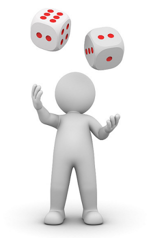


Notas de aula

um tanto bagunçada ainda, de

Introdução à Probabilidade

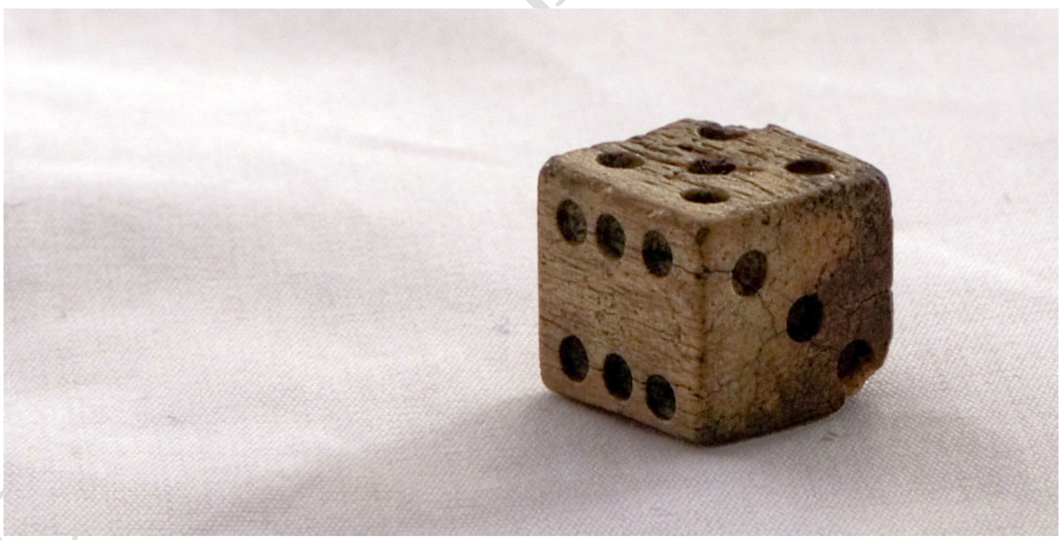
por J Donadelli

Estas são notas de aula que escrevo para a parte de probabilidade da disciplina *Introdução à probabilidade e à estatística* e da disciplina *Introdução à modelagem e processos estocásticos*. O texto abrange mais tópicos de probabilidade do que é ensinado nessas disciplinas — as seções com (∗) podem ser deixadas pra uma segunda leitura e as seções com (∗∗) podem ser deixadas para uma terceira leitura.

Sumário

§0 Prólogo

Ingenuamente *probabilidade* é uma forma quantitativa de expressar o conhecimento ou crença de que um evento ocorra e está vinculada a interpretações como a **clássica**, que trata os eventos elementares como igualmente prováveis, a **frequentista** que trata de experimentos que são imprevisíveis, bem definidos e podem ser repetidos, de modo que a probabilidade é a frequência relativa (limite) de ocorrência do evento e a interpretação **subjetiva** que atribui uma probabilidade a qualquer evento, mesmo quando não há aleatoriedade envolvida, a qual representa um grau (subjetivo) de crença e quanto maior a probabilidade mais plausível é sua ocorrência. Mais sobre interpretações pode se lido neste link[1]. Felizmente as mesmas propriedades matemáticas valem para quase todas as interpretações importantes, independentemente de interpretação, embora a escolha tenha grandes implicações pelo modo em que a probabilidade é usada para modelar o mundo real. Nós estamos interessados, em primeiro lugar, nessas propriedades que independem de interpretações; essa abordagem moderna não constrói conceitos de probabilidade e, sim, fundamentos matemáticos para o cálculo de probabilidade, trata probabilidade axiomaticamente e as propriedades são deduzidas dos axiomas.



0.1 Breve histórico. Um jogo de azar marca o começo do estudo sistemático de probabilidades pelos famosos matemáticos franceses Blaise Pascal (1623–1662) e Pierre de Fermat (1601–1665) por volta de 1654. Antoine Gombaud (1607–1684), o *Chevalier de Méré*, um escritor francês com interesse em jogos e apostas, chamou a atenção de Pascal para uma aparente contradição num jogo de dados que será explicado abaixo, no exemplo 1. Tal problema e outros colocados por de Méré levou

[1] http://plato.stanford.edu/entries/probability-interpret/

a uma troca de cartas entre Pascal e Fermat em que os princípios fundamentais da Teoria das Probabilidades foram formuladas pela primeira vez. Vários historiadores definem esse momento como o início da Teoria de Probabilidade apesar de alguns problemas específicos sobre os jogos de azar terem sido resolvidos por alguns matemáticos italianos nos séculos 15 e 16 usando probabilidade; em 1494 Fra Luca Paccioli (1447–1517) escreveu o primeiro trabalho impresso sobre probabilidade, *Summa de Arithmetica, geometria, proportioni e proportionalita* e em 1550 Geronimo Cardano (1501–1576), inspirado na *Summa* de Paccioli, escreveu um livro sobre jogos de azar chamado *Liber de Ludo Aleae*. Cálculos de probabilidades também aparecem nas obras do astrônomo e matemático alemão Johannes Kepler (1571–1630) e do astrônomo e matemático italiano Galileo Galilei (1564–1642).

O holandês Christian Huygens (1629–1695) estudou as correspondências citas acima e pouco depois, em 1657, publicou o primeiro livro sobre probabilidade, intitulado *De ratiociniis no ludo aleae* (Sobre o raciocínio em jogos de dados), um tratado sobre problemas relacionados jogos que estende o trabalho de Pascal e Fermat e foi a referência mais profunda até o final do século 17 sobre cálculos de probabilidades. Por causa do apelo inerente aos jogos de azar a Teoria de Probabilidades logo se tornou popular e o assunto se desenvolveu rapidamente durante o século 18. Os principais contribuintes durante este período foram Jakob Bernoulli (1654–1705), que estabeleceu a lei dos grandes números em sua forma mais simples, e Abraham de Moivre (1667–1754) que forneceu muitas ferramentas para tornar o método clássico mais útil, seu livro *The Doctrine of Chances*, de 1718, foi popular e passou por três edições. Thomas Bayes (1671–1746) introduziu o importantíssimo conceito de probabilidade condicional.

Ao longo do século 18 a aplicação de probabilidade passou de jogos de azar para problemas científicos. Em 1812 Pierre de Laplace introduziu uma série de novas ideias e técnicas matemáticas em seu livro *Théorie des Analytique Probabilités* que permitiu a disciplina ultrapassar sua (primeira) fase combinatória. Laplace aplicou ideias probabilísticas em muitos problemas científicos e práticos, ele destaca o papel da lei normal e mostra uma versão do teorema central do limite. Após a publicação do livro de Laplace o desenvolvimento matemático da teoria de probabilidades não teve grandes avanços por muitos anos. Em 1850 muitos matemáticos consideraram o método clássico não realista e redefiniram probabilidade em termos de frequências.

O desenvolvimento da teoria utilizando os métodos de análise ocupa o século 19 e início do século 20, notadamente com base no trabalho de Émile Borel (1871–1956) e Henri Lebesgue (1875–1941) sobre a Teoria da Medida. A Teoria dos Erros e a Mecânica Estatística (por James Maxwell (1831–1879) e Ludwig Boltzmann (1844–1906)) são exemplos de aplicações importantes desenvolvidas no século 19. Começam a surgir no final do século 19 os modelos para fenômenos aleatórios que evoluem ao longo do tempo; Francis Galton (1822–1911) e Henry Watson (1827–1903) estudam a evolução do número de indivíduos de uma população durante suas gerações, um exemplo de processo aleatório que será introduzido em toda a sua generalidade por Andrei Markov (1856–1922). Por volta de 1905

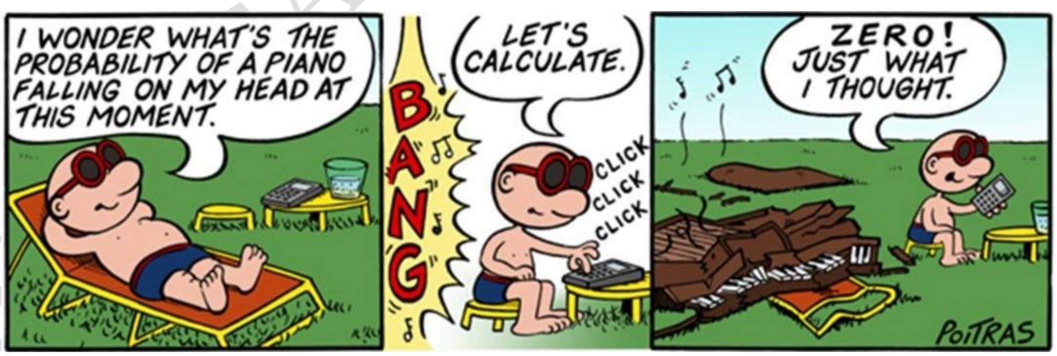
Albert Einstein (1879–1955) está interessado no conceito de movimento browniano; o botânico Robert Brown (1773–1858) tinha observado em 1827 o movimento desorganizado de uma partícula de pólen na superfície da água. Louis Bachelier (1870-1946), cinco anos antes de Einstein, introduziu o movimento Browniano para modelar a dinâmica do mercado de ações.

Uma das dificuldades no desenvolvimento de uma teoria matemática foi chegar a uma definição de probabilidade que é precisa o suficiente para uso em matemática, mas abrangente o suficiente para ser aplicável a uma grande variedade de fenômenos. A busca por uma definição amplamente aceitável levou quase três séculos e foi marcada por muita controvérsia. A questão foi finalmente resolvida no século 20, tratando a teoria de probabilidades em uma base axiomática, apresentada em 1933 numa monografia escrita pelo matemático russo Andrei Kolmogorov (1903–1987), que forma a base para a teoria moderna. Após o trabalho fundamental de Kolmogorov, Paul Lévy (1886–1971) deu o tom para o trabalho moderno sobre processos estocásticos.

Como tantos outros ramos da matemática, o desenvolvimento da Teoria de Probabilidades tem sido estimulado pela variedade das suas aplicações. Muitos matemáticos têm contribuído para a teoria desde a época de Laplace; dentre os mais importantes estão os russos Pafnuti Chebyshev (1821–1894), Andrei Markov (1856–1922) e Andrei Kolmogorov. Em 2006 Wendelin Werner foi o primeiro matemático da área de probabilidade a receber a medalha Fields e em 2007 Srinivasa Varadhan foi o primeiro probabilista a ganhar o Prêmio Abel. A medalha Fields e o Prêmio Abel são consideradas as maiores honrarias por mérito que um matemático pode receber.

0.2 Motivação.

Antes de começarmos nossos estudos vejamos exemplos simplificados da utilidade de modelos probabilísticos para tratar de problemas reais, alguns que afetam o nosso modo de vida.



Exemplo 1 (*Chevalier de Méré*). O escritor Antoine Gombaud, o *Chevalier de Méré*, viveu na França de 1607 a 1684 e é mais conhecido por sua contribuiçã na história da Teoria da Probabilidade. *Chevalier de Méré* era jogador que pensou ter descoberto uma maneira de ganhar dinheiro em um jogo de dados. A sua aposta era a seguinte: *em 4 lançamentos de um dado o* 6 *ocorre pelo menos uma vez*; a ideia é que se em um lançamento o 6 ocorre com probabilidade 1/6 então em 4 lançamentos a chance é

4 vezes maior, ou seja, $4/6 = 2/3$. Assim, no longo prazo, a cada 3 apostas ele vence 2. Essa essa estratégia o fez prosperar, até que resolveu testar outra aposta: *em 24 lançamentos de dois dados ocorre um par de* 6, *pelo menos uma vez*. A ideia é a mesma, a chance de um par de 6 em um lançamento é $1/36$ e em 24 lançamentos $24/36 = 2/3$. Entretanto, com essa nova estratégia Chevalier de Méré começou a perder dinheiro. Chevalier de Méré levou seu problema para o amigo matemático Blaise Pascal, que conjuntamente com Pierre de Fermat respondeu o problema lançando os fundamentos Teoria da Probabilidade. $\lozenge$

Rapidamente, uma explicação para o fenômeno relatado no exemplo acima é que a probabilidade de sair 6 pelo menos uma vez em 4 lançamentos é $1-(5/6)^4 \approx 0{,}51$, mas a probabilidade de sair par de 6 pelo menos uma vez em 24 lançamentos é $1-(35/36)^{24} \approx 0{,}49$. A razão da prosperidade no primeiro caso é a mesma que os cassinos usam em seu favor, apesar dos jogos envolverem aleatoriedade (não há dúvida de que os cassinos sempre lucram com as apostas), e a explicação probabilística disso é a Lei dos Grandes Números.

Exemplo 2 (*testes de primalidade*). Como algoritmos que testam se um dado número é primo são muito lentos na prática, os *softwares* que implementam essa função usam testes probabilísticos. Tais testes têm uma probabilidade pequena de decir erroneamente se um número é primo.

Por exemplo, para o *Mathematica* encontramos

> *Mathematica versions 2.2 and later have implemented the multiple Rabin-Miller test in bases 2 and 3 combined with a Lucas pseudoprime test as the primality test used by the function PrimeQ[n]*

e para o *Maxima* encontramos

> *Função: primep (n)*
>
> *Teste de primalidade. Se primep(n) retorna false, n é um número composto, e se ele retorna true, n é um número primo com grande probabilidade. Para n menor que* 10^{16} *uma versão deterministica do teste de Miller-Rabin é usada. Se primep(n) retorna true, então n é um número primo.*
>
> *Para n maior do que* 10^{16} *primep realiza primep_number_of_tests testes de pseudo-primalidade de Miller-Rabin e um teste de pseudo-primalidade de Lucas. A probabilidade com que n passe por um teste de Miller-Rabin é inferior a 1/4. Usando o valor padrão 25 para primep_number_of_tests, a probabilidade de n ser composto é muito menor do que* 10^{-15}.

Em *Python* a biblioteca *SymPy* implementa um teste de primalidade descrito como

> *Test if* n *is a prime number (True) or not (False). For* $n < 10^{16}$ *the answer is accurate; greater* n *values have a small probability of actually being pseudoprimes.*

Negative primes (e.g. -2) are not considered prime.

The function first looks for trivial factors, and if none is found, performs a safe Miller-Rabin strong pseudoprime test with bases that are known to prove a number prime. Finally, a general Miller-Rabin test is done with the first k bases which, which will report a pseudoprime as a prime with an error of about 4^{-k}. The current value of k is 46 so the error is about 2×10^{-28}.

O *Mathematica* e o *Maxima* são sistemas de computação algébrica e *Python* é uma linguagem de programação. O próximo exemplo contextualiza a importância dos testes de primalidade. ◊

Exemplo 3 (*Comércio eletrônico*). Numa transação eletrônica, pela internet, entre um Consumidor (**C**) e o comércio eletrônico em `www.vendoenaoentrego.com.br` (**S**) há troca de vários dados, como o número do cartão de crédito, *CPF*, endereço e outros que o consumidor espera estarem protegidos sabendo que o tráfego de informação na rede pode ser facilmente capturado por terceiros. A comunicação enter **C** e **S** se dá como segue, colocando de modo simplificado,

1. **S** *sorteia* dois números primos P e Q, determina $N := P \cdot Q$ e $M := (P-1)(Q-1)$;

2. **S** escolhe um par de números inteiros $1 < e, d < M$, ditos *chaves*, com o propriedades de que $\mathrm{mdc}(e, M) = 1$ e $e \cdot d \bmod M = 1$; os números e e N podem ser tornados públicos (ambos são de conhecimento público, qualquer um pode conhecer), mas os números P, Q e d devem ser mantidos em segredo;

3. **S** envia (N, e) para **C**;

4. se X é o número do cartão de crédito de **C**, então **C** envia o número $X^e \bmod N$ para **S**;

5. **S** recebe o número enviado e recupera o número X, do cartão de crédito, computando: $X = M^d \bmod N$.

Um espião que conhece os parâmetros públicos (N, e) e que captura o texto codificado M poderia tentar descobrir d, ou tentar fatorar N, ou tentar determinar $\sqrt[d]{M} \pmod N$, mas levaria mais de uma vida para conseguir fazê-lo com "força bruta" e um computador potente. Isto é, assumindo que os números escolhidos e sorteados no procedimento acima sigam alguns parâmetros de segurança, com a tecnologa atual é impossível descobrir os segredos e isso torna o procedimento de criptografia seguro. ◊

O que foi descrito acima é o protocolo para usar o algoritmo conhecido como RSA para criptografar e enviar X a um terceiro; mais explicações sobre esse algoritmo e porque ele funciona estão neste link[2]. O fato que merece alguma explicação agora é onde entra probabilidade nessa história. São dois pontos: (1) por razões de segurança P e Q devem ser inteiros escolhidos aleatoriamente; esses

[2] http://en.wikipedia.org/wiki/RSA

números têm que ser grandes e qualquer pista de como eles são escolhidos pode ajudar a fatorar N rapidamente, o que quebra a segurança do protocolo; (2) os números escolhidos aleatoriamente devem ser primos e testes primalidade[3] determinísticos são lentos o suficiente para tornarem esse método de criptografia inviável. Na prática são usados testes probabilísticos que podem errar ao afirmar que um número é primo com uma probabilidade muito pequena (como descrito no exemplo 2).

O leitor com acesso a um computador pode gerar um par de chaves (e, d) usando, por exemplo, o programa openssl[4] da seguinte forma: num terminal de linha de comando do sistema operacional, o comando

```
openssl genrsa -out private_key.pem 32
```

para gerar um par de chaves de 32 bits (não serve pra uso real por ser um número muito pequeno), tal comando devolve

```
Generating RSA private key, 32 bit long modulus
.+++++++++++++++++++++++++++
.+++++++++++++++++++++++++++
e is 65537 (0x10001)
```

cada + significa que o número *sorteado* passou no teste de Miller–Rabin[5], um teste probabilístico de primalidade. Nesse caso, foram realizados 27 testes e os números (P e Q) foram declarados primos em todos eles. A probabilidade de um deles não ser primo é menor que $0{,}555 \times 10^{-18}$ (é mais provável ganhar na mega-sena duas vezes seguidas). Para conhecer os parâmetros resultantes dessa operação basta digitar

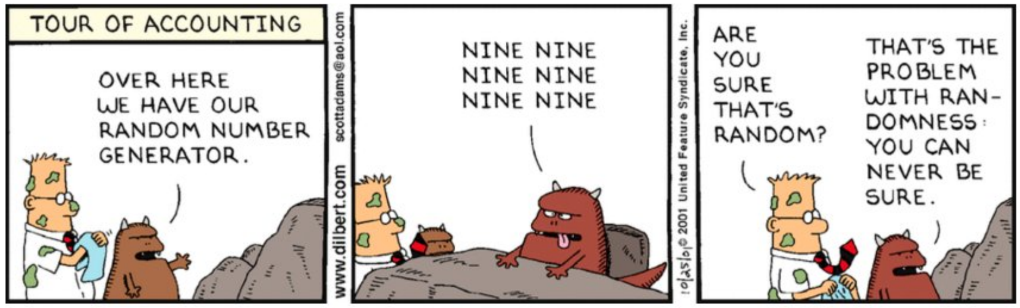
```
openssl rsa -text -in private_key.pem
```

que devolve

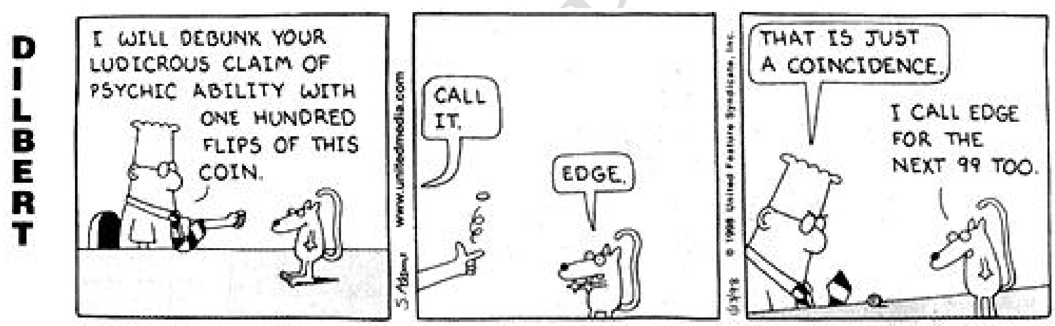
```
Private-Key: (32 bit)
modulus: 2883180901 (0xabd9d965)
publicExponent: 65537 (0x10001)
privateExponent: 386553605 (0x170a5705)
prime1: 58403 (0xe423)
prime2: 49367 (0xc0d7)
```

ou seja, $N = PQ = 2.883.180.901$, com $P = 58.403$ e $Q = 49.367$,e $e = 65537$ e $d = 386553605$.

[3] http://en.wikipedia.org/wiki/Primality_test
[4] http://en.wikipedia.org/wiki/OpenSSL
[5] http://en.wikipedia.org/wiki/Miller-Rabin_primality_test



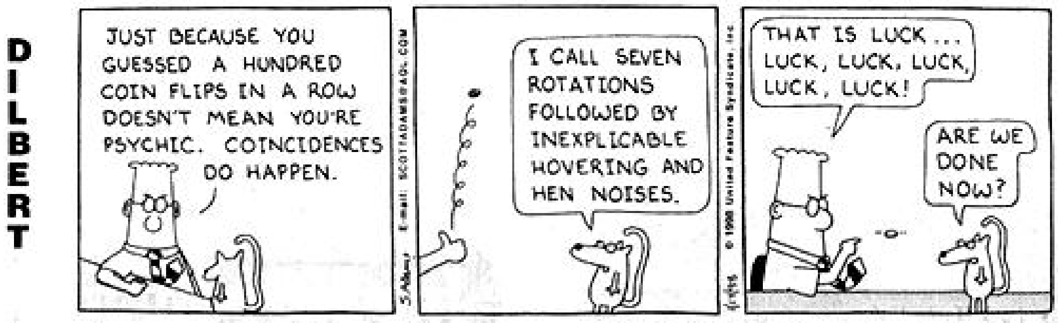
Exemplo 4 (*p-valor*). Para testar a eficácia de uma nova droga é realizado um experimento com dois grupos de sujeitos, um toma a droga e outro toma um placebo. O resultado do grupo que toma droga ser melhor que o resultado do outro grupo não é suficiente para constatar a eficácia da droga, assim como não é suficiente eu alegar que o sol nasce todo dia no leste para comprovar que tenho poderes paranormais. É preciso descartar a negação dessa hipótese, isto é, precisamos descartar a hipótese da droga não fazer efeito. Infelizmente, para os pesquisadores, há sempre a possibilidade de que a droga não seja efetiva, ou seja, que não seja observada uma diferença entre os grupos. Essa possibilidade é chamada *hipótese nula*. Se pela hipótese nula eu não tenho poderes especiais e eu faço o sol não nascer então temo um resultado bastante improvável pela hipótese nula.



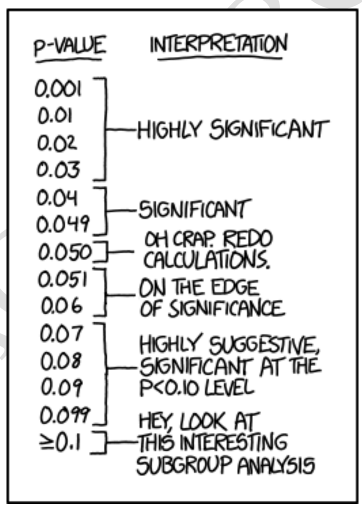
Consideremos um teste para uma droga que nós sabemos é totalmente ineficaz, a hipótese nula é verdade: não existe nenhuma diferença de resultado entre os grupos experimentais. Apesar da hipótese nula ser verdadeira é possível ocorrer um efeito diferencial nos dados da amostra devido a aleatoriedade na amostra. De fato, é extremamente improvável que os grupos se comportem igual a hipótese nula. Suponhamos que a chance de não haver cura é 10% nos dois grupos, cada um com 50 pessoas. Então, a probabilidade de nos dois grupos ocorrer o mesmo número de pessoas não curadas é, aproximadamente, 13%. A chance de um grupo de dar melhor que o outro é, aproximadamente, 43%. Consequentemente, a posição da hipótese nula é que a diferença observada na amostra não reflete a verdadeira diferença de efeito nas populações.

Agora, suponha que todas as pessoas que tomaram a droga foram curadas enquanto que 5 do grupo do placebo não foram curadas. A probabilidade de que todas as pessoas tenhas sido curadas é 0,005, ou 1 chance em 200, o que é bastante improvável para a hipótese da droga não fazer efeito.

As probabilidades calculadas acima medem o quanto compatível os dados estão com a hipótese nula. Qual a probabilidade de o efeito observado se a hipótese nula é verdadeira? Esse valor é dito *p*-valor. Um *p*-valor baixo sugere que a amostra fornece evidência suficiente para podermos rejeitar a hipótese nula.



Um *p*-valor é a probabilidade de se obter um resultado pelo menos tão extremo como o seus assumindo a verdade da hipótese nula. Por exemplo, se no estudo da droga obtém-se um *p*-valor de 0,04, ele indica que se a vacina não teve efeito então esses resultados seriam observados em 4% dos estudos devido a amostragem aleatória. ◊



A teoria de probabilidades é frequentemente ilustrada com dispositivos simples de jogos de azar: moedas, dados, cartas, urnas com bolas e assim por diante. Exemplos baseadas em tais dispositivos são valiosos por causa de sua simplicidade e clareza conceitual. Problemas envolvendo moedas, dados, etc. são metáforas para os problemas mais complexos e realistas.

[*Rosencrantz and Guildenstern are riding horses down a path - they pause*]
R: Umm, uh...
[*Guildenstern rides away, and Rosencrantz follows. Rosencrantz spots a gold coin on the ground*]
R: Whoa - whoa, whoa.
[*Gets off horse and starts flipping the coin*] **R**: Hmmm. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads. Heads.

[*Guildenstern grabs the coin, checks both sides, then tosses it back to Rosencrantz*]
R: Heads.
[*Guildenstern pulls a coin out of his own pocket and flips it*]
R: Bet? Heads I win?
[*Guildenstern looks at coin and tosses it to Rosencrantz*]
R: Again? Heads.

[...]

R: Heads
G: A weaker man might be moved to re-examine his faith, if in nothing else at least in the law of probability.
R: Heads
G: Consider. One, probability is a factor which operates within natural forces. Two, probability is not operating as a factor. Three, we are now held within um...sub or supernatural forces. Discuss!
R: What?
[...]
R: Heads, getting a bit of a bore, isn't it?

[...]

R: 78 in a row. A new record, I imagine.
G: Is that what you imagine? A new record?
R: Well...
G: No questions? Not a flicker of doubt?
R: I could be wrong.
G: No fear?
R: Fear?
G: Fear!
R: Seventy nine.

[...]

G: I don't suppose either of us was more than a couple of gold pieces up or down. I hope that doesn't sound surprising because its very unsurprisingness is something I am trying to keep hold of. The equanimity of your average tosser of coins depends upon a law, or rather a tendency, or let us say a probability, or at any rate a mathematically calculable chance, which ensures that he will not upset himself by losing too much nor upset his opponent by winning too often. This made for a kind of harmony and a kind of confidence. It related the fortuitous and the ordained into a reassuring union which we recognized as nature. The sun came up about as often as it went down, in the long run, and a coin showed heads about as often as it showed tails.

Tom Stoppard, *Rosenkrantz and Guildenstern are dead* (1996).

0.3 Generalidades sobre conjuntos. A teoria dos conjuntos é uma linguagem que é adequada para descrever e explicar as estruturas matemáticas; de fato, toda a matemática pode ser descrita na teoria dos conjuntos. É possível desenvolver a teoria de conjuntos de maneira axiomática, como foi feito por Ernest Zermelo (1908) e Abraham Fraenkel (1922). A abordagem aqui é intuitiva e será suficiente para nossos propósitos.

