forme um conjunto com a imagem inversa, por $X : \Omega \rightarrow \mathbb{R}$, de cada elemento de Σ . O resultado é o nosso espaço de eventos E para um curso de engenharia que dispõe somente da integral de Riemann para fazer integrações. Isto é,

$$E = \{X^{-1}(s) \mid s \in \Sigma\}.$$

Com as ferramentas de integração de que dispomos não seria possível colocarmos em Σ elementos da forma $\{a\}$, $a \in \mathbb{R}$, pois então inevitavelmente $\mathbb{Q} \in \Sigma$, e não poderemos calcular a probabilidade eventos do tipo, por exemplo $[0, 1] \cap \mathbb{Q}$ com a integral de Riemann, como vimos em (5.3). O estudante poderia perguntar por que desejamos colocar uniões enumeráveis em Σ . A razão é que uma das condições para que uma função seja uma função de probabilidade $P : E \rightarrow [0, 1]$ é que

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n),$$

quando $m \neq n \Rightarrow A_n \cap A_m = \emptyset$, então devemos ser capazes de calcular probabilidades de uniões enumeráveis. O mesmo tipo de comentário vale para complementações.

Confessamos que E não tem um aspecto elegante, mas como disse Hardy, é o suficiente.

Esses formam o conjunto para os quais estamos somos capazes de calcular probabilidades referentes à variável aleatória X , por exemplo $P(X \leq a)$ (que, lembramos, é um abuso de linguagem para $P(X^{-1}(] - \infty, a]))$), ou $P(X \in]a, b])$, quando $a, b \in A$.

Infelizmente, rigorosamente, neste curso não podemos falar de $P(X = a)$, quando X for uma variável aleatória contínua e a for um número qualquer em \mathbb{R} . O máximo de resolução que podemos ter é perguntar por $P(X = a)$, quando a é um número racional cuja representação decimal tem apenas N_{prec} casas decimais, ou ainda *interpretar* a probabilidade que X seja a , como “aproximadamente a ”, dentro da precisão dada por N_{prec} . Esse é o preço que pagamos por manter a simplicidade da integral de Riemann.

5.4.2 Função probabilidade

O próximo passo é determinarmos a função probabilidade $P : E \rightarrow [0, 1]$ para a variável aleatória $X : \Omega \rightarrow \mathbb{R}$, que assume valores em um conjunto não enumerável. Como já vimos na seção 5.3.2, como $X(\Omega) = \{X(\omega) | \omega \in \Omega\}$ não é enumerável, não é possível atribuímos uma probabilidade não nula a cada um dos valores de $X(\Omega)$.

Vamos trabalhar com o caso especialíssimo em que isso é feito por meio da integração de uma função. Supomos que exista uma função contínua por trechos $f_X(x)$ definida em \mathbb{R} , que atribui a cada valor x da imagem da função (variável aleatória) $X : \Omega \rightarrow \mathbb{R}$ o valor não negativo $f_X(x)$. Fixado $f_X : \Omega \rightarrow [0, +\infty[$, a probabilidade de qualquer evento de $V \in E$ é calculado pela integral

$$P(V) = \int_V f_X(x) dx.$$

Em particular, a probabilidade do evento $P(X \leq a)$ para $a \in A$ deve ser dada por

$$P(X \leq a) = \int_{-\infty}^a f_X(x) dx, \quad a \in A. \quad (5.4)$$

Naturalmente, $f_X : \mathbb{R} \rightarrow [0, +\infty[$ deve ser tal que

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

A função $f_X : \mathbb{R} \rightarrow [0, +\infty[$ é a *função densidade de probabilidade* (f.d.p.) da variável aleatória $X : \Omega \rightarrow \mathbb{R}$. Note que f é definida em todo o \mathbb{R} .

A função $F_X : \mathbb{R} \rightarrow [0, 1]$ dada por (5.4) é a *função distribuição de probabilidade acumulada* (f.d.) da variável aleatória $X : \Omega \rightarrow \mathbb{R}$. Para deixarmos registrado,

$$F_X(a) = \int_{-\infty}^a f_X(x) dx, \quad a \in A. \quad (5.5)$$

Tendo fixado $f_X : \mathbb{R} \rightarrow [0, +\infty[$, usamos a integral em (5.5) para estender F_X a todo \mathbb{R} . Assim,

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi, \quad x \in \mathbb{R}. \quad (5.6)$$

A fórmula (5.6) dá sentido à probabilidade estendida $P(X \leq x)$, mesmo quando $x \notin A$, isto é, quando o evento $(X \leq x)$ não é elemento de E , que é o domínio original da função P . O mesmo pode ser dito com relação à probabilidade estendida do evento $[a, b] \subset \mathbb{R}$ dada por

$$P([a, b]) = \int_{[a, b]} f_X(\xi) d\xi,$$

que também é obtida por extensão, uma vez que $f_X : \mathbb{R} \rightarrow [0, 1]$ é fixo.

Para deixarmos claro, fizemos uma extensão de (5.5) para obter (5.6), mas o espaço de eventos E continua fixo, desde que N_{prec} foi escolhido, por exemplo $N_{prec} = 10$, sugerido por Hardy para os pragmáticos engenheiros. Formalmente não estamos autorizados a calcular a probabilidade de subconjuntos de Ω a não ser aqueles que são elementos de E . *Somente* para os casos especiais $F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$, e $P([a, b])$, $a, b \in \mathbb{R}$, é que a probabilidade estendida pode ser calculada, uma vez que $f_X : \mathbb{R} \rightarrow [0, +\infty[$ é fixo. Infelizmente há uma sutileza de nomenclatura aqui, a que o estudante deve estar atento. Temos duas funções: uma de probabilidade definida em E , e outra, a probabilidade estendida (que não é propriamente uma função probabilidade!)

Todas essas restrições seriam removidas, e teríamos uma teoria mais elegante se tivéssemos à disposição uma teoria de integração mais completa, como a de Lebesgue.

Dos cursos de cálculo, temos uma relação adicional que envolve f_X e a função de distribuição de probabilidade acumulada estendida F_X , que é

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

nos pontos $x \in \mathbb{R}$ em que f_X é contínua.

A razão da designação “variável aleatória *contínua*”. Uma função real g definida em um intervalo $[a, b] \subset \mathbb{R}$ é absolutamente contínua em $[a, b]$ se dado $\epsilon > 0$, existe $\delta > 0$ tal que para toda coleção de intervalos disjuntos $\{a_k, b_k\}_{k=1}^n$,

$$\sum_{k=1}^n (b_k - a_k) < \delta \Rightarrow \sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon.$$

Pode-se demonstrar que a função $F_X : \mathbb{R} \rightarrow [0, 1]$ é absolutamente contínua em todo intervalo fechado $[a, b] \subset \mathbb{R}$, mesmo quando f_X é apenas integrável, sem precisar ser contínua.

A razão para a designação “contínua” para a variável aleatória vem do fato que a função distribuição de probabilidade acumulada é dada por uma função absolutamente contínua. Temos um contraste ofuscante com a função de probabilidade acumulada de uma variável aleatória discreta, que não pode ser sequer contínua, pois no caso das discretas, $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$, F_X não decrescente, com subidas dadas em saltos.

Observe que se X for uma variável aleatória discreta, então a função de probabilidade $P(X \leq n)$ a ela associada não pode ser dada por uma fórmula do tipo (5.5), devido à discussão acima. Para começar, o domínio de uma função densidade de probabilidade contínua por trechos deve ser o conjunto Ω que no caso das discretas é um conjunto enumerável, então a integral em (5.5) seria sempre nula.

5.5 Função de densidade de probabilidade multivariada

Dado um espaço amostral Ω de um experimento, assim como no caso de variáveis aleatórias discretas, podemos tratar de várias variáveis aleatórias X_1, X_2, \dots, X_n , simultaneamente. Uma função de densidade de probabilidade multivariada $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow [0, +\infty]$ fornece a probabilidade $P(X_1 \leq x_1, \dots, X_n \leq x_n)$ através da fórmula

$$P(X_1 \leq a_1, \dots, X_n \leq a_n) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

com $a_1, \dots, a_n \in A$.

Novamente devido às limitações impostas pela integral de Riemann, f_{X_1, \dots, X_n} deve ser contínua por trechos neste curso.

A função distribuição de probabilidade acumulada estendida $F_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$ é simplesmente dada por $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq$

$x_1, \dots, X_n \leq x_n$). Para deixarmos registrado, fazendo uma extensão como a mencionada acima,

$$F_{X_1, \dots, X_n}(a_1, \dots, a_n) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

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

5.5.1 Marginalização

Assim como no caso discreto, podemos usar o processo de marginalização para obtermos F_{X_j} , $j = 1, \dots, n$ a partir de F_{X_1, \dots, X_n} . De fato,

$$F_{X_j}(x_j) = \underbrace{\int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} f_{X_1, \dots, X_n}(x_1, \dots, x_j, \dots, x_n) dx_1 \dots dx_n}_{\text{“Sem a integral em } x_j\text{”}}$$

5.6 Variáveis aleatórias independentes

Nada muda com relação ao caso das variáveis aleatórias discretas. Assim, X e Y são discretas se $P(X \leq x | Y \leq y) = P(X \leq x)$, para todos $x, y \in \mathbb{R}$.

De forma geral, X_1, \dots, X_n são independentes se

$$F_{X_1, \dots, X_n}(a_1, a_2, \dots, a_n) = F_{X_1}(a_1)F_{X_2}(a_2) \dots F_{X_n}(a_n),$$

para todos $a_1, \dots, a_n \in \mathbb{R}$.

5.7 Função densidade de probabilidade condicionada

Chamamos atenção para o fato que para eventos do tipo $\{a\} \in E$, com $a \in A$, vale $P(\{a\}) = 0$. Assim como esse evento tem probabilidade nula, muitos outros eventos em E também têm. Isso significa que a fórmula

$$P(X \in A | Y \in B) = \frac{P((X \in A) \cap (Y \in B))}{P(Y \in B)}$$

não faz sentido, a não ser que $P(Y \in B) \neq 0$.

5.7. FUNÇÃO DENSIDADE DE PROBABILIDADE CONDICIONADA 93

Definimos a função densidade de probabilidade $f_{X|Y} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ por

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad (5.7)$$

e calculamos, por exemplo, $P(X \leq a | Y = y)$, com $a, b \in A$, pela fórmula

$$F_{X|Y}(a|Y=y) = P(X \leq a | Y = y) = \int_{-\infty}^a f_{X|Y}(x|y) dx.$$

Como feito anteriormente, a fórmula acima permite a extensão a todos $a, b \in \mathbb{R}$. Assim,

$$F_{X|Y}(x|Y=y) = P(X \leq x | Y = y) = \int_{-\infty}^x f_{X|Y}(\xi|y) d\xi, \quad \forall x, y \in \mathbb{R}. \quad (5.8)$$

A fórmula (5.8) permite a obtenção de outras fórmulas que nos serão úteis. A partir dela, podemos obter a função densidade de probabilidade f_X pela manipulação abaixo, usando a função de densidade de probabilidade estendida.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy \\ &= \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy, \quad \forall x \in \mathbb{R}, \end{aligned}$$

e então

$$\begin{aligned} P(X \leq x) &= \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^x f_{X|Y}(x|y) f_Y(y) dx dy \\ &= \int_{-\infty}^{+\infty} F_{X|Y}(x|y) f_Y(y) dx dy, \quad \forall x \in \mathbb{R}. \end{aligned}$$

Dada a importância da fórmula acima, vamos deixá-la registrada. Como usamos a função de densidade de probabilidade estendida para obtê-la, se quisermos usa-la propriamente no contexto do espaço probabilístico original (Ω, E, P) , devemos formalizá-la como uma definição. Aqui está.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{+\infty} F_{X|Y}(x|Y=y) f_Y(y) dx dy, \quad \forall x \in A,$$

e novamente por extensão:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{+\infty} F_{X|Y}(x|Y=y) f_Y(y) dx dy, \quad \forall x \in \mathbb{R}, \quad (5.9)$$

5.8 Soma de variáveis aleatórias

Dadas duas variáveis aleatórias X e Y dentro de um espaço probabilístico, podemos definir uma terceira variável aleatória por $Z = X + Y$.

Quando conhecemos as funções densidade de probabilidade f_X e f_Y de X e Y é possível obtermos a correspondente a Z de forma muito simples através da fórmula (5.9). De fato,

$$\begin{aligned} P(Z \leq z) &= P(X + Y \leq z) = P(X \leq z - Y) \\ &= \int_{-\infty}^{+\infty} P(X \leq z - Y | Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} F_{X|Y}(z - Y | Y = y) f_Y(y) dy. \end{aligned}$$

A expressão da última linha pode ainda ser escrita de uma forma mais simples. Daí ficamos com a fórmula

$$P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{+\infty} F_X(z - y) f_Y(y) dy, \quad (5.10)$$

que é de grande importância, válida para todo $z \in \mathbb{R}$. Entretanto, apesar do cálculo de $P(Z \leq z)$ ser possível para todo $z \in \mathbb{R}$, formalmente estamos autorizados a calcular somente $P(Z \leq z)$ para $z \in A$, isto é somente para o caso em que o evento $(Z \leq z)$ é um elemento de E , uma vez N_{prec} ter sido fixado. Devemos usar (5.10) como uma definição de $P(X + Y \leq z)$, para $z \in A$, e encará-la como uma extensão, quando $z \in \mathbb{R}$.