0.3.1 Conjunto: é um conceito primitivo[6] da teoria dos conjuntos, que informalmente pode ser entendido como uma coleção não ordenada de entidades distintas, chamadas de **elementos** do conjunto. De fato, essa tentativa de definir conjuntos é circular pois usa o termo *coleção* que é quase sinônimo de conjunto. Assumimos que todos entendem a concepção intuitiva de conjuntos. Discorreremos sobre algumas propriedades que caracterizam o conceito de conjuntos. Eventualmente, no contexto de Probabilidade, é conveniente fixarmos um conjunto $\mathcal{U}$ donde são tomados os elementos que definem um conjunto, chamaremos $\mathcal{U}$ de **universo**[7]. Usamos as letras maiúsculas do início do alfabeto $A, B, C, \dots$ para denotar conjuntos.

Pertinência: dizemos que um elemento x *pertence* a um conjunto A se x é um elemento de A. Escrevemos $x \in A$ se x *é elemento de* A e escrevemos $x \notin A$ se x *não é elemento de* A.

Vazio: há um conjunto especial sem elementos chamado de conjunto vazio; o símbolo $\emptyset$ denota o conjunto vazio.

Igualdade de conjuntos: dois conjuntos A e B são iguais se têm os mesmos elementos.

Descrição: da igualdade de conjuntos podemos inferir que descrever todos os elementos de um conjunto é suficiente para defini-lo. Podemos descrever um conjunto de diversas formas. Se um conjunto tem poucos elementos, podemos listá-los entre chaves "$\{\}$". Por exemplo, o conjunto do dígitos primos é formado pelos números inteiros 2, 3, 5 e 7 e escrevemos $\{2, 3, 5\}$. Quando os conjuntos têm muitos elementos não é viável escrever todos seus elementos e uma solução comum, mas que só usamos quando o contexto não dá margem a ambiguidade sobre seu significado, é o uso de reticências $(\dots)$. Por exemplo, o conjunto dos naturais menores que 2.015 é descrito por $\{0, 1, \dots, 2.014\}$; o conjunto dos naturais pares $\{0, 2, 4, 6, \dots\}$. Podemos especificar um conjunto através de uma ou mais propriedade de seus elementos e, nesse caso, usamos a notação

$$\{a \in \mathcal{U} : P(a)\}$$

ou mesmo $\{a : P(a)\}$ quando o universo está subentendido, em que a é uma variável ligada ao universo $\mathcal{U}$ e $P(a)$ uma afirmação que pode ser verdadeira ou falsa dependendo do valor de a, o conjunto é formado por aqueles valores que têm a propriedade, ou seja, pelos $a \in \mathcal{U}$ tais que $P(a)$ é verdadeiro. Por exemplo, $\{x \in \mathbb{R} : x^2 \leqslant 2\} = [-\sqrt{2}, \sqrt{2}]$.

Existem alguns conjuntos de números que são muito usados em matemática e têm notações convencionais bem estabelecidas: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ e $\mathbb{R}$ denotam, respectivamente, os conjuntos dos números

[6] i.e., não tem uma definição explícita.

[7] Na teoria dos conjuntos "não existe um conjunto universo" é um teorema. Em probabilidade os conjuntos representam eventos aleatórios os quais representam alguns resultados de um fenômeno aleatório. Para cada fenômeno temos um "universo" de resultados possíveis

naturais, inteiros, racionais e reais. Ademais

$$\begin{aligned}
\mathbb{Z}^+ &:= \{x \in \mathbb{Z} : x > 0\} \\
\mathbb{Q}^+ &:= \{x \in \mathbb{Q} : x > 0\} \\
\mathbb{R}^+ &:= \{x \in \mathbb{R} : x > 0\} \\
\mathbb{Z}^* &:= \{x \in \mathbb{Z} : x \neq 0\} \\
\mathbb{Q}^* &:= \{x \in \mathbb{Q} : x \neq 0\} \\
\mathbb{R}^* &:= \{x \in \mathbb{R} : x \neq 0\}
\end{aligned}$$

Relação de inclusão: para um determinado universo $\mathcal{U}$, o conjunto A é *subconjunto* de um conjunto B, fato denotado por $A \subseteq B$, se todo elemento de A pertence a B, ou seja, $(\forall a \in \mathcal{U})(a \in A \Rightarrow a \in B)$.

Se A não é *subconjunto* de B denotamos $A \not\subseteq B$,

$$\begin{aligned}
A \not\subseteq B &\Leftrightarrow \text{não}\,((\forall a \in \mathcal{U})(a \in A \Rightarrow a \in B)) \\
&\Leftrightarrow (\exists a \in \mathcal{U})\text{não}(a \in A \Rightarrow a \in B) \\
&\Leftrightarrow (\exists a \in \mathcal{U})(a \in A \text{ ou } a \notin B)
\end{aligned}$$

Observemos que

$$A = B \Leftrightarrow A \subseteq B \text{ e } B \subseteq A.$$

e, assim, $A \neq B \Leftrightarrow A \not\subseteq B$ ou $B \not\subseteq A$.

0.3.2 Proposição. *Para qualquer conjunto* A, $\emptyset \subseteq A$.

Demonstração. Para provar que $\emptyset \subset A$ precisamos provar que para todo a, $a \in \emptyset \Rightarrow a \in A$. Mas como a hipótese $a \in \emptyset$ é falsa, a implicação é verdadeira. □

Usamos a notação $A \subset B$ com o mesmo significado de $A \subseteq B$ e, usaremos $A \subsetneq B$ para expressar $A \subset B$ e $A \neq B$.

0.3.3 Conjunto das partes: 2^A denota o conjunto formado por todos os subconjuntos de A, isto é,

$$B \in 2^A \Leftrightarrow B \subseteq A$$

conjunto das partes de A.

Algumas referências usam $\wp(A)$ ou $\mathcal{P}(A)$.

0.3.4 Operações sobre conjuntos: as operações sobre conjuntos definem novos conjuntos. A seguir descrevemos as operações mais usuais e suas propriedades.

União: $A \cup B$ denota a união dos conjuntos A e B que é o conjunto dos elementos x do universo que pertencem a A ou a B

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

Intersecção: $A \cap B$ denota a intersecção dos conjuntos A e B que é o conjunto dos elementos x do universo que pertencem a A e a B

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

A e B são **disjuntos** se $A \cap B = \emptyset$. Notemos que $A \cap B \subset A \subset A \cup B$ para quaisquer conjuntos A e B pois

$$x \in A \cap B \Rightarrow x \in A \text{ e } x \in B \Rightarrow x \in A \cup B.$$

Diferença: $A \setminus B$ denota o conjunto dos elementos x do universo que pertencem a A e não a B

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

Diferença simétrica: $A \triangle B$ denota o conjunto dos elementos x do universo que pertencem exclusivamente a A ou a B, não a ambos

$$A \triangle B = \{x \in \mathcal{U} : x \in A \cup B \text{ e } x \notin A \cap B\}.$$

Complemento: se $A \subset \mathcal{U}$, o símbolo $\overline{A}^{\mathcal{U}}$ denota complemento de com relação ao universo $\mathcal{U}$,

$$\overline{A}^{\mathcal{U}} = \mathcal{U} \setminus A = \{x \in \mathcal{U} : x \notin A\}$$

e se não há perigo de confusão (ambiguidade) usamos simplesmente $\overline{A}$. Algumas referências usam a notação $A^{\complement}$ para denotar $\overline{A}$.

Propriedades das operações em conjuntos: deixamos para o leitor a verificação das seguintes propriedades

Leis de identidade

$$A \cap \mathcal{U} = A$$
$$A \cup \emptyset = A$$

Leis de dominação

$$A \cup \mathcal{U} = \mathcal{U}$$
$$A \cap \emptyset = \emptyset$$

Leis de idempotência

$$A \cup A = A$$
$$A \cap A = A$$

Leis do inverso

$$A \cup \overline{A} = \mathcal{U}$$
$$A \cap \overline{A} = \emptyset$$

Duplo complemento

$$\overline{\overline{A}} = A$$

Leis distributivas

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Leis comutativas

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

Leis associativas

$$A \cap (B \cap C) = (A \cap B) \cap C$$
$$A \cup (B \cup C) = (A \cup B) \cup C$$

Leis de De Morgan

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Leis de absorção

$$A \cup (A \cap B) = A$$
$$A \cap (A \cup B) = A$$

Prova de uma das leis de De Morgan. Sejam A e B conjuntos e vamos provar que $\overline{A \cup B} = \overline{A} \cap \overline{B}$. Para provar essa igualdade, precisamos provar que (i) $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$ e que (ii) $\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}$.

Para provar (i), seja $x \in \overline{A \cup B}$

$$\begin{aligned}
x \in \overline{A \cup B} \Rightarrow &\quad x \notin A \cup B && \text{por definição de complemento}\\
\Rightarrow &\quad \text{não}(x \in A \cup B) && \text{por definição de } \notin\\
\Rightarrow &\quad \text{não}(x \in A \text{ ou } x \in B) && \text{por definição de } \cup\\
\Rightarrow &\quad \text{não}(x \in A) \text{ e não}(x \in B) && \text{por De Morgan (proposicional)}\\
\Rightarrow &\quad x \notin A \text{ e } x \notin B && \text{por definição de complemento}\\
\Rightarrow &\quad x \in \overline{A} \cap \overline{B} && \text{por definição de } \cap
\end{aligned}$$

Portanto $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$.

Para provar (ii) basta notar que a recíproca de todas as implicações no argumento acima são verdadeiras, portanto $\overline{A} \cap \overline{B} \subseteq \overline{A \cup B}$.

Das duas inclusões segue o teorema. □

Podemos simplificar $\overline{\overline{(A \cup B) \cap C} \cup \overline{B}}$ usando as propriedades acima

$$\begin{aligned}
\overline{\overline{(A \cup B) \cap C} \cup \overline{B}} &= (\overline{\overline{(A \cup B) \cap C}}) \cap \overline{\overline{B}} && \text{De Morgan}\\
&= ((A \cup B) \cap C) \cap B && \text{duplo complemento}\\
&= (A \cup B) \cap (C \cap B) && \text{associativa}\\
&= (A \cup B) \cap (B \cap C) && \text{comutativa}\\
&= ((A \cup B) \cap B) \cap C && \text{associativa}\\
&= B \cap C && \text{absorção}
\end{aligned}$$

0.3.5 União e interseção com mais de dois conjuntos: Sejam $A_0, A_1, A_2, \ldots$ conjuntos

(a) $\displaystyle\bigcup_{i=1}^{n} A_i := \{x \in \mathcal{U} : x \in A_i \text{ para algum } i \in \{1, \ldots, n\}\}$;

(b) $\displaystyle\bigcap_{i=1}^{n} A_i := \{x \in \mathcal{U} : x \in A_i \text{ para todo } i \in \{1, \ldots, n\}\}$.

0.3.6 Partição: o conjunto $\{A_1, A_2, \ldots, A_n\}$ é uma *partição* do conjunto A se seus elementos são subconjuntos não-vazio de A, disjuntos e a união deles é A, isto é,

(a) $\emptyset \neq A_i \subset A$ para todo i;

(b) $\forall i \, \forall j \, (j \neq i \Rightarrow A_i \cap A_j = \emptyset)$;

(c) $\displaystyle\bigcup_{i=1}^{n} A_i = A$.

Exemplo 5. Sejam R_0, R_1 e R_2 subconjuntos de $\mathbb{N}$ definidos por

$$R_i = \{n \in \mathbb{N} : n \text{ dividido por 3 deixa resto } i\}$$

$\{R_0, R_1, R_2\}$ é uma partição de $\mathbb{N}$. $\Diamond$

0.3.7 Produto cartesiano: denotamos por (a, b) um par ordenado de elementos, no qual a é o primeiro elemento e b é o segundo elemento. O produto cartesiano dos conjuntos não vazios A e B é escrito $A \times B$ e denota o conjunto dos pares ordenados com o primeiro elemento em A e o segundo em B

$$A \times B = \{(a, b) : a \in A \text{ e } b \in B\}.$$

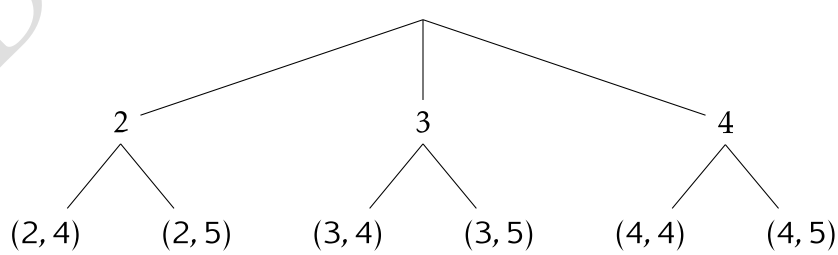
Par ordenado: o matemático polonês Kazimierz Kuratowski definiu par ordenado em termos de conjunto como

$$(a, b) := \{\{a\}, \{a, b\}\}.$$

Daí é claro que um par ordenado é diferente de um conjunto de dois elementos, pois a ordem é importante (por exemplo, o par $(1, 2)$ é diferente do par $(2, 1)$, verifique usando a definição acima). Dois pares ordenados (a, b) e (c, d) são iguais se, e somente se, $a = c$ e $b = d$ (deduza esse fato da definição dada acima).

Como no produto cartesiano os pares são ordenados, temos que $A \times B \neq B \times A$ (exceto quando $A = B$ ou $A = \emptyset$ ou $B = \emptyset$).

Uma maneira de representar produtos cartesianos, que é viável quando envolve conjuntos pequenos, é chamado de *diagrama de árvore*. Por exemplo, no caso $A = \{2, 3, 4\}$ e $B = \{4, 5\}$, o produto cartesiano é representado pela árvore



Na raiz temos uma ramificação para cada elemento na primeira coordenada, em cada ramo temos uma ramificação para cada elemento na segunda coordenada. No fim de cada ramo da árvore há um elemento do produto cartesiano.

De um modo geral, se $A_1, A_2, \dots, A_n$ são conjuntos não vazios

$$\bigtimes_{i=1}^{n} A_i = A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \dots, a_n) : a_i \in A_i (\forall i)\}.$$

No caso em que os conjuntos $A_1, A_2, \dots, A_n$ são iguais a A denotamos por A^n o produto cartesiano $\bigtimes_{i=1}^{n} A_i$.

0.4 Cardinalidade.

Cardinalidade é um conceito da Teoria dos Conjuntos que estende para qualquer conjunto a noção quantidade de elementos de um conjunto, a qual é intuitivamente clara no caso de conjuntos finitos. A cardinalidade do conjunto A é denotada por $|A|$. Na verdade, a ideia de cardinalidade torna-se bastante sutil quando os conjuntos são infinitos. O ponto principal aqui é explicar como existem tipos diferentes infinito e alguns infinitos são maiores do que outros.

0.4.1 Conjunto finito: nos casos em que o conjunto A é vazio ou existe uma bijeção $f : \{1, 2, \dots, n\} \to A$ nós dizemos que A é finito. Uma tal bijeção é chamada de **enumeração**. Desse modo, $A = \{f(1), f(2), \dots, f(n)\}$ e dizemos que A tem n elementos. Observemos que se $g : \{1, 2, \dots, m\} \to A$ é uma outra bijeção então $m = n$ (prove). Assim, definimos que a cardinalidade de A é n e denotamos esse fato por $|A| = n$. Ademais, $|A| = 0$ se e só se $A = \emptyset$.

0.4.2 Teorema (princípio aditivo). *Se A e B são conjuntos finitos e disjuntos, então $|A \cup B| = |A| + |B|$.*

Demonstração. Sejam A e B conjuntos disjuntos com cardinalidade n e m, respectivamente. Se pelo menos um deles for vazio então o teorema claramente vale. Vamos supor $m, n > 0$ e vamos mostrar uma bijeção $h : \{1, 2, \dots, n + m\} \to A \cup B$.

Se $f : \{1, 2, \dots, n\} \to A$ e $g : \{1, 2, \dots, m\} \to B$ são bijeções então definimos h por

$$h(x) = \begin{cases} f(x), & \text{se } 1 \le x \le n \\ g(x - n), & \text{se } n + 1 \le x \le n + m. \end{cases} \tag{0.4.1}$$

h é sobrejetora: se $y \in A \cup B$, então $y \in A$ ou $y \in B$, mas não em ambos já que são disjuntos. Se $y \in A$ então $f(x) = y$ para algum $x \in \{1 \dots, n\}$, portanto $h(x) = y$. Se $y \in B$ então $g(x) = y$ para algum $x \in \{1 \dots, m\}$, portanto, $h(x + n) = g(x)$. Ainda, h é injetora: como A e B são disjuntos, se $h(x) = h(y)$ então $f(x) = f(y)$ ou $g(x) = g(y)$, o que é uma contradição. □

A notação 2^A para o conjunto das partes de A é em parcialmente inspirada no fato de que $|2^A| = 2^{|A|}$ para qualquer A finito.

0.4.3 Teorema. *Todo conjunto A de cardinalidade $n \in \mathbb{N}$ tem 2^n subconjuntos distintos.*

Demonstração. Seja A um conjunto de cardinalidade n. Se $n = 0$ então $A = \emptyset$ é o único subconjunto dele mesmo e $2^0 = 1$. Se $n > 1$ então existe uma bijeção $f : \{1, 2, \dots, n\} \to A$. Como

$A = \{f(1), f(2), f(3), \dots, f(n)\}$, cada subconjunto $B \subset A$ corresponde a uma, e só uma, sequência $\mathbf{b}(B) = (b_1, b_2, \dots, b_n) \in \{0,1\}^n$ dada por

$$b_i = 1 \Leftrightarrow f(i) \in B$$

para cada $i \in \{1, 2, \dots, n\}$, ou seja

$$\begin{aligned} \mathbf{b} : 2^A &\to \{0,1\}^n \\ B &\mapsto \mathbf{b}(B) \end{aligned}$$

assim definida é bijetiva, de modo que $|2^A| = |\{0,1\}^n| = 2^n$ (prove a segunda igualdade usando indução em n). □

0.4.4 Teorema. *Se A e B são conjuntos finitos não vazios, então $|A \times B| = |A| \cdot |B|$.*

Demonstração. Seja $n = |A|$ e $A = \{f(1), f(2), \dots, f(n)\}$ para alguma enumeração f de A. Definimos os conjuntos dois-a-dois disjuntos $E_i = \{(f(i), b) \in A \times B : b \in B\}$ para todo i. Então $\{E_1, \dots, E_n\}$ é uma partição de $A \times B$ e

$$|A \times B| = \left| \bigcup_{i=1}^{n} E_i \right| = \sum_{i=1}^{n} |B| = |A||B|$$

onde a segunda igualdade segue da generalização do teorema 0.4.2 para o caso de união disjunta de $n \geq 2$ conjuntos. □

0.4.5 Conjunto infinito: se A não é finito então A é infinito. O conjunto dos naturais não é finito. De fato, se houvesse uma bijeção $f : \{1, 2, \dots, n\} \to \mathbb{N}$ então tomariamos $m = f(1) + f(2) + \cdots + f(n)$ de modo que m pertenceria à imagem de f contradizendo que $m > f(i)$ para todo $i \in \{1, \dots, n\}$.

No caso de conjuntos infinitos não se pode falar em quantidade de elementos e, além disso, dizer simplesmente que são infinitos elementos não diz muita coisa desde que Cantor nos mostrou a possibilidade de vários "tamanhos" de infinito, como veremos a seguir. Podemos, no entanto, comparar os "tamanhos" de conjunto da seguinte forma: os conjuntos A e B têm a *mesma cardinalidade*, e escrevemos $|A| = |B|$, se existe uma bijeção $f : A \to B$. Ademais, escrevemos $|A| \leqslant |B|$ se existe $f : A \to B$ injetiva. Também, escrevemos $|A| < |B|$ se $|A| \leqslant |B|$ e $|A| \neq |B|$, ou seja, se existe $f : A \to B$ injetiva mas não existe uma bijeção $A \to B$.

0.4.6 Teorema (Teorema de Cantor). *Para todo conjunto A, $|A| < |2^A|$.*

Demonstração. Se $A = \emptyset$ então $|A| = 0 < |2^A| = |\{\emptyset\}| = 1$. Seja A um conjunto não vazio. A função

$$\begin{aligned} f : A &\to 2^A \\ a &\to \{a\} \end{aligned}$$

é injetiva, portanto $|A| \leqslant |2^A|$. Pra mostrar que $|A| \neq |2^A|$ provaremos que não há sobrejeção $g: A \to 2^A$.

Suponhamos que $g: A \to 2^A$ é sobrejetiva. Para todo $a \in A$, $g(a) \subset A$. Definamos

$$B := \{a \in A : a \notin g(a)\}.$$

$B \subset A$ e g sobrejetiva implica que $B = g(b)$ para algum b. Daí $b \in B \Rightarrow b \notin g(b)$, pela definição do conjunto B, também, $b \notin B \Rightarrow b \in g(b)$, uma contradição. Isso completa a prova. □

O seguinte resultado é bastante famoso e não trivial no caso de conjuntos infinitos. A utilidade deste resultado vem do fato que, em geral, estabelecer uma bijeção que comprove $|A| = |B|$ pode ser muito difícil enquanto que estabelecer funções injetivas que comprovem $|A| \leqslant |B|$ e $|B| \leqslant |A|$ é mais fácil.

Teorema de Cantor–Schröder–Bernstein. *Se $|A| \leqslant |B|$ e $|B| \leqslant |A|$ então $|A| = |B|$.*

Uma demonstração do teorema será apresentada adiante.

Exemplo 6. Alguns exemplos importantes são:

(a) $|\mathbb{N}| = |\mathbb{Z}|$;

(b) $|\mathbb{N}| = |\mathbb{Q}|$;

(c) $|\mathbb{N}| < |\mathbb{R}|$;

(d) $|\mathbb{R}| = |(0,1)|$;

(e) $|\mathbb{R}^2| = |\mathbb{R}|$;

(f) $|2^{\mathbb{N}}| = |\mathbb{R}|$.

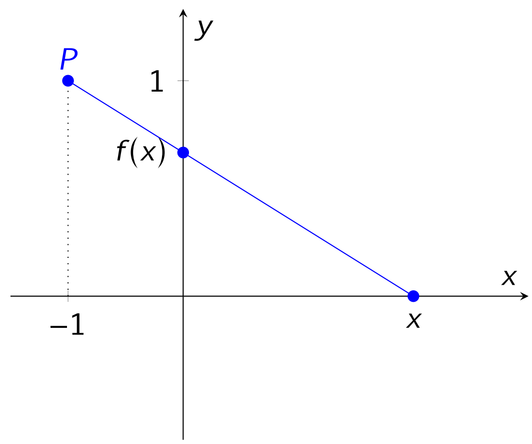
No item (d) podemos trocar $(0,1)$ por qualquer intervalo aberto da reta. ◊

Considerações a respeito do exemplo 6: para mostrar que $|\mathbb{N}| = |\mathbb{Z}|$ definimos a função $f: \mathbb{Z} \to \mathbb{N}$ dada por

$$f(z) = \begin{cases} 2z, & \text{se } z \geqslant 0 \\ 2(-z) - 1, & \text{se } z < 0. \end{cases}$$

Dado $n \in \mathbb{N}$, se n é par então $n = 2z$ para algum $z \in \mathbb{N}$, portanto $f(z) = n$; senão n é ímpar, $n = 2z - 1$ para algum $z \in \mathbb{Z}^+$, portanto $f(-z) = 2(-(-z)) - 1 = n$. Assim f é sobrejetora. Agora, se $f(z_1) = f(z_2)$ então $2z_1 = 2z_2$ ou $2(-z_1) + 1 = 2(-z_2) - 1$ e em ambos os casos $z_1 = z_2$. Portanto a função é bijetora.

Os exercícios 22 e 23 pedem para verificar bijeções $f: \mathbb{R}^+ \to (0,1)$, dada por $f(x) = \frac{x}{x+1}$, e $g: \mathbb{R} \to \mathbb{R}^+$, dada por $g(x) = 2^x$, que estabelecem que $|\mathbb{R}| = |(0,1)|$ (item (d) do exemplo) pois $f \circ g$ também é bijeção. A função f tem a seguinte interpretação gráfica



Para cada $x \in (0, +\infty)$ o valor $f(x)$ é dado pela intersecção da reta que passa por x e por $P = (-1, 1)$ com o eixo y. Usando semelhança de triângulos temos

$$\frac{1}{x+1} = \frac{f(x)}{x}$$

donde tiramos a expressão para $f(x)$.

Usando o Teorema de Cantor–Schröder–Bernstein podemos mostrar de modo fácil algumas das igualdades. Por exemplo, claramente há uma função injetiva $f : \mathbb{N} \to \mathbb{Q}$ pois $\mathbb{N} \subset \mathbb{Q}$, logo $|\mathbb{N}| \leq |\mathbb{Q}|$. Para mostrar que $|\mathbb{Q}| \leq |\mathbb{N}|$ consideremos os racionais não-nulo dados pelas frações da forma

$$\frac{p}{q},\ p \in \mathbb{Z} \text{ e } q \in \mathbb{N},\ \mathrm{mdc}(p, q) = 1$$

agora, definimos $g : \mathbb{Q} \to \mathbb{N}$ por $g(0) = 0$ e

$$g\left(\frac{p}{q}\right) = \begin{cases} 2^p 3^q, & \text{se } p > 0 \\ 5^p 3^q, & \text{se } p < 0 \end{cases}$$

que é injetiva (verifique). É possível exibir um bijeção entre $\mathbb{Q}$ e $\mathbb{N}$ mas isso também é bastante trabalhoso.

O item (c) do exemplo tem a famosa demonstração de Cantor por diagonalização. Como $\mathbb{N} \subset \mathbb{R}$, temos $|\mathbb{N}| \leqslant |\mathbb{R}|$ logo precisamos mostrar que $|\mathbb{N}| \neq |\mathbb{R}|$. Suponha que exista $f : \mathbb{N} \to (0, 1)$ bijetiva, então podemos enumerar (todos) os elementos do intervalo

$$\begin{aligned} f(0) &= 0{,}d_{0,0}d_{0,1}d_{0,2}d_{0,3}d_{0,4}\dots d_{0,n}\dots \\ f(1) &= 0{,}d_{1,0}d_{1,1}d_{1,2}d_{1,3}d_{1,4}\dots d_{1,n}\dots \\ f(2) &= 0{,}d_{2,0}d_{2,1}d_{2,2}d_{2,3}d_{2,4}\dots d_{2,n}\dots \\ &\ \vdots \\ f(n) &= 0{,}d_{n,0}d_{n,1}d_{n,2}d_{n,3}d_{n,4}\dots d_{n,n}\dots \\ &\ \vdots \end{aligned}$$

com $d_{i,j} \in \{0,1,2,3,4,5,6,7,8,9\}$. Consideremos o número real

$$\alpha = 0{,}d_0 d_1 d_2 d_3 \dots d_n \dots \text{ com } d_i \in \{0,1,2,3,4,5,6,7,8,9\} \setminus \{0,9,d_{i,i}\}\ (\forall i \in \mathbb{N}).$$

Esse número α pertence ao intervalo $(0,1)$ pois $d_i \neq 0$, logo α é diferente de $0 = 0{,}00000\dots$, e $d_i \neq 9$ logo α é diferente de $1 = 0{,}9999\dots$. Ademais, $\alpha \neq f(i)$ pois $d_i \neq d_{i,i}$ para todo $n \in \mathbb{N}$, uma contradição. Portanto, não existe $f : \mathbb{N} \to (0,1)$ bijetiva, tampouco $f : \mathbb{N} \to \mathbb{R}$ bijetiva.