5.9 Transformação de variáveis aleatórias

Suponha um experimento com espaço amostral Ω , variável aleatória $X : \Omega \rightarrow \mathbb{R}$, espaço de eventos E como definido na seção 5.4.1 e com função distribuição de probabilidade acumulada $F : \mathbb{R} \rightarrow [0, 1]$ dada pela integral do tipo (5.5), com $f_X : \mathbb{R} \rightarrow [0, +\infty[$ contínua por trechos, como explicado acima. As probabilidades $P : E \rightarrow [0, 1]$ dos eventos em E são calculadas por meio de f_X , por exemplo,

$$P(X \in]a, b]) = F_X(b) - F_X(a).$$

Suponha que definamos uma variável aleatória Y a partir de X através de uma função contínua $g : Im(X) \rightarrow \mathbb{R}$ por $Y = g \circ X$. A notação em

curso de probabilidade é simplesmente $Y = g(X)$. Por exemplo, Y poderia ser $Y = X^2$, $Y = \sin(X)$.

Essa construção é útil e essencial em Engenharia, pois é muito comum desejarmos conhecer as propriedades de funções de variáveis aleatórias, como por exemplo, a potência $W = rI^2$ elétrica dissipada sobre um resistor de resistência r quando passa por ela uma corrente I .

As propriedades probabilísticas de Y são conhecidas uma vez conhecida F_Y , e daí f_Y , a sua derivada.

A probabilidade $P(Y \leq y)$ é dada por $P(g(X) \leq y)$, pois $Y \leq y$ e $g(X) \leq y$ correspondem exatamente aos mesmos eventos. Lembre-se da discussão sobre o abuso de linguagem feito no capítulo anterior, e da extensão do cálculo de $P(Y \leq y)$ para $y \in \mathbb{R}$, mesmo quando $y \notin A$. Em particular, $g(X) \leq y$ ou $g(X) \in]-\infty, y]$, significa o evento $X \in g^{-1}(]-\infty, y])$, para $y \in \mathbb{R}$. Perceba que g não precisa ser inversível. $g^{-1}(]-\infty, y])$ é simplesmente a imagem inversa do conjunto $]-\infty, y]$ através da função g .

$$g^{-1}(]-\infty, y]) = \{x \in X \mid g(x) \in]-\infty, y]\}.$$

Podemos escrever então como consequência de $P(Y \leq y)$ é dada por $P(g(X) \leq y)$, que

$$\int_{]-\infty, y]} f_Y(y) dy = \int_{\{x \in X \mid g(x) \in]-\infty, y]\}} f_X(x) dx. \quad (5.11)$$

Quando $g : Im(X) \rightarrow \mathbb{R}$ contínua tem forma simples, o cálculo acima é fácil. Por exemplo, suponha que $Y = X^3$, isto é, $g(x) = x^3$. Nesse caso,

$$\{x \in X \mid g(x) \in]-\infty, y]\} = \{x \in X \mid x^3 \leq y\} = \{x \in X \mid x \leq \sqrt[3]{y}\},$$

e então (5.11) assume a forma

$$P(Y \leq y) = \int_{-\infty}^y f_Y(y) dy = \int_{-\infty}^{\sqrt[3]{y}} f_X(x) dx.$$

A função densidade de probabilidade f_Y agora pode ser facilmente obtida quando derivamos a expressão acima com relação a y .

$$f_Y(y) = \frac{1}{3\sqrt[3]{y^2}} f_X(\sqrt[3]{y}). \quad (5.12)$$

O estudante deve estar consciente do grande poder que ele possui agora. Na seção anterior (sec. 5.8) ele obteve uma fórmula para a obtenção de funções distribuição acumulada de probabilidade de somas de variáveis aleatórias. Na presente seção, ele se apoderou de técnicas que permitem o cálculo de probabilidades de funções de variáveis aleatórias. Em suma, agora ele é capaz de somar, subtrair e compor variáveis aleatórias.

5.10 Valor esperado e variância de variáveis aleatórias contínuas

Quando temos à disposição a função de densidade de probabilidade $f_X : \mathbb{R} \rightarrow [0, +\infty[$, o valor esperado da variável aleatória $X : \Omega \rightarrow \mathbb{R}$ é definida por

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx, \quad (5.13)$$

quando $\int_{-\infty}^{+\infty} |x f_X(x)| dx < +\infty$. Caso essa condição não for satisfeita, diremos que o valor esperado de X não existe.

A variância da variável aleatória X é definida por

$$V[X] = E[X^2] - (E[X])^2,$$

se $E[X^2]$ existir.

As interpretações de valor esperado e variância são as mesmas do caso discreto. As propriedades do valor esperado e da variância são também as mesmas do caso discreto. Para deixarmos registrado,

$$E[\alpha X + Y] = \alpha E[X] + E[Y],$$

para quaisquer variáveis aleatórias X e Y que possuam valor esperado, não precisando ser independentes, e qualquer $\alpha \in \mathbb{R}$.

Quanto à variância, valem

$$V[\alpha X] = \alpha^2 V[X]$$

e

$$V[X + Y] = V[X] + V[Y],$$

quando X e Y forem variáveis aleatórias independentes. As provas das propriedades do valor esperado e da variância podem ser encontradas em [Dan13].

Uma fórmula útil para o valor esperado. Um fórmula muito útil, quando existe $a \in \mathbb{R}$ tal que $f_X(x) = 0$ para $x > a$, é dada por

$$E[X] = a - \int_{-\infty}^a F_X(x) dx. \quad (5.14)$$

A prova para o caso contínuo é muito simples. De fato, como para nós f_X é sempre contínua por trechos (e há somente um número finito deles aqui), podemos usar a integração por partes.

$$\begin{aligned}
E[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^a x f_X(x) dx \\
&= \int_{-\infty}^a x \frac{dF_X}{dx}(x) dx \\
&= a - \int_{-\infty}^a F_X(x) dx.
\end{aligned}$$

Nas manipulações acima usamos que $F_X(a) = 1$ e que $\lim_{x \rightarrow -\infty} x F_X(x) = 0$.

A fórmula (5.14) vale também para o caso discreto.

Valor esperado da função de uma variável aleatória. Seja $X : \Omega \rightarrow \mathbb{R}$ uma variável aleatória contínua, com função densidade de probabilidade f_X definida em \mathbb{R} . Seja também uma função contínua $g : \text{Im}(X) \rightarrow \mathbb{R}$.

O valor esperado da variável aleatória $Y \doteq g(X)$ é dada por

$$E[Y] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

Em geral,

$$E[Y] \neq g(E[X]).$$

Somente em casos especiais vale $E[Y] = g(E[X])$.

5.11 Modelos de variáveis aleatórias contínuas

Assim como no caso de variáveis aleatórias discretas, as contínuas dão modelos prontos, e tendem a esconder o formalismo técnico. Assim, tendem a esconder o espaço amostral e o espaço de eventos.

5.11.1 Variável aleatória uniforme

A variável aleatória contínua uniforme X tem imagem em um intervalo $[a, b] \subset \mathbb{R}$ e a função densidade de probabilidade é dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \forall x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

O valor esperado $E[X]$ e a variância $V[X]$ são dadas por

$$E[X] = \frac{b-a}{2}, \quad V[X] = \frac{(b-a)^2}{12}.$$

A utilidade de variável aleatória uniforme é a sua simplicidade. Não se trata da distribuição com a menor informação sobre um problema. Isso pode ser apreciado por um exemplo. Suponha que tenhamos esferas com raio aleatório variando no intervalo $[1, 2]$. Se não soubermos mais nada a respeito, o estudante poderia imaginar que na falta de maiores informações, a distribuição para o raio que daria a maior incerteza seria a especificação de uma distribuição uniforme para a variável aleatória raio R . Dessa forma,

$$f_R(r) = \begin{cases} 1, & \forall x \in [1, 2], \\ 0, & x \notin [1, 2]. \end{cases}$$

Entretanto, não saber nada sobre as esfera, poderia também ser traduzido em não saber nada sobre seu volume, e assim, nada sobre a variável aleatória $Y = R^3$. Entretanto, usando a fórmula (5.12), obtemos a distribuição de probabilidade do volume Y , distribuído no intervalo $[1, 2^3]$, a partir da de R .

$$f_Y(y) = \frac{1}{3\sqrt[3]{y^2}}, \quad y \in [1, 2^3],$$

Seu gráfico é mostrado na figura 5.2. Uma olhada rápida nessa figura revela que não saber nada sobre R , conforme o critério da especificação de uma distribuição uniforme para ela, indica que sabemos bastante sobre o volume Y . Em particular, que é mais provável que ele esteja próximo de 1 que de 8.

Em conclusão, a distribuição uniforme é útil porque é simples, mas certamente não porque representa algum estado de desconhecimento sobre o problema.

Para deixar registrado:

1. A soma de uniformes independentes não é uniforme, como se pode constatar usando a fórmula (5.10).
2. O produto de uma uniforme com um número real produz uma outra variável aleatória uniforme, como se pode constatar usando as técnicas da seção 5.9.

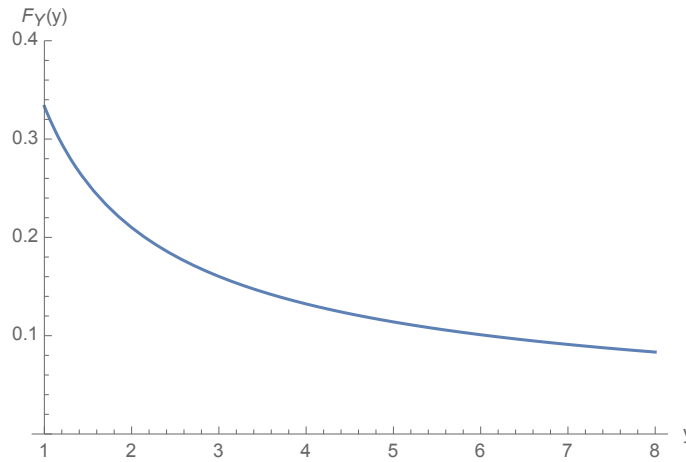


Figura 5.2: Função densidade de probabilidade do volume de uma esfera.

5.11.2 Variável aleatória exponencial

A variável aleatória contínua exponencial X com parâmetro $\lambda > 0$ tem imagem no intervalo $[0, +\infty[$ real e a função densidade de probabilidade é dada por

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

Fazendo a integração de f_X , obtemos a função distribuição acumulada de probabilidade.

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

O valor esperado $E[X]$ e a variância $V[X]$ são dadas por

$$E[X] = \frac{1}{\lambda}, \quad V[X] = \frac{1}{\lambda^2}.$$

Assim como no caso da variável aleatória discreta geométrica, a exponencial também tem a propriedade de falta de memória, caracterizada por

$$P(X > t + s | X > s) = P(X > t). \quad (5.15)$$

Podemos perguntar se existem outras variáveis aleatórias contínuas que tem a propriedade de falta de memória, isto é, que satisfazem (5.15).

Vamos provar que a distribuição exponencial é a única. Seja uma variável aleatória T com imagem em $[0, +\infty[$ que tem função distribuição de probabilidade acumulada absolutamente contínua $F_T : [0, +\infty[\rightarrow [0, 1]$ e função densidade de probabilidade f_T .

De (5.15), obtemos

$$\frac{1 - F_T(t + s)}{1 - F_T(s)} = 1 - F_T(t)$$

e então

$$1 - F_T(t + s) = (1 - F_T(t))(1 - F_T(s)).$$

Derivando essa relação com relação à t , obtemos

$$f_T(t + s) = f_T(t) \int_s^{+\infty} f_T(\tau) d\tau, \forall t, s \geq 0.$$

Derivando com relação a s , (e renomeando as variáveis) obtemos

$$\frac{df_T}{dt}(t + s) = -f_T(t) f_T(s), \forall t, s \geq 0.$$

Para $s = 0$, temos

$$\frac{df_T}{dt}(t) = -f_T(t) f_T(0), \forall t, s \geq 0.$$

Como pode ser verificado, a solução dessa equação que também satisfaz $\int_0^{+\infty} f_T = 1$ é a função

$$f_T(t) = \lambda e^{-\lambda t},$$

que é a exponencial. Logo, a única variável aleatória contínua que tem a propriedade de falta de memória é mesmo a exponencial.

Para deixar registrado:

1. A soma de exponenciais independentes não é uma variável aleatória exponencial, como se pode constatar usando a fórmula (5.10).
2. O produto de uma exponencial com um número real produz uma outra variável aleatória exponencial, como se pode constatar usando as técnicas da seção 5.9.

5.11.3 Variável aleatória normal

A variável aleatória normal X , ou gaussiana, assume valores em toda a reta real e tem dois parâmetros $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ que são relacionados com $E[X]$ e $V[X]$ por

$$E[X] = \mu, \quad V[X] = \sigma^2.$$

A sua função densidade de probabilidade é dada por

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

A distribuição normal, denotada por $X \sim N(\mu, \sigma^2)$ é sem dúvida, a mais importante das distribuições pelos seus aspectos teóricos, práticos e históricos.

Nesse nível, a propriedade mais notável de variáveis aleatórias normais é que a soma de variáveis aleatórias normais, independentes ou não, resulta em uma outra variável aleatória normal.