Nos itens (e) e (f) basta mostrarmos que $|(0,1)\times(0,1)| \leqslant |(0,1)|$ e que $|2^{\mathbb{N}}| \leqslant |(0,1)|$. No primeiro caso, um ponto no quadrado $(0,1)\times(0,1)$ é da forma (x,y) com $x = 0{,}a_1 a_2 a_3 \dots$ e $y = 0{,}b_1 b_2 b_3 \dots$, e uma função injetiva sobre $(0,1)$ é dada quando mapeamos tal ponto em $0{,}a_1 b_1 a_2 b_2 a_3 b_3 \dots$ de $(0,1)$. No segundo caso, um subconjunto B de $\mathbb{N}$ pode ser representado por uma sequência binária infinita $b_0 b_1 b_2 \dots$ em que $b_i = 1 \Leftrightarrow i \in B$ para todo $i \in \mathbb{N}$. Essa sequência é mapeada na representação binária $0{,}b_0 b_1 b_2 \dots$ de um real do intervalo $(0,1)$; tal função é injetora (verifique).

0.4.7 Conjunto enumerável: o conjunto A é dito *enumerável* se é finito ou se tem a mesma cardinalidade de $\mathbb{N}$, isto é $|A| = |\mathbb{N}|$ de modo que $A = \{f(1), f(2), \dots\}$. $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{Q}$ são enumeráveis. $\mathbb{R}$ não é enumerável.

União enumerável de conjuntos: é definida por

$$\bigcup_{i\geqslant 1} A_i = \bigcup_{i=1}^{\infty} A_i := \{x \in \mathcal{U} : x \in A_i \text{ para algum inteiro } i \geqslant 1\}$$

Intersecção enumerável de conjuntos:

$$\bigcap_{i\geqslant 1} A_i = \bigcap_{i=1}^{\infty} A_i := \{x \in \mathcal{U} : x \in A_i \text{ para todo inteiro } i \geqslant 1\}.$$

Também, temos o produto cartesiano $A^{\mathbb{N}}$ formado por todas as sequências $(a_i : i \geqslant 1)$ de elementos de A.

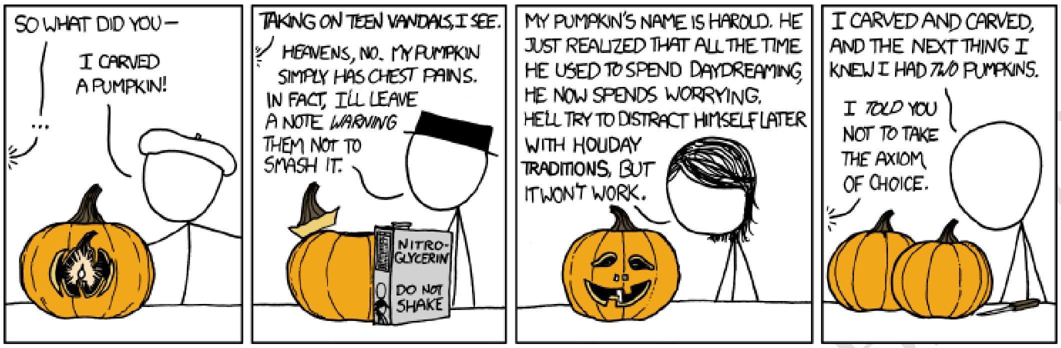
Uma dúvida que pode surgir nesse momento é saber se vale a lei de tricotomia para cardinalidades, ou seja, para quaisquer A e B, ou $|A| < |B|$, ou $|A| = |B|$, ou $|B| < |A|$. De fato, vale tal lei se assumirmos que vale o **axioma da escolha**: *seja C uma coleção de conjuntos não vazios. Então, podemos escolher um membro de cada conjunto nessa coleção.* Nesse caso, vale que para qualquer conjunto A

(a) se $|A| < |\mathbb{N}|$ então A é finito;

(b) se $|A| = |\mathbb{N}|$ então A é enumerável;

(c) se $|A| > |\mathbb{N}|$ então A é infinito e não enumerável.

Embora ingênuo e usado sem qualquer reserva pela maioria dos matemáticos, o axioma da escolha é independente na axiomática de Zermelo–Fraenkel para a Teoria dos Conjuntos e tem algumas consequências que já geraram controvérsia, uma dessas consequências é o famoso teorema (algumas vezes dito paradoxo) de Banach–Tarski: *é possível decompor uma esfera em seis partes que podem ser reagrupados por movimentos rígidas de modo a formar duas esferas da mesma dimensão que o original.*



0.4.8 Conjunto discreto e conjunto contínuo: chamamos de discreto qualquer conjunto enumerável, finito ou infinito, e chamamos de *conjunto contínuo* qualquer conjunto A com $|A|=|\mathbb{R}|$.

Acima, provamos que $|\mathbb{N}|<|\mathbb{R}|$ e que $|\mathbb{R}|=|2^{\mathbb{N}}|$. Por quase um século após a descoberta de Cantor de que há diferentes infinitos muitos matemáticos atacaram o problema de descobrir se existe um conjunto A tal que $|\mathbb{N}|<|A|<|\mathbb{R}|=|2^{\mathbb{N}}|$. Suspeitava-se que tal conjunto não existiria e a proposição que *não existe tal* A é conhecida como **hipótese do contínuo**. Gödel, nos anos 1930, provou que a negação da hipótese do contínuo não pode ser provada dentro dos axiomas padrão (i.e., axiomas de Zermelo–Fraenkel) da teoria dos conjuntos. Em 1964, Paul Cohen descobriu que nenhuma prova pode deduzir a hipótese do contínuo. Tomados em conjunto, os resultados de Gödel e Cohen significa que dos axiomas padrão da Teoria dos Conjuntos não se pode decidir se a hipótese do contínuo é verdadeira ou falsa; nenhum conflito lógico surge a partir da afirmação ou negação da hipótese do contínuo.

Demonstração do Teorema de Cantor–Schröder–Bernstein: Sejam A e B conjuntos tais que $|A|\leqslant|B|$ e $|B|\leqslant|A|$ e vamos mostrar que $|A|=|B|$. Sejam $f\colon A\to B$ e $g\colon B\to A$ funções injetivas, que existem por hipótese. Vamos mostrar que existe uma bijeção $h\colon A\to B$.

Definamos, para todo $X\subset A$

$$F(X):=A\setminus g(B\setminus f(X))=A\setminus g\left(\overline{f(X)}^{B}\right)=\overline{g\left(\overline{f(X)}^{B}\right)}^{A}$$

onde $f(X)$ é o subconjunto de B formado pela imagem dos elementos de X. Vamos mostrar que existe $A_0\subset A$ tal que $F(A_0)=A_0$. Primeiro, notemos que par uma sequência qualquer $(A_i : i\geqslant 1)$ de subconjuntos de A temos

$$
\begin{aligned}
F\left(\bigcap_{i\geqslant 1} A_i\right) &= \overline{g\left(\overline{f(\bigcap_{i\geqslant 1} A_i)}^{B}\right)}^{A} && \text{por definição}\\
&= \overline{g\left(\overline{\bigcap_{i\geqslant 1} f(A_i)}^{B}\right)}^{A} && \text{pois } f \text{ é injetiva}\\
&= \overline{g\left(\bigcup_{i\geqslant 1} \overline{f(A_i)}^{B}\right)}^{A} && \text{por De Morgan}\\
&= \overline{\bigcup_{i\geqslant 1} g\left(\overline{f(A_i)}^{B}\right)}^{A} && \text{pois } g \text{ é injetiva}\\
&= \bigcap_{i\geqslant 1} \overline{g\left(\overline{f(A_i)}^{B}\right)}^{A} && \text{por De Morgan}\\
&= \bigcap_{i\geqslant 1} F(A_i) && \text{por definição de } F.
\end{aligned}
$$

Tomemos

$$A_0 := A \cap F(A) \cap F^2(A) \cap F^3(A) \cap \cdots$$

onde $F^n(A) = F(F^{n-1}(A))$ donde temos

$$F(A_0) = F\Big(A \cap F(A) \cap F^2(A) \cap F^3(A) \cap \cdots\Big) = F(A) \cap F(F(A)) \cap F(F^2(A)) \cap F(F^3(A)) \cap \cdots$$

logo $F(A_0) = F(A) \cap F^2(A) \cap F^3(A) \cap F^4(A) \cap \cdots = A_0$ pois $A \supset F(A) \supset F^2(A) \supset \cdots$.

Desse modo $h : A \to B$ dado por

$$h(x) = \begin{cases} f(x), & \text{se } x \in A_0 \\ g^{-1}(x), & \text{caso contrário, isto é } x \in g\Big(\overline{f(A_0)}^{B}\Big) \end{cases}$$

é uma bijeção. Que é sobrejetiva: seja $y \in B$. Se $y \in f(A_0)$, então $y = f(x)$ para $x \in A_0$, portanto $y = h(x)$; senão, $y \notin f(A_0)$, ou seja $y \in \overline{f(A_0)}^{B}$, logo $g(y) \notin A_0$ logo $h(g(y)) = g^{-1}(g(y)) = y$, portanto h é sobrejetora. Que é injetiva: sejam $x, y \in A$ com $x \neq y$. A demonstração segue em três casos; (i) se $x, y \in A_0$ então $h(x) = f(x) \neq f(y) = h(y)$; (ii) se $x \in A_0$, então $h(x) = f(x) \in f(A_0)$, e se $y \notin A_0$, ou seja $y \in g\Big(\overline{f(A_0)}^{B}\Big)$, então $h(y) = g^{-1}(y) \in g^{-1}\Big(g\Big(\overline{f(A_0)}^{B}\Big)\Big) = \overline{f(A_0)}^{B}$, portanto $h(x) \neq h(y)$; (iii) se $x, y \notin A_0$ então $h(x) = g^{-1}(x) \neq g^{-1}(y) = h(y)$. Em todos os casos $h(x) \neq h(y)$, logo h é injetora.

Exercícios.

Exercício 1. Descreva o conjunto $2^{\emptyset}$.

Exercício 2. Seja $A = \{1, \{1\}, \{2\}\}$. Quais das afirmações a seguir são verdadeiras?

(a) $\emptyset \subseteq \{\emptyset\}$;

(b) $\emptyset \subseteq \emptyset$;

(c) $\emptyset \in \{\emptyset\}$;

(d) $\emptyset = \{\emptyset\}$;

(e) $\{\emptyset\} \subset \{\emptyset, \{\emptyset\}\}$;

(f) $\{\emptyset\} \in \{\emptyset, \{\emptyset\}\}$;

(g) $\{\emptyset\} = \{\emptyset, \{\emptyset\}\}$;

(h) $1 \in A$;

(i) $\{1\} \in A$;

(j) $\{2\} \in A$;

(k) $\{1\} \subseteq A$;

(l) $\{2\} \subseteq A$;

(m) $\{\{1\}\} \subseteq A$;

(n) $\{\{2\}\} \subseteq A$;

(o) $\{2\} \subsetneq A$;

Exercício 3. Prove cada uma das propriedades das operações de conjunto listas no texto.

Exercício 4. Use as definições ou as propriedades das operações para escrever uma dedução para:

(a) $A \subset A \cup B$;

(b) $B \subset A \cup B$;

(c) $A \cap B \subset A$;

(d) $A \cap B \subset B$;

(e) $B = (B \cap A) \cup (B \cap \overline{A})$;

(f) $A \triangle B = B \triangle A$;

(g) $A \triangle A = \emptyset$;

(h) $\overline{A \triangle B} = A \triangle \overline{B}$.

Exercício 5. Simplifique as expressões abaixo usando as propriedades de operações em conjuntos

(a) $A \cap (B \setminus A)$;

(b) $(A \cap B) \cup (A \cap B \cap \overline{C} \cap D) \cup (\overline{A} \cap B)$;

(c) $(A \setminus B) \cup (A \cap B)$;

(d) $\overline{A} \cup \overline{B} \cup (A \cap B \cap \overline{C})$.

Exercício 6. Enuncie e prove as leis de De Morgan no caso de união e interseção com mais de dois conjuntos.

Exercício 7. Prove que A e B são disjuntos se, e somente se, $A \triangle B = A \cup B$.

Exercício 8. Escreva o conjunto das partes de

(a) $\{1, 2, 3\}$;

(b) $\{\{1\}, \{2\}, \{3\}\}$.

(c) $\{\{1, 2\}, \{3\}\}$.

Exercício 9. Seja $q > 1$ um número inteiro. Para cada $i \in \{0, 1, \ldots, q-1\}$ denote por R_i o subconjunto dos números inteiros que deixam resto i quando divididos por q. Prove que $\{R_0, R_1, \ldots, R_{q-1}\}$ é uma partição de $\mathbb{Z}$.

Exercício 10. Sejam U um conjunto, $E \subset U$ e $\{A_1, A_2, \ldots, A_n\}$ uma partição de U. Prove que

$$E = \bigcup_{i=1}^{n} (A_i \cap E)$$

Exercício 11. (**Princípio de Inclusão–Exclusão**) Sejam $A_1, A_2, \dots, A_n$ conjuntos finitos. Prove que

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n}|A_i| - \sum_{1\leqslant i<j\leqslant n}|A_i\cap A_j| + \sum_{1\leqslant i<j<k\leqslant n}|A_i\cap A_j\cap A_k| + \cdots + (-1)^{k-1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}|A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}| + \cdots + (-1)^{n-1}|A_1\cap A_2\cap\cdots\cap A_n|$$

Exercício 12. Sejam $A_1, A_2, \dots, A_n$ conjuntos finitos. Prove que

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1|\cdot|A_2|\cdots|A_n|$$

(Dica: use o princípio da indução finita)

Exercício 13. Sejam A, B, C e D conjuntos não-vazios. Prove

(a) $A\times(B\cap C) = (A\times B)\cap(A\times C)$

(b) $A\times(B\cup C) = (A\times B)\cup(A\times C)$

(c) $(B\cap C)\times A = (B\times A)\cap(C\times A)$

(d) $(B\cup C)\times A = (B\times A)\cup(C\times A)$

(e) $A\times\emptyset = \emptyset$

(f) $A\times B\subseteq C\times D \Leftrightarrow A\subseteq C$ e $B\subseteq D$.

Exercício 14. Seja $A\subset\mathcal{U}$ tal que $|2^A| = n$. Determine $|2^B|$ se

(a) $B = A\cup\{x\}$ para algum $x\in\mathcal{U}\setminus A$;

(b) $B = A\cup\{x,y\}$ para algum $x,y\in\mathcal{U}\setminus A$;

(c) $B = A\cup\{x_1,x_2,\dots,x_k\}$ para $x_1,x_2,\dots,x_k\in\mathcal{U}\setminus A$.

Exercício 15. Mostre que o conjunto dos inteiros positivos ímpares tem a mesma cardinalidade que $\mathbb{N}$.

Exercício 16. Verifique se é verdadeira ou falsa cada uma das afirmações a seguir. No caso de ser falsa, apresente um contra-exemplo.

(a) se A e B são infinitos então $A\cap B$ é infinito;

(b) se B é infinito $A\subseteq B$ então A é infinito;

(c) se B é finito $A\subseteq B$ então A é finito;

(d) se A é finito $A\subseteq B$ então B é finito.

Exercício 17. Prove que acrescentar um novo elemento a um conjunto finito resulta num conjunto finito.

Exercício 18. Prove que remover um elemento de um conjunto infinito resulta num conjunto infinito

Exercício 19. Prove que todo conjunto infinito contem um subconjunto infinito enumerável.

Exercício 20. Prove que todo subconjunto de um conjunto finito também é finito.

Exercício 21. Prove que todo conjunto infinito contém um subconjunto próprio de mesma cardinalidade.

Exercício 22. Mostre que $f : \mathbb{R}^+ \to (0,1)$ dada por $f(x) = \frac{x}{x+1}$ é uma bijeção.

Exercício 23. Mostre que $g : \mathbb{R} \to \mathbb{R}^+$ dada por $g(x) = 2^x$ é uma bijeção.

Exercício 24. Constatamos, facilmente, que $|[0,1)| \leqslant |(0,1)|$ pois $f : (0,1) \to [0,1)$ dada por $f(x)$ é injetiva. Determine $g : [0,1) \to (0,1)$ injetiva de modo que $|[0,1)| = |(0,1)|$.

0.5 Limites de sequências de conjuntos.

Seja $(A_n : n \geqslant 1)$, uma sequência de conjuntos.

Sequências monótonas: a sequência é dita monótona se vale um dos casos

- **crescente**: $A_1 \subset A_2 \subset \cdots \subset A_n \subset A_{n+1} \subset \cdots$ e definimos

(0.5.1)
$$\lim_{n\to\infty} A_n := \bigcup_{n\geqslant 1} A_n.$$

- **decrescente**: $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots$ e definimos

$$\lim_{n\to\infty} A_n := \bigcap_{n\geqslant 1} A_n.$$

Por exemplo

$$\lim_{n\to\infty} [0,1/n) = \bigcap_{n\geqslant 1} [0,1/n) = \{0\}$$
$$\lim_{n\to\infty} (0,1/n) = \bigcap_{n\geqslant 1} (0,1/n) = \emptyset$$
$$\lim_{n\to\infty} (a-1/n, b+1/n) = \bigcap_{n\geqslant 1} (a-1/n, b+1/n) = [a,b]$$
$$\lim_{n\to\infty} [a+1/n, b-1/n] = \bigcup_{n\geqslant 1} [a+1/n, b-1/n] = (a,b)$$

Sequências arbitrárias: um sequência de conjuntos que não é monótona pode não ter limite (exiba um exemplo). Nesse caso, sempre existe o limite superior da sequência $(A_n : n \geqslant 1)$ que é o conjunto de todos elementos que pertencem a A_n's para infinitos valores de n

$$\limsup_{n\to\infty} A_n := \bigcap_{n\geqslant 1} \bigcup_{m\geqslant n} A_m.$$

de modo que

$$\omega \in \limsup_{n\to\infty} A_n \Leftrightarrow \forall n \geqslant 1, \exists m \geqslant n, (\omega \in A_m).$$

O limite inferior da sequência $(A_n : n \geqslant 1)$ é o conjunto de todos os elementos que pertencem a todos os A_n's exceto um número finito deles, ou seja, pertence a todo A_n para todo n suficientemente grande

$$\liminf_{n\to\infty} A_n := \bigcup_{n\geqslant 1} \bigcap_{m\geqslant n} A_m$$

de modo que

$$\omega \in \liminf_{n\to\infty} A_n \Leftrightarrow \exists n \geqslant 1, \forall m \geqslant n, (\omega \in A_m).$$

Naturalmente, a sequência é convergente se

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$$

e nesse caso o resultado comum é o limite da sequência

$$\lim_{n\to\infty} A_n. \tag{0.5.2}$$

0.6 Famílias estruturadas de conjuntos (∗).

Consideremos um conjunto não vazio S e um conjunto $\mathcal{A} \subset 2^S$ formado por subconjuntos de S, que chamamos de família ou classe de conjuntos.

0.6.1 Semiálgebra: $\mathcal{A}$ é uma semiálgebra de subconjuntos de S se

SA1 – $\emptyset \in \mathcal{A}$;

SA2 – para quaisquer $A, B \in \mathcal{A}$ vale que $A \cap B \in \mathcal{A}$;

SA3 – para qualquer $A \in \mathcal{A}$ vale que $\overline{A}$ é uma união finita de elementos de $\mathcal{A}$.

Como $\emptyset$ é um elemento da semiálgebra o próprio conjunto S é uma união finita de elementos da semiálgebra.

Semiálgebra dos intervalos semiabertos da reta: um exemplo importante de semiálgebra é dado pelos intervalos semiabertos da reta, isto é, os intervalos da forma $(a, b]$ com $a, b \in \mathbb{R}$ e $a \leqslant b$ (que é vazio no caso da igualdade), juntamente com os intervalos $(-\infty, +\infty) = \mathbb{R}$, $(-\infty, b]$ e $(a, +\infty)$ e com as convenções que $(a, +\infty]$ deve ser entendido como $(a, +\infty)$, que $x < +\infty$ para todo real x e que $x > -\infty$

para todo real x, temos: $\emptyset$ está na semiálgebra por definição,

$$(a,b]\cap(c,d]=\begin{cases}\emptyset & \text{se } b\leqslant c \text{ ou } d\leqslant a\\ (c,d] & \text{se } a\leqslant c\leqslant d\leqslant b\\ (a,b] & \text{se } c\leqslant a\leqslant b\leqslant d\\ (c,b] & \text{se } a\leqslant c\leqslant b\leqslant d\\ (a,d] & \text{se } a\leqslant c\leqslant b\leqslant d\end{cases}$$

também $\overline{(a,b]}=(-\infty,a]\cup(b,+\infty)$; nos outros casos deixamos a verificação como exercício.

0.6.2 Álgebra: de uma álgebra de subconjuntos de S se

A1 – $\emptyset\in\mathcal{A}$;

A2 – para quaisquer $A,B\in\mathcal{A}$ vale que $A\cap B\in\mathcal{A}$;

A3 – para qualquer $A\in\mathcal{A}$ vale que $\overline{A}\in\mathcal{A}$.

Notemos que para $A,B\in\mathcal{A}$ temos $A\cup B=\overline{\overline{A}\cap\overline{B}}$, logo $A\cup B\in\mathcal{A}$. São exemplos de álgebras de S: $\{\emptyset,S\}$, 2^S, $\{\emptyset,A,\overline{A},S\}$, $\{A\subset S : A \text{ é finito ou } \overline{A} \text{ é finito}\}$.

A família de todos os subconjuntos $I\subset\mathbb{R}$ que são união finita de intervalos disjuntos tomados na semiálgebra dos semiabertos da reta, digamos $I=\bigcup_{i=1}^{n}(a_i,b_i]$ com $a_1\leqslant b_1\leqslant a_2\leqslant b_2\leqslant a_3\leqslant\cdots\leqslant a_n\leqslant b_n$, formam uma álgebra. Claramente, $\emptyset$ é dessa forma. Ademais, para a interseção de duas uniões disjuntas de semiabertos temos

$$\left(\bigcup_{i=1}^{n}I_i\right)\cap\left(\bigcup_{j=1}^{m}J_j\right)=\bigcup_{i=1}^{n}\bigcup_{j=1}^{m}(I_i\cap J_j)$$

em que $I_i\cap J_j$ é um intervalo semiaberto; $I_i\cap J_j$ e $I_k\cap J_\ell$ são disjuntos para $i\neq k$ ou $j\neq\ell$. Também

$$\overline{\bigcup_{i=1}^{n}I_i}=\bigcap_{i=1}^{n}\overline{I_i}$$

que, como vimos acima, pode ser escrito como união disjunta de semiabertos.

Como **A3** implica (trivialmente) **SA3** temos que toda álgebra é uma semiálgebra.

0.6.3 Observação. *Uma álgebra de subconjuntos de S é uma álgebra no sentido usual sobre o corpo finito com dois elementos, em que interseção é a operação multiplicativa, a diferença simétrica a operação aditiva e a multipicação por escalar é a natural:* $0\cdot A=\emptyset$ *e* $1\cdot A=A$.

Se $\mathcal{A}$ é uma álgebra de subconjuntos de S então $\{\emptyset,S\}\subset\mathcal{A}\subset 2^S$, nesse sentido $\{\emptyset,S\}$ é a menor e 2^S a maior álgebra de subconjuntos de S.

0.6.4 Álgebra gerada por uma semiálgebra: se $A, B \subset S$ são elementos das álgebras $\mathcal{A}_1$ e $\mathcal{A}_2$ então $A \cap B$, $\overline{B}$ e $\overline{A}$ também são elementos das duas álgebras, ou seja, a interseção $\mathcal{A}_1 \cap \mathcal{A}_2$ de duas álgebras de subconjuntos de S também é uma álgebra de subconjuntos de S.

Se $\mathcal{B} \subset 2^S$ é uma família de subconjuntos de S então podemos tomar a interseção de todas as álgebras que contém a família $\mathcal{B}$, essa álgebra denotada por $\mathfrak{a}(\mathcal{B})$, é a menor álgebra que contém $\mathcal{B}$.

Particularmente interessante é o caso quando $\mathcal{B}$ semiálgebra.

0.6.5 Proposição. *Para toda semiálgebra $\mathcal{B}$, a álgebra $\mathfrak{a}(\mathcal{B})$ consiste exatamente de todas as uniões finitas de elementos disjuntos de $\mathcal{B}$.*

Demonstração. Seja $\mathcal{C}$ a família de todos os conjuntos expressos por uniões finitas de elementos disjuntos de $\mathcal{B}$. Se C é um elemento de $\mathcal{C}$, então $C = B_1 \cup \cdots \cup B_n$, com $B_i \in \mathcal{B}$ para todo i. Contudo $B_i \in \mathfrak{a}(B)$ que, por ser álgebra, é estável pra uniões finitas, ou seja, $C \in \mathfrak{a}(\mathcal{B})$. Portanto, $\mathcal{C} \subset \mathfrak{a}(\mathcal{B})$.

Agora, vamos mostrar que $\mathcal{C}$ é uma álgebra, donde segue que $\mathcal{C} = \mathfrak{a}(\mathcal{B})$. Seja $C = B_1 \cup \cdots \cup B_n$ um elemento de $\mathcal{C}$. Então $\overline{C} = \overline{B_1} \cap \cdots \cap \overline{B_n}$. Ainda, cada B_i é união finita de subconjuntos disjuntos de S porque $\mathcal{B}$ é uma semiálgebra. □

0.6.6 σ-álgebra: vimos que uma família de subconjuntos fechada para uma quantidade finita de operações sobre seus elementos é chamada de álgebra. Agora, pedimos que a família seja fechada para uma quantidade enumerável de operações sobre seus elementos. A família $\mathcal{A}$ é uma σ-álgebra de subconjuntos de S se

σ**A1 –** $\emptyset \in \mathcal{A}$;

σ**A2 –** para quaisquer $A_1, A_2, \ldots \in \mathcal{A}$ vale que $\bigcup_{i \geqslant 1} A_i \in \mathcal{A}$;

σ**A3 –** para qualquer $A \in \mathcal{A}$ vale que $\overline{A} \in \mathcal{A}$.

Também, a interseção de uma sequência de elementos da σ-álgebra está na σ-álgebra. De fato, para qualquer sequência $(A_i : i \geqslant 1)$ temos que

$$\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}}$$

e como $\overline{A_i} \in \mathcal{A}$ por σ**A3**, segue que $\bigcup_{i=1}^{\infty} \overline{A_i} \in \mathcal{A}$ por σ**A2** e tomando o complemento asseguramos, novamente σ**A3**, a pertinência da interseção.

Toda σ-álgebra é uma álgebra. Por outro lado, uma álgebra será uma σ-álgebra se for finita.

Um conjunto S pode ter diferentes σ-álgebra associadas a ele, por exemplo: a σ-álgebra *trivial* dada por $\{\emptyset, S\}$; a σ-álgebra *discreta* dada pelo conjunto das partes 2^S.

Exemplo 7. A σ-álgebra *gerada por* $A \subset S$ é a família $\{\emptyset, A, \overline{A}, S\}$. ◊

Notemos que para qualquer σ-álgebra $\mathcal{A}$ de S temos que $\{\emptyset, S\} \subset \mathcal{A} \subset 2^S$. Nesse sentido a σ-álgebra trivial é a menor e a discreta a maior σ-álgebra de subconjuntos de S.

Um caso importante é a σ-álgebra gerada por uma família qualquer de subconjuntos de S.

0.6.7 σ-álgebra gerada por uma família de subconjuntos: a interseção de duas σ-álgebras de subconjuntos de S também é uma σ-álgebra de subconjuntos de S. Assim, se $\mathcal{B}$ é uma família de subconjuntos de S podemos tomar a interseção de todas as σ-álgebras que contém $\mathcal{B}$, essa σ-álgebra é denotada por $\sigma(\mathcal{B})$, contém $\mathcal{B}$ e está contida em toda outra que contém $\mathcal{B}$, portanto é a menor σ-álgebra que contém.

Família monótona: quando a família $\mathcal{B}$ do parágrafo acima é uma álgebra, temos a seguinte descrição interessante para $\sigma(\mathcal{B})$: uma família $\mathcal{M}$ de subconjuntos de S é dita *monótona* se $\lim_{n\to\infty} B_n$ pertence a $\mathcal{M}$ sempre que $(B_n : n \geqslant 1)$ é uma sequência monótona de elementos de $\mathcal{M}$. A interseção de duas famílias monótonas é uma família monótona de modo que podemos definir a **famílila monótona gerada** por uma família $\mathcal{B}$ qualquer, denotada por $\mathfrak{m}(\mathcal{B})$.

0.6.8 Proposição. *Para toda álgebra $\mathcal{B}$ de subconjuntos de S, $\sigma(\mathcal{B}) = \mathfrak{m}(\mathcal{B})$.*

Demonstração. Se $(B_n : n \geqslant 1)$ é uma sequência monótona de elementos de $\sigma(\mathcal{B})$ então $\lim_{n\to\infty} B_n$ é uma intersecção ou uma união enumerável de elementos da σ-álgebra de modo que tal limite pertence a $\sigma(\mathcal{B})$, portanto a σ-álgebra é uma família monótona, isto é, $\sigma(\mathcal{B}) \supset \mathfrak{m}(\mathcal{B})$.

Agora, vamos mostrar que $\mathfrak{m}(\mathcal{B})$ é uma σ-álgebra.

Tomemos $M_1 := \{B \in \mathfrak{m}(\mathcal{B}) : \overline{B} \in \mathfrak{m}(\mathcal{B})\}$. Se $(B_i : i \geqslant 1)$ é uma sequência monótona de elementos de M_1 e $\lim_{i\to\infty} B_i = B$ então $(\overline{B}_i : i \geqslant 1)$ é uma sequência monótona de elementos de $\mathfrak{m}(\mathcal{B})$ e $\lim_{i\to\infty} \overline{B}_i = \overline{B} \in \mathfrak{m}(\mathcal{B})$, pois $\mathfrak{m}(\mathcal{B})$ é monótona, portanto $B \in M_1$, ou seja, M_1 é monótona. Assim $M_1 = \mathfrak{m}(\mathcal{B})$ pois $\mathfrak{m}(\mathcal{B})$ é a menor família monótona que contém $\mathcal{B}$. Disso nós concluímos que $\mathfrak{m}(\mathcal{B})$ é fechada para o complemento.

Fixemos $B \in \mathcal{B}$ e tomemos $M_2 := \{A \in \mathfrak{m}(\mathcal{B}) : A \cap B \in \mathfrak{m}(\mathcal{B})\}$. Notemos que $\mathcal{B} \subset M_2 \subset \mathfrak{m}(\mathcal{B})$.

Seja $(A_i : i \geqslant 1)$ uma sequência monótona de elementos de M_2 com $\lim_{i\to\infty} A_i = A$. Se $B \in \mathcal{B}$ então $\lim_{i\to\infty}(A_i \cap B) = A \cap B$, pois $\mathfrak{m}(\mathcal{B})$ é uma família monótona, logo $\lim_{i\to\infty} A_i = A$ é um elemento de M_2. Portanto, $M_2 = \mathfrak{m}(\mathcal{B})$, isto é, como B foi tomado arbitrariamente, se $A \in \mathfrak{m}(\mathcal{B})$ e $B \in \mathcal{B}$ então $A \cap B \in \mathfrak{m}(\mathcal{B})$.

Agora, fixemos $C \in \mathfrak{m}(\mathcal{B})$ e tomemos $M_3 := \{A \in \mathfrak{m}(\mathcal{B}) : A \cap C \in \mathfrak{m}(\mathcal{B})\}$. Notemos que, como antes, essa família é monótona e $\mathcal{B} \subset M_3 \subset \mathfrak{m}(\mathcal{B})$. Como C é um elemento arbitrário de $\mathfrak{m}(\mathcal{B})$ nós temos que se $A \in \mathfrak{m}(\mathcal{B})$ e $C \in \mathfrak{m}(\mathcal{B})$ então $A \cap C \in \mathfrak{m}(\mathcal{B})$.

Concluímos assim que $\mathfrak{m}(\mathcal{B})$ é fechada para o complemento e a interseção, portanto é uma álgebra de subconjuntos. Para terminar, seja $(B_i : i \geqslant 1)$ uma sequência monótona de elementos de $\mathfrak{m}(\mathcal{B})$, definimos

$$A_n = \bigcup_{i=1}^{n} B_i$$

que é um elemento de $\mathfrak{m}(\mathcal{B})$ pois essa é uma álgebra que por ser monónota $\lim_{n\to\infty} A_n \in \mathfrak{m}(\mathcal{B})$, ou seja, $\mathfrak{m}(\mathcal{B})$ é uma σ-álgebra. □

0.6.9 Uma σ-álgebra de intervalos da reta – os borelianos: denotemos por $\mathcal{S}_0$ a semiálgebra dos intervalos semiabertos da reta construída em 0.6.1. A σ-álgebra gerada por $\mathcal{S}_0$ é denotada por $\mathcal{B}(\mathbb{R})$ é chamada σ-álgebra de Borel e seus elementos são chamados de borelianos.

Para quaisquer reais x, a, b, tal que $a < b$, são exemplos de borelianos os intervalos

(a) (x, ∞) que é o complemento de $(-\infty, x]$;

(b) $(-\infty, x)$ que é união enumerável dos intervalos $(-\infty, x - \frac{1}{n}]$;

(c) $[x, \infty)$ pois é o complemento de $(-\infty, x)$;

(d) $(a, b]$ pois pode ser escrito como $\overline{(-\infty, a] \cup (b, \infty)}$;

(e) $\{x\}$ que é dado pela interseção enumerável dos intervalos $(x - \frac{1}{n}, x]$;

(f) (a, b) pois é o mesmo que $(a, b] \setminus \{b\}$;

(g) qualquer conjunto enumerável, pois é união enumerável de unitários;

(h) $[a, b)$ que pode ser escrito como $(a, b) \cup \{a\}$.

Alternativamente, podemos definir $\mathcal{B}(\mathbb{R})$ como a σ-álgebra gerada pela família

$$\pi(\mathbb{R}) := \{(-\infty, x] : x \in \mathbb{R}\}.$$

De fato, $\pi(\mathbb{R}) \subset \mathcal{S}_0 \subset \sigma(\mathcal{S}_0) = \mathcal{B}(\mathbb{R})$, portanto, $\sigma(\pi(\mathbb{R})) \subset \mathcal{B}(\mathbb{R})$. Por outro lado $\mathcal{S}_0 \subset \sigma(\pi(\mathbb{R}))$ (verifique), portanto $\sigma(\mathcal{S}_0) \subset \sigma(\pi(\mathbb{R}))$. Logo $\sigma(\mathcal{S}_0) = \sigma(\pi(\mathbb{R}))$.

Exercício 25. Prove que a σ-álgebra dos borelianos da reta é gerada pelas famílias

(a) $\{(a, b) : a, b \in \mathbb{R}\}$.

(b) $\{(a, b] : a, b \in \mathbb{R}\}$.

(c) $\{[a, b] : a, b \in \mathbb{R}\}$.

(d) $\{(a, \infty) : a \in \mathbb{R}\}$.

Podemos provar que essa σ-álgebra coincide com a σ-álgebra gerada pelos abertos da reta. De fato, denotemos por $\mathcal{O}$ a família dos abertos da reta, e provemos que

$$\sigma(\pi(\mathbb{R})) = \sigma(\mathcal{O}). \tag{0.6.1}$$

Para todo real x

$$(-\infty, x] = \bigcap_{n\geqslant 1}\left(-\infty, x+\frac{1}{n}\right)$$

ou seja, $(-\infty, x] \in \sigma(\mathcal{O})$ pois é uma interseção enumerável de abertos. Portanto $\sigma(\pi(\mathbb{R})) \subset \sigma(\mathcal{O})$. Também, todo intervalo aberto (a, b), com $a < b$, é um elemento de $\sigma(\pi(\mathbb{R}))$ pois

$$(a, b) = \bigcup_{n\geqslant 2}\left(a, b-\frac{b-a}{n}\right]$$

de modo que, como todo aberto $G \in \mathcal{O}$ é uma união enumerável de intervalos abertos, $G \in \sigma(\pi(\mathbb{R}))$ para todo aberto G.

Observamos que em todo espaço topológico podemos definir a σ-álgebra gerada pelos abertos desse espaço; essa σ-álgebra é cahamda σ-álgebra de Borel do espaço topológico. A σ-álgebra de Borel da reta real é a mais importante σ-álgebra em Probabilidade, é muito difícil ocorrer alguma situação na qual lidamos com um subconjunto da reta que não seja um boreliano; embora tais conjuntos existam é muito difícil construí-los explicitamente. Ademais, os elementos de $\sigma(\mathcal{O})$ são muito complicados e na maioria dos casos nos valemos da equação (0.6.1) pois $\pi(\mathbb{R})$ é mais fácil de descrever.

0.6.10 π-sistemas: um π-sistema de subconjuntos de S é qualquer família de subconjuntos de S que é fechada para intersecção finita de seus elementos. Por exemplo, $\{(-\infty, x] : x \in \mathbb{R}\}$ é um π-sistema. Notemos que uma semiálgebra é um π-sistema.

Exercícios.

Exercício 26. Verifique que a definição na equação (0.5.2) coincide com a definição dada acima na equação (0.5.1) e equação (0.5.1) no caso de sequência monótona.

Exercício 27. Prove que a interseção de duas álgebras é uma álgebra.

Exercício 28. Prove que a interseção de duas σ-álgebras é uma σ-álgebra.

Exercício 29. Prove que a a interseção de duas famílias monótonas é uma família monótona.

Exercício 30. Seja $\mathcal{P}$ uma partição finita de S. Mostre que a família dada pelas uniões (finitas) de partes de $\mathcal{P}$ é uma álgebra.

Exercício 31. Dada uma sequência $(A_n : n \geqslant 1)$ de elementos de uma álgebra, mostre que existe uma sequência $(B_n : n \geqslant 1)$ de elementos disjuntos da álgebra tal que $B_n \subset A_n$ para todo n e $\bigcup_{n\geqslant 1} A_n = \bigcup_{n\geqslant 1} B_n$.

Exercício 32. Seja $\mathcal{A}$ uma semiálgebra de subconjuntos de S. Prove ou dê um contraexemplo para a igualdade $\sigma(\mathfrak{a}(\mathcal{A})) = \sigma(\mathcal{A})$.

Exercício 33. Dê um exemplo de um conjunto M e duas σ-álgebras $\mathcal{A}_1$ e $\mathcal{A}_2$ de subconjuntos de M tais que $\mathcal{A}_1 \cup \mathcal{A}_2$ não é σ-álgebra de subconjuntos de M.

Exercício 34. Dê um exemplo de um conjunto M e uma família monótona $\mathcal{M}$ de subconjuntos de M tal que $\emptyset,\ M \in \mathcal{M}$ mas que não é σ-álgebra de subconjuntos de M.

Exercício 35. Sejam $M,\ N$ conjuntos não vazios, $\mathcal{M}$ uma σ-álgebra de subconjuntos de M e $f : N \to M$ uma função sobrejetiva. Prove que $\{f^{-1}(A) : A \in \mathcal{M}\}$ é uma σ-álgebra de subconjuntos de N.

Exercício 36. Prove que se $\mathcal{A}$ é uma σ-álgebra de subconjuntos de M e $S \subset M$ então

$$\sigma(\mathcal{A} \cup \{S\}) = \Big\{(A \cap S) \cup (B \cap \overline{S}) : A, B \in \mathcal{A}\Big\}.$$

Exercício 37. Seja $\mathcal{S}$ uma semiálgebra de subconjuntos. Vale que $\sigma(\mathfrak{a}(\mathcal{S})) = \sigma(\mathcal{S})$?

§1 Probabilidade

Problema de Monty Hall: esse notório problema surgiu a partir de um concurso televisivo dos Estados Unidos chamado *Let's Make a Deal*, exibido na década de 1970. O jogo consiste no seguinte: Monty Hall (o apresentador) apresentava 3 portas a um concorrente, sabendo que atrás de uma delas, escolhida ao acaso, está um carro e que as outras duas têm um bode. O protocolo da brincadeira é:

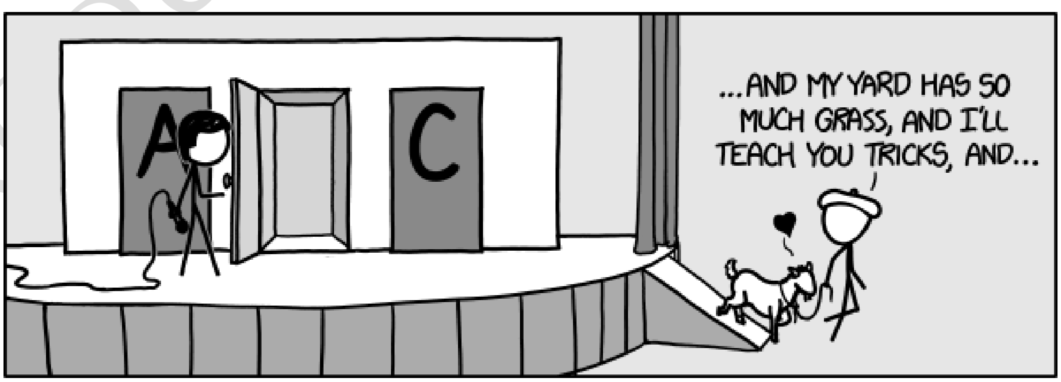
(a) Na 1ª etapa o concorrente escolhe uma porta ao acaso (que ainda não é aberta);

(b) em seguida Monty Hall abre uma das outras duas portas que o concorrente não escolheu, sabendo que ela esconde um bode. Se são duas possibilidades, ele escolhe uma ao acaso;

(c) em seguida, com duas portas fechadas apenas, e sabendo que o carro está atrás de uma delas, o apresentador oferece ao concorrente a oportunidade de trocar de porta. O concorrente tem que se decidir se permanece com a porta que escolheu no início do jogo ou se muda para a outra porta que ainda está fechada;

(d) feita a escolha, o apresentador abre a porta escolhida e o concorrente leva o prêmio escondido pela porta.

O problema é determinar a estratégia (trocar ou não trocar no passo (c)) que maximiza a chance de ganhar o carro. Assista ao show neste link[1]. Teste o jogo neste link[2]. Para responder essa pergunta, vamos descrever um modelo matemático para o problema, que chamamos de *modelo probabilístico*.



[1] http://www.youtube.com/watch?gl=BR&v=WKR6dNDvHYQ
[2] http://math.ucsd.edu/~crypto/Monty/monty.html

1.1 Modelo matemático para experimento aleatório.

Um modelo é um objeto ou conceito usado para representar aspectos do mundo real em uma forma que possamos compreender e explicar a situação estudada. Um modelo matemático é um modelo envolvendo conceitos matemáticos, isto é, conjunto de símbolos e relações matemáticas que representam de alguma forma o sistema estudado. Em muitas situações é desejável ter uma estrutura matemática na qual pode-se representar proposições não determinísticas e em que variáveis não tomam um ou outro determinado valor com certeza definitiva.

A teoria da probabilidade é a disciplina da matemática mais amplamente adotada para capturar formalmente o conceito de não-determinismo/aleatoriedade. Se o fenômeno estudado inclui elementos de aleatoriedade, como as escolhas feitas pelos jogadores em Monty Hall, então dizemos que o modelo é probabilístico.

Experimento aleatório: intuitivamente, um experimento aleatório é qualquer processo que nos fornece um resultado que não sabemos qual é até que o processo termine. Vários processos se encaixam nessa descrição vaga: lançamento de uma moeda, lançamento de um dado, o tempo de vida de uma lâmpada e muitos outros. Não nos esforçaremos para definir mais precisamente esse conceito e contamos com a intuição do leitor para sua compreensão. Um *modelo probabilístico* é um modelo matemático idealizado para o estudo de um experimento aleatório.

Um modelo probabilístico para um experimento aleatório é caracterizado por um *espaço amostral* — conjunto dos resultados possíveis — um *espaço de eventos* — família dos subconjuntos de resultados de interesse — e uma *(medida de) probabilidade* — uma função que associa um valor numérico a cada evento que representa a ideia de chance, possibilidade, perspectiva, credibilidade.

1.1.1 Espaço amostral: o espaço amostral de um experimento, denotado por Ω, é um conjunto não vazio que representa todos os resultados possíveis do experimento. Um elemento de Ω é chamado de *ponto amostral* e a escolha de algum ponto amostral representa uma *realização* do experimento.

Exemplo 8. São experimentos com respectivos espaços amostrais

(a) um dado é lançado e observamos a face para cima, $\Omega = \{1, 2, 3, 4, 5, 6\}$;

(b) uma moeda é lançada e observamos sua face para cima, $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}$;

(c) uma moeda é lançada até sair Coroa $\Omega = \{(\mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), \dots, (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca} \dots)\}$;

(d) uma moeda é lançada indefinidamente e os resultados são observados, $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}^{\mathbb{N}}$;

(e) observamos tempo de vida de uma lâmpada, $\Omega = \{t \in \mathbb{R} : t \geqslant 0\}$;

(f) um dardo é lançado num alvo circular de raio 1 e observamos o ponto atingido $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant 1\}$. ◊

O espaço amostral de um experimento aleatório não é único; no item (f) nós podemos escrever o espaço amostral em coordenadas polares como $\{(r,\theta) \in \mathbb{R}^2 : 0 \leqslant r \leqslant 1 \text{ e } -\pi < \theta \leqslant \pi\}$. Ainda, o espaço amostral reflete o interesse da observação.

Classificamos, de acordo com a cardinalidade do espaço amostral, os modelos probabilísticos em discretos e contínuos, os quais têm ferramentas diferentes, o primeiro é combinatorial e o segundo precisa de teoria da medida. Um *espaço amostral discreto* é um espaço amostral finito ou infinito enumerável. Um *espaço amostral contínuo* é um espaço não enumerável de cardinalidade do contínuo. Por exemplo, são espaços discretos os espaços amostrais dos experimentos (a), (b) e (c) do exemplo 8 acima. Os experimentos (d), (e) e (f) têm espaços contínuos.

Evento aleatório: se os resultados de um experimento aleatório são modelados pelo espaço amostral Ω, então os eventos aleatórios são modelados por subconjuntos de Ω. O evento $A \subset \Omega$ *ocorre* se o resultado do experimento é um elemento de A e para qualquer outro resultado dizemos que A *não ocorre*. Em especial $\emptyset$ é o *evento impossível*, o *complemento* do evento A é o evento *não-*A dado por $\overline{A} = \Omega \setminus A$ e o complemento do evento impossível, i.e. Ω, é o *evento certo*.

1.1.2 Espaço de eventos: quando da realização de um experimento aleatório, os vários resultados possíveis descritos por um predicado é um evento aleatório, o qual pode ocorrer ou não ocorrer. Por exemplo, no lançamento de um dado "número primo" é um evento e esse evento ocorre quando resultado de um lançamento da dado for um número primo (3 ou 5).

O espaço de eventos é a família $\mathcal{A}$ dos eventos aleatórios de um experimento aleatório aos quais conseguimos atribuir uma probabilidade de ocorrer. Quais são as famílias de subconjuntos de Ω que podem ser tomadas como espaço de eventos é um assunto delicado que não trataremos agora. Uma escolha óbvia é o conjunto 2^Ω das partes de Ω, mas acontece que em muitos casos é preciso restringir essa família a um subconjunto de 2^Ω para que questões probabilísticas façam sentido. Os motivos para isso são deixados para uma seção de conteúdo mais avançado.

Ao espaço de eventos pedimos, por enquanto, que satisfaça ao menos o seguinte: o evento certo pertence ao espaço de eventos e quaisquer operações elementares (complemento, união e interseção) entre eventos de $\mathcal{A}$ resulta num evento de $\mathcal{A}$, ou sejam $\mathcal{A}$ deve ser uma álgebra de subconjuntos de Ω (adiante, pediremos mais, que seja uma σ-álgebra; veja as definições nas páginas 31 e seguintes).

A Probabilidade tem um linguagem peculiar para os eventos:

Notação	*Probabilidade*	*Conjunto*
Ω	espaço amostral, evento certo	universo
$\emptyset$	evento impossível	vazio
ω	ponto amostral	elemento
$\{\omega\}$	evento elementar	conjunto unitário
A	ocorre A	subconjunto
$\overline{A}$	não ocorre A	complemento
$A \cap B$	ocorre A e B	intersecção
$A \cup B$	ocorre A ou B	união
$A \setminus B$	ocorre A e não ocorre B	diferença
$A \triangle B$	ocorre A ou B, não ambos	diferença simétrica
$A \subset B$	se ocorre A, então ocorre B	inclusão
$A \setminus B$	ocorre A, mas não ocorre B	diferença
$A \cap B = \emptyset$	eventos exclusivos	conjuntos disjuntos

Uma sequência $A_1, A_2, \dots$ de eventos é dita de eventos mutuamente exclusivos se os eventos são *mutuamente exclusivos quando tomados dois-a-dois*.

Exemplo 9. No lançamento de dados, $\Omega = \{1, 2, 3, 4, 5, 6\}$

- $A = \{2, 4, 6\}$ representa o evento "número par";
- $\overline{A} = \{1, 3, 5\}$ representa o evento "não é número par";
- $B = \{4, 5, 6\}$ representa o evento "resultado > 3";
- $C = \{4\}$ representa o evento "o resultado 4";
- $B \cap C = \{4\}$ representa o evento "número > 3 **e** 4";
- $B \cap A = \{4, 6\}$ representa o evento "número > 3 **e** número par". ◊

1.2 Probabilidade clássica: espaços finitos. No modelo probabilístico clássico para espaços amostrais finitos (não vazios) todos os eventos elementares são igualmente prováveis de ocorrer como resultado da realização do experimento

$$\mathbb{P}(\{\omega\}) := \frac{1}{|\Omega|}.$$

Ademais, se um experimento tem n desfechos dos quais m correspondem a um determinado evento aleatório, então a probabilidade de ocorrer o evento é proporção de resultados favoráveis m/n, isto é, o subconjunto $A \subset \Omega$ ocorre com probabilidade

$$\mathbb{P}(A) := \frac{|A|}{|\Omega|} = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) \tag{1.2.1}$$

a qual chamamos de **probabilidade uniforme**. Usualmente, escrevemos $\mathbb{P}(\omega)$ para denotar $\mathbb{P}(\{\omega\})$.

Nesse modelo adotamos o espaço de eventos $\mathcal{A} = 2^{\Omega}$. O modelo probabilístico em espaços equiprováveis finitos para um experimento aleatório é um dado pelo conjunto Ω, que representa os resultados possíveis do experimento e a medida de probabilidade $\mathbb{P} : 2^{\Omega} \to \mathbb{R}$ como definida na equação (1.2.1), que associa um número real não-negativo a cada evento.

Exemplo 10 (**lançamento de uma moeda equilibrada**). Uma moeda equilibrada é lançada e observamos a face para cima. Um modelo probabilístico clássico para o experimento é $\Omega = \{\text{Ca}, \text{Co}\}$ com $\mathbb{P}(\text{Ca}) = \mathbb{P}(\text{Co}) = 1/2$. ◊

Exemplo 11 (n **lançamentos de uma moeda equilibrada**). Uma moeda equilibrada é lançada n vezes, para algum natural $n \geqslant 1$, e observamos a face para cima em cada lançamento. Um modelo probabilístico clássico para o experimento é $\Omega = \{\text{Ca}, \text{Co}\}^n$ com $\mathbb{P}((x_1, \dots, x_n)) = (1/2)^n$. ◊

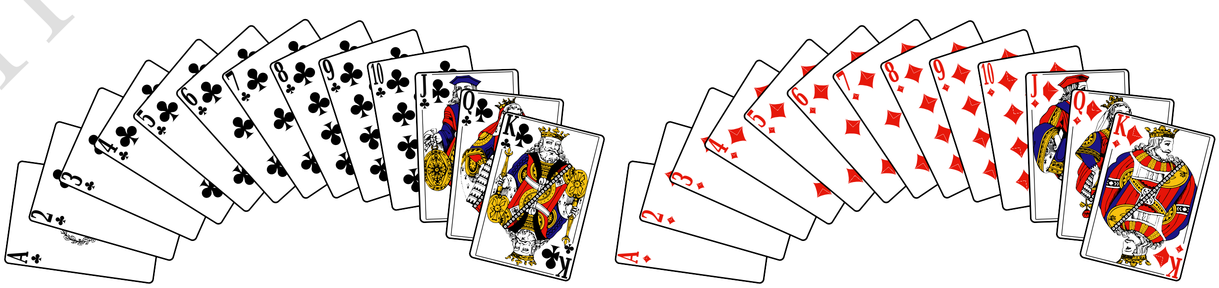
Exemplo 12 (**lançamento de um dado equilibrado**). Um dado equilibrado é lançado e observamos a face para cima:

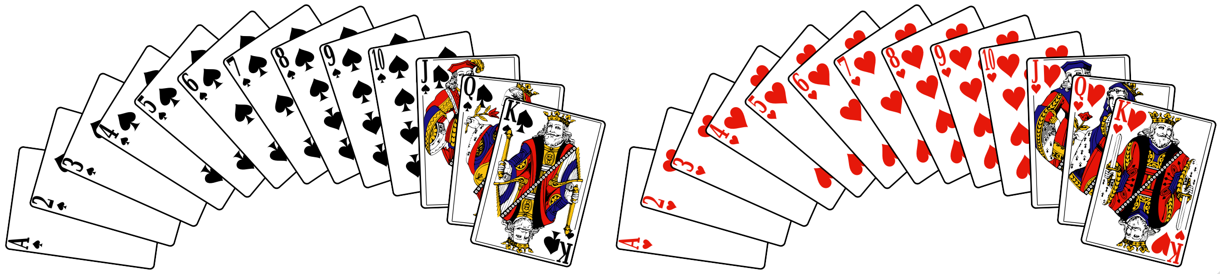
⚀ ⚁ ⚂ ⚃ ⚄ ⚅

Um modelo probabilístico clássico para o experimento é $\Omega = \{1, 2, 3, 4, 5, 6\}$ com $\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \mathbb{P}(4) = \mathbb{P}(5) = P(6) = 1/6$. A probabilidade do evento "número par" é a probabilidade do subconjunto $\{2, 4, 6\}$ que é $1/2$. ◊

Exemplo 13 (**escolha de uma carta de baralho tradicional**). Um baralho tem 52 cartas divididas em quatro famílias (ou naipes) representadas pelas figuras: ♦ (ouros), ♣ (paus), ♥ (copas) e ♠ (espadas). Cada família traz os números de 2 a 10 e as letras A (ás), J (valete), Q (dama), K (rei); uma maneira de modelar os resultados é descrever cada uma das 52 cartas A♣, 2♣, 3♣, …, A♦, 2♦, 3♦ …e assim por diante. Uma maneira mais compacta é

$$\Omega = \{A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\} \times \{\diamondsuit, \clubsuit, \heartsuit, \spadesuit\}$$





A probabilidade de uma escolha aleatória de uma carta do baralho resultar em 5♣ é 1/52; a probabilidade de uma escolha aleatória resultar em ♠ é 13/52 = 1/4; a probabilidade de uma escolha aleatória resultar em 4 é 1/13. ◊

1.2.1 Observação (escolha aleatória/sorteio aleatório). *Num espaço amostral finito* há uma única maneira *de definir probabilidade para os eventos de modo que os eventos elementares do espaço amostral sejam igualmente prováveis. Essa probabilidade define de maneira precisa a expressão intuitiva "aleatório", como ocorre em* escolha aleatória *de uma carta de baralho,* sorteio aleatório *de uma bola numa urna,* resultado aleatório *do lançamento de um dado,* escolha aleatória *de um indivíduo de uma população. Uma* escolha aleatória *é um resultado de um experimento idealizado com respostas equiprováveis. Essa não ambiguidade nem sempre é possível, como veremos na seção 1.3.3.*

Decorre das definições dadas acima que valem as seguintes propriedades:

da definição na equação (1.2.1) temos que

$$\mathbb{P}(A) \geqslant 0, \quad \text{para todo evento } A \tag{1.2.2}$$

$$\mathbb{P}(\Omega) = 1, \tag{1.2.3}$$

para eventos mutuamente exclusivos A e B

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{1.2.4}$$

segue da equação (1.2.1) e do teorema 0.4.2.