Para se verificar esse fato, basta calcular

$$f_Z(Z \leq z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y) dy.$$

Essa fórmula foi obtida de (5.10) simplesmente derivando-a com relação a z .

O resultado é que se duas variáveis aleatórias $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ são somadas, obtemos

$$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2).$$

Uma verificação imediata nos leva à conclusão que

$$X \sim N(\mu, \sigma^2) \Rightarrow \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Uma variável aleatória normal com média $\mu = 0$ e $\sigma^2 = 1$ é dita *padrão*. Devido à sua importância histórica, a normal padrão é tabelada e impressa na maioria dos livros de probabilidades para a engenharia. Aqui não faremos isso. Preferimos que os estudantes usem programas ou a própria internet como material de consulta da distribuição normal (ou qualquer outra distribuição).

5.12 Exemplos de fixação

Para fixarmos as ideias, elencamos agora alguns exemplos de fixação.

Eventos com probabilidade nula e o vazio. “A probabilidade do evento \emptyset é zero”. “Um evento de probabilidade nula ocorreu”. Essas duas frases merecem atenção, por causa dos detalhes que elas contêm.

Devemos notar que o conjunto vazio, isto é \emptyset é *sempre* um evento em qualquer espaço amostral, e *sempre* $P(\emptyset) = 0$ para qualquer definição de função probabilidade, não importando nem o experimento e portanto nem o espaço amostral e espaço de eventos.

O fato de um evento com probabilidade nula ter ocorrido não deve causar surpresa, em vista do que já discutimos na página 83 no tópico “A probabilidade de se sortear um racional dentro do intervalo $[0, 1]$ é zero”. De fato, eventos com probabilidade nula acontecem o tempo todo.

O evento impossível (esse sim nunca ocorre) é o conjunto \emptyset , que tem $P(\emptyset) = 0$, mas $P(A) = 0$ não significa que $A = \emptyset$.

A vacuidade de sentido na locução “completamente ao acaso”. Esse exemplo de fixação é discutido no vídeo que acompanha esse capítulo. Você pode vê-lo escaneando o QR code abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Em geral, as pessoas leigas atribuem a locução “completamente ao acaso” o sentido de distribuição uniforme no caso de variáveis aleatórias contínuas, ou que qualquer evento do tipo $\{a\}$, com $a \in \Omega$ discreto tem a mesma probabilidade, como no caso do lançamento de moedas e dados não viciados.

Mas já vimos na página 81, quando discutimos o tópico “A probabilidade de sortearmos um número par “completamente ao acaso” dentre os naturais não é $1/2$ ”, que quando se trata de Ω com infinitos elementos, a afirmação é problemática.

Vamos ver agora mais um exemplo, este célebre, o do Paradoxo de Bertrand proposto em 1889. O problema é determinar-se a probabilidade de que uma corda traçada de forma totalmente aleatória em uma circunferência seja maior do que o comprimento de um lado do triângulo equilátero inscrito em uma circunferência de raio unitário. A figura 5.3 ilustra o triângulo com suas dimensões principais. A altura é $h = 3/2$, o lado vale $L = \sqrt{3}$ e o ângulo que lado inclinado faz com a horizontal vale $\pi/6$.

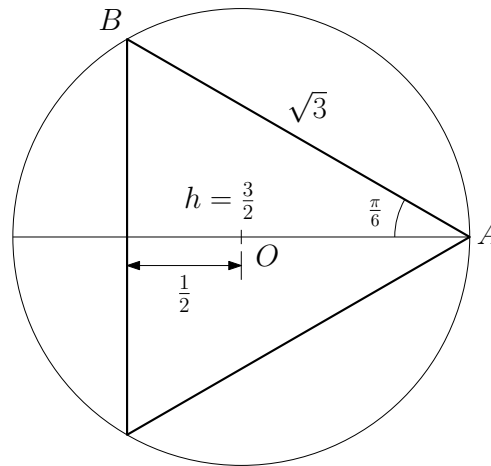


Figura 5.3: Triângulo equilátero inscrito.

Vamos apresentar apenas duas soluções que dão resultados diferentes.

Na primeira solução, fixamos uma extremidade do segmento, o ponto A , que deve estar posicionado no ponto de coordenadas $(1, 0)$ conforme a figura 5.4, enquanto B varre os pontos da circunferência. Então, θ , que define a posição do ponto B , é tal que $\theta \in]-\pi/2, \pi/2[$, e a corda AB terá comprimento maior que o do lado de um triângulo equilátero, se a variável aleatória $\Theta \in]-\pi/6, \pi/6[$. Como a corda é traçada de forma “totalmente aleatória”, a distribuição de probabilidade para Θ seria a uniforme dentro do intervalo $] -\pi/2, \pi/2[$. Isso significa que a probabilidade que a corda seja maior que o lado de um triângulo equilátero inscrito é simplesmente $\frac{2\pi/6}{\pi}$, ou seja, $1/3$.

A segunda solução olha o problema de outra forma, colocando a corda sempre na horizontal, como na figura 5.5. Isso não afeta a análise, pois é como sempre girássemos a circunferência.

Agora, vemos que a corda terá comprimento maior que o lado de um triângulo equilátero inscrito, se o ponto x da figura 5.5, que passeia no intervalo $[-1, 1]$, estiver a uma distância menor que $1/2$ do centro O . Em outros

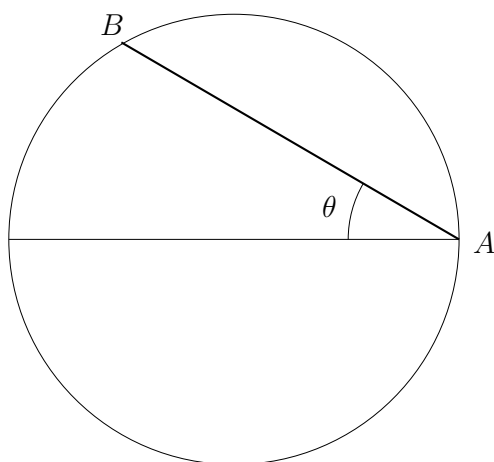


Figura 5.4: Lado com uma extremidade fixa.

termos, se a variável aleatória X uniformemente distribuída no intervalo $[-1, 1]$ satisfizer $X \in [-1/2, 1/2]$. Como a variável aleatória é uniformemente distribuída, a probabilidade que a corda é maior que o lado de um triângulo equilátero inscrito é simplesmente $\frac{1}{2}$.

Como vemos, as respostas são bem diferentes. Na primeira abordagem, a probabilidade é $\frac{1}{3}$ e na segunda é $\frac{1}{2}$. Em ambas, o experimento ocorre de forma “totalmente aleatória”. O estudante pode perguntar, com razão, qual das duas respostas é a correta. O que podemos dizer é que o que não está correto é o enunciado, pois não a locução “totalmente aleatória” é desprovida de sentido.

As variáveis aleatórias usadas, Θ e X têm naturezas diferentes. Inclusive têm unidades distintas! Apliquemos o método que temos aplicado até aqui para analisarmos problemas probabilísticos. Começamos pelo experimento. Do que se trata ele? Em essência, o traçado aleatório de cordas na circunferência de raio 1. Os resultados experimentais não são distâncias ou ângulos, mas sim cordas! Então o espaço amostral é dado por cordas do tipo AB , como na figura 5.4. Como discutido na seção 5.4.1, o espaço de eventos é construído a partir das variáveis aleatórias eleitas. As variáveis aleatórias são $X : \Omega \rightarrow \mathbb{R}$ e $\Theta : \Omega \rightarrow \mathbb{R}$. Essencialmente, as imagens inversas por X e Θ de intervalos reais nos dão os elementos do espaço de eventos E . Aqui é importante notar que os eventos tratados são os mesmos, embora as variáveis aleatórias sejam distintas. A função probabilidade $P : E \rightarrow [0, 1]$ é definida sobre o mesmo conjunto E . Como E é o mesmo para X e Θ estamos autorizados a usar o *mesmo* símbolo P para, por exemplo $P(X \in [0, x])$ e $P(\Theta \in [0, \theta])$. A questão é saber qual probabilidade atribuir aos eventos

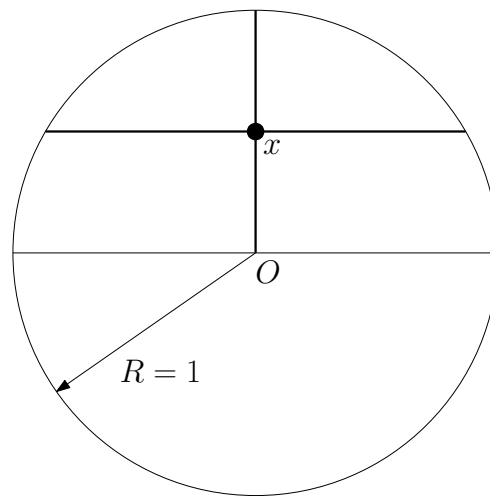


Figura 5.5: Corda horizontal.

de E . Essa é a chave do paradoxo, que não estava disponível em 1889, quando o paradoxo foi proposto. A atribuição de probabilidades aos eventos de E pode ser feita através das variáveis aleatórias, mas como vimos, ao se especificar a distribuição para X , digamos, segundo uma uniforme, ela força a especificação de uma outra distribuição de probabilidades para Θ . No próximo exemplo de fixação vamos explorar esse ponto.

Pode-se perguntar o que acontece quando o problema é colocado de forma física e concreta e não de forma puramente matemática, como feito acima. Nesse caso, o problema é inverso. Não se especifica a função probabilidade, mas sim obtém-a-na. Nesse caso, o que se deve ser precisamente especificado são as condições experimentais.

Transformações entre variáveis aleatórias usando a função distribuição de probabilidade acumulada F . No exemplo de fixação anterior, o do paradoxo de Bertrand, suponha que a variável aleatória X seja uniformemente distribuída no intervalo $[0, 1]$. Da discussão já feita, a distribuição para X induz uma distribuição de probabilidades para a variável aleatória Θ , que assume valores no intervalo $[0, \pi/2]$. Isso não é de se estranhar, uma vez que pela figura 5.6 X e Θ estão relacionadas por

$$X = \sin \Theta, \quad \text{com } X \in [0, 1], \Theta \in [0, \frac{\pi}{2}].$$

Surpreendentemente, a função densidade de probabilidade de Θ é fácil de ser obtida a partir da de X . De fato, a probabilidade de que $X \in$

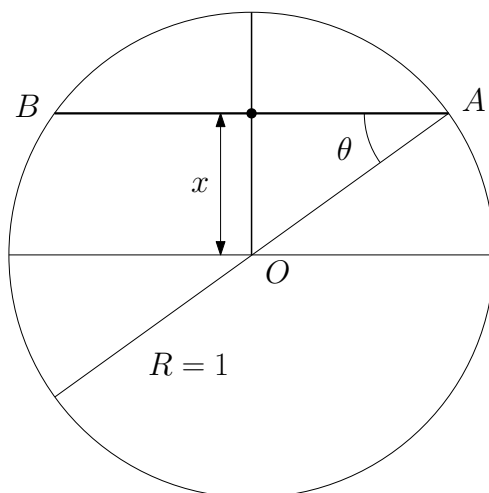


Figura 5.6: Relação entre as variáveis aleatórias X e Θ .

$[0, x]$ deve ser exatamente a mesma que $\Theta \in [0, \arcsin(x)]$, afinal a corda AB , que é o resultado experimental, define simultaneamente $X(AB) = x$ e $\Theta(AB) = \theta$. Lembre-se que variáveis aleatórias são funções definidas no espaço amostral Ω . $X : \Omega \rightarrow \mathbb{R}$, $\Theta : \Omega \rightarrow \mathbb{R}$. Uma vez feito o experimento, ambas as funções X e Θ tem seus valores definidos. Além disso, $X^{-1}([0, x])$ e $\Theta^{-1}([0, \arcsin(x)])$ são exatamente o mesmo evento definido no espaço E , e portanto devem ter exatamente a mesma probabilidade. Perceba como todas as peças da teoria se encaixam!

Então, é inevitável que

$$\int_0^x f_X(x) dx = \int_0^x 1 dx = \int_0^{\arcsin(x)} f_\Theta(\theta) d\theta.$$

Agora basta derivamos a expressão acima (cuidado com a regra da cadeia) com relação a x para obtermos f_Θ .

$$1 = f_\Theta(\arcsin(x)) \arcsin'(x) = f_\Theta(\arcsin(x)) \frac{1}{\sqrt{1-x^2}}.$$

Substituímos $x = \sin \theta$ para obter

$$1 = f_\Theta(\theta) \frac{1}{\cos \theta}.$$

Assim,

$$f_\Theta(\theta) = \cos \theta.$$

Note que o resultado é uma função densidade de probabilidade legítima, pois é positiva e a integral em seu domínio vale 1, isto é, $\int_0^{\pi/2} \cos \theta d\theta = 1$.

Paradoxo do ponto de ônibus – Distribuição uniforme. Suponha uma linha circular de ônibus, em que percorrem três ônibus, um após o outro sem ultrapassagem e velocidade constante. Eles passam por um ponto, onde devemos pegar um deles, em intervalos de tempo estritamente regulares. Δ_i é o intervalo de tempo entre o i -ésimo ônibus e o próximo. $\Delta_1 = 10$ min, $\Delta_2 = 20$ min, $\Delta_3 = 30$ min. Chegamos ao ponto. É claro que a média dos intervalos de tempo de espera é $(10/2 + 20/2 + 30/2)/3 = 10$ minutos. Então poderíamos imaginar que o valor esperado para o tempo de espera seja 10 minutos. Mas o valor exato não é esse. Ele é mais que 10% maior. Vamos analisar o problema.

Denominamos por T a variável aleatória que dá o tempo que devemos esperar até passar o primeiro ônibus. Nosso objetivo é o cálculo de $E[T]$.

O resultado do experimento é a posição dos ônibus no momento em que você chega ao ponto. Uma configuração possível está mostrada na figura 5.7. A posição do ponto de ônibus está marcada pelo ponto P .

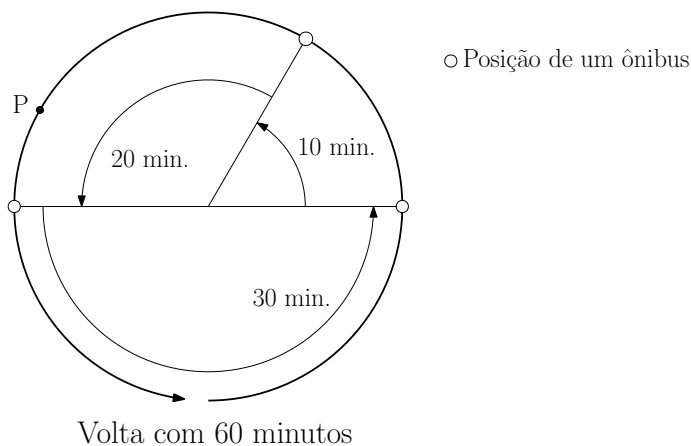


Figura 5.7: Um resultado experimental no paradoxo do ponto de ônibus.

Como o que realmente interessa é a posição temporal relativa entre os ônibus e o ponto, podemos fixar a posição dos ônibus e imaginar que o ponto P é que é aleatório. A figura 5.8 retifica a circunferência da figura 5.7. Agora, o ponto P pode cair em qualquer instante do intervalo $[0, 60]$.

A posição do ponto P é agora quantificada pela variável aleatória X . Devido à total simetria do problema, o uso da distribuição uniforme para X é perfeitamente justificável. Assim, assumimos que X seja uniformemente distribuída no intervalo $[0, 60]$.

A partir de X , construímos a variável aleatória $T(X)$ que dá o tempo de espera em função de X . Pela figura 5.8, vemos que se $X = 15$, o tempo

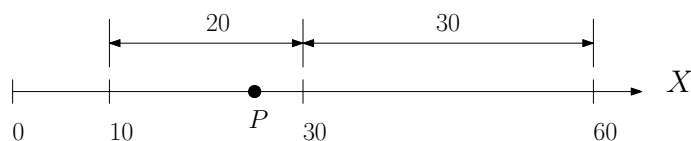


Figura 5.8: Posição relativa entre o ponto e os ônibus. O ponto P agora é que se move no intervalo $[0, 60]$.

de espera é 5 minutos, se $X = 33$, então o tempo de espera é 3 minutos e assim por diante. O gráfico da figura 5.9 mostra o tempo de espera em função de $X = x$.

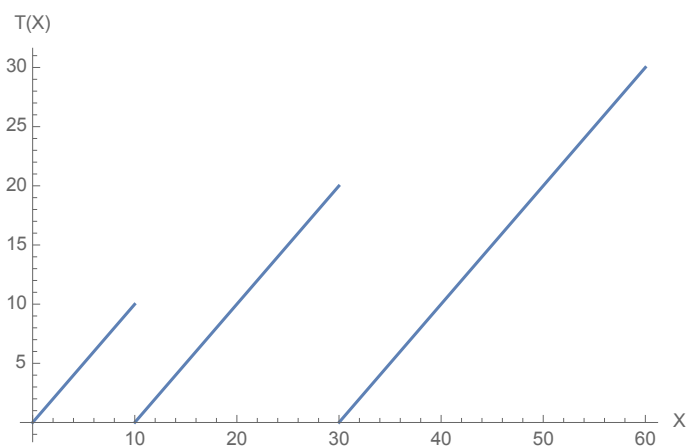


Figura 5.9: Tempo de espera em função de X .

Tendo a função $T(X)$, que formalmente é dada por

$$T(X) = \begin{cases} X, & X \in [0, 10[, & (5.16) \\ X - 10, & X \in [10, 30[, & (5.17) \\ X - 30, & X \in [30, 60[& (5.18) \end{cases}$$

e usando a técnica descrita na seção 5.9 na página 94, podemos obter a função distribuição acumulada de probabilidade F_T de T , que é mostrada na figura 5.10.

Formalmente, F_T é dada por

$$F_T(t) = \begin{cases} \frac{3t}{60}, & t \in [0, 10[, & (5.19) \\ \frac{2t}{60} + \frac{1}{6}, & t \in [10, 20[, & (5.20) \\ \frac{t}{60} + \frac{1}{2}, & t \in [20, 30[. & (5.21) \end{cases}$$

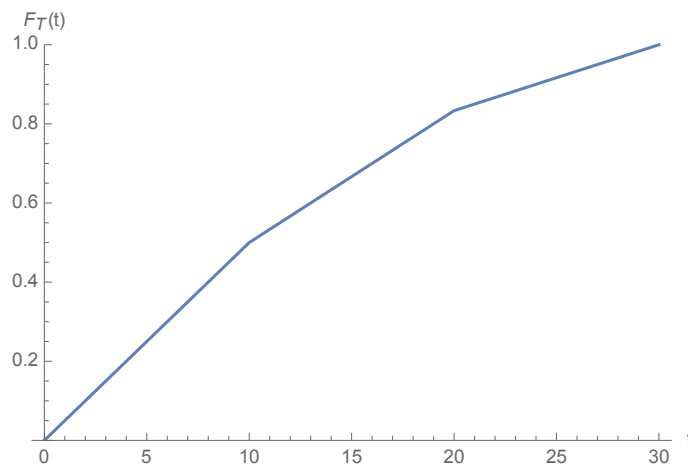


Figura 5.10: função distribuição acumulada de probabilidade da variável aleatória $F(T(X))$.

Finalmente, aplicando a fórmula (5.14), facilmente chegamos ao resultado

$$E[T] = \frac{35}{3}.$$

Chegamos à solução, mas o estudante ainda pode ter ficado com a sensação que o paradoxo não foi resolvido. O que ocorre é que a probabilidade de o ponto P cair em um intervalo maior (digamos o de 30 minutos) é, naturalmente, maior que cair em um menor (digamos no de 10 minutos). Assim, o valor esperado é certamente maior que o valor médio todos tempos de espera. Aliás, o valor esperado de T é mais que 10% que o tempo médio.

Esse raciocínio leva a um outro modo de solução que é usar a ideia de probabilidade condicionada:

$$\begin{aligned} E[T] &= E[T \mid \text{“cair no intervalo de 10 min”}] \times P(\text{“cair no intervalo de 10 min”}) \\ &\quad + E[T \mid \text{“cair no intervalo de 20 min”}] \times P(\text{“cair no intervalo de 20 min”}) \\ &\quad + E[T \mid \text{“cair no intervalo de 30 min”}] \times P(\text{“cair no intervalo de 30 min”}) \\ &= 5\frac{1}{6} + 10\frac{2}{6} + 15\frac{3}{6} = \frac{35}{3}, \end{aligned}$$

como antes. O problema aqui é conceitual. Não definimos o que seja valor esperado condicionado.

Quando o estudante for pegar o próximo circular, sendo ele azarado, ele vai se lembrar dolorosamente desse problema.

A soma de exponenciais e a distribuição de Poisson. Esse exemplo de fixação é objeto de um vídeo que pode ser visto no link acessível pelo QR code abaixo. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como tudo se passou durante a pandemia.



Soma de variáveis aleatórias

Suponha que T seja distribuída de acordo com a distribuição exponencial com $E[T] = \frac{1}{\lambda}$. Tome X distribuída de acordo com uma Poisson com $E[X] = \lambda$. Então a probabilidade que em um intervalo de comprimento a tenhamos $X = 0$ é a mesma que a probabilidade que $T > a$.

Por definição, $E[X] = \lambda$ significa que em um intervalo de comprimento unitário, o valor esperado do número X de ocorrências é λ . Então em um intervalo de comprimento a , o valor esperado de ocorrências é simplesmente $a\lambda$. Note que a palavra intervalo é usada em um sentido bem genérico, como já discutido no capítulo anterior, quando discutimos a distribuição de Poisson.

A probabilidade $P_T(T > a)$ que $T > a$ é dada por

$$P_T(T > a) = 1 - F_T(a) = e^{-\lambda a},$$

pois T é distribuída de acordo com uma exponencial de parâmetro λ ($E[T] = 1/\lambda$).

Por outro lado, a distribuição de Poisson com parâmetro λ nos fornece que a probabilidade $P_X(X = 0)$ que $X = 0$ em um intervalo de comprimento a é

$$P_X(X = 0) = e^{-\lambda a}.$$

Então, $P_X(X = 0) = P_T(T > a)$.

Se nesse exemplo o intervalo se refere de fato a um intervalo temporal, podemos interpretar o resultado da seguinte forma. A exponencial fornece o tempo entre as ocorrências que são contadas pela variável aleatória de Poisson. A figura 5.11 fornece um esquema de relação entre a Poisson e a Exponencial.

A função densidade de probabilidade da exponencial é simples o suficiente para que façamos as contas diretamente.

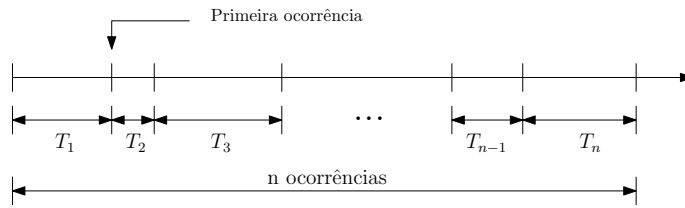


Figura 5.11: Esquema da relação entre a Exponencial e a Poisson.

Se T_1 e T_2 são independentes e distribuídas de acordo com a exponencial de parâmetro λ e $T = T_1 + T_2$, podemos fazer facilmente a integral em (5.10) para obter a função densidade de probabilidade f_T .

$$f_T(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & t \geq 0 \\ 0 & t < 0. \end{cases} \quad (5.22)$$

$$(5.23)$$

A partir de f_T obtemos a probabilidade que $T > a$.

$$P_T(T > a) = 1 - \int_0^a f_T(t) dt = e^{-a\lambda}(a\lambda + 1).$$

Comparamos esse resultado com aquele obtido através da Poisson com parâmetro $a\lambda$. Calculamos a probabilidade que $X \leq 1$, isto é que $X \in \{0, 1\}$.

$$P(X = 0) + P(X = 1) = e^{-a\lambda}(a\lambda + 1),$$

que é o mesmo resultado obtido anteriormente. Isto é,

$$P_X(X = 0) + P_X(X = 1) = P_T(T_1 + T_2 > a).$$

Podemos proceder indutivamente, para concluir que

$$\sum_{j=0}^{n-1} P_X(X = j) = P_T(T_1 + \dots + T_n > a). \quad (5.24)$$

Aproveitamos para mostrar graficamente a soma de exponenciais independentes. Aparentemente a função distribuição de probabilidade da soma converge para uma função com formato de sino, como a normal. Na figura X estão mostradas os gráficos da função densidade de probabilidade da exponencial e da soma de apenas duas exponenciais. Note que mesmo com apenas duas parcelas o resultado já é bem diferente da exponencial.

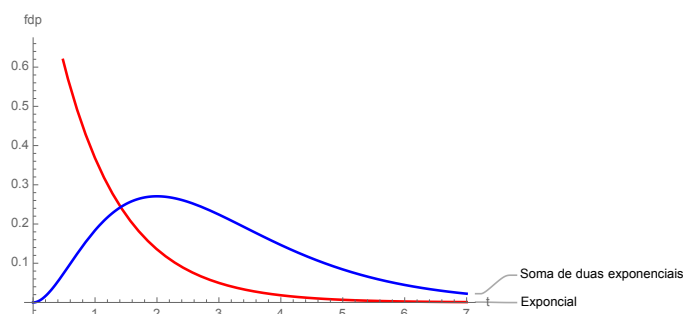


Figura 5.12: A soma de duas exponenciais pode ser bem diferente da própria exponencial.

Uma competição de natação entre amigos – Distribuição normal. Três amigos nadadores da categoria 55+ (entre 55 e 59 anos de idade) vão ver quem é o mais rápido na prova de 100m nado peito. Podemos assumir que os tempos T_A, T_B, T_C dos três são distribuídos de acordo com a normal. $T_A \sim N(1m27.54s, 0.25^2)$, $T_B \sim N(1m27.24s, 0.23^2)$, $T_C \sim N(1m27.82s, 0.21^2)$. Desejamos saber qual é a probabilidade que o nadador A seja o vencedor.

A solução desse problema nos indicará o quanto é difícil vencer o favorito. A diferença de tempo entre B, que é o favorito, e A não parece ser grande. Podemos ter uma noção melhor ao observarmos um gráfico das funções densidade de probabilidade normais relacionadas aos três nadadores. A figura 5.13 as mostra.