Notação: em vista da equação (1.2.4), a união disjunta tem um papel importante em cálculo de probabilidades e por isso enfatizamos o fato dos conjuntos envolvidos serem disjuntos usando

$$A \uplus B$$

assim escrevemos (1.2.4) como $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Das três propriedades de $\mathbb{P}$ dadas acima podemos deduzir várias outras, por exemplo, temos que $\Omega = A \cup \overline{A}$, com A e $\overline{A}$ mutuamente exclusivos, portanto de equação (1.2.3) e equação (1.2.4)

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) \tag{1.2.5}$$

portanto

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

Notemos que, embora a equação (1.2.5) possa ser deduzido imediatamente da definição

$$\mathbb{P}(\overline{A}) = \frac{|\overline{A}|}{|\Omega|} = \frac{|\Omega|-|A|}{|\Omega|} = 1 - \mathbb{P}(A)$$

é interessante ressaltar que qualquer outra definição para a função $\mathbb{P} : \Omega \to \mathbb{R}$ para a qual vale a equação (1.2.2), equação (1.2.3) e equação (1.2.4), também valerá a equação (1.2.5). O mesmo vale para as seguintes propriedades, isto é, apesar de serem consequência da definição de $\mathbb{P}$ dada na equação (1.2.1) elas podem ser inferidas para qualquer função $\mathbb{P}$ para a qual vala a equação (1.2.2), equação (1.2.3) e equação (1.2.4):

(i) $\mathbb{P}(\emptyset) = 0$

(ii) $0 \leqslant \mathbb{P}(A) \leqslant 1$

(iii) se $A \subset B$ então $\mathbb{P}(A) \leqslant \mathbb{P}(B)$

(iv) $\mathbb{P}(A \cup B) \leqslant \mathbb{P}(A) + \mathbb{P}(B)$

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

(vi) $\mathbb{P}\left(\biguplus_{i=1}^{t} A_i\right) = \sum_{i=1}^{t} \mathbb{P}(A_i)$.

1.2.2 Exemplos:

1. Um dado equilibrado é lançado e observamos a face para cima; o modelo probabilístico é dado por $\Omega = \{1, 2, 3, 4, 5, 6\}$ com $\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \mathbb{P}(4) = \mathbb{P}(5) = P(6) = 1/6$.

   A probabilidade de ocorrer "número par" é $\mathbb{P}(\{2, 4, 6\}) = 1/2$; a probabilidade de ocorrer "número ímpar" é a mesma de *não* ocorrer "número par" $1 - 1/2 = 1/2$.

   A probabilidade de ocorrer "número primo" é $\mathbb{P}(\{2, 3, 5\}) = 1/2$; a probabilidade de ocorrer "número composto" é $\mathbb{P}(\{4, 6\}) = 1/3$. A probabilidade de ocorrer "primo *ou* composto" é

   $$\mathbb{P}(\{2, 3, 5\} \uplus \{4, 6\}) = \frac{5}{6}$$

2. Um dado equilibrado é lançado duas vezes e observamos o par de faces resultante; o modelo probabilístico é dado por $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), (2, 3) \dots, (6, 6)\}$ com $\mathbb{P}((i, j)) = 1/36$ para todos $i, j \in \{1, 2, 3, 4, 5, 6\}$.

A probabilidade dos resultados terem sido iguais é $6/36 = 1/2$; a probabilidade da soma ser 7 é a probabilidade do evento definido pelos pares $(i, 7-i)$ para $i \in \{1, 2, \dots, 6\}$ que é $6/36 = 1/6$.

Qual a probabilidade do primeiro resultado ser maior que o segundo? Se o resultado do primeiro lançamento foi i então restam para o segundo lançamento $6-i$ resultados maiores que i, para $i \in \{1, 2, \dots, 6\}$, portanto a probabilidade é $(5+4+3+2+1+0)/36 = 15/36$.

Com que probabilidade ocorre um par de números coprimos? Os pares não coprimos são

$$\{(2,2),(2,4),(2,6),(3,3),(3,6),(4,2),(4,4),(4,6),(5,5),(6,2),(6,3),(6,4),(6,6)\}$$

portanto, a probabilidade procurada é $1 - 13/36$.

3. Um número de $\{1, 2, 3, \dots, 10.000\}$ é sorteado. Com que probabilidade ocorre um múltiplo de 2 ou múltiplo de 3? Denotamos por A_p o evento "múltiplo de p", assim procuramos pela probabilidade de $A_2 \cup A_3$. Lembremos que $A_2 \cap A_3$ são, exatamente, os múltiplos de 6, logo

$$\begin{aligned}\mathbb{P}(A_2 \cup A_3) = \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_2 \cap A_3) &= \frac{\frac{10.000}{2}}{10.000} + \frac{\left\lfloor \frac{10.000}{3} \right\rfloor}{10.000} - \frac{\left\lfloor \frac{10.000}{6} \right\rfloor}{10.000} \\ &= \frac{1}{2} + \frac{3.333}{10.000} - \frac{1.666}{10.000} \approx 0{,}66.\end{aligned}$$

4. Num baralho três cartas são selecionadas aleatoriamente. Com que probabilidade pelo menos uma delas é uma Ás? Podemos representar os resultados do experimento por triplas (c_1, c_2, c_3) das cartas, na ordem, retiradas do baralho. O número total de triplas é $52 \cdot 51 \cdot 50$. Contar o número de triplas nas quais pelo menos uma das cartas é algum dos quatro áses é relativamente trabalhoso, podemos contar o complemento e usar a equação (1.2.5). São 48 cartas que não são ás e o total de triplas sem ás é $48 \cdot 47 \cdot 46$, portanto a probabilidade de pelo menos uma das cartas ser um ás é

$$1 - \frac{48 \cdot 47 \cdot 46}{52 \cdot 51 \cdot 50} \approx 0{,}217.$$

5. (**paradoxo do aniversário**) Assuma que o ano tem 365 dias (esqueça os bissextos) e assuma que cada pessoa nasce em algum desses dias com igual probabilidade. Consideremos uma seleção de 20 pessoas. Com que probabilidade há, nesse grupo, pelo menos duas pessoas que nasceram no mesmo dia?

O espaço amostral pode ser dado por sequências de datas de aniversário $(d_1, d_2, \dots, d_{20})$ da pessoa mais velha (d_1) para a mais nova (d_{20}) (não há nada especial em tomar essa ordenação específica). Esse espaço amostral tem cardinalidade 365^{20}. O evento "todas as pessoas nasceram em dias diferentes" tem cardinalidade $365 \cdot 364 \cdot 363 \cdot \dots \cdot 346$ (um produto com vinte termos). A probabilidade de "todas as pessoas nasceram em dias diferentes" é

$$\frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot 346}{365^{20}} \approx 0{,}588$$

logo, pelo menos duas pessoas nasceram no mesmo dia com probabilidade aproximadamente 0,412. Se forem 23 pessoas a probabilidade de duas terem aniversário no mesmo dia é 0,5073, ou seja, com pelo menos 23 pessoas há mais que 50% de chance de acontecer de pelo menos duas delas terem nascido no mesmo dia; esse fato é conhecido como paradoxo do aniversário que, de fato, não é um paradoxo e sim um resultado contrário a intuição.

1.2.3 Princípios de contagem: notemos que a probabilidade de ocorrer $A \subset \Omega$ não depende da sua natureza como subconjunto, depende apenas da sua cardinalidade $|A|$. O problema de determinar a probabilidade resume-se num problema de contagem, que é estudado no ramo da matemática conhecido como Combinatória.

Princípio aditivo: queremos comprar um computador de um dos dois fabricantes mais comuns de processadores: Intel e AMD. Suponha também que nosso orçamento nos faz ter 3 opções de modelos Intel e 4 opções de modelos AMD. Então, existem no total $3 + 4 = 7$ opções diferentes de modelos de computador que podemos comprar.

Princípio aditivo: suponha que o evento E pode ocorrer n maneiras e o evento F de m maneiras distintas das outras n. Então, o número de maneiras de ocorrer o evento "E ou F" é $n + m$.

Em termos de conjuntos, o princípio aditivo é formulado da seguinte maneira

se E e F são conjuntos finitos disjuntos então $|E \uplus F| = |E| + |F|$.

Exemplo 14. Com que probabilidade um inteiro entre 1 e 16 escolhido aleatoriamente é múltiplo de 3 ou de 7? Para determinar a probabilidade devemos determinar a cardinalidade do conjunto de inteiros entre 1 e 16 que são múltiplos de 3 ou múltiplos de 7, tais conjuntos são disjuntos pois $\mathrm{mmc}(3, 7) = 21$. Os múltiplos de 3 são cinco, os múltiplos de 7 são dois, portanto, os múltiplos de ambos são $5 + 2 = 7$. O evento *múltiplo de* 3 *ou múltiplo de* 7 ocorre se for sorteado um elemento dos 7 elementos, cuja probabilidade é $7/16$. ◊

De modo geral, se $E_1, E_2, \dots, E_k$ são conjuntos finitos disjuntos dois-a-dois então

$$\left|\biguplus_{i=1}^{k} E_i\right| = \sum_{i=1}^{k} |E_i|$$

que pode ser demonstrado usando princípio da indução.

1.2.4 Teorema (**Princípio de Inclusão–Exclusão**). *Se E e F são conjuntos finitos (não necessariamente disjuntos) então*

$$|E \cup F| = |E| + |F| - |E \cap F|. \tag{1.2.6}$$

Demonstração. Podemos escrever $E \cup F$ como a união $E \cup F = E \uplus (F \setminus E)$, portanto

$$|E \cup F| = |E| + |F \setminus E|$$

e podemos escrever F como a união $F = (E \cap F) \uplus (F \setminus E)$, portanto

$$|F| = |E \cap F| + |F \setminus E|.$$

Isolando $|F \setminus E|$ na segunda equação e substituindo na primeira, deduzimos a equação (1.2.6). $\square$

1.2.5 Corolário. *Para quaisquer eventos E e F de um modelo probabilístico*

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F). \tag{1.2.7}$$

Exemplo 15. Com que probabilidade o lançamento de um dado resulta num múltiplo de 2 ou de 3? Os múltiplos de 2 definem o subconjunto $E = \{2, 4, 6\}$, os múltiplos de 3 definem o subconjunto $F = \{3, 6\}$, portanto,

$$\mathbb{P}(E \cup F) = \frac{|E \cup F|}{6} = \frac{|E| + |F| - |E \cap F|}{6} = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}.$$

◊

Exercício 38. Com que probabilidade um inteiro entre 1 e 16 escolhido aleatoriamente é múltiplo de 3 ou de 5?

Esboço da solução. $E = \{3, 6, 9, 12, 15\}$ e $F = \{5, 10, 15\}$. E ou $F = \{3, 6, 7, 9, 12, 14, 15\}$. $\mathbb{P}(E \cup F) = 5/16 + 3/16 - 1/16$. $\square$

Exercício 39. Prove que se E_1, E_2 e E_3 são conjuntos então

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_2 \cap E_3) - \mathbb{P}(E_1 \cap E_3) + \mathbb{P}(E_1 \cap E_2 \cap E_3).$$

Exercício 40. Com que probabilidade um número escolhido aleatoriamente em $\{1, 2, \ldots, 100\}$ não é divisível por 2, 3 ou 5?

Princípio multiplicativo: digamos que as únicas roupas limpas que você tem são duas camisetas e quatro calças. Quantos pares diferentes de camiseta e calça você pode escolher? São duas camisetas e pra cada escolha de camiseta é possível escolher uma de quatro calças, portanto, são $2 \times 4 = 8$ pares possíveis.

> Princípio multiplicativo: se um evento pode ser descrito em duas etapas de modo que há n desfechos possíveis para a 1ª etapa e há m desfechos possíveis para a 2ª etapa, então o número de possíveis desfechos para o evento é $n \cdot m$.

Em termos de conjuntos, se E_1 e E_2 são conjuntos que representam os resultados das etapas, então o evento composto é representado por $E_1 \times E_2$

se E_1 e E_2 são conjuntos finitos então $|E_1 \times E_2| = |E_1| \cdot |E_2|$.

De um modo geral, se $E_1, \dots E_r$ representam r etapas de evento composto, então o números de modos distintos de realizar o evento é

$$\left| \bigtimes_{i=1}^{r} E_i \right| = |E_1| \cdot |E_2| \cdots |E_r|$$

que pode ser demonstrado usando princípio da indução.

Exercício 41. Uma placa de carro é uma sequência de 3 letras seguidas por 4 dígitos. Qual é a quantidade de placas distintas que podemos obter?

Esboço de solução. $E_i = \{A, B, \dots, Z\}$ para $i = 1, 2, 3$.

$E_i = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ para $i = 4, 5, 6, 7$.

$\left| \bigtimes_{i=1}^{8} E_i \right| = 26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 26^3 10^4 = 175.760.000.$ □

Exemplo 16. Cada posição da memória (célula de memória) de um computador tem um endereço que é uma sequência binária. As arquiteturas com processadores 32-bits tem capacidade para endereçamento de $2^{32} = 4.294.967.296$ posições de memória, aproximadamente 4 Gigabytes. As arquiteturas com processadores 64-bits tem capacidade para endereçamento de

$$2^{64} = 18.446.744.073.709.551.616$$

posições de memória, aproximadamente 16 Exabytes (16 milhões de Terabytes). Suponhamos que um dispositivo de 1 Gigabyte ocupe um dispositivo de dimensões 1 mm × 1 mm × 1 mm. Para guardar 16 Exabytes precisaríamos de uma quarto de dimensões 2,5 m × 2,5 m × 2,5 m. ◊

Exemplo 17 (**o problema de Méré**). Com que probabilidade o 6 ocorre pelo menos uma vez em 4 lançamentos de um dado? O número de resultados possíveis em 4 lançamentos é 6^4 e o número de resultados possíveis onde não ocorre o 6 é 5^4, portanto a probabilidade de não ocorrer 6 é $(5/6)^4$. Usando a equação (1.2.5) a probabilidade de sair pelo menos um seis é $1-(5/6)^4$, que é aproximadamente $0{,}51$.

E, com que probabilidade um par de 6 ocorre pelo menos uma vez em 24 lançamentos de um par de dados? Dos 36^{24} resultados possíveis do experimento 35^{24} são pela não ocorrência de um par de 6, portanto, a probabilidade de sair par de 6 pelo menos uma vez em 24 lançamentos é $1-(35/36)^{24} \approx 0{,}49$. ◊

A seguir destacamos alguns casos particulares do Princípio Multiplicativo. Essencialmente são dois tipos de listas: *arranjos* e *combinações*. Nos arranjos a ordem dos elementos importa e nas combinações a ordem não importa.

1.2.6 Arranjo: quantas palavras (sequências) de 3 letras *distintas* do alfabeto latino podem ser formadas? Como o alfabeto tem 26 letras, pelo princípio multiplicativo são $26 \cdot 25 \cdot 24$ palavras.

> Um <u>arranjo simples</u> de r elementos tomados de um conjunto A de n elementos ($r \leq n$) é uma sequência $(a_1, a_2, \dots, a_r)$ de elementos não repetidos de A.

O número de arranjos simples de r elementos tomados de um conjunto de n elementos ($r \leq n$) é

$$(n)_r := n(n-1)(n-2)\cdots(n-r+1)$$

Quando é permitido repetição dizemos <u>*arranjo com repetição*</u>, que é um caso particular do Princípio Multiplicativo com todos os conjuntos iguais. Por *arranjo* nos referimos a arranjo simples.

Alguns textos usam a notação $A(n, r)$ para $(n)_r$.

Exemplo 18. Se sorteamos um inteiro entre 000 e 999 (inclusive e com 3 dígitos) qual a probabilidade do resultado possuir todos os dígitos distintos? O espaço amostral tem 1.000 resultados possíveis. A quantidade deles sem dígitos repetidos é $(10)_3 = 10 \cdot 9 \cdot 8 = 720$, portanto, a probabilidade procurada é $720/1.000 = 0{,}72$. ◊

Exemplo 19 (**paradoxo do aniversário**). Com que probabilidade ocorre que num grupo com 25 pessoas 2 ou mais pessoas façam aniversário no mesmo dia? O aniversário de 25 pessoas pode ocorrer de 365^{25} modos diferentes. O aniversário de 25 pessoas sem que nenhum deles coincida pode ocorrer de $(365)_{25}$ modos diferentes. Portanto, há $365^{25} - (365)_{25}$ possibilidades diferentes para o aniversário de 25 pessoas com pelo menos duas aniversariando no mesmo dia; a probabilidade desse evento é

$$\frac{365^{25} - (365)_{25}}{365^{25}} = 1 - \frac{(365)_{25}}{365^{25}} > 0{,}56.$$

Com 55 pessoas a probabilidade é maior que 98%. ◊

Permutação: o caso de arranjo simples com $r = n$ é uma permutação. Quantas palavras com letras *distintas* podem ser formadas com as letras a, b e c? Pelo princípio multiplicativo são $3 \cdot 2 \cdot 1 = 6$ palavras.

> Uma <u>permutação</u> de elementos de um conjuntos A é uma sequência de elementos de A.

O número de permutações dos elementos de um conjunto de $n \geqslant 0$ elementos é

$$0! := 1$$
$$n! := n(n-1)(n-2)\cdots 1, \text{ se } n \geqslant 1.$$

Notemos que

$$(n)_r = \frac{n!}{(n-r)!}.$$

Por exemplo, o número de permutações possíveis com as letras a, b, c, d e e é $5! = 120$. O número de permutações possíveis com as letras a, b, c, d, e e f é $6! = 720$.

Exemplo 20. A quantidade de permutações que podem ser formadas com as letras da palavra $livros$ é $6! = 720$. A quantidade de permutações que podem ser formadas com as letras da palavra $teclado$ é $7! = 5.040$. A quantidade de permutações que podem ser formadas com as letras da palavra $discreta$ é $8! = 40.320$. A quantidade de permutações que podem ser formadas com as letras da palavra $universal$ é $9! = 362.880$. A quantidade de permutações que podem ser formadas com as letras da palavra $pernambuco$ é $10! = 3.628.800$. A quantidade de permutações que podem ser formadas com as letras da palavra $seminublado$ é $11! = 39.916.800$. A quantidade de permutações que podem ser formadas com as letras da palavra $configuravel$ é $12! = 479.001.600$. ◊

Perceba que o fatorial cresce bastante rápido com n. 11! é mais que a quantidade de segundos que passam em 1 ano. 12! é mais que a quantidade de segundos que passam em 12 anos. 13! é mais que a quantidade de segundos que passam em 100 anos.

1.2.7 Fórmula de Stirling: duas sequências de números a_n e b_n são *assintoticamente iguais* e escrevemos $a_n \sim b_n$, se

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

Frequentemente, é muito útil quanto trabalhamos com fatoriais a seguinte igualdade assintótica conhecida como fórmula de Stirling

$$\boxed{n! \sim n^n e^{-n}\sqrt{2\pi n}.}$$

n	$n!$	stirling
1	1	0.922137
3	6	5.83621
7	5040	4980.396
10	3628800	3598696
20	$2.432902e+18$	$2.422787e+18$
50	$3.041409e+64$	$3.036345e+64$
75	$2.480914e+109$	$2.478159e+109$
100	$9.332622e+157$	$9.324848e+157$
142	$2.695364e+245$	$2.693783e+245$

Uma demonstração (probabilística) dessa aproximação para o valor de $n!$ é dada em 3.0.27 na página 213.

Exemplo 21 (**o problema dos chapéus**). Três convidados chegaram numa festa vestindo chapéu e os entregaram na recepção. O funcionário, pouco cuidadoso, não identificou os chapéus e no final da festa os entregou aleatoriamente para as mesmas três pessoas. Com que probabilidade ninguém recebeu o próprio chapéu?

O número de maneiras diferentes com que os chapéus podem ser devolvidos aos convidados é 3!. Seja E_i o conjunto de todas as permutações nas quais o i-ésimo convidado recebe o próprio chapéu. A probabilidade de algum convidado receber o próprio chapéu é, pelo princípio de inclusão–exclusão (veja exercício 39 acima)

$$\mathbb{P}\left(\bigcup_{i=1}^{3} E_i\right) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_2 \cap E_3) - \mathbb{P}(E_1 \cap E_3) + \mathbb{P}(E_1 \cap E_2 \cap E_3).$$

Para qualquer i temos $\mathbb{P}(E_i) = 2!/3!$ e temos $\mathbb{P}(E_i \cap E_j) = 1/3!$ sempre que $i \neq j$. Ainda, $\mathbb{P}(E_1 \cap E_2 \cap E_3) = 1/3!$, logo

$$\mathbb{P}\left(\bigcup_{i=1}^{3} E_i\right) = 3\frac{2!}{3!} - 3\frac{1}{3!} + \frac{1}{3!} = \frac{2}{3}.$$

de modo que ninguém pega o próprio chapéu com probabilidade 1/3.

Essa dedução pode ser generalizada para qualquer quantidade maior de convidados. Se são $n > 3$ convidados e ocorre a mesma confusão, então a probabilidade de k deles receberem o próprio chapéu é

$$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = \frac{(n-k)!}{n!}$$

pois $(n-k)!$ permutações distintas têm os mesmos k chapéus fixos; do princípio de inclusão–exclusão (exercício 11, página 28) temos

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{1 \leqslant i < j \leqslant n} \mathbb{P}(E_i \cap E_j) + \sum_{1 \leqslant i < j < k \leqslant n} \mathbb{P}(E_i \cap E_j \cap E_k) + \cdots \\ &\quad + (-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) + \cdots + (-1)^{n-1}\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) \\ &= \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!}\end{aligned}$$

portanto, a probabilidade de ninguém pegar o próprio chapéu é

$$1 - \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!} = 1 + \sum_{k=1}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Um fato curioso é que da conhecida série

$$\sum_{j\geqslant 0} \frac{(-1)^j}{j!} = \frac{1}{\mathrm{e}}$$

concluímos que a probabilidade de ninguém pegar o próprio chapéu

$$1 - \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \xrightarrow[n\to\infty]{} \frac{1}{\mathrm{e}}.$$

◊

Exercício 42. Qual é o número de permutações distintas com as letras da palavra ana?

Solução. São 3! permutações de 3 símbolos, mas há permutações que dão origem a mesma sequência. As seis permutações de ana são:

ana ana aan aan naa naa

em cada duas permutações a palavra é a mesma, só muda a ordem da letra repetida, portanto são

$$\frac{3!}{2!} = 3$$

permutações distintas. □

Exercício 43. Qual é o número de permutações distintas com as letras da palavra $bala$?

Solução. da mesma meneira, das 4! permutações as 2! permutações que troca a ordem da letra igual e deixam as outras letras na mesma posição da sequência geram a mesma palavra, portanto são

$$\frac{4!}{2!} = 12$$

permutações distintas. □

Exercício 44. Qual é o número de permutações distintas com as letras da palavra $banana$?

Solução. São 6! permutações de 6 símbolos. Mas agora há permutações que diferem apenas na ordem das letras a ou das letras n e dão origem a mesma sequência, por exemplo:

banana banana banana banana

são permutações diferentes que determinam a mesma sequência. As 3! permutações das letras a são indistinguíveis, assim como as 2! da letras n, portanto, há

$$\frac{6!}{3!\,2!} = 60$$

permutações distintas. □

Exercício 45. Um sinal é composto por nove bandeiras alinhadas. Quantos sinais diferentes é possível formar quando há disponíveis 4 bandeiras brancas, três bandeiras vermelhas e duas bandeiras azuis? Bandeiras da mesma cor são idênticas.

Esboço de solução. $\frac{9!}{4!3!2!} = 1.260$. □

No caso de permutações com repetição de objetos, se são n objetos no total, com r tipos de objetos distintos e k_i objetos do tipo i ($1 \leq i \leq r$, $n = k_1 + \cdots + k_r$) então temos $n!$ permutações donde descontamos as $k_i!$ permutações de objetos do mesmo tipo resultando

$$\binom{n}{k_1,\ k_2, \dots,\ k_r} := \frac{n!}{k_1!\, k_2! \cdots k_r!}$$

Exemplo 22. Numa mão de Bridge as 52 cartas de um baralho embaralhado são divididas igualmente entre 4 jogadores. O número de modos distintos com que isso é feito pode calculado da seguinte forma: uma distribuição de cartas corresponde a uma sequência de 52 objetos, os 13 primeiros objetos da sequência são as cartas do primeiro jogador, os 13 seguintes do segundo jogador, os próximos 13 do terceiro e os 13 últimos objetos da sequência são as cartas do quarto jogador. Há 52! sequências distintas de cartas. Entretanto, dessas 52! temos que, para cada jogador, 13! permutações correspondem a mesma sequência de cartas em sua mão, portanto, são

$$\binom{52}{13, 13, 13, 13} = 53.644.737.765.488.792.839.237.440.000$$

modos distintos de distribuir as cartas, ou mãos diferentes de início de jogo. ◊

Exemplo 23. Se numa mão de Bridge as 52 cartas de um baralho são divididas igualmente e aleatoriamente entre 4 jogadores. Com que probabilidade cada jogador recebe um ás? ◊

Solução. Os 4 ases podem ser distribuídos de 4! modos distintos entregando um para cada jogador. As 48 cartas restantes são distribuídas pelos jogadores de $\binom{48}{12,12,12,12}$ maneiras distintas. Pelo princípio multiplicativo são $4!\binom{48}{12,12,12,12}$ modos distintos de os jogadores receberem um ás cada. Portanto a probabilidade é

$$\frac{4!\binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}}$$

que vale aproximadamente 0,0044. □

Exercício 46. Numa mão de Bridge as 52 cartas de um baralho são divididas aleatoriamente entre 4 jogadores. Com que probabilidade um jogador recebe todas as cartas de paus?

Solução. Denotamos por E_1 o evento "o primeiro jogador receba todas as cartas de paus". As 13 cartas de paus são entregues ao primeiro jogador e as $52-13=39$ cartas restantes são distribuídas aleatoriamente entre os outros 3 jogadores. O número de maneiras que E_1 ocorre é $\binom{39}{13,13,13}$. Portanto, a probabilidade do primeiro jogador receber todas as cartas de paus é

$$\frac{\binom{39}{13,13,13}}{\binom{52}{13,13,13,13}} = \frac{39!\,13!}{52!}$$

que é aproximadamente 2×10^{-22}. Se E_2, E_3, E_4 denotam os eventos "o segundo jogador receba todas as cartas de paus", "o terceiro jogador receba todas as cartas de paus"e "o quarto jogador receba todas as cartas de paus", então a probabilidade que algum jogador recebe todas as cartas de paus é a probabilidade da união de quatro subconjuntos disjuntos entre si, cada um com a probabilidade dada acima, ou seja,

$$\mathbb{P}(E_1 \uplus E_2 \uplus E_3 \uplus E_4) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) = 4 \cdot \frac{39!\,13!}{52!}$$

que é aproximadamente 1×10^{-21}. Agora, qual a probabilidade de algum jogador receber as mesmas cartas de qualquer um dos quatro naipes? (veja exerc. 65, pág. 86) □

1.2.8 Combinações: tomemos um arranjo de r elementos escolhidos de um conjunto com n elementos. A quantidade de arranjos que têm os mesmos r elementos é $r!$ pois a única diferença entre eles é a ordem com que se apresentam os r elementos. Por exemplo, se selecionamos sequencialmente e sem reposição 3 cartas de um baralho então temos $52 \cdot 51 \cdot 50$ arranjos distintos, um dos quais é $(K\heartsuit, 5\clubsuit, Q\diamondsuit)$. Agora, se selecionamos três cartas de uma só vez as 3! permutações de $(K\heartsuit, 5\clubsuit, Q\diamondsuit)$ correspondem a mesma seleção. A quantidade de seleções distintas é

$$\frac{52 \cdot 51 \cdots 50}{3!} = \frac{52!}{49!\,3!}.$$

Quando essa ordem não importa o que temos são combinações.

> Uma combinação de r elementos escolhidos de um conjunto A com n elementos é simplesmente um subconjunto com r elementos de A.