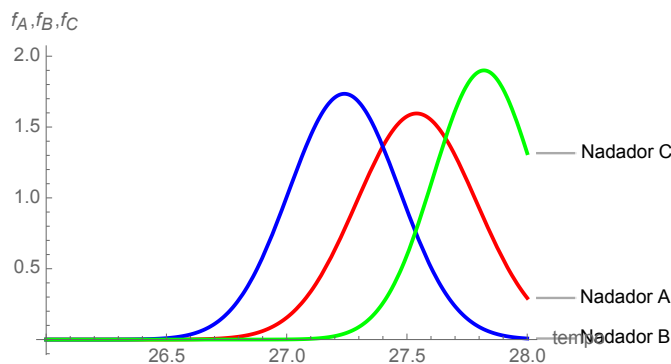


Figura 5.13: Função densidade de probabilidade dos tempos dos nadadores.

A probabilidade que o nadador A ganhe de B é a mesma que $P(T_A < T_B)$. Definimos uma nova variável aleatória T_{A-B} por $T_{A-B} = T_A - T_B$.

Essa nova variável aleatória é também normalmente distribuída conforme discutido na seção 5.11.3.

$$E[T_{A-B}] = E[T_A - T_B] = E[T_A] - E[T_B].$$

Como assumimos que T_A e T_B são independentes,

$$V[T_{A-B}] = V[T_A - T_B] = V[T_A] + V[T_B].$$

Assim, $E[T_{A-B}] = 0.30$ e $V[T_{A-B}] = 0.1154$. Com esses dados podemos calcular a probabilidade (se quiser usando uma tabela da normal padrão, ou um programa),

$$P(T_{A-B} < 0) = 0.188586.$$

Da mesma forma, podemos calcular a probabilidade que o tempo do nadador A seja menor que o de C .

$$P(T_A - T_C < 0) = 0.80444.$$

Ambos os números nos dão uma ideia de como é difícil o resultado de uma prova ser difícil do balizamento inicial. Em particular, como é difícil vencer o favorito.

5.13 Problemas

1. O objetivo desse problema é fazer com o que o estudante atente para a construção da função de probabilidade a partir da função distribuição, para os casos discreto e contínuo. As figuras 5.14 e 5.15 correspondem a funções de distribuição de probabilidade acumulada. A segunda função é descrita por

$$F_X(x) = \begin{cases} 0, & x < 0, & (5.25) \\ \frac{x^3}{0.424}, & x \in [0, 0.4[, & (5.26) \\ \frac{(x - 0.4)^2 + 0.4^3}{0.424}, & x \in [0.4, 1[, & (5.27) \\ 1, & x \geq 1. & (5.28) \end{cases}$$

- a) Quais correspondem a variáveis discretas contínuas e discretas?

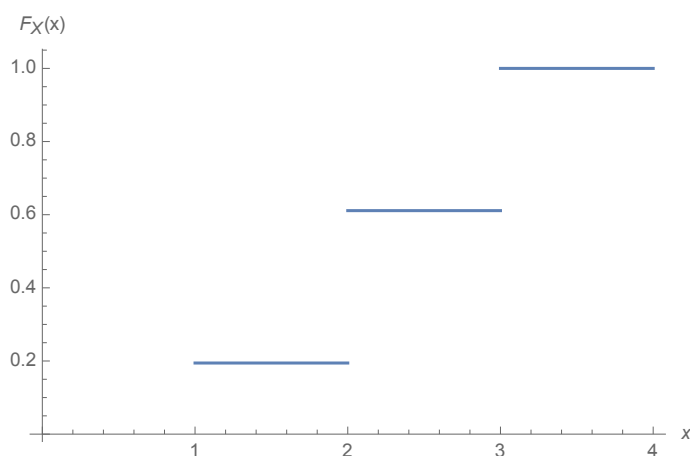


Figura 5.14: Função distribuição probabilidade acumulada.

- b) Complete a figura 5.14 colocando bolas abertas e fechadas nos extremos dos segmentos de reta horizontais para que a figura efetivamente seja de uma função distribuição probabilidade acumulada.
 - c) A partir das funções distribuição probabilidade acumuladas, determinar as distribuições de probabilidade das respectivas variáveis aleatórias.
2. O objetivo desse problema é continuar a discussão sobre ser “totalmente aleatório”. Uma fábrica produz azulejos exatamente quadrados com lado L de forma “totalmente aleatória”. Os tamanhos dos azulejos são distribuídos entre 220 cm^2 e 230 cm^2 . Qual é a probabilidade que a área que o próximo azulejo produzido seja maior que 225 cm^2 ?
 - a) Assumindo que ser “totalmente aleatório” signifique que L seja distribuído uniformemente entre $\sqrt{220} \text{ cm}$ e $\sqrt{230} \text{ cm}$.
 - b) Assumindo que ser “totalmente aleatório” signifique que L^2 seja distribuído uniformemente entre 220 cm^2 e 230 cm^2 .
 3. No problema 2, suponha que L seja distribuído uniformemente entre $\sqrt{220} \text{ cm}$ e $\sqrt{230} \text{ cm}$. Nesse caso:
 - a) Qual é a função densidade de probabilidade para a variável aleatória que fornece a área de um azulejo?
 - b) Qual é a função distribuição de probabilidade acumulada para a variável aleatória que fornece a área de um azulejo?

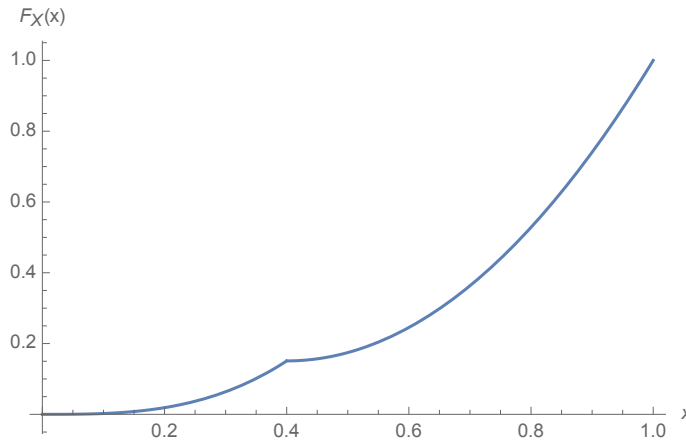


Figura 5.15: Função distribuição probabilidade acumulada.

- c) Qual é o valor esperado para a área de um azulejo? Use as fórmulas (5.13) e (5.14).
4. O objetivo desse problema é fortalecer o conceito de independência de variáveis aleatórias. Sejam as variáveis aleatórias X e Y contínuas, que assumem valores somente no intervalo $[0, 1] \times [0, 1]$, correspondentes a um experimento probabilístico. A função de distribuição de probabilidade acumulada $F_{X,Y} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ é dada por

$$F_{X,Y}(x, y) = \int_0^x \int_0^y (\sin(2\pi\xi) \sin(2\pi\psi))^2 d\psi d\xi,$$

- a) Verifique que $F_{X,Y} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ é mesmo uma função distribuição de probabilidade acumulada.
- b) As variáveis aleatórias X e Y são independentes?
5. Verifique rapidamente que no problema do paradoxo do ponto de ônibus, se na linha circular de ônibus, em que percorrem três ônibus, um após o outro sem ultrapassagem e velocidade constante, os três passam em intervalos precisamente regulares, a cada 20 minutos, o valor esperado para o seu tempo de espera no ponto é exatamente 10 minutos.
6. Volte ao exemplo de fixação “A soma de exponenciais e a distribuição de Poisson” que está na página 110 e faça as contas para comprovar a fórmula (5.24) para $n = 3$.

7. Sejam as variáveis aleatórias T_1, T_2 independentes e distribuídas de acordo com uma exponencial de parâmetro $\lambda > 0$. Determine a função densidade de probabilidade de $T_1 - T_2$.
8. Três amigos nadadores da categoria 55+ (entre 55 e 59 anos de idade) vão ver quem é o mais rápido na prova de 100m nado peito. Podemos assumir que os tempos T_A, T_B, T_C dos três são distribuídos de acordo com a normal. $T_A \sim N(1m26.21s, 0.18^2)$, $T_B \sim N(1m26.94s, 0.33^2)$, $T_C \sim N(1m26.19s, 0.21^2)$.

Qual é a probabilidade que o nadador A seja o vencedor em 2 de 3 tiros?

Capítulo 6

Teorema do Limite Central

6.1 Introdução

Neste capítulo temos como objetivo primário a apresentação do Teorema do Limite Central, que essencialmente diz que a soma de variáveis aleatórias independentes converge em distribuição para uma variável aleatória normal. Para que esse objetivo seja atingido e o teorema seja apreciado, apresentamos todos os elementos teóricos necessários. Muito já foi visto nos capítulos anteriores, desde espaço amostral até o que são variáveis aleatórias. Agora exploramos o significado de convergência, e em particular o que é convergência em distribuição. Veremos que convergência em distribuição é muito fraca, no sentido que ela nem mesmo implica convergência de funções densidade de probabilidade.

Aproveitamos a oportunidade para explorar a desigualdade de Chebyshev e a lei fraca dos grandes números. A vivência do estudante nos laboratórios de física também é aproveitada, e comentamos um experimento objetivando a medida da aceleração da gravidade na Terra, lembrando que a fórmula relevante não envolve somas, mas sim produtos e razões.

6.2 Vídeo que acompanha este capítulo

Este texto acompanha a aula sobre o Teorema do Limite Central do reorientamento do curso de Probabilidades para a Escola Politécnica da Universidade de São Paulo em 2020. Acompanha o texto um vídeo, que pode se visto escaneando-se o código QR abaixo. Ele é relacionado ao assunto do problema proposto 4 que está na página 131. A qualidade pode ser um pouco aquém da ideal, porém trata-se de um registro histórico de como

tudo se passou durante a pandemia.



QRCode para o vídeo que acompanha este capítulo.

6.3 Contexto histórico – Biografia de Patnuty Chebyshev

Para compreendermos o contexto histórico científico, mostramos algumas passagens da biografia do matemático Russo Чебышев [But99, Enc10, Sen13].

Chebyshev, Patnuty Lvovich (1821– Okatova – 1894 – São Petersburgo/Russia) é considerado o fundador da escola de matemática de São Petersburgo, também conhecida como escola de Chebyshev, que é lembrada principalmente pelos seus trabalhos na teoria dos números primos.

Na sua infância, em Okatova – Rússia, Chebyshev teve a educação básica em casa, sendo alfabetizado pela mãe e recebendo lições de francês e aritmética de um primo. A sua fluência em francês muito o beneficiou na sua futura interação com os grandes matemáticos europeus.

Em 1832, a família mudou-se para Moscou, onde Patnuty complementou a sua educação secundária, ainda em casa, sob tutoria de Pogorelski um autor de livros de matemática muito populares. A sua educação formal foi iniciada em 1837, na Universidade de Moscou, tendo se bacharelado em matemática em 1841 e o mestrado em 1843. Neste período, foi muito influenciado pelo professor Brashman, seu orientador de mestrado, a quem devotava muito respeito e atribuía o seu desenvolvimento em matemática.

O mestrado *Essay in Elementary Analysis of the Theory of Probabilities* foi o seu primeiro trabalho em Teoria das Probabilidades. Revisitava a lei fraca dos grandes números de Poisson, a partir de uma análise matemática rigorosa, apesar de elementar. No entanto, este trabalho inovador praticamente não repercutiu no meio matemático.

Em 1947, Chebychev ingressou como professor assistente na Universidade de São Petersburgo, obtendo a cátedra em 1860 e realizando as suas pesquisas, mesmo depois da aposentadoria em 1882. Manteve forte con-

tato com matemáticos e pesquisadores da Europa, especialmente com os franceses, especialmente com Liouville.

Na época, o interesse por probabilidades era motivado principalmente pela aplicação desta na análise e na elaboração de tábuas de vida (expectativa de vida) para fundos de pensão e seguros. Neste contexto, Irénée – Jules Bienaymé, professor de probabilidade da Sorbonne, então conhecido pelo trabalho *De la durée de la vie em France*, realizou estudos na mesma linha que Chebyshev, tendo obtido resultados mais abrangentes e posteriormente reconhecidos por Chebyshev, culminando assim com a famosa desigualdade de Bienaymé – Chebyshev, que ficou mais conhecida como a desigualdade de Chebyshev .

Por fim, em 1891, Chebyshev publicou um artigo propondo uma prova do teorema do limite central para a soma de variáveis independentes não identicamente distribuídas, empregando o método dos momentos. Este trabalho suscitou discussões e questionamentos e foi colocado em bons termos por Markov e Liapunov.

Chebyshev foi professor de probabilidade de 1860 a 1882, além de ministrar cursos de álgebra e teoria dos números. Foi também um educador ativo, trabalhando com outros cientista russos na melhoria do ensino de matemática, física e astronomia nas escolas secundárias.

Destacamos neste texto as contribuições na Teoria de Probabilidades, mas ressalta-se o ecletismo e abrangência de seus trabalhos e pesquisas em matemática e mecânica, ressaltando-se: teoria dos números primos, formas quadráticas, funções, ortogonais, teoria das integrais, polinômios de Chebyshev, teoria das Congruências, construção de mapas geográficos e mecanismos geradores de trajetórias retilíneas.