Coeficiente binomial: o número de combinações de r elementos tomados de um conjunto de n, para $0 \leqslant r \leqslant n$, elementos é o coeficiente binomial

$$\binom{n}{r} := \frac{(n)_r}{r!} = \frac{n!}{r!\,(n-r)!}.$$

Convencionamos que $\binom{a}{b} = 0$ se $a < b$.

Exemplo 24 (**Mega-Sena**). O jogo de apostas consiste em acertar 6 dezenas escolhidas dentre 60. O número de possíveis resultados para a Mega-Sena é $\binom{60}{6} = 50.063.860$. Se uma aposta em seis números demorar 1 segundo para ser registrada, então registrar 50.063.860 demoraria um ano e meio, aproximadamente. A probabilidade de acertar os seis números é

$$\frac{1}{\binom{60}{6}} = \frac{1}{50.063.860} \approx 2 \times 10^{-8}.$$

A chance[3] de morrer por raio no Brasil em 2010 era $0{,}8 \times 10^{-6}$ (40 vezes maior). ◊

Exemplo 25. Numa população com n elementos, n_1 são azuis e $n_2 = n - n_1$ são verdes. De quantas maneiras podemos escolher k elementos com r deles azuis? ($0 \leq r \leq \min\{n_1, k\}$) O número de maneiras de escolher $k - r$ verdes é $\binom{n_2}{k-r}$. O número de maneiras de escolher r azuis é $\binom{n_1}{r}$. Pelo Princípio Multiplicativo, o número de maneiras de escolher r azuis e $k - r$ verdes é $\binom{n_2}{k-r}\binom{n_1}{r}$. ◊

Exercício 47. Prove que a seguinte identidade

$$i\binom{n}{i} = n\binom{n-1}{i-1}.$$

Exercício 48. Três bolas são retiradas aleatoriamente de uma caixa com 6 bolas brancas e 5 bolas pretas. Com que probabilidade a escolha resulta em 1 branca e 2 pretas?

Solução. No total são 13 bolas das quais escolhemos 3. O número de possíveis resultados é $\binom{13}{3}$. "6 bolas brancas e 5 bolas pretas" ocorre de $\binom{6}{1}\binom{5}{2}$ modos diferentes, pelo exercício anterior. Portanto a probabilidade é $\frac{\binom{6}{1}\binom{5}{2}}{\binom{13}{3}}$. □

Exercício 49. Cinco pessoas são escolhidas aleatoriamente dentre 6 homens e 9 mulheres. Com que probabilidade tal comitê é composto por 3 homens e 2 mulheres?

Solução. Seguindo a mesma linha de raciocínio do exemplo anterior, a probabilidade é $\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}}$. Explique os detalhadamente como se chega a esse resultado. □

Exercício 50. Cinco pessoas são escolhidas aleatoriamente dentre 6 homens e 9 mulheres. Com que probabilidade a maioria é homem?

Se A é um conjunto com n elementos então a quantidade de subconjuntos de A de cardinalidade r é o número de maneiras distintas que podemos selecionar r elementos dentre os n do conjunto, isto é, são $\binom{n}{r}$ subconjuntos, como há 2^n subconjuntos de A concluímos que

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n \tag{1.2.8}$$

[3] esse número é uma média no sentido de que considera que todos têm a mesma chance de ser atingido, o que não é real. [http://www2.uol.com.br/sciam/reportagens/os_numeros__surpreendentes__de_mortes_por_raios_no_brasil.html]

Esse fato é consequência, também, de um resultado mais geral conhecido como Teorema Binomial.

Teorema Binomial: Para todo $n > 0$, vale $(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$.

Fazendo $x = y = 1$ temos a equação (1.2.8).

Vejamos como o produto de desenvolve para valores pequenos de n

$$\begin{aligned}
(x+y)(x+y) &= (x+y)x + (x+y)y \\
&= xx + yx + xy + yy \\
(x+y)(x+y)(x+y) &= (x+y)(xx + yx + xy + yy) \\
&= (x+y)xx + (x+y)yx + (x+y)xy + (x+y)yy \\
&= xxx + yxx + xyx + yyx + xxy + yxy + xyy + yyy \\
(x+y)(x+y)(x+y)(x+y) &= (x+y)(xxx + yxx + xyx + yyx + xxy + yxy + xyy + yyy) \\
&= xxxx + xyxx + xxyx + xyyx + xxxy + xyxy + xxyy + xyyy \\
&+ yxxx + yyxx + yxyx + yyyx + yxxy + yyxy + yxyy + yyyy
\end{aligned}$$

Assim, $(x+y)^n =$

$$(x+y)(x+y)\cdots(x+y) \tag{1.2.9}$$

com n ocorrências de $(x+y)$. Desenvolvendo o produto temos uma soma em que cada termo é da forma $x^r y^{n-r}$, para $0 \leq r \leq n$. Para cada r o coeficiente de $x^r y^{n-r}$ é o número de maneiras de escolher o x de r fatores da equação (1.2.9) (para o y, dos $n-r$ fatores restantes). O número de maneiras de escolher r fatores dentre n é $\binom{n}{r}$, portanto

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}.$$

Exemplo 26. Se A, B e C são três subconjuntos aleatórios de $X = \{1, 2, \dots, n\}$, com que probabilidade $A \cap B \subset C$?

O espaço amostral é formado pelas ternas $(A, B, C) \in S \times S \times S$, em que S o conjunto das partes de $\{1, 2, \dots, n\}$, de modo que queremos a probabilidade dos eventos formados pelas ternas tais que $A \cap B \subset C$ assumindo que cada terna tem a mesma probabilidade de ser escolhida. Vamos contar o número de ternas (A, B, C) com $A \cap B \subset C$.

Primeiro, fixamos o conjunto $C \in S$ e o conjunto $A \in S$. Digamos que C tem cardinalidade c e $A \setminus C$ tem cardinalidade a, com $0 \leqslant c \leqslant n$ e $0 \leqslant a \leqslant n - c$. O número de possíveis $B \in S$ tais que $A \cap B \subset C$ é contado da seguinte maneira: para que $A \cap B$ seja subconjunto de C o conjunto B não pode ter elementos de $A \setminus C$, caso contrário teríamos $x \in A \cap B$ e $x \notin C$. Qualquer outro elemento de X pode estar em B logo temos 2^{n-a} possibilidades em S para o conjunto B.

Deixemos C fixado e vamos contar quantos $A \in S$ são possíveis com a elementos em $A \setminus C$. Um conjunto A com a elementos em $A \setminus C$ é formado por elementos de um subconjunto de a elementos escolhidos em $n-c$, ou seja $\binom{n-c}{a}$ possibilidades para $A \setminus C$ aos quais unimos um dos 2^c subconjuntos de C para formar A, pelo princípio multiplicativo são $2^{n-a}\binom{n-c}{a}2^c$ possibilidades para formar A. Assim o número de pares (A, B) cuja interseção está contida em C (fixo) é

$$\sum_{a=0}^{n-c} 2^{n-a}\binom{n-c}{a}2^c = 2^c 2^c \sum_{a=0}^{n-c}\binom{n-c}{a}2^{n-c-a} = 4^c(1+2)^{n-c} = 4^c 3^{n-c} \tag{1.2.10}$$

na segunda igualdade usamos o teorema binomial.

Da equação (1.2.10) acima, são $4^c 3^{n-c}$ pares (A, B) cuja interseção está contida em qualquer $C \in S$ cuja cardinalidade é c. A quantidade de tais C é $\binom{n}{c}$, portanto, o número de ternas $(A, B, C) \in S \times S \times S$ tais que $A \cap B \subset C$ é

$$\sum_{c=0}^{n}\binom{n}{c}4^c 3^{n-c} = (4+3)^n \tag{1.2.11}$$

novamente, na equação (1.2.11) usamos o teorema binomial para concluir o resultado.

Portanto

$$\mathbb{P}\Big(\{(A, B, C) \in S \times S \times S : A \cap B \subset C\}\Big) = \left(\frac{7}{8}\right)^n$$

pois $|\{(A, B, C) \in S \times S \times S : A \cap B \subset C\}| = 7^n$ e são 8^n ternas em S. $\lozenge$

Coeficiente multinomial: de volta ao exemplo 22, se dividimos a tarefa de distribuir as cartas em 4 etapas, cada etapa seleciona 13 cartas para um jogador, então pelo princípio multiplicativo o número de maneiras distintas de distribuir as cartas no jogo de bridge é

$$\binom{52}{13}\binom{52-13}{13}\binom{52-13-13}{13}\binom{52-13-13-13}{13} = \binom{52}{13, 13, 13, 13}$$

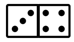
Com raciocínio análogo ao feito para o Teorema Binomial,

$$(x+y+z)^n = \sum_{r_1+r_2+r_3=n}\binom{n}{r_1, r_2, r_3}x^{r_1}y^{r_2}z^{r_3}.$$

e, de modo geral,

Teorema Multinomial: Para todo $n > 0$, vale $(x_1 + x_2 + \cdots + x_k)^n = \displaystyle\sum_{r_1+r_2+\cdots+r_k=n}\binom{n}{r_1, r_2, \ldots, r_k}x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$.

de modo que $\binom{n}{r_1, r_2, \ldots, r_k}$ é conhecido como **coeficiente multinomial**.

Combinação com repetição: se escolhemos uma peça de dómino ao acaso, com que probabilidade obtemos  ?

As peças de dominós são formadas por dois números tomados dos números de 0 a 6 podendo haver repetição. Os dominós com pares de números diferentes são $\binom{7}{2} = 21$, mais os 7 pares repetidos resultam em 28 peças de dominós, portanto, são 28 combinações de 2 objetos tomados dentre 7 com repetição. Essa estratégia de contagem não é facilmente generalizável, o leitor pode tentar contar o número de peças de dominós de 5 pontas com 16 possíveis números diferentes, o resultado deverá ser 15.504.

A resposta para o caso geral: dentre n objetos, queremos selecionar r podendo haver repetição. Isso pode ser feito de

$$\binom{n+r-1}{r} \tag{1.2.12}$$

maneiras diferentes. No caso dos dominós, por exemplo, são 7 números dos quais selecionamos 2, podendo repetir número

$$\binom{7+2-1}{2} = \binom{8}{2} = \frac{8!}{6!\,2!} = 36.$$

Uma maneira de modelar combinação com repetição para deduzir equação (1.2.12) é escrever uma equação com uma indeterminada para cada um dos n objetos, $x_1, x_2, \dots, x_n$. A variável x_i indica quantas vezes o i-ésimo objeto será selecionado, portanto $x_1 + x_2 + \cdots + x_n = r$. Assim, o número combinações com repetição é a quantidade de soluções inteiras de

$$x_1 + x_2 + x_3 + \cdots + x_n = r \tag{1.2.13}$$

1.2.9 Soluções inteiras de equações lineares: vamos começar com um caso simples, estudaremos o número de soluções de

$$x_1 + x_2 + x_3 = 6, \quad \text{com } x_i \in \{1, 2, \dots\} \text{ para todo } i. \tag{1.2.14}$$

Escrevemos

$$1 + 1 + 1 + 1 + 1 + 1 = 6$$

e uma solução da equação (1.2.14) corresponde a escolha de 2 operadores + dentre os 5 escritos na equação acima; por exemplo, se usamos $\oplus$ para representar as escolhas

$$\underbrace{1+1}_{x_1} \oplus \underbrace{1+1+1}_{x_2} \oplus \underbrace{\;1\;}_{x_3} = 6$$

corresponde a $x_1 = 2$, $x_2 = 3$ e $x_3 = 1$, e

$$\underbrace{1+1}_{x_1} \oplus \underbrace{1+1}_{x_2} \oplus \underbrace{1+1}_{x_3} = 6$$

corresponde a $x_1 = 2$, $x_2 = 2$ e $x_3 = 2$. Portanto essa equação tem $\binom{5}{2}$ soluções em $\mathbb{Z}^+$.

Agora, estudaremos o número de soluções de

$$x_1 + x_2 + x_3 = 6, \quad \text{com } x_i \in \{0, 1, 2, \dots\} \text{ para todo } i.$$

Notemos que uma solução $(x_1, x_2, x_3) = (x, y, z)$ inteira e *positiva* da equação $x_1 + x_2 + x_3 = 6 + 3$ determina uma única solução inteira e *não-negativa* $(x - 1, y - 1, z - 1)$ da equação $x_1 + x_2 + x_3 = 6$ e vice-versa, ou seja, as equações

$$x_1 + x_2 + x_3 = 6 + 3, \quad \text{com } x_i \in \{1, 2, 3, \dots\} \text{ para todo } i.$$
$$x_1 + x_2 + x_3 = 6, \quad \text{com } x_i \in \{0, 1, 2, \dots\} \text{ para todo } i.$$

têm o mesmo número de soluções.

De volta ao problema que gerou essa discussão: o número de maneiras de selecionar r objetos, podendo haver repetição, dentre n objetos é igual ao número de soluções inteiras da equação (1.2.13), que é o mesmo número de soluções inteiras de

$$x_1 + x_2 + x_3 + \dots + x_n = n + r, \quad \text{com } x_i \in \{1, 2, \dots\} \text{ para todo } i.$$

consideramos $1 + 1 + 1 + \dots + 1 = n + r$ e dos $n + r - 1$ operadores + escolhemos $n - 1$, ou seja, são $\binom{n+r-1}{n-1}$ soluções inteiras. Finalmente, a equação (1.2.12) segue do seguinte exercício

Exercício 51. Verifique que vale

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$$

para todo n e todo r para os quais os coeficientes binomiais estão definidos.

A quantidade de maneiras diferentes de selecionarmos r elementos de um conjunto de n elementos é,

	com repetição	sem repetição
com ordem	n^r	$(n)_r$
sem ordem	$\binom{n+r-1}{r}$	$\binom{n}{r}$

Tabela 1.1: seleção de r elementos dentre n.

1.2.10 Bolas em caixas: um modelo para problemas de contagem é considerar a distribuição de bolas em caixas. Consideramos os casos: *caixas distinguíveis* ou *caixas idênticas* e *bolas distinguíveis* ou *bolas idênticas*.

Por exemplo, há 8 modos de distribuir as bolas distinguíveis a, b, c em duas caixas distinguíveis, *esquerda* e *direita*

$$\begin{array}{cccc} abc| & ab|c & a|bc & |abc \\ ac|b & bc|a & b|ac & c|ab \end{array}$$

Se as caixas fossem indistinguíveis então $abc|$ e $|abc$ definem a mesma configuração, assim como $ac|b$ e $b|ac$, como $ab|c$ e $c|ab$, como $bc|a$ e $a|bc$. Portanto, há 4 modos de distribuir as bolas distinguíveis a, b, c em duas caixas indistinguíveis. Se as bolas são indistinguíveis e as caixas não

$$* * * \mid \qquad ** \mid * \qquad * \mid ** \qquad \mid * * *$$

são 4 modos e se caixas e bolas são indistinguíveis são 2 modos (por quê?).

Exemplo 27. De quantas maneiras distintas podemos distribuir 3 bolas distinguíveis em 4 caixas distinguíveis?

Para cada uma das 3 bolas bola há 4 possíveis caixas, portanto temos um evento composto por uma sequência de 3 etapas em que cada um tem 4 possíveis resultados, portanto 4^3 maneiras distintas para distribuir as bolas nas caixas. ◊

Exercício 52 (**distribuição de bolas em caixas — arranjo com repetição**). Prove que há n^r maneiras distintas de distribuir r bolas distinguíveis em n caixas distinguíveis.

Exercício 53 (**distribuição de bolas em caixas — arranjo**). De quantas maneiras distintas podemos distribuir r bolas distinguíveis em n caixas distinguíveis ($r \leq n$) de modo que cada caixa receba no máximo 1 bola?

Solução. Para cada uma das r bolas, escolhemos uma caixa dentre as disponíveis, a qual se torna indisponível a partir de então. Temos o evento composto pela sequência $E_1, E_2, \dots E_r$ de etapas com $|E_i| = |E_{i-1}| - 1$ $(1 < i \leq n)$ e $E_1 = n$, portanto

$$\left| \bigtimes_{i=1}^{r} E_i \right| = (n)_r$$

maneiras distintas de distribuir r bolas distinguíveis em n caixas distinguíveis de modo que cada caixa receba no máximo 1 bola. □

Exercício 54 (**distribuição de bolas em caixas — permutação com repetição**). De quantas maneiras distintas podemos distribuir n bolas distinguíveis em r caixas distinguíveis de modo que a caixa 1 receba k_1 bolas, a caixa 2 receba k_2 bolas, e assim por diante até que a caixa r receba k_r bolas.

Solução. Para cada uma das $n!$ permutações das bolas, colocamos as k_1 primeiras da sequência na caixa 1, as k_2 próximas na caixa 2 e assim por diante. As k_1 primeiras podem ocorrer de $k_1!$ maneiras distintas e aqui ordem não importa pois vão todas para a mesma caixa independentemente da ordem, as k_2 próximas de $k_2!$ maneiras distintas, e assim por diante. Cada maneira de distribuir as bolas nas caixas de acordo com a restrição enunciada pode ser obtida por uma permutação como acima, portanto o número de maneiras é $\binom{n}{k_1,\ k_2,\dots,\ k_r}$. □

Exercício 55 (**distribuição de bolas em caixas — combinação**). De quantas maneiras distintas podemos distribuir r bolas idênticas em n caixas distinguíveis ($r \leq n$) de modo que cada caixa receba no máximo 1 bola?

Solução. Das n caixas podemos escolher r de $\binom{n}{r}$ maneiras diferentes, como as bolas são indistinguíveis distribuímos arbitrariamente uma por caixa. □

Exemplo 28 (**distribuição de bolas em caixas — combinação com repetição**). Uma seleção de r objetos dentre n podendo haver repetição equivale a distribuir r bolas indistinguíveis em n caixas distinguíveis. Uma caixa vazia corresponde a um objeto não selecionado, uma caixa com k bolas corresponde a um mesmo objeto repetido k vezes. ◊

Exemplo 29. Numa mão de Bridge as 52 cartas de um baralho são divididas entre 4 jogadores. O número de maneiras distintas de distribuir as cartas é o número de maneiras de distribuir 52 bolas distinguíveis em 4 caixas distinguíveis de modo que cada caixa receba 13 bolas. Vimos que isso pode ser feito de $\binom{52}{13,13,13,13}$ maneiras distintas. ◊

1.2.11 Probabilidade condicional e experimentos compostos: Consideremos uma escolha aleatória de uma família com dois filhos da qual observamos o sexo dos filhos. Qual é a probabilidade do evento "ambos são meninos" se é sabido que ocorre o evento "pelo menos um deles é menino"?

Se descrevemos o espaço amostral como $\Omega = \{(♂,♂),(♂,♀),(♀,♂),(♀,♀)\}$ e sabendo-se que um dos filhos é menino então a resposta é um elemento de $\{(♂,♂),(♂,♀),(♀,♂)\}$ (agora o evento certo), de modo que a probabilidade de ambos serem meninos é $1/3$. Agora, suponhamos que alguém bate à porta de uma família que tem dois filhos, um menino abre a porta e diz "eu sou o filho mais velho". Qual é a probabilidade de que ele tem um irmão? A resposta é um elemento de $\{(♂,♂),(♀,♂)\}$, portanto a probabilidade de ambos serem meninos é $1/2$. Agora, se um dos dois filhos, escolhido aleatoriamente, abre a porta e não diz nada, qual é a probabilidade de que ele tem um irmão dado que foi um menino que abriu a porta?

Seja E evento de Ω com $E \neq \emptyset$. Se ocorre algum dos $|E|$ pontos amostrais de E, então o evento A ocorre se um dos $|A \cap E|$ pontos amostrais de $A \cap E$ ocorre, de modo que *a probabilidade com que A*

ocorre dado que E ocorre é

$$\mathbb{P}(A \mid E) := \frac{|A \cap E|}{|E|} = \frac{\frac{|A\cap E|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}. \tag{1.2.15}$$

Exemplo 30. Uma urna tem 20 bolas azuis e 10 bolas brancas. Das bolas azuis, 5 têm a letra X e 15 têm a letra Y; das bolas brancas, 1 têm a letra X e 9 tem a letra Y. Uma bola é escolhida e sua cor é azul. Qual é a probabilidade dessa bola ter a letra X?

O número de bolas azuis é 20 e dessas bolas 5 têm a letra X, portanto a probabilidade da bola azul sorteada ter a letra X é $5/20 = 1/4$. Se E representa o evento "ocorre bola azul" e A o evento "tem a letra X" então calculamos que $\mathbb{P}(A \mid E) = 1/4$. Notemos que da equação (**??**) podemos deduzir que a probabilidade de sortear uma bola azul e com a letra X é $\mathbb{P}(A \cap E) = \mathbb{P}(A \mid E) \cdot \mathbb{P}(E) = (1/4) \cdot (20/30) = 1/6$. ◊

1.2.12 Regra da multiplicação: é como é conhecido o fato usado no exemplo acima, a saber $\mathbb{P}(A \cap E) = \mathbb{P}(A \mid E) \cdot \mathbb{P}(E)$, o qual é consequência direta de equação (1.2.15).

O conceito de probabilidade condicional é fundamental e um dos mais importantes Teoria de Probabilidades, tanto que retomaremos o estudo com mais profundidade mais a frente, a partir da página 99. Por ora, chamamos a atenção ao fato de que não precisa haver uma relação temporal ou causal entre os eventos A e C.

Exercício 56. Fixado um evento E como acima, a função que associa $A \mapsto \mathbb{P}(A \mid E)$, para todo evento A, satisfaz as propriedades da equação (1.2.2), da equação (1.2.3) e da equação (1.2.4) na página 43. Verifique.

Como consequência do exercício acima, todas as outras propriedades que decorrem dessas três, equação (1.2.2), equação (1.2.3) e equação (1.2.4), também são válidas para $\mathbb{P}(\cdot \mid E)$, em particular valem

(i) $\mathbb{P}\big(\overline{A} \bigm| E\big) = 1 - \mathbb{P}(A \mid E)$;

(ii) $\mathbb{P}(\emptyset \mid E) = 0$

(iii) $0 \leqslant \mathbb{P}(A \mid E) \leqslant 1$

(iv) se $A \subset B$ então $\mathbb{P}(A \mid E) \leqslant \mathbb{P}(B \mid E)$

(v) $\mathbb{P}(A \cup B \mid E) \leqslant \mathbb{P}(A \mid E) + \mathbb{P}(B \mid E)$

(vi) $\mathbb{P}(A \cup B \mid E) = \mathbb{P}(A \mid E) + \mathbb{P}(B \mid E) - \mathbb{P}(A \cap B \mid E)$

(vii) $\mathbb{P}\big(\biguplus_{i=1}^{t} A_i \bigm| E\big) = \sum_{i=1}^{t} \mathbb{P}(A_i \mid E)$.

1.2.13 Experimentos compostos: o conceito de probabilidade condicionada nos fornece uma técnica para definir modelos probabilísticos para experimentos compostos nos quais certas etapas do experimento dependem dos resultados de outras etapas. Suponhamos que temos n experimentos $\mathcal{E}_1, \dots, \mathcal{E}_n$ com espaços amostrais finitos e eventos elementares equiprováveis. Podemos formar experimento novo, composto, realizando os n experimentos em sequência: primeiro $\mathcal{E}_1$ em seguida $\mathcal{E}_2$ e assim por diante. Se os n experimetos têm, respectivamente, espaços amostras $\Omega_1, \Omega_2, \dots, \Omega_n$ então o produto cartesiano $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ é o espaço de amostral natural para o experimento composto que consiste em realizar as experiências na sequência.

Exemplo 31. Consideremos três urnas, cada uma com a mesma probabilidade de ser escolhida, 1/3. Em cada uma das urnas há seis bolas, cada uma com a mesma probabilidade de ser escolhida, 1/6. Na urna A temos três bolas brancas e três bolas azuis; na urna B temos duas bolas brancas e quatro azuis; na urna do tipo C todas as bolas nela são brancas. Uma urna é escolhida aleatoriamente e, em seguida, uma bola é escolhida aleatoriamente e observamos a cor dessa bola.

Essa descrição de experimento se encaixa no modelo descrito acima. Temos dois experimentos: $\mathcal{E}_1$ consiste de sortear uma urna e $\mathcal{E}_2$ de sortear uma bola da urna escolhida. Para $\mathcal{E}_1$ descrevemos o espaço amostral $\Omega_1 = \{A, B, C\}$ e denotamos a probabilidade uniforme nesse espaço por $\mathbb{P}_1$ Para $\mathcal{E}_2$ descrevemos o espaço amostral $\Omega_2 = \{\mathrm{b}_1, \dots, \mathrm{b}_6\}$ e denotamos a probabilidade uniforme nesse espaço por $\mathbb{P}_2$. Notemos que o interesse, de fato, no experimento 2 é a cor da bola, portanto teremos os eventos de interesse $E_A \subset \Omega_2$ definido pelas bolas azuis e $E_B \subset \Omega_2$ definido pelas bolas brancas e passamos a considerar $\Omega_2' = \{E_A, E_B\}$; a probabilidade desses eventos depende da urna em Ω_1 escolhida na primeira fase porque a quantidade de bolas em cada cor depende da urna.

O espaço amostral do experimento composto dessas duas etapas é definido pelos pares de $\Omega_1 \times \Omega_2' = \{A, B, C\} \times \{E_A, E_B\}$. Por exemplo, se F representa o evento "urna do tipo A" no experimento composto, i.e.,

$$F = \{A\} \times \Omega_2' = \{(A, y) : y = E_A \text{ ou } y = E_B\}$$

então $\mathbb{P}(F) = \mathbb{P}_1(A) = 1/3$. O evento "bola branca" no experimento composto é

$$E = \Omega_1 \times \{E_B\} = \{(x, E_B) : x = A, \text{ ou } x = B, \text{ ou } x = C\}. \tag{1.2.16}$$

A medida de probabilidade de um par (x, y) não é uniforme e é dada pela regra de multiplicação: vejamos o caso de (A, E_B), i.e., sorteamos "urna A e bola branca". Primeiro, fazemos $\mathbb{P}(F \times \Omega_2) = \mathbb{P}_1(F)$, para todo $F \subset \Omega_1$, ou seja, a probabilidade de ocorrer um determinado evento F no primeiro experimento e ocorrer qualquer resultado no segundo experimento é probabilidade do evento F no primeiro experimento. Agora, a probabilidade de sortear uma bola branca no segundo experimento depende do resultado do primeiro experimento; na urna A temos $\mathbb{P}_2(E_B) = 1/2$. No experimento composto essas são as probabilidades condicionais $\mathbb{P}(\text{"bola branca"} \mid \text{"urna A"}) = \mathbb{P}(E \mid F) = \mathbb{P}_2(E_B) = 1/2$, em que "bola branca" representa o evento E definido na equação (1.2.16). Com isso, "urna A e bola branca" que é o evento

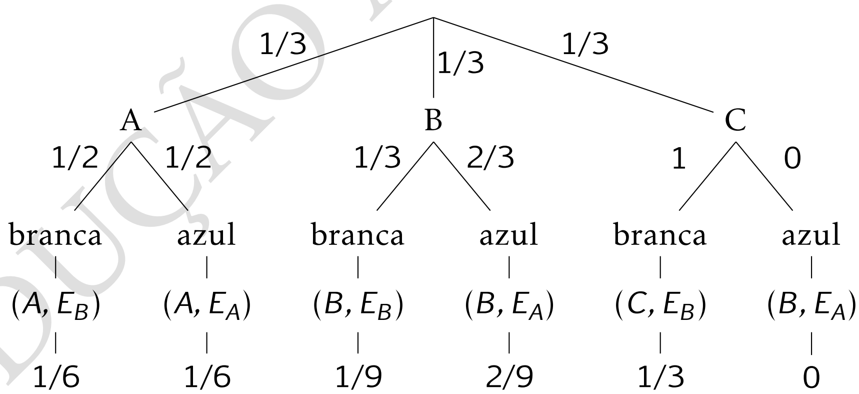
$$\{(A, E_B)\} = (\Omega_1 \times E_B) \cap (\{A\} \times \Omega_2) \tag{1.2.17}$$

tem probabilidade

$$\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F)\mathbb{P}(F) = \frac{1}{2} \cdot \frac{1}{3}. \tag{1.2.18}$$

Podemos determinar do mesmo modo a probabilidade de todo ponto amostral a partir das informações (1) probabilidade $\mathbb{P}$ dos eventos no primeiro experimento e (2) a probabilidade $\mathbb{P}(\cdot \mid \cdot)$ dos eventos do segundo experimento condicionada ao resultado do primeiro experimento. Isso porque cada ponto amostral pode ser escrito como uma interseção, como na equação (1.2.17) acima, e a probabilidade é calculada pelo produto de probabilidades como na equação (1.2.18). Porém, nesse exemplo é mais conveniente descrevermos o modelo probabilístico pictoricamente através de um diagrama de árvore.