Os méritos de Chebyshev foram reconhecidos desde o início da sua carreira. Quando em um congresso alguém o descreveu como um “matemático russo esplêndido”, Chebyshev, ofendido, questionou “Por que russo, e não em escala mundial?” .

6.4 Para compreender o Teorema do Limite Central – Noções de convergência

O Teorema do Limite Central versa sobre convergência de variáveis aleatórias. Em termos bem gerais, ele diz que a soma de variáveis aleatórias independentes converge para uma variável aleatória distribuída normalmente.

Os estudantes que acompanham este curso já sabem que uma variável aleatória é uma função $X : \Omega \rightarrow \mathbb{R}$ que tem por domínio o espaço amostral Ω e cuja imagem inversa dos elementos da menor sigma-álgebra que contém $] - \infty, a]$ e $] - \infty, a[$, $a \in A_{prec}$, A_{prec} como definido no capítulo anterior, define o espaço de eventos E . Quando X é uma variável aleatória discreta, que no nosso caso significa que Ω é discreto, não há problema em se tomar E como sendo o conjunto das partes de Ω .

Quando há variáveis aleatórias contínuas X no problema, dada a função densidade de probabilidade f_X , que no nosso caso é necessariamente contínua por trechos, fazemos a extensão da definição da função probabilidade $P : E \rightarrow [0, 1]$ para conjuntos do tipo $] - \infty, x]$ e $[x, y]$ para $x, y \in \mathbb{R}$ usando a integral de Riemann.

$$P(X \leq x) = \int_{-\infty}^x f_X(\xi) d\xi, \quad P(X \in [x, y]) = \int_x^y f_X(\xi) d\xi.$$

O estudante também já sabe o que é uma variável aleatória distribuída de acordo com uma normal. Designaremos por

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

a situação em que a variável aleatória X tem distribuição normal com média μ e variância σ^2 .

O ingrediente que falta é compreender o que significa *convergência* de variáveis aleatórias.

O estudante certamente conhece o Teorema de Taylor, que diz em uma de suas versões que se $f : \mathbb{R} \rightarrow \mathbb{R}$ é k vezes derivável no ponto $a \in \mathbb{R}$, então existe uma função $e_k : \mathbb{R} \rightarrow \mathbb{R}$ tal que

$$f(x) = f(a) + \sum_{n=1}^k \frac{f^{(n)}(a)}{n!} (x-a)^n + e_k(x)(x-a)^k,$$

com $\lim_{x \rightarrow a} e_k(x) = 0$.

Esse teorema nos dá uma informação importante sobre a convergência do polinômio de Taylor de ordem k de f à função f . Na figura 6.1 está mostrada a aproximação da função $\sin : \mathbb{R} \rightarrow \mathbb{R}$ no intervalo $[0, 2\pi]$ por polinômios de Taylor.

No contexto probabilístico, temos vários modos de convergência, que são um pouco mais abstratos e sofisticados. Entre eles as chamadas

- convergência quase sempre,

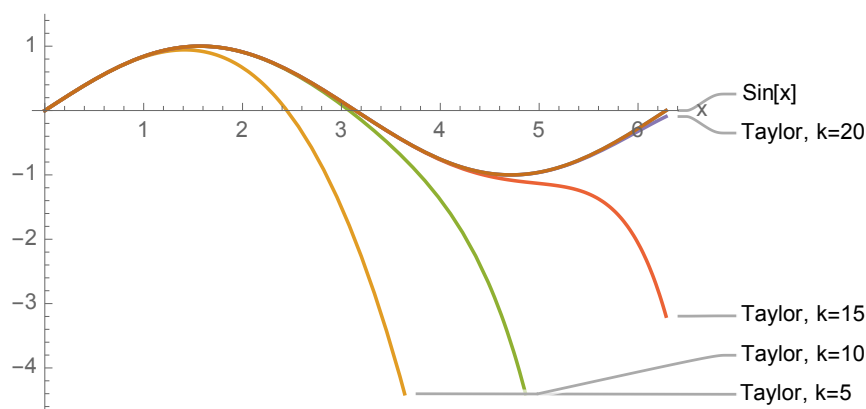


Figura 6.1: Aproximação por polinômios de Taylor.

- convergência em probabilidade,
- convergência em distribuição.

Pela limitação auto-imposta de usarmos somente a integral de Riemann para definirmos a função probabilidade $P : E \rightarrow [0, 1]$, não podemos discutir a contento as duas primeiras formas de convergência quando as variáveis aleatórias envolvidas são contínuas.

Neste texto discutiremos apenas a convergência em probabilidade para o caso em que há somente variáveis aleatórias discretas envolvidas, e a convergência em distribuição para o caso de variáveis aleatórias contínuas, que é essencial para a compreensão do enunciado do Teorema do Limite Central.

Convergência em distribuição. Dizemos que a sequência de variáveis aleatórias $(X_n)_{n \in \mathbb{N}}$, $X_n : \Omega \rightarrow \mathbb{R}$ converge em distribuição para a variável aleatória X quando a sequência de funções distribuição de probabilidade acumulada (F_{X_n}) associada a $(X_n)_{n \in \mathbb{N}}$ converge para F_X , a função distribuição acumulada de probabilidade de X .

Esse modo de convergência é bem fraco, no sentido que $\lim_{n \rightarrow +\infty} X_n$ tem a mesma F de F_X , mas não que $\lim_{n \rightarrow +\infty} X_n = X$, e *nem sequer* que a variável aleatória que resulta de $\lim_{n \rightarrow +\infty} X_n$ tenha a mesma função densidade de probabilidade f_X de X .

Tomemos um exemplo. Considere duas variáveis aleatórias independentes Y_1 e Y_2 uniformemente distribuídas em $[0, 1]$. Montamos uma sequência $(X_n)_{n \in \mathbb{N}}$ com $X_n = Y_1$, para todo $n \in \mathbb{N}$. É claro que $X_n \rightarrow Y_1$ e $X_n \rightarrow Y_2$

em distribuição, pois todos os X_n tem a mesma distribuição, mas note que Y_1 e Y_2 são variáveis aleatórias distintas (e independentes).

Para ver que mesmo que $(X_n)_{n \in \mathbb{N}}$ convirja em distribuição a X , isso não signifique que a variável aleatória resultante de $\lim_{n \rightarrow +\infty} X_n$ tenha a mesma função densidade de probabilidade de X , é um pouco mais difícil. Vamos dar um exemplo de como isso pode acontecer.

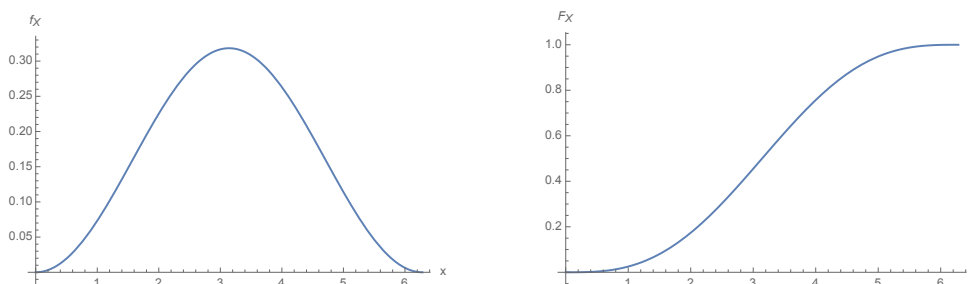
Exemplo 7. *Seja uma sequência $(X_n)_{n \in \mathbb{N}}$ de variáveis aleatórias. Suponha que X_n tenha função densidade de probabilidade f_n dada por*

$$f_{X_n}(x) = \begin{cases} \frac{1 - \cos(nx)}{2\pi}, & x \in [0, 2\pi], \\ 0, & x \in \mathbb{R} \setminus [0, 2\pi]. \end{cases} \quad (6.1)$$

Fazendo a integração de f_{X_n} , obtemos a distribuição acumulada de probabilidade F_{X_n} correspondente.

$$F_{X_n}(x) = \begin{cases} \frac{x}{2\pi} - \frac{\sin(nx)}{2\pi n}, & x \in [0, 2\pi], \\ 0, & x \in \mathbb{R} \setminus [0, 2\pi]. \end{cases} \quad (6.2)$$

As funções f_{X_n} e F_{X_n} para $n = 1$ estão mostradas na figura 6.2.



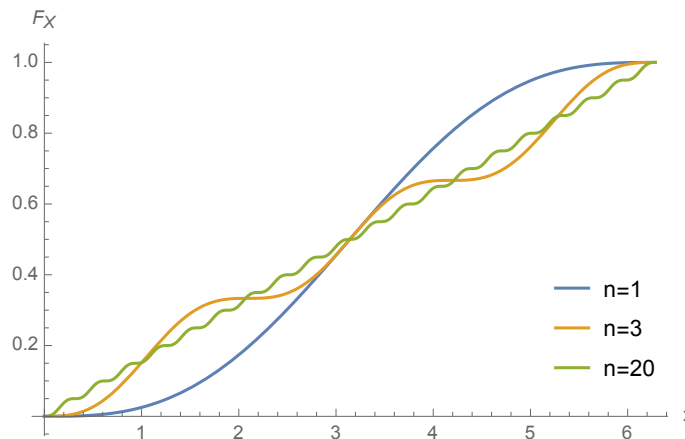
(a) Função densidade de probabilidade f_{X_n}

(b) Função distribuição acumulada F_{X_n} .

Figura 6.2: Função distribuição e densidade.

Observando (6.2) é muito claro que $F_{X_n} \rightarrow F_X$, sendo F_X a função distribuição acumulada de probabilidade de uma variável aleatória uniformemente distribuída em $[0, 2\pi]$.

$$F_{X_n}(x) \rightarrow \begin{cases} \frac{x}{2\pi}, & x \in [0, 2\pi], \\ 0, & x \notin [0, 2\pi]. \end{cases} \quad (6.3)$$

Figura 6.3: Convergência de F_{X_n} .

A figura 6.3 mostra graficamente a convergência.

Entretanto, observe que f_{X_n} não converge a f_X da variável aleatória uniformemente distribuída em $[0, 2\pi]$. Para compreender isso, basta observar a integral

$$\int_0^{2\pi} |f_{X_n}(\xi) - f_X(\xi)| d\xi = \frac{2}{\pi}, \forall n \in \mathbb{N}.$$

Esse exemplo mostra claramente que a convergência em distribuição, que é aquela que aparece no Teorema do Limite Central é um modo de convergência muito fraco. Isso indica que seu uso deve ser usado com cuidado. \triangle

Convergência em probabilidade. Após ter visto acima a convergência em distribuição, veremos agora o que é a convergência em probabilidade, mas somente para o caso de variáveis aleatórias discretas, pois não é possível analisar o caso das variáveis aleatórias contínuas com apenas funções Probabilidade definidas usando a integral de Riemann, que é a disponível no ciclo básico de um curso de Engenharia.

Dizemos que uma sequência $(X_n)_{n \in \mathbb{N}}$, $X_n : \Omega \rightarrow \mathbb{R}$, de variáveis aleatórias discretas convergem em probabilidade para a variável aleatória $X : \Omega \rightarrow \mathbb{R}$ se para qualquer $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \rightarrow 0.$$

O grande exemplo desse modo de convergência é dado pelo Lei fraca dos Grandes números (Jacob Bernoulli, 1713), cujo enunciado é o seguinte.

Teorema 6.4.1. *Seja X_1, \dots uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média μ e variância $\sigma^2 < \infty$. Seja $S_n = \sum_{j=1}^n X_j$. Então dado qualquer $\epsilon > 0$,*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| \geq \epsilon \right) = 0.$$

Vamos ver a prova desse teorema quando X_1, \dots são variáveis aleatórias discretas. Devido à sua importância, uma seção inteira deste texto será dedicada a ela.

6.5 Lei fraca dos grandes números para variáveis aleatórias discretas

Desigualdade de Chebyshev. O teorema a seguir é importante porque dá uma estimativa para probabilidades que dependem *somente* dos parâmetros mais básicos de uma variável aleatória.

Teorema 6.5.1. *Para toda variável aleatória X que possui valor esperado $E[X]$ e variância $V[X]$, dado $\epsilon > 0$, vale a desigualdade*

$$\mathbb{P}(|X - E[X]| \geq \epsilon) \leq \frac{V[X]}{\epsilon^2}. \quad (6.4)$$

Vamos ver a prova para o caso de variáveis aleatórias discretas, que é essencialmente a mesma para o caso de variáveis aleatórias contínuas, mas evita questões técnicas ligadas à integração.

Suponha que X pode assumir valores em um conjunto enumerável $\{x_1, x_2, \dots\} \subset \mathbb{R}$, e que tenha valor esperado $E[X]$ e variância $V[X]$ bem definidas.

$$\begin{aligned} V[X] &= \sum_{j \in \mathbb{N}} (X_j - E[X])^2 P(X_j) \\ &\geq \sum_{\{j \in \mathbb{N} \mid |X_j - E[X]| \geq \epsilon\}} (X_j - E[X])^2 P(X_j) \\ &\geq \sum_{\{j \in \mathbb{N} \mid |X_j - E[X]| \geq \epsilon\}} \epsilon^2 P(X_j) \\ &= \epsilon^2 \mathbb{P}(\{|X - E[X]| \geq \epsilon\}). \end{aligned}$$

Logo,

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

Quando $\epsilon > 0$ é pequeno, é claro que o resultado é inócuo, mas quando $\epsilon > 0$ é tal que $\frac{V[X]}{\epsilon^2} < 1$, a estimativa (6.4) passa a ser importante. Note que essa estimativa depende *somente* dos parâmetros mais básicos de uma variável aleatória, que são o seu valor esperado e a sua variância, e de mais nada.

Ensaio de Bernoulli Realizamos um experimento que consiste em verificar-se o resultado de uma sucessão de ensaios. Cada ensaio pode ser considerado um sucesso com probabilidade p , ou não sucesso com probabilidade $1 - p$. Então, os elementos do espaço amostral Ω são sequências do tipo $\omega = 110001110010\dots$, com $1 =$ sucesso e $0 =$ não sucesso. Aqui, cada algarismo indica o resultado de um ensaio.

Seja X_i a variável aleatória que dá o resultado do i -ésimo ensaio.

Seja a variável aleatória $S_n = \sum_{i=1}^n X_i$. Note que S_n conta o número de sucessos no experimento até o n -ésimo ensaio.

Intuitivamente, espera-se que a sequência de variáveis aleatórias $(Y_n)_{n \in \mathbb{N}}$ definida por $Y_n = \frac{S_n}{n}$ convirja para p , isto é, $Y_n = \frac{S_n}{n} \rightarrow p$. De fato, é o que ocorre, pois

$$E[Y_n] = \frac{nE[X_1]}{n} = E[X_1] = p,$$

em que usamos o fato que $E[X_j] = E[X_1]$, pois todas as variáveis aleatórias são identicamente distribuídas, e

$$V[Y_n] = \frac{nV[X_1]}{n^2} = \frac{p(1-p)}{n}.$$

Assim, pela desigualdade de Chebyshev, temos, para todo $\epsilon > 0$ arbitrário, que

$$P(\{|Y_n - p| \geq \epsilon\}) \leq \frac{p(1-p)}{n\epsilon^2}, \quad (6.5)$$

isto é,

$$\lim_{n \rightarrow +\infty} \frac{S_n}{n} \rightarrow p$$

em probabilidade.

O que vimos nessa seção dá uma prova para a Lei fraca dos grandes números para o caso muito particular em que X_j são variáveis aleatórias discretas, e além disso advém de ensaios de Bernoulli. Entretanto, o resultado dado pelo Teorema 6.4.1 vale em todos os casos cobertos pelo seu enunciado. O mesmo comentário vale para a Desigualdade de Chebyshev, Teorema 6.5.1.

6.6 Teorema do Limite Central

O Teorema do Limite Central fortalece a Lei Fraca dos Grandes Números, pois não só diz para onde uma sequência de variáveis aleatórias converge, mas dá também sua distribuição.

Há várias versões. Damos aqui apenas uma, e sem prova.

Teorema 6.6.1. *Seja X_1, \dots uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2 . Seja $S_n = \sum_{j=1}^n X_j$. Então a variável aleatória $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ converge em distribuição para $Z \sim N(0, 1)$.*

Um caso particular, muito importante, de aplicação do Teorema do Limite Central, é o próprio experimento dos ensaios de Bernoulli, visto acima.

Exemplo 8. *Lembramos que em ensaios de Bernoulli, $X_i = 1$ com probabilidade p e $X_i = 0$ com probabilidade $(1 - p)$. Definimos $S_n = \sum_{i=1}^n X_i$.*

Temos

$$E[S_n] = np, \quad V[S_n] = np(1 - p).$$

Daí

$$Z_n = \frac{S_n - np}{\sqrt{np(1 - p)}} \rightarrow Z \sim N(0, 1),$$

em distribuição. △

6.7 Exemplos de fixação

Para fixarmos as ideias, elencamos agora alguns exemplos de fixação.

Convergência de uniformes. Podemos ver como a soma de variáveis aleatórias independentes converge rapidamente para uma normal.

Consideremos a soma das variáveis aleatórias X_1, \dots, X_6 independentes e uniformemente distribuídas no intervalo $[0, 1]$.

Sabemos que cada uma delas tem valor esperado $E[X_1] = \frac{1}{2}$ e variância $V[X_1] = \frac{1}{12}$.

A soma de duas, três e de seis delas está mostrada na figura 6.4.

Em particular, mostramos na figura 6.5 a comparação de $S_4 = \sum_{n=1}^4 X_n$ com a distribuição normal $N(2, 4\sqrt{\frac{1}{12}})$.

Observe como a aproximação acontece muito rapidamente.

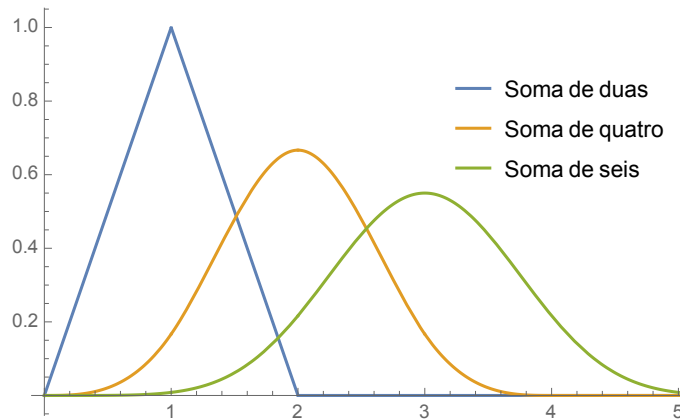


Figura 6.4: Soma de duas, quatro e seis variáveis aleatórias uniformes no intervalo $[0, 1]$ indedentes.

No laboratório de física. Muito se se diz que a distribuição normal é importante na prática, porque muito do que observamos experimentalmente é resultado da soma de muitas parcelas independentes. Dá-se o exemplo da soma de erros. Entretanto, muitas grandezas físicas são calculadas através de razões, como a velocidade, aceleração, pressão, etc. Podemos então indagar se produtos e divisões de variáveis aleatórias independentes se aproximam de alguma forma de uma normal.

De maneira geral, a resposta é não, mas em situações particulares a aproximação pela normal pode ser razoável. Considere a distribuição Lognormal. Ela é obtida através de uma variável aleatória normalmente distribuída da seguinte forma. Suponha que $X \sim N(\mu, \sigma^2)$. Então $Y = \log_e X$ tem distribuição Lognormal de parâmetros μ e σ . $E[Y]$ e $V[Y]$ são relacionados com μ e σ^2 por

$$E[Y] = \mu + \frac{\sigma^2}{2}, \quad V[Y] = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2}.$$

Quando a razão σ/μ é pequena, digamos, menor que 0.02, a Lognormal se aproxima de uma normal. Na figura 6.6 está mostrado um exemplo, que compara Y lognormal de parâmetros $\mu = 5$ e $\sigma^2 = 0.1^2$ com uma normal de mesmo valor esperado e variância.

Como

$$\log_e \prod_{j=1}^n X_j = \sum_{j=1}^n \log_e X_j,$$

vemos, pelo Teorema do Limite Central, que o logaritmo do resultado de um produto de fatores independentes tem distribuição normal, e pelo que

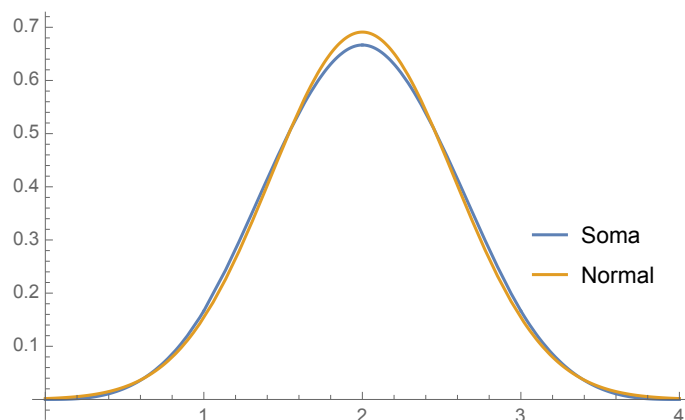


Figura 6.5: Comparação com a normal.

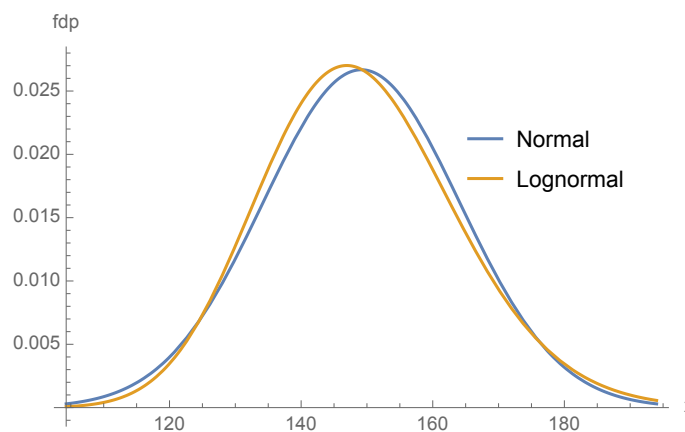


Figura 6.6: Comparação entre uma lognormal e uma normal.

discutimos acima, $\prod_{j=1}^n X_j$ pode ser aproximado por uma normal, desde que $\prod_{j=1}^n X_j$ tenha variância suficientemente pequena.

Por esse motivo, quando o estudante estiver no laboratório de física e estiver medindo experimentalmente a aceleração da gravidade, podemos supor que o resultado experimental segue uma normal, se as medidas de tempo e de distância forem suficientemente boas, isto é, tiverem variância pequena.

Façamos uma simulação numérica, supondo que cada grupo de alunos faça uma medida independente de tempo e de distância percorrida por um corpo com momento de inércia desprezível rolando um plano inclinado, conforme a figura 6.7. As distâncias ideais são $d = 2\text{m}$ e $h = 0.10\text{m}$. Cada

grupo de alunos pega uma trena, mede Δ , H e solta o corpo que rola até a base e mede o tempo até que isso ocorra. Supomos que o momento de inércia e o atrito são desprezíveis. Vamos supor que tanto os tempos com as distâncias sejam distribuídas uniformemente. Δ se distribui no intervalo $[2 - 0.01, 2 + 0.01]$, H se distribui em $[0.5 - 0.005, 0.5 + 0.005]$ e o tempo cronometrado no intervalo $[T - 0.03, T + 0.03]$.

Definimos uma nova variável aleatória G , que se tudo fosse medido com precisão absoluta, daria o valor exato g da aceleração da gravidade no laboratório.

$$G = \frac{2\Delta^2}{HT^2},$$

em que Δ é a distância percorrida, H é a altura do plano inclinado e T é o tempo gasto no percurso.

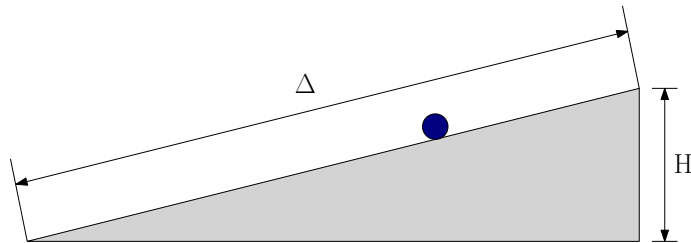


Figura 6.7: Experiência do plano inclinado.

Os resultados (simulados com o programa *Mathematica*) estão mostrados na figura 6.8, quando supomos a participação de 100 mil grupos.

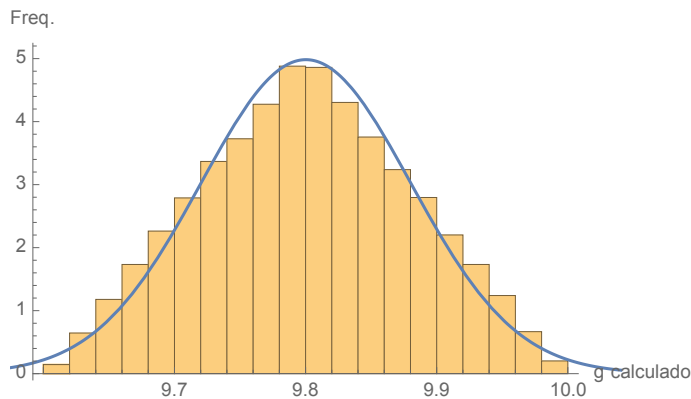


Figura 6.8: Resultados experimentais.

Note que mesmo que Δ , H e T sejam todas distribuídas uniformemente, a aceleração da gravidade calculada a partir dos dados experimentais segue

de forma muito próxima uma normal. Com os dados experimentais acima, calculamos os valores numéricos da média e do desvio padrão experimentais. Eles são respectivamente

$$\bar{G} = 9.8, \quad s^2 = 6.4 \times 10^{-3}.$$

Podemos usar s como a estimativa da variância σ^2 da distribuição normal subjacente e calcular a probabilidade que o valor de uma medida \bar{G} fique no intervalo $[g - \frac{s}{2}, g + \frac{s}{2}]$ em que $g = 9.8$ é a aceleração da gravidade terrestre no laboratório. Fazendo o cálculo, temos:

$$P(G \in [g - \frac{s}{2}, g + \frac{s}{2}]) = 0.382925.$$

Se o grupo de estudantes fizer $N = 3$ medidas independentes da aceleração da gravidade, obtendo 3 valores da variável aleatória G , a probabilidade que $\bar{G} = \frac{1}{N} \sum_{j=1}^N G_j$ se situe no mesmo intervalo acima, $[g - \frac{s}{2}, g + \frac{s}{2}]$, vale

$$P(\bar{G} \in [g - \frac{s}{2}, g + \frac{s}{2}]) = 0.866386,$$

pois

$$E[\bar{G}] = g = 9.8, \quad V[\bar{G}] = \frac{1}{9}3V[G] = 0.0266855.$$

Observe a melhoria do poder do experimento, quando usamos uma média com três valores apenas.