O diagrama de árvore abaixo representa cada etapa do experimento em cada nível da árvore, com as respectivas probabilidades nas ramificações correspondentes aos resultados de cada etapa. A partir do segundo nível essas probabilidades são condicionadas ao que ocorreu na etapas anteriores. Uma maneira prática de atribuir probabilidade a um ponto amostral é tomar o produto das probabilidades no caminho até ele nessa árvore, por exemplo, $\mathbb{P}((A, E_B)) = 1/3 \cdot 1/2$,



e estendemos a probabilidade para qualquer evento $G \subset \Omega_1 \times \Omega_2'$ somando a probabilidade de seus elementos

$$\mathbb{P}(G) = \sum_{\omega \in G} \mathbb{P}(\omega).$$

O evento G definido por "a bola sorteada é branca" ocorre com probabilidade

$$\mathbb{P}\Big((A, E_B)\Big) + \mathbb{P}\Big((B, E_B)\Big) + \mathbb{P}\Big((C, E_B)\Big) = \frac{11}{18}.$$

Observemos que essa probabilidade não depende do número de bolas brancas na urna C, portanto, *a probabilidade de B não é a quantidade de bolas brancas dividido pelo número total de bolas*, que é um erro cometido frequentemente nesse caso. ◊

Exercício 57. Verifique que Ω e $\mathbb{P}$ nesse exemplo dado acima verificam as propriedades na equação (1.2.2), equação (1.2.3) e equação (1.2.4) na página 43.

Como vimos, nos experimentos compostos podemos montar um diagrama de árvore e cada resultado corresponde a um caminho nessa árvore. Em cada etapa, a partir da segunda, a probabilidade depende das respostas das etapas anteriores, assim é uma probabilidade condicional, e o produto das probabilidades de cada etapa corresponde à probabilidade do resultado final.

Exemplo 32. Numa cômoda há três gavetas iguais e em cada gaveta um par de meias. Na primeira gaveta há um par de meias brancas, na segunda um par de meias pretas e na terceira gaveta um par com um pé de meia de cada cor, preta e branca. Uma das três gavetas é escolhida ao acaso e em seguida essa gaveta é aberta e, sem olhar para o interior da gaveta escolhida, um pé de meia é escolhido aleatoriamente dentre os dois pés daquela gaveta que é então fechada. O pé de meia retirado é branco. Qual a probabilidade de o outro pé que ficou sozinho na gaveta ser branco?

Tomemos o espaço amostral formado pelos pares (gaveta escolhida, cor da meia retirada) no qual denotaremos as gavetas por 1, 2 e 3

$$\Omega = \{1,2,3\} \times \{\text{branca}, \text{preta}\} = \{(1,\text{branca}), (1,\text{preta}), (2,\text{branca}), (2,\text{preta}), (3,\text{preta}), (3,\text{branca})\}$$

e em cada etapa temos modelos probabilísticos clássicos $\left(\{1,2,3\}, 2^{\{1,2,3\}}, \mathbb{P}_1\right)$ para o primeiro experimento e $\left(\{\text{branca}, \text{preta}\}, 2^{\{\text{branca},\text{preta}\}}, \mathbb{P}_2\right)$ para o segundo, com $\mathbb{P}_1$ e $\mathbb{P}_2$ uniformes.

Queremos determinar $\mathbb{P}(\text{"gaveta 1"} \mid \text{"meia branca"})$, que corresponde a probabilidade do evento "sobrar meia branca dado que sorteou meia branca" e onde "gaveta 1"e "meia branca"são os seguintes eventos de Ω

$$\begin{aligned} A &= \text{"meia branca"} = \Omega_1 \times \{\text{branca}\} \\ E &= \text{"gaveta 1"} = \{1\} \times \Omega_2 \end{aligned}$$

Sabemos que $\mathbb{P}(E) = \mathbb{P}_1(1) = 1/3$ e que $\mathbb{P}(A \mid E) = \mathbb{P}_2(\text{branca}) = 1$, portanto, pela regra da multiplicação

$$\mathbb{P}\big((1,\text{branca})\big) = \mathbb{P}(A \cap E) = \mathbb{P}(A \mid E) \cdot \mathbb{P}(E) = \frac{1}{3}$$

repetindo para todo ponto amostral de Ω temos

$$\begin{aligned} \mathbb{P}\big((1,\text{branca})\big) = \mathbb{P}\big((2,\text{preta})\big) &= \frac{1}{3} \\ \mathbb{P}\big((3,\text{preta})\big) = \mathbb{P}\big((3,\text{branca})\big) &= \frac{1}{6} \\ \mathbb{P}\big((1,\text{preta})\big) = \mathbb{P}\big((2,\text{branca})\big) &= 0 \end{aligned}$$

que estendemos para todo $A \subset \Omega$ tomando

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega)$$

de modo que $\mathbb{P}(A) = \mathbb{P}(\text{"meia branca"}) = \mathbb{P}\Big((1, \text{branca}), (2, \text{branca}), (3, \text{branca})\Big) = 1/2$ e

$$\mathbb{P}(E \mid A) = \mathbb{P}(\text{"gaveta 1"} \mid \text{"meia branca"}) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

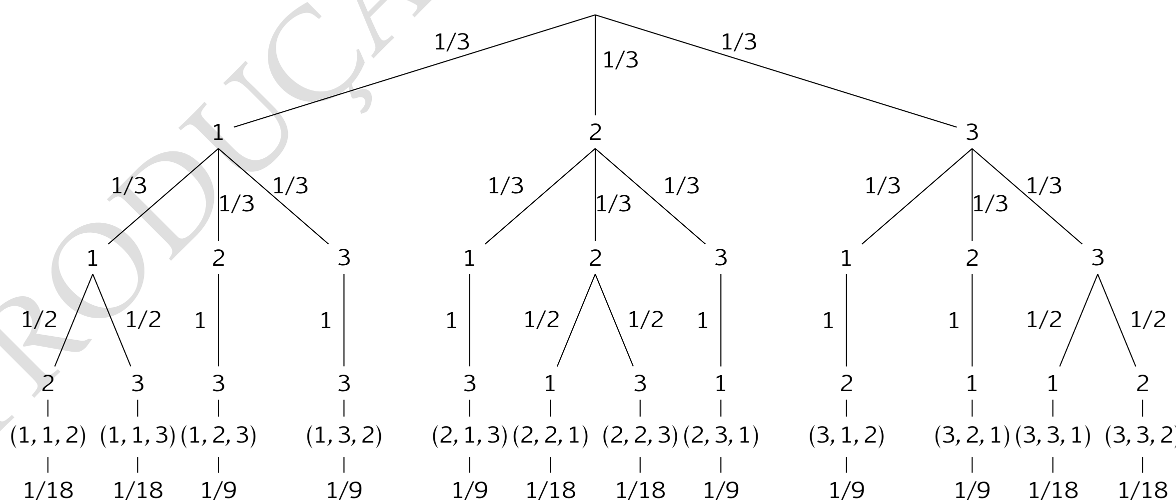
◊

Exercício 58. Verifique que Ω e $\mathbb{P}$ nesse exemplo verificam as propriedades da equação (1.2.2), equação (1.2.3) e equação (1.2.4) na página 43.

Monty Hall: No caso do problema de Monty Hall, consideremos o experimento que consiste das seguintes três etapas

1. o apresentador esconde o carro atrás de uma das portas escolhida com probabilidade 1/3;
2. com probabilidade 1/3, uma porta é escolhida pelo jogador;
3. o apresentador revela, dentre as duas que o jogador não escolheu, aquela que não esconde o carro. Se houver duas possibilidades então o apresentador escolhe uma delas com probabilidade 1/2.

O espaço amostral é definido pelas ternas (porta com carro, escolha inicial, porta revelada) e se a portas estão numeradas por 1,2,3, então $\Omega = \{1, 2, 3\}^3$ e definimos uma medida de probabilidade de acordo com o diagrama de árvore a seguir, os pontos amostrais omitidos têm probabilidade 0



a primeira ramificação corresponde a escolha de porta para esconder o carro, as segundas ramificações correspondem a escolha do jogador e as terceiras ramificações correspondem a escolha de porta para abrir feita pelo apresentador. Estendemos a probabilidade a qualquer evento somando a probabilidade de seus elementos.

Os eventos de interesse são A dado por “o jogador vence trocando de porta” e B dado por “o carro está na porta escolhida inicialmente”

$$\begin{aligned} A &= \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\} \\ B &= \{(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\} \end{aligned}$$

O jogador ganha sem trocar de porta se ocorre B

$$\mathbb{P}(B) = \mathbb{P}(\{(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\}) = \frac{1}{3}$$

e ganha quando troca de porta se ocorre A

$$\mathbb{P}(A) = \mathbb{P}(\{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}) = \frac{2}{3}.$$

Exercício 59. Verifique que Ω e $\mathbb{P}$ nesse exemplo verificam as propriedades na equação (1.2.2), equação (1.2.3) e equação (1.2.4) na página 43.

Exercício 60. Verifique a seguinte igualdade para a regra da multiplicação com três eventos

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

e identifique seu uso na solução para o problema de Monty Hall.

Exercício 61. Consideremos o seguinte experimento composto genérico: executamos um primeiro experimento aleatório cujo modelo probabilístico é $(\Omega_0, \mathcal{A}_0, \mathbb{P}_0)$ em que o espaço amostral tem cardinalidade n. Se o resultado desse primeiro experimento foi $\omega_j \in \Omega_0$, com $1 \leqslant j \leqslant n$, então executamos um experimento aleatório cujo modelo probabilístico (discreto) é $(\Omega_j, \mathcal{A}_j, \mathbb{P}_j)$. Construa um modelo probabilístico que corresponda ao experimento composto.

1.2.14 Independência de eventos: consideremos eventos A e B com probabilidade positiva tais que $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Notemos que isso implica que $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. Se A e B são independentes então $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Vimos, no exercício 41, que o número de placas diferentes para carros no atual modelo com três letras e quatro dígitos é 175.760.000. Assim, se no emplacamento o órgão competente sorteasse um placa, seu carro teria a placa IPE2015 com probabilidade $1/175.760.000 \approx 5{,}7 \times 10^{-9}$. Suponha que num recanto longínquo da galaxia o emplacador tenha apenas uma caixa com as letras do alfabeto e outra caixa com os dez algarismos arábicos. O processo de sortear uma placa é composto pelos processo (1) sortear uma letra da caixa de letras e depois devolvê-la para a caixa; (2) sortear uma letra da caixa de letras e depois devolvê-la; (3) sortear uma letra da caixa de letras e depois devolvê-la; (4) sortear um dígito da caixa de algarismos e depois devolvê-lo; (5) sortear um dígito da caixa de algarismos e depois devolvê-lo; (6) sortear um dígito da caixa de algarismos e depois devolvê-lo; (7) sortear um dígito da caixa de algarismos e depois devolvê-lo. Qual a probabilidade desse experimento composto por sete etapas, cada uma sendo um experimento aleatório, terminar com a placa IPE2015? Nesse caso, um modo de atribuir probabilidade a ponto amostral é considerar o produto das probabilidades dos resultados em cada etapa, assim, IPE2015 tem probabilidade

$$\frac{1}{26}\cdot\frac{1}{26}\cdot\frac{1}{26}\cdot\frac{1}{10}\cdot\frac{1}{10}\cdot\frac{1}{10}\cdot\frac{1}{10}$$

e daí estendemos a probabilidade a qualquer evento somando a probabilidade de seus elementos.

Notemos que, de modo genérico, nesse caso são n etapas $\Omega_1,\dots,\Omega_n$ e a probabilidade

$$\frac{1}{|\Omega_1|}\cdot\frac{1}{|\Omega_2|}\cdots\frac{1}{|\Omega_n|}=\frac{1}{\left|\bigtimes_{i=1}^{n}\Omega_i\right|}$$

coincide com o modelo clássico em $\Omega=\bigtimes_{i=1}^{n}\Omega_i$.

Exercício 62. Verifique que Ω e $\mathbb{P}$ nesse exemplo verificam as propriedades na equação (1.2.2), equação (1.2.3) e equação (1.2.4) na página 43.

1.2.15 Espaços finitos de probabilidade: tendo em vista os últimos exemplos anteriores deve estar claro que a medida uniforme, que trata equiprovavelmente todos os pontos amostrais, não é o único modo de atribuir probabilidades a fenômenos aleatórios.

Seja Ω um conjunto finito. Uma família $\mathcal{A}$ de subconjuntos de Ω é uma álgebra se (a) $\emptyset\in\mathcal{A}$; (b) para quaisquer $A,B\in\mathcal{A}$ vale que $A\cup B\in\mathcal{A}$; (c) para qualquer $A\in\mathcal{A}$ vale que $\overline{A}\in\mathcal{A}$. Notemos que para $A,B\in\mathcal{A}$ temos $A\cap B\in\mathcal{A}$. São exemplos de álgebras de S: $\{\emptyset,S\}$, 2^S, $\{\emptyset,A,\overline{A},S\}$, $\{A\subset S : A \text{ é finito ou } \overline{A} \text{ é finito}\}$.

Uma função real $\mathbb{P}$ definida em $\mathcal{A}$ é uma **medida de probabilidade** se valem a equação (1.2.2), a equação (1.2.3) e a equação (1.2.4), ou seja,

$$\begin{aligned}\mathbb{P}(A)&\geqslant 0,\quad \text{para todo evento } A\\ \mathbb{P}(\Omega)&=1\\ \mathbb{P}(A\uplus B)&=\mathbb{P}(A)+\mathbb{P}(B).\end{aligned}$$

A terna $(\Omega,\mathcal{A},\mathbb{P})$ é um **espaço finito de probabilidade**

Como consequência dessa definição todas as outras propriedades que decorrem das três condições em $\mathbb{P}$ valem, em particular valem $\mathbb{P}(\overline{A})=1-\mathbb{P}(A)$ e as propriedades (i)–(vi) descritas na página 44.

Notemos que no caso particular em que $\mathcal{A}=2^\Omega$, o chamado **espaço de probabilidade discreto**, temos que a medida de probabilidae $\mathbb{P}$ fica determinada pelo valor nos pontos amostrais $\mathbb{P}(\omega)=\mathbb{P}(\{\omega\})$, para todo $\omega\in\Omega$, pois como cada $A\subset\Omega$ é dado pela união disjunto do unitários de seus elementos, segue da aditividade de $\mathbb{P}$ que $\mathbb{P}(A)=\sum_\omega\mathbb{P}(\omega)$.

Exemplo 33 (**lançamento de uma moeda enviesada**). Uma moeda não equilibrada com probabilidade $p \in (0,1)$ de resultar cara é lançada. Um modelo probabilístico para o experimento fica definido por $\Omega = \{\text{Ca}, \text{Co}\}$ com $\mathbb{P}(\text{Ca}) = p$ e $\mathbb{P}(\text{Co}) = 1 - p$. ◊

Exercícios.

1. Proponha um espaço amostral para os seguintes experimentos

   a) Uma moeda é lançada duas vezes.

   b) Um dado e uma moeda são lançados simultaneamente.

   c) Duas cartas são retiradas de um baralho de 52 cartas.

   d) Um pacote de seis cartas numeradas é embaralhado e os números são revelados um a um.

   e) Lançar uma moeda até sair cara.

   f) Uma urna contém 3 bolas, uma vermelha, uma verde e uma azul. Retire uma bola da urna, devolva-a e retire uma segunda bola.

   g) Uma urna contém 3 bolas, uma vermelha, uma verde e uma azul. Retire uma bola da urna e retire uma segunda bola.

2. Sejam A, B e C eventos. Determine expressões para

   a) somente A ocorre;

   b) A e B mas não C ocorrem;

   c) os três eventos ocorrem;

   d) pelo menos um evento ocorre;

   e) pelo menos dois eventos ocorrem;

   f) exatamente um evento ocorre;

   g) exatamente dois eventos ocorrem;

   h) nenhum evento ocorre;

   i) no máximo três eventos ocorrem;

   j) não mais que dois eventos ocorrem.

3. A união de dois eventos $A \cup B$ pode ser escrita como a união de dois eventos mutuamente exclusivos da seguinte forma $A \cup B = A \cup (B \setminus A \cap B))$. Expresse a união de três eventos $A \cup B \cup C$ como a união de três eventos mutuamente exclusivos.

4. Dois dados são lançados. Seja E o evento em que a soma dos dados é impar; seja F o evento em que pelo menos um dos números na face virada para cima seja 1; e seja G o evento em que a soma é 5. Descreva os eventos $E \cap F$, $E \cup F$, $F \cap G$, $E \cap \overline{F}$, e $E \cap F \cap G$.

5. Considere o espaço amostral decorrente de distribuir r bolas em n caixas enumeradas de 1 a n. Esta classe de problemas aparece, por exemplo, na Física Estatística quando é estudada a distribuição de partículas entre estados (que podem ser níveis de energia). Na física estatística dizemos que:

a) as partículas obedecem a estatística de Maxwell-Boltzmann se são distinguíveis.

b) as partículas obedecem a estatística de Fermi-Dirac se são indistinguíveis e se estão sujeitas ao princípio de exclusão de Pauli de no máximo uma partícula por estado.

c) as partículas obedecem a estatística de Bose-Einstein se são indistinguíveis mas não estão sujeitas ao princípio de exclusão de Pauli.

Descreva um espaço amostral para estes modelos de alocação de partículas.

6. Um dado é lançado duas vezes, qual a probabilidade de que o resultado do primeiro lançamento seja maior que o do segundo lançamento?

7. Se uma moeda for lançada sete vezes

   a) qual a probabilidade que não saia nenhuma cara?

   b) Qual a probabilidade que saia 3 caras?

   c) Qual a probabilidade que saia pelo menos 3 caras?

8. Uma *mão* de pôquer consiste de 5 cartas escolhidas aleatoriamente de um baralho tradicional de 52 cartas. Assumindo que todas as possíveis mãos são igualmente prováveis qual é a probabilidade de

   a) um *flush* (as 5 cartas do mesmo naipe); 0,002

   b) um par (exatamente duas cartas do mesmo valor); 0,4226

   c) dois pares; 0,0475

   d) uma trinca (exatamente três cartas do mesmo valor); 0,0211

   e) uma quadra (exatamente quatro cartas do mesmo valor). 0,00024

9. Calcule a probabilidade de que uma mão de 13 cartas contenha um ás e um rei de cada naipe.

10. Calcule a probabilidade de que uma mão de 13 cartas contenha um um grupo de quatro cartas do mesmo valor dos treze valores possíveis?

11. Prove que no modelo probabilístico clássico vale as seguintes propriedades

    a) $\mathbb{P}(\emptyset) = 0$;

    b) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$;

    c) $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$;

    d) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;

12. Uma urna contém 3 bolas vermelhas e 7 pretas. Começando com A, os jogadores A e B retiram uma bola, alternadamente, até que ocorra uma bola vermelha.

Qual é a probabilidade de A retirar uma bola vermelha?

$$\frac{3 \cdot 9! + 7 \cdot 6 \cdot 3 \cdot 7! + (7)_5 5! + (7)_6 3!\, 3}{10!}$$

13. Sete bolas são retiradas ao acaso de uma urna que contém 12 bolas vermelhas, 16 azuis e 18 verdes. Determine a probabilidade de que ocorra

    a) pelo menos 2 vermelhas;

    b) todas da mesma cor;

    c) 3 vermelhas, 2 azuis e 2 verdes;

    d) exatamente 3 vermelhas ou exatamente 3 azuis.

14. Prove que

$$(x+y+z)^n = \sum_{\substack{k_1,\ k_2,\ k_3 \in \mathbb{N} \\ k_1+k_2+k_3=n}} \binom{n}{k_1,\ k_2,\ k_3} x^{k_1} y^{k_2} z^{k_3}.$$

15. Prove que

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{n-k_1-k_2-\cdots-k_{r-1}}{k_r} = \frac{n!}{k_1!\,k_2!\cdots k_r!}.$$

16. Prove por indução os casos gerais dos Princípios Aditivo e Multiplicativo.

17. Uma sala tem 6 portas. De quantas maneiras é possível entrar e sair dessa sala? 36

    De quantas formas é possível entrar e sair da sala por portas distintas? 30

18. Quantos inteiros entre 10000 e 100000 existem cujos dígitos são somente $6, 7$ ou 8? 243 E quantos são os inteiros cujos dígitos são somente $0, 6, 7$ ou 8? 768

19. Quantos inteiros entre 1000 e 9999 (inclusive) existem com todos os dígitos distintos? Desses quantos são ímpares? Desses quantos são pares? 4536,2240,2296

20. Uma carta é escolhida aleatoriamente de um maço de cartas de um baralho tradicional. Com que probabilidade escolhemos

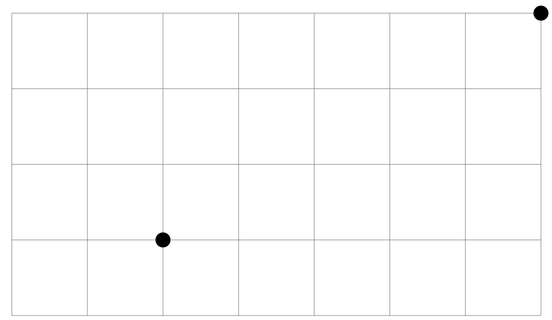
    a) um rei ou uma rainha;

    b) uma carta de copa, ou de ouro, ou de pau;

    c) um número par ou uma espada;

    d) um rei ou uma carta preta.

21. Na figura abaixo, de quantas maneiras pode-se caminhar do ponto $(2, 1)$ até o ponto $(7, 4)$ se cada passo só pode ser para a direita ou para cima (ou seja, a partir do ponto (i, j) só se chega, em um passo, ao $(i + 1, j)$ ou ao $(i, j + 1)$)?



De quantas maneiras pode-se caminhar do ponto $(0, 0)$ até o ponto $(7, 4)$? De quantas maneiras pode-se caminhar do ponto $(0, 0)$ até o ponto $(7, 4)$ passando pelo ponto $(2, 1)$?

22. Prove que para quaisquer inteiros positivos n e k, $n \geqslant 2k$, a fração $n!/2^k$ resulta num inteiro.

23. Numa estante temos 13 livros: 6 de cálculo, 3 de geometria analítica e 4 de probabilidade. De quantas maneiras distintas podemos dispor os livros numa prateleira? E se pusermos as seguintes restrições, quantas são as maniras?

    a) Os livros são dispostos em ordem alfabética do assunto, os de cálculo primeiro, depois os de geometria analítica e por fim os de probabilidade.

    b) Os livros do mesmo assunto fiquem juntos.

    c) Considerando agora os livros do mesmo assunto são iguais, responda novamente cada um dos itens anteriores.

24. Quantas placas de carro podem ser feitas se, ao invés de utilizar 3 letras e 4 números, forem utilizados 2 letras seguidas de 4 números? E se nenhuma letra ou número possa se repetir?

25. Quantos anagramas (permutações distintas das letras) podem ser criados com as letras das palavras: MISSISSIPPI e PROBABILIDADE?

26. Quantas são as funções monótonas[4] $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$? (dica: $x_0 = f(1), x_i = f(i + 1) - f(i), x_n = n - f(n)$, reformule em termos de soluções inteiras de equações lineares)

27. Para jogar uma partida de futebol, 22 crianças dividem-se em dois times de 11 cada. Quantas divisões diferentes são possíveis?

[4] $i < j \Rightarrow f(i) \leq f(j)$

28. Em uma caixa há 100 bolas enumeradas de 1 a 100. Cinco bolas são escolhidas ao acaso. Qual a probabilidade de que os números correspondentes as cinco bolas escolhidas sejam consecutivos?

29. Verifique as identidades

a) $(n+1)_r = \frac{n+1}{n+1-r}(n)_r$　　b) $\binom{n}{r} = \binom{n}{n-r}$　　c) $n\binom{m+n}{m} = (m+1)\binom{m+n}{m+1}$

30. Um *byte* é uma sequência de 8 *bits*. Quantos *bytes* contém

a) exatamente dois 1's;

b) exatamente quatro 1's;

c) exatamente seis 1's;

d) pelo menos seis 1's.

31. Se retirarmos 5 cartas de um baralho, de quantas maneiras distintas podem ocorrer:

a) 5 cartas do mesmo naipe;

b) 4 ás;

c) 3 ás e 2 valetes;

d) 1 par;

e) 1 trinca;

f) 1 trinca e 1 par;

g) 4 do mesmo valor.

32. Um apostador possui 18 fichas e quer aposta-las em 4 cavalos, de modo que a aposta em cada cavalo seja de pelo menos uma ficha, de quantos modo o apostador pode realizar sua aposta?

33. De quantas maneiras podemos

a) distribuir r bolas distintas em n caixas distintas com qualquer número de bolas por caixa;

b) distribuir r bolas distintas em n caixas distintas com no máximo uma bola por caixa;

c) distribuir r bolas idênticas em n caixas distintas com no máximo uma bola por caixa;

d) distribuir r bolas idênticas em n caixas distintas com qualquer número de bolas por caixa.

34. Formule os seguintes problemas em termos de soluções inteiras de equações:

a) O número de maneiras de distribuir r bolas idênticas em n caixas distintas com pelo menos k bolas na primeira caixa.

b) O número de maneiras de distribuir r bolas idênticas em n caixas distintas com nenhuma caixa com mais de duas bolas.

c) O número de subconjuntos de $\{A, B, C, D, E\}$ com 3 elementos.

d) O número de maneiras de distribuir r bolas idênticas em n caixas distintas tal que as duas primeiras caixas tenham juntas p bolas.

35. Formule os seguintes problemas em termos de soluções inteiras de equações e distribuição de bolas em caixas:

    a) Seleção de seis sorvetes a partir de 31 sabores
    b) Seleção de cinco camisas de um grupo de cinco vermelhas, quatro azuis e duas amarelas.
    c) Seleção de 12 cervejas de 4 tipos com pelo menos duas de cada tipo.
    d) Seleção de 20 refrigerantes de 4 tipos com número par de cada tipo e não mais que oito do mesmo tipo.

36. Determine o número de soluções da equação $x_1 + x_2 + x_3 + x_4 + x_5 = 28$ nos casos

    a) $x_i \geqslant 0$;
    b) $x_i > 0$; e
    c) $x_i > i$.

37. para que valores inteiros de n as equações

$$x_1 + x_2 + \cdots + x_{19} = n$$
$$y_1 + y_2 + \cdots + y_{64} = n$$

    têm o mesmo número de soluções inteiras positivas?

38. Sejam n e m inteiros positivos, $m \geqslant n$. Prove que o número de maneiras de distribuir m bolas idênticas em n caixas distinguíveis sem que alguma caixa fique vazia é

$$\binom{m-1}{n-1}.$$

    Prove que se cada caixa tiver que receber pelo menos r objetos, $m \geqslant rn$, então são

$$\binom{m-1+(1-r)n}{n-1}.$$

    maneiras.

39. Uma caixa tem um número desconhecido de bolas idênticas, que denotamos por n, e queremos estimá-lo. Para tal, selecionamos aleatoriamente n_1 bolas, cada uma recebe uma marca e é devolvida para a caixa.

    a) De quantas maneiras podemos extrair r bolas de modo que k estarão marcadas?
    b) Qual é o número de seleções de r bolas.
    c) Defina o índice de acerto por $p_k(n) = \dfrac{\text{resultado do item (a)}}{\text{resultado do item (b)}}$. Mostre que $p_k(n)$ é
        i. crescente para os valores de n tais que $p_k(n) > p_k(n-1)$,

ii. decrescente para os valores de n tais que $p_k(n) < p_k(n-1)$.