Esse exemplo suscita o questionamento filosófico do que seja probabilidade. Esse assunto é explorado no problema 4 da próxima seção.

6.8 Problemas

1. O objetivo deste problema é mostrar que a convergência da função distribuição acumulada de probabilidade não implica a convergência da função densidade de probabilidade. Faça um gráfico da função densidade de probabilidade f_{X_n} do exemplo 7 da página 122. Note que apesar de F_{X_n} convergir, a sequência de f_{X_n} não converge, no sentido que não existe nenhuma função densidade de probabilidade f tal que $\int |f(x) - f_{X_n}(x)| dx \rightarrow 0$.
2. O objetivo deste problema é comparar numericamente a lei fraca dos grandes números com o resultado do teorema do limite central. O teorema da lei fraca dos grandes números (1867 - Chebyshev) é

anterior ao teorema do limite central (entre outros, 1901 - Lyapunov). Faça uma estimativa no número de lances de uma moeda não viciada, usando (6.5), para que a média $m_n = \frac{\sum_{j=1}^n X_j}{n}$, satisfaça $P(|m_n - \frac{1}{2}| \leq \frac{1}{10}) \leq \frac{1}{100}$, com $X_j = 1$ se a moeda der cara e $X_j = 0$, caso contrário. Refaça a estimativa usando o teorema do limite central.

3. O objetivo deste problema é aplicar a desigualdade de Chebyshev. Só a imaginação é o limite para a criação de distribuições de probabilidade para variáveis aleatórias discretas ou para funções densidade de probabilidade para variáveis aleatórias contínuas, mas existe alguma variável aleatória X com valor esperado nulo e variância unitária, tal que a probabilidade $P(|X| \geq 2)$ seja maior que $1/4$?
4. Esse problema é objeto do vídeo que acompanha este texto. Os objetivos são explorar a dificuldade da interpretação do que seja probabilidade, apontar para um erro comum quando se usa probabilidades condicionadas e fornecer uma estimativa numérica usando a desigualdade de Chebyshev aplicada à lei dos grandes números.

Suponha que temos uma moeda viciada, que dá cara com probabilidade p e coroa com probabilidade $(1 - p)$. Lançamos N vezes a moeda independentemente e contamos o número X de ocorrências de caras e calculamos a razão $Y = \frac{X}{N}$, que é uma variável aleatória. Dado p , podemos calcular a probabilidade $P(|Y - p| \leq \epsilon | p)$. De fato, como X é uma variável aleatória binomial, as contas são bem simples. Por exemplo, se $N = 1000$, $p = \frac{1}{2}$ e $\epsilon = 10^{-2}$, então $|Y - p| \leq \epsilon$ é o mesmo que pedir que $N(1/2 - \epsilon) \leq X \leq N(1/2 + \epsilon)$, ou concretamente $490 \leq X \leq 510$, que corresponde a uma probabilidade de

$$\sum_{k=490}^{510} P(X = k|p) = \sum_{k=490}^{510} \binom{N}{k} p^k (1-p)^{(N-k)} = 0.49334.$$

Note que, na situação que estamos descrevendo, p é um valor conhecido, e calculamos a probabilidade que o *estimador* Y fique em torno do valor justo p . Explicitamente, calculamos o valor de

$$P(|Y - p| \leq \epsilon | p) \tag{6.6}$$

que vale, como vimos acima, 0.49334. Aqui a notação do condicionamento $P(\cdot | p)$ sobre p é um abuso de linguagem, uma vez que p não é uma variável aleatória, mas é uma notação comum para explicitar o fato que p é conhecido.

É interessante notar que a probabilidade acima calculada $P(N(1/2 - \epsilon) \leq X \leq N(1/2 + \epsilon)) = 0.49334$, poderia ter sido obtida aproximadamente aplicando-se o Teorema do Limite Central. X tem aproximadamente uma distribuição normal com média $\mu = Np$ e variância $\sigma^2 = N(1 - p)p$, isto é, $X \approx \tilde{X} \sim N(Np, N(1 - p)p)$. Nesse caso, $P(N(1/2 - \epsilon) \leq X \leq N(1/2 + \epsilon)) = 0.49334$ é aproximada por

$$P(N(1/2 - \epsilon) \leq \tilde{X} \leq N(1/2 + \epsilon)) = 0.472911,$$

cujo valor é calculado por uma rápida verificação consultando-se uma tabela da distribuição normal, ou melhor ainda, através de um programa qualquer.

Agora, suponhamos que realizamos o experimento como acima, obtendo $x = 497$ caras, e portanto $y = 497/1000 = 0.497$ é uma realização concreta da variável aleatória Y . No nosso contexto, uma pergunta sobre

$$P(|y - p| \leq \epsilon | Y = y)$$

não tem muito sentido, pois ou a resposta é 1 ou é 0. Estamos exatamente na mesma situação em que se pergunta qual é a probabilidade de ter saído uma cara ao mesmo tempo que olhamos para a moeda. Para o estudante reticente, temos duas questões que podem ajudar a esclarecer a situação. Qual é o espaço amostral de p usado? Qual é o espaço de eventos? Note que na teoria desenvolvida, p é um parâmetro e não uma variável aleatória.

Suponha que não conheçamos o valor de p , mas temos em mãos a realização concreta de $x = 497$ caras, e portanto $y = 497/1000 = 0.497$, como anteriormente. Já que p é desconhecido, poderíamos encará-la como uma variável aleatória. Essa é a abordagem Bayesiana. Podemos perguntar quanto vale

$$P(|y - p| \leq \epsilon | Y = y). \quad (6.7)$$

Isto é, estamos formulando uma pergunta sobre p , que agora é desconhecida e é tratada como uma variável aleatória, em uma interpretação não clássica, Bayesiana, em que probabilidades são vistas subjetivamente como graus de certeza.

A interpretação de (6.6) deve ser muito clara para o estudante neste ponto do curso. Dada $p = 1/2$, uma interpretação frequentista diria que no longo prazo, em 49.334% das vezes, o intervalo $[X/N - \epsilon, X/N + \epsilon]$ contém o valor p , pois

$$\left| \frac{X}{N} - p \right| \leq \epsilon \quad \Leftrightarrow \quad \frac{X}{N} - \epsilon \leq p \leq \frac{X}{N} + \epsilon.$$

Porém, quando p não é conhecido, uma interpretação mais natural seria a Bayesiana. Entretanto, uma visão clássica ainda é possível para (6.7), mesmo quando p é tratada como uma variável aleatória. Pode-se nesse caso imaginar um supra experimento em que a moeda com que se faz o experimento é resultado de um sorteio anterior, em que moedas honestas e viciadas estão em uma caixa, mas nem sempre isso é possível. Por exemplo, se p fosse um parâmetro físico como o a massa da Terra e Y algum resultado experimental como uma medida da aceleração da gravidade g . Nesse caso, há somente uma Terra.

Em termos práticos, estamos interessados em estabelecer um valor, o mais preciso possível, para p , quando ele é desconhecido, dentro do contexto clássico, não Bayesiano. Como proceder? Podemos fixar ϵ bem pequeno, digamos $\epsilon = 10^{-4}$ e determinar N para que tenhamos

$$P(|Y - p| \geq \epsilon) \leq \delta, \quad (6.8)$$

para δ também bem pequeno, digamos $\delta = 10^{-6}$. Assim, ao realizarmos o experimento, quando obtivermos uma realização y da variável aleatória Y , ele será bastante próximo de p .

- i. Usando $\epsilon = 10^{-4}$ e $\delta = 10^{-6}$, determine usando a desigualdade de Chebyshev, o número N de lançamentos independentes da moeda para que (6.8) seja satisfeita, para qualquer valor de p (sugestão: majore a variância, que é função de p). Na prática, Y é usado como estimador de p .

Feitas as contas, voltemos a refletir.

O intervalo $[Y - \epsilon, Y + \epsilon]$ contém p com probabilidade $1 - \delta$. Esse intervalo é conhecido na Estatística como intervalo de confiança. Como δ é muito pequeno, na grande maioria das vezes em que o experimento for feito, o intervalo de confiança vai conter p . Essa é a interpretação que devemos dar. Logo, as probabilidades envolvidas se referem *ao método* e *não* à qualidade do valor estimado para p .

- ii. Faça um programa simples (com qualquer software como o Mathematica, Matlab, Excel, etc.) para simular a determinação de um intervalo de confiança para p , usando ϵ e δ convenientes. Isto, é, fixe p , obtenha uma realização y de Y , e sem seguida obtenha $[y - \epsilon, y + \epsilon]$.
5. Seja uma sequência de variáveis aleatórias independentes $(X_n)_{n \in \mathbb{N}}$, cada uma distribuída uniformemente entre 1 e 7. Constrói-se a soma $Y = \sum_{n=1}^M X_n$. Tome $M = 5$.

- i) Calcule a probabilidade que Y fique compreendido entre M e $7M$ de forma exata.
 - ii) Calcule a probabilidade que Y fique compreendido entre M e $7M$ usando a aproximação pela normal.
 - iii) Calcule probabilidade que Y se situe entre $\frac{M}{2}$ e M .
 - iv) Calcule a probabilidade Y se situe entre $8M$ e $9M$.
 - v) A distribuição de Y é uma Normal ou uma uniforme, estritamente falando?
6. Infelizmente, com o tempo, os alunos começam a fazer as experiências no laboratório de física de maneira inversa. Por exemplo, no experimento do plano inclinado, como a aceleração da gravidade deve dar $g = 9.8 \text{ m/s}^2$, os alunos passam a fazer as contas ao contrário para determinar as distâncias e tempos a serem observados, para que os resultados “experimentais” deem certo. Convidamos os alunos a pensarem do ponto de vista do professor e atribuir uma nota em função da distância até o valor de $g = 9.8 \text{ m/s}^2$. Na sua opinião, quais seriam os valores de experimentais para g admissíveis quando os alunos fazem $N = 5$ medidas?

Bibliografia

- [Bar14] Robert G. Bartle. *The Elements of Integration and Lebesgue Measure*. John Wiley & Sons, 2014.
- [Bel04] D. R. Bellhouse. The reverend thomas bayes, FRS: A biography to celebrate the tercentenary of his birth. *Statistical Science*, 19(1):3–43, 2004.
- [Bel14] E.T. Bell. *Men of Mathematics*. Touchstone, 2014.
- [But99] Paul Butzer. P. L. Chebyshev (1821 -1894) -A Guide to his Life and Work. *Journal of Approximation Theory*, 96:111–138, 1999.
- [Dan13] Carlos Alberto Barbosa Dantas. *Probabilidade – Um Curso Introdutório*. EDUSP, 3rd. edition, 2013.
- [dL14] P.S. de Laplace. *Essai philosophique sur les probabilités*;. Mme. Ve. Courcier, 1814.
- [DL16] S. S. Demidov and B. V. Levshin. *The caso of academician Nikolaya Nikolaevich Luzin*, volume 20. American Mathematical Society, 2016.
- [Enc10] *Encyclopaedia Britannica*. Encyclopaedia Britannica, Inc., 2010.
- [Eva13] Lawrence C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, 2013.
- [Gre18] P. Greco. *La scienza e l'Europa. Il primo Novecento*. Le gerle. L'Asino d'Oro, 2018.
- [GS01] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, 3rd. edition, 2001.

- [Hac06] I. Hacking. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge Series on Statistical And Probabilistic Mathematics. Cambridge University Press, 2006.
- [Har00] G. H. Hardy. *Em defesa de um matemático*. Martins Fontes - selo Martins, 2000.
- [KM19] A.N. Kolmogorov and N. Morrison. *Foundations of the Theory of Probability*. AMS Chelsea Publishing. American Mathematical Soc., 2019.
- [Kol20] *Kolmogorov in perspective – History of Mathematics*, volume 20. American Mathematical Society and London Mathematical Society, 2020.
- [Kop99] Marek Capinski & Ekkehard Kopp. *Measure, integral and probability*. Springer, 1999.
- [Mcg15] Sharon Bertsch Mcgrayne. *A teoria que não morreria*. Ed.Perspectiva, 2015.
- [OR21] J. J. O’Connor and E. F. Robertson. Pierre – Simon Laplace. <https://mathshistory.st-andrews.ac.uk/Biographies/Laplace/>, Acessado em 31 de janeiro de 2021.
- [Sen13] Eugene Senete. A Tricentenary history of the Law of Large Numbers. *Bernoulli*, 4:1088 –1121, 2013.
- [Tod65] I. Todhunter. “A” *History of the Mathematical Theory of Probability: From the Time of Pascal to that of Laplace*. Macmillan, 1865.
- [TP93] I. Todhunter and K. Pearson. *A History of the Theory of Elasticity and of the Strength of Materials: From Galilei to the Present Time*. Number v. 2, no. 1. University Press, 1893.
- [Zab89] S. L. Zabell. Reflections on revolution. *Journal of the History of the Behavioral Sciences*, 25:391–395, 1989.