Conclua que $\lfloor n_1 r/k \rfloor$ é ponto de máximo de $p_k(n)$.

40. Atribuindo um valor apropriado para as variáveis na expansão binomial de $(x+y)^n$ ou em alguma de suas derivadas, mostre as identidades

    a) $\displaystyle\sum_{k=0}^{n}(-1)^k k\binom{n}{k} = 0.$

    b) $\displaystyle\sum_{k=0}^{n}k(k-1)\binom{n}{k} = n(n-1)2^{n-2}.$

41. Descreva um modelo probabilístico para o seguinte experimento: há três urnas A, B e C com seis bolas em cada uma delas. Na urna A há uma 1 bola branca e 5 azuis, na B há 2 brancas e 4 azuis e na C há 3 de cada. Um dado é lançado, se o resultado foi 1,2, ou 3 uma bola é sorteada da urna A, se o resultado foi 3 ou 4 uma bola é sorteada da urna B, caso contrário é sorteada uma bola da urna C.

42. Três amigos Villa, Laerton e Glauquito quanto saem pra tomar umas cervejas decidem quem paga a conta da seguinte maneira: um saco tem três bolas duas pretas e uma branca. Villa, Laerton e Glauquito retiram uma bola cada, nessa ordem até que saia a bola branca. Com que probabilidade Glauquito fica com a bola branca? 1/3 Esse processo de decisão é honesto, isto é, todos têm a mesma probabilidade de pegar a bola branca?

43. No meu bolso há 3 cartas. Uma é verde em ambas as faces, outra é laranja em ambas as faces, a última tem uma face de cada uma dessas cores. Retiro uma das três cartas escolhida ao acaso e somente um dos lados é exibido; é uma face laranja. Qual é a probabilidade de que o outro lado, que está oculto, seja laranja?

    Um argumento comum diz: "Não pode ser a carta verde-verde. Se for verde-laranja, então o outro lado é verde, enquanto que se for o laranja-laranja, o outro lado é laranja. Uma vez que essas possibilidades são igualmente prováveis, o outro lado é igualmente provável de ser verde laranja."

    Essa conclusão está errada, faça uma análise cuidadosa do problema.

44. Suponha que A e B são eventos de um experimento aleatório. Prove que

    a) (correlação positiva) $\mathbb{P}(A \mid B) > \mathbb{P}(A)$ se e somente se $\mathbb{P}(B \mid A) > \mathbb{P}(B)$ se e somente se $\mathbb{P}(A \cap B) > \mathbb{P}(A)\mathbb{P}(B)$;

    b) (correlação negativa) $\mathbb{P}(A \mid B) < \mathbb{P}(A)$ se e só se $\mathbb{P}(B \mid A) < \mathbb{P}(B)$ se e só se $\mathbb{P}(A \cap B) < \mathbb{P}(A)\mathbb{P}(B)$;

    c) (independência) $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ se e só se $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ se e só se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

45. Uma empresa tem 200 empregados: 120 são mulheres e 80 são homens. Dos 120 trabalhadores do sexo feminino, 30 são classificados como gerentes, enquanto 20 dos 80 funcionários do sexo masculino são gerentes. Suponha que um empregado é escolhido aleatoriamente.

    a) Encontre a probabilidade de que o empregado é do sexo feminino.
    b) Encontre a probabilidade de que o empregado é um gerente.
    c) Encontre a probabilidade condicional de que o empregado é um gerente, uma vez que o empregado é do sexo feminino.
    d) Encontre a probabilidade condicional de que o empregado é do sexo feminino, uma vez que o empregado é um gerente.
    e) São os eventos feminino e gerente positivamente ou negativamente correlacionados ou independentes ?

46. Suponha que A e B são eventos de um experimento aleatório $\mathbb{P}(A) = 1/4$; $\mathbb{P}(B) = 1/3$ e $\mathbb{P}(A \cap B) = 1/10$. Calcule

    a) $\mathbb{P}(A \mid B)$;
    b) $\mathbb{P}(B \mid A)$;
    c) $\mathbb{P}\Big(\overline{A} \;\Big|\; B\Big)$;
    d) $\mathbb{P}\Big(\overline{B} \;\Big|\; A\Big)$;
    e) $\mathbb{P}\Big(\overline{B} \;\Big|\; \overline{A}\Big)$;
    f) $\mathbb{P}(A \cup B)$;
    g) $\mathbb{P}(\overline{A} \cup B)$;

47. Uma urna contém duas moedas: uma moeda comum e uma moeda de duas faces cara. Uma moeda é escolhida ao acaso e lançada duas vezes. Defina os seguintes eventos.

    - A dado por "o primeiro lançamento é cara";
    - B dado por "o segundo lançamento é cara";
    - C dado por "a moeda regular foi escolhida".

    Determine $PCAC$, $\mathbb{P}(B \mid C)$, $\mathbb{P}(A \cap B \mid C)$, $\mathbb{P}(A)$, $\mathbb{P}(B)$, e $\mathbb{P}(A \cap B)$.

    Verifique que A e B não são independentes, mas eles são condicionalmente independentes dado C, isto é, $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C) \cdot \mathbb{P}(B \mid C)$.

1.3 Probabilidade geométrica — o modelo clássico em espaços contínuos.

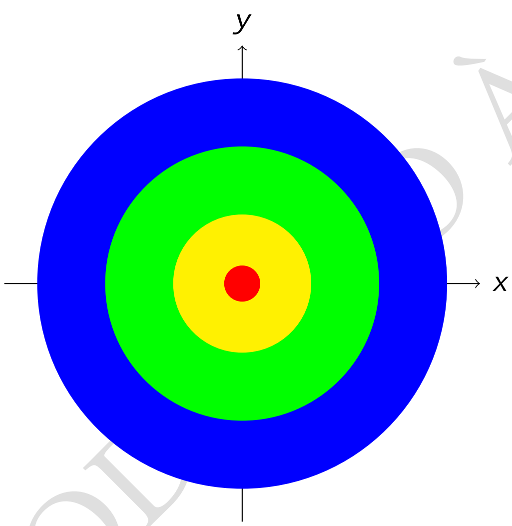
"If we were allowed to rename the field of geometric probability - sometimes already renamed integral geometry - then we would be tempted to choose the oxymoron 'continuous combinatorics."' Daniel Klain, Gian-Carlo Rota, *Introduction to geometric probability*, Lezioni Lincee, cambridge Univ. Press, 1997.

Até aqui consideramos espaços amostrais finitos. A partir de agora estudarmos espaços amostrais contínuos $\Omega \subset \mathbb{R}^n$ referidos por modelos geométricos pois envolvem eventos num espaço euclidiano. No caso de espaço amostral contínuo temos um pouco mais de trabalho pra definir um modelo precisamente. O principal objetivo dessa seção é chamar a atenção para alguns problemas de modelagem probabilística quando tratamos o caso contínuo.

No caso finito a probabilidade de evento é proporcional a cardinalidade do subconjunto que o representa, num certo sentido é a porcentagem do espaço amostral ocupada pelo evento. A tradução desse fenômeno no caso contínuo que, em certo sentido, preserva a noção de eventos elementares equiprováveis, é considerar que a probabilidade de um evento é a "proporção de ocupação" desse evento no espaço amostral; num intervalo da reta ou numa região do plano ou do espaço a probabilidade é proporcional ao volume (comprimento ou área) do subconjunto.

Podemos propor uma medida de probabilidade num intervalo da reta com sendo proporcional ao seu comprimento. Por exemplo, um ponto escolhido aleatoriamente numa corda de 1 metro está a 10 centímetros de um de seus extremos com probabilidade $10/100 + 10/100 = 1/5$. No plano, podemos definir a probabilidade de uma região A proporcional a área de A relativa ao espaço amostral.

Exemplo 34. Consideremos que um dardo acerta aleatoriamente um alvo composto de círculos concêntricos de raios $1/4$, 1, 2 e 3



$$
\begin{aligned}
\Omega &= \left\{(x,y)\colon x^2+y^2 \leqslant 9\right\}\\
\text{Azul} &= \left\{(x,y)\colon 4 < x^2+y^2 \leqslant 9\right\}\\
\text{Verde} &= \left\{(x,y)\colon 1 < x^2+y^2 \leqslant 4\right\}\\
\text{Amarelo} &= \left\{(x,y)\colon \frac{1}{16} < x^2+y^2 \leqslant 1\right\}\\
\text{Vermelho} &= \left\{(x,y)\colon x^2+y^2 \leqslant \frac{1}{16}\right\}
\end{aligned}
$$

A probabilidade de atingir qualquer ponto de uma região A é

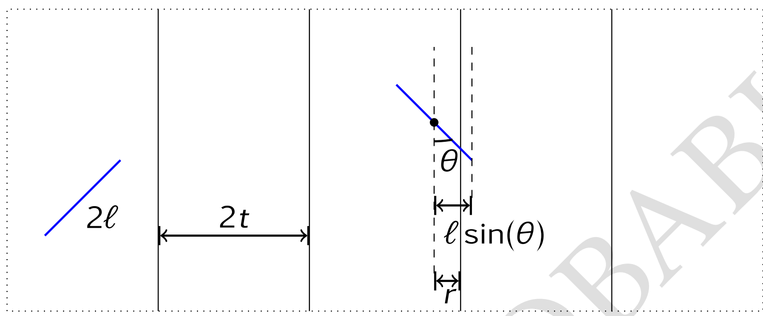
$$\mathbb{P}(A) = \frac{\text{Área}(A)}{\text{Área}(\Omega)} = \frac{\text{Área}(A)}{9\pi}. \tag{1.3.1}$$

Por exemplo, a probabilidade do dardo atingir a região azul é $5/9$, a região verde $1/3$, a região amarela aproximadamente $1/10$ e a probabilidade do dado cair na região vermelha é $\frac{1/16}{9}$ que é, aproximadamente, 0,007. ◊

Exercício 63. Verifique que Ω e $\mathbb{P}$ nesse exemplo verificam as propriedades na equação (1.2.2), equação (1.2.3) e equação (1.2.4) na página 43.

1.3.1 As agulhas de Buffon: atribui-se a Georges-Louis Leclerc, conhecido como Conde de Buffon (1707–1788), o início do estudo da probabilidade geométrica com o conhecido problema das Agulhas de Buffon: Uma agulha cai aleatoriamente num piso com linhas paralelas que distam $2t$ cm, a agulha tem comprimento 2ℓ cm, com $\ell < t$. Qual é a probabilidade de que a agulha irá cruzar uma linha?

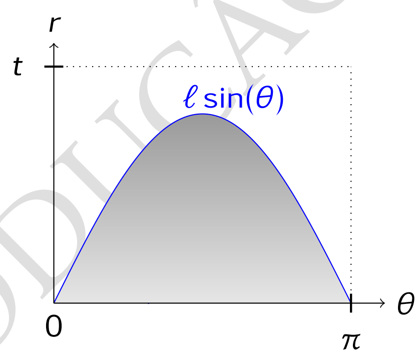
Sejam $r \in [0,t)$ a distância do centro da agulha até a divisória entre tábuas mais próxima, $\theta \in [0,\pi)$ o ângulo com que a agulha cai em relação às linhas paralelas.



O espaço amostral é dado pelos pares $(r,\theta) \in [0,t) \times [0,\pi)$ e os eventos incluem o produto $D \times A$ de intervalos $D = (d_1,d_2) \subset [0,t)$ (que têm área definida); a probabilidade de $E = D \times A \subset [0,t) \times [0,\pi)$ é

$$\mathbb{P}(E) = \frac{\text{Area}(E)}{\text{Area}([0,t) \times [0,\pi))} = \frac{|d_2 - d_1||a_2 - a_1|}{t\pi}.$$

A agulha cruza uma linha paralela se $r < \ell\sin(\theta)$ e a probabilidade de E é a fração $1/t\pi$ da área da região demarcada no gráfico abaixo



$$\mathbb{P}(E) = \frac{1}{t\pi}\int_0^{\pi} \ell\sin\theta \, d\theta = \frac{2\ell}{t\pi}$$

Veja neste link uma simulação usada para determinar o valor aproximado de π.

Exemplo 35. Um palito é quebrado em dois pontos escolhidos aleatoriamente, com que probabilidade as três partes do palito formam um triangulo? Sem perda de generalidade, identificamos o palito com o intervalo $[0,1]$; o espaço amostral é

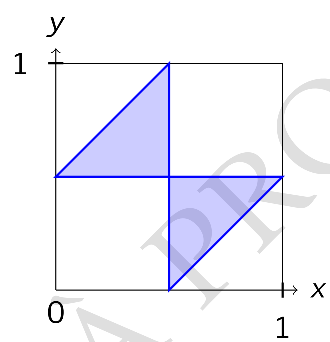
$$\Omega = \{(a,b) : 0 \leqslant a,\ b \leqslant 1\}.$$

As partes têm comprimento a, $b-a$ e $1-b$, caso $b > a$ e têm comprimento b, $a-b$ e $1-a$, caso $b \leqslant a$.

É sabido, da Geometria Euclidiana Plana, que uma terna de números corresponde aos lados de um triângulo se, e somente se, cada um deles é menor que a soma dos outros dois. No caso $a < b$ temos de

$$\begin{aligned} a &< b-a+1-b \\ b-a &< a+1-b \\ 1-b &< a+b-a \end{aligned}$$

que $a < \frac{1}{2}$ e $b-a < \frac{1}{2}$ e $b > \frac{1}{2}$. Analogamente, no caso $b < a$ temos $a > \frac{1}{2}$ e $a-b < \frac{1}{2}$ e $b < \frac{1}{2}$. Cada um desses casos define um triângulo em Ω de área $1/8$, a região azul abaixo é o subconjunto que interessa



portanto o evento "(a,b) define um triângulo" tem probabilidade $1/4$. ◊

1.3.2 Observação (propriedades da probabilidade). *Notemos que como no caso finito valem as propriedades listadas na página 43, equação (1.2.2), equação (1.2.3) e equação (1.2.4), também valem a equação (1.2.5)* $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$ *e a equação equação (1.2.7)* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Um problema dessa abordagem geométrica, a qual é bastante intuitiva, é que *não é possível definir área para todo subconjunto (limitado) do plano*, portanto, alguns subconjuntos não tem uma probabilidade associada[5].

Qual é a área da figura abaixo?

O fato importante para o qual chamamos a atenção é de que

pode ocorrer que nem todo subconjunto admita uma medida de probabilidade

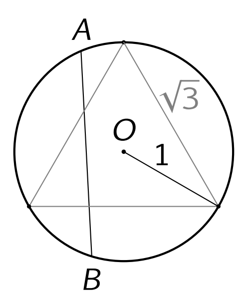
[5]B. R. Gelbaum, J. M. H. Olmsted, *Counterexamples in Analysis*, capítulo 11

quando pedimos que essa medida satisfaça algumas propriedades naturais, satisfeitas por comprimento, área e volume, por exemplo.

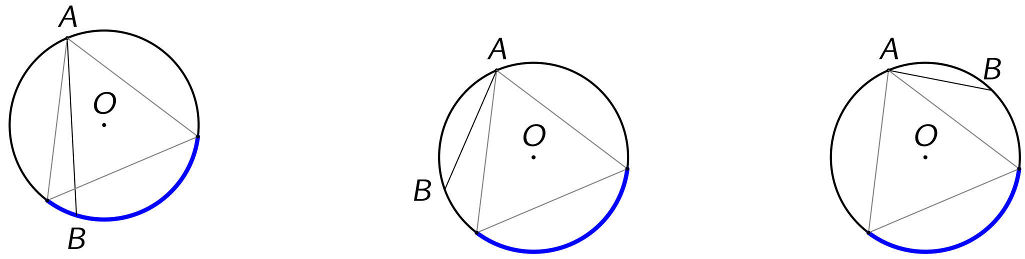
Esse problema só ocorre quando o espaço amostral é infinito e não-enumerável. Há casos em que não é possível definir $\mathbb{P}(E)$ para todo $E \subset \Omega$ quando Ω é infinito e não-enumerável (leia neste link se está curioso, com a advertência de que é, tecnicamente, bastante difícil, ou veja a seção 1.6.7, "Probabilidade uniforme no intervalo $[0,1]$").

1.3.3 Paradoxo de Bertrand: Ao contrário do caso finito, no caso contínuo uma *escolha aleatória* não define unicamente o modelo probabilístico, como exemplifica o fato conhecido como *Paradoxo de Bertrand*, descrito abaixo, no qual três interpretações diferentes para *escolha aleatória* leva a três resultados distintos.

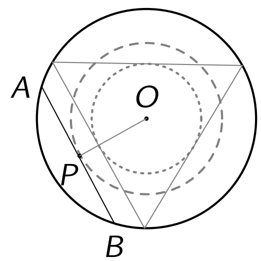
O seguinte problema, conhecido como Paradoxo de Bertrand, que a rigor não é um paradoxo, é passível de mais de uma interpretação para a palavra *aleatório*, ao contrário do caso finito equiprovável. Numa circunferência de raio 1, um triângulo equilátero inscrito tem lado $\sqrt{3}$. Qual é a probabilidade de que uma corda AB escolhida ao acaso tem comprimento maior que $\sqrt{3}$?



1ª interpretação: a escolha da corda é por tomarmos A e B escolhidos dentre os pontos da circunferência. Imaginemos, s.p.g., o triângulo rotacionado de modo que um de seus vértices coincida com o ponto A. A corda tem comprimento maior que o lado do triângulo se B está no arco da circunferência entre os dois outros vértices do triângulo, o que ocorre com probabilidade $1/3$ (os vértices dividem a circunferência em três arcos de mesmo comprimento).

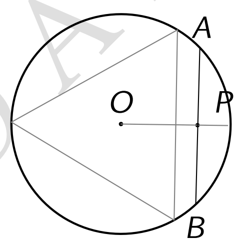


2ª interpretação: a corda é obtido por uma escolha de P no interior da circunferência e AB é a corda cujo ponto médio é P.



A corda é maior que o lado do triângulo se P está no interior da circunferência de centro O e raio $1/2$, o que ocorre com probabilidade $1/4$.

3ª interpretação: Fixamos um raio. A corda é obtida escolhendo um ponto P no raio e tomando a corda que passa por P e perpendicular ao raio.



A corda é maior do que um lado do triângulo, se o ponto escolhido está mais próximo do centro do círculo, que o ponto onde o lado do triângulo intersecta o raio, logo se $|OP| \in (0, 1/2)$ o que ocorre com probabilidade $1/2$.

Exercício 64. Se a é um ponto sorteado no intervalo $(-1, 1)$, com que probabilidade $ax^2 + x + 1 = 0$ tem raízes reais?

1.4 Probabilidade axiomática. Observamos, nas seção anterior, que pode ocorrer que nem todo subconjunto de um espaço amostral contínuo admita uma medida de probabilidade consistente com as propriedades esperadas para tal função, entretanto, eventualmente, tais conjuntos não expressam eventos com interesse, são casos patológicos em certo sentido. Na prática podemos nos preocupar somente com os eventos admitem uma medida de probabilidade.

Uma solução é considerarmos um espaço de eventos $\mathcal{A}$ restrito aos subconjuntos de Ω para quais podemos atribuir uma probabilidade. Da nossa experiência até o momento podemos propor que probabilidade é qualquer função $\mathbb{P}$ definida num espaço de eventos aleatórios $\mathcal{A} \subset 2^{\Omega}$ que seja estável para as operações elementares de conjuntos, isto é, $\mathcal{A}$ é uma *álgebra*[6] de subconjuntos de Ω, que assuma valores reais e que satisfaça

não-negatividade – $\mathbb{P}(A) \geqslant 0$ para todo $A \in \mathcal{A}$,

normalização – $\mathbb{P}(\Omega) = 1$,

aditividade finita – $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$ para todos $A, B \in \mathcal{A}$ disjuntos,

em que por *álgebra* de subconjuntos de Ω entendemos: (i) $\Omega \in \mathcal{A}$, (ii) se $A \in \mathcal{A}$ então $\overline{A} \in \mathcal{A}$ e (iii) se $A, B \in \mathcal{A}$ então $A \cup B \in \mathcal{A}$.

Para corroborar com essa proposta, vejamos outra interpretação (antiga) de probabilidade sem entrar em detalhes: a interpretação frequentista, ou estatística, foi proposta como alternativa à clássica devido a problemas e paradoxos derivados no modelo clássico.

Exemplo 36 (**Probabilidade frequentista**). Se o experimento pode ser repetido potencialmente infinitas vezes, então a probabilidade de um evento pode ser definida através de frequências relativas, isto é, se em n realizações do experimento o evento A ocorreu n_A vezes então $\mathbb{P}(A) \approx n_A/n$.

Por exemplo, se nós lançássemos um dado repetidamente, poderíamos construir um tabela da distribuição de frequência, que mostra quantas vezes cada face veio ocorreu; se a face 6 ocorre em 107 de 600 lançamentos, a frequência relativa desse resultado é $107/600 = 0,178$. Com mais lançamentos espera-se que a proporção de resultados iguais a 6 fique cada vez mais próximo de 1/6. Quanto maior o valor de n melhor é a aproximação da probabilidade pela frequência relativa; intuitivamente $\mathbb{P}(A) = \lim_{n\to\infty} n_A/n$. Nesse caso, pela frequência relativa temos $\mathbb{P}(A) \geqslant 0$, $\mathbb{P}(\Omega) = 1$ e para eventos mutuamente exclusivos $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$.

É conhecido que Buffon (o mesmo que apareceu na seção 1.3) jogou uma moeda de 4.040 vezes. Cara apareceu 2.048 vezes. Pearson jogou moeda 12.000 vezes e 24.000 vezes. Cara ocorreu 6.019 vezes e 12.012 vezes, respectivamente. Para estas três jogadas as frequências relativas de cara são 0,5049, 0,5016 e 0,5005. ◊

Com esses três axiomas temos o bastante pra modelos probabilísticos que resolvem uma grande quantidade de problemas, entretanto, o axioma de aditividade finita é uma limitação para outro tanto de problemas. g

Vejamos um exemplo em que precisamos de um pouco mais do que descrevemos até o momento. Consideremos, novamente, o experimento conceitual de lançar uma moeda infinitas vezes. O espaço

[6] seção 0.6.2, página 32.

amostral do experimento é $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}^{\mathbb{N}}$ e o evento A que nos interessa é dado por "jamais ocorre o resultado cara". Seja A_n representa o evento "não ocorre cara nos n primeiros lançamentos" cuja probabilidade é, pela independência do resultado dos lançamentos, 2^{-n} (convença-se de que essa é a probabilidade do evento, na página 114, exemplo 50, pode ser lida uma justificativa). É natural esperarmos que

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) = 0$$

Esse fato não pode ser demonstrado a partir dos axiomas e para expressar esse fato precisamos de mais um axioma. Antes de enunciar esse axioma, notemos que, no caso do lançamento de moedas descrito acima, a sequência de eventos $(A_n : n \geqslant 1)$ é decrescente, isto é, $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots$ e "jamais ocorre o resultado cara" é dado por $A = \bigcap_{n\geqslant 1} A_n$. O axioma da continuidade da medida de probabilidade é:

continuidade – para qualquer sequência decrescente de eventos $(A_n : n \geqslant 1)$

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

ou

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

escrevendo $\lim_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} A_n$ para uma sequência decrescente de eventos.

Ao invés de tomarmos não-negatividade, normalização, aditividade finita e continuidade como axiomas para uma medida de probabilidade nós (assim como a maioria dos textos) usaremos a equivalência[7]: *aditividade finita juntamente com a continuidade equivalem a aditividade enumerável:*

para qualquer família $(A_i : i \geqslant 1)$ de eventos mutuamente exclusivos $\mathbb{P}\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Nesse caso, o espaço de eventos deve incluir, além de uma álgebra de subconjuntos de Ω, os limites das sequências crescentes e o limite das sequências decrescentes de eventos, ou seja, deve ser uma σ-álgebra (veja a proposição 0.6.8, página 34).

1.4.1 Espaço de eventos: consideramos como espaço de eventos de um espaço amostral Ω qualquer família $\mathcal{A} \subset 2^{\Omega}$ tal que: (i) $\Omega \in \mathcal{A}$, (ii) se $A \in \mathcal{A}$ então $\overline{A} \in \mathcal{A}$ e (iii) se $(A_n : n \geqslant 1) \in \mathcal{A}$ então $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$. Nesse caso, $\mathcal{A}$ é uma *σ-álgebra* de subconjuntos de Ω.

[7]uma demosntração é dada em 1.4.12

1.4.2 Medida de probabilidade: é qualquer função que atribui para cada *evento aleatório* A de um espaço de eventos $\mathcal{A}$ de Ω um número real $\mathbb{P}(A)$ satisfazendo

(P1) não-negatividade $\mathbb{P}(A) \geqslant 0$;

(P2) normalização $\mathbb{P}(\Omega) = 1$;

(P3) σ**-aditividade** $\mathbb{P}\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ para qualquer família $\{A_i \in \mathcal{A} : i \geqslant 1\}$, de eventos mutuamente exclusivos.

Notemos que toda σ-álgebra é uma álgebra de subconjuntos logo se A e B são eventos aleatórios de uma σ-álgebra então também o são: $A \cup B$, $A \cap B$ e $\overline{A}$, assim como a união e a interseção de qualquer quantidade finita de elementos $\mathcal{A}$.

1.4.3 Consequências dos axiomas: algumas consequências desses axiomas são dadas a seguir. A maioria delas são válidas se considerarmos $\mathbb{P}$ finitamente (ao invés de σ) aditiva.

1.4.4 Proposição (probabilidade do evento impossível). *A probabilidade do evento impossível* $\emptyset$ *é* 0.

Demonstração. Escolhendo $A_1 = \Omega$ e $A_i = \emptyset$ para todo $i \geqslant 2$ temos pelo axioma **P3** que

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset \cup \emptyset \cup \cdots \cup \emptyset \cup \cdots) = \mathbb{P}(\Omega) + \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$$

portanto, resta que $\mathbb{P}(\emptyset) = 0$. □

1.4.5 Proposição (probabilidade de uma união finita de eventos disjuntos). *Para quaisquer* $A_1, A_2, \ldots, A_n$ *eventos mutuamente exclusivos*

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

Demonstração. Dados $A_1, A_2, \ldots, A_n$ eventos mutuamente exclusivos, podemos fazer $A_i = \emptyset$ para todo $i > n$ e por **P3**

$$\mathbb{P}\left(\biguplus_{i=1}^{n} A_i\right) = \mathbb{P}\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{i=n+1}^{\infty} \mathbb{P}(\emptyset) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

pois $\mathbb{P}(\emptyset) = 0$. □

1.4.6 Proposição (probabilidade do complemento). *Se* A *é um evento então* $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.

Demonstração. Como A e $\overline{A}$ são mutuamente exclusivos, segue da proposição 1.4.5 que $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \uplus \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A})$. □

1.4.7 Proposição (probabilidade é monótona). *Se A e B são eventos então* $A \subset B \Rightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$.

Demonstração. Sejam A e B eventos tais que $A \subseteq B$. Usamos que B pode ser escrito como $A \uplus (\overline{A} \cap B)$, da proposição 1.4.5 $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(\overline{A} \cap B)$ e como $\mathbb{P}(\overline{A} \cap B) \geqslant 0$ temos $\mathbb{P}(B) \geqslant \mathbb{P}(A)$. □

1.4.8 Corolário. *Para todo evento A, $\mathbb{P}(A) \leqslant 1$.*

Demonstração. Basta notar que $A \subset \Omega$ e usar a proposição anterior. □

1.4.9 Proposição (teorema da inclusão–exclusão). *Se A e B são eventos* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Demonstração. Sejam A e B eventos quaisquer. Vamos novamente recorrer a proposição 1.4.5. A união $A \cup B$ pode ser escrita como a $(A \setminus B) \uplus (B \setminus A) \uplus (A \cap B)$ donde concluímos que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(\overline{A} \cap B). \tag{1.4.1}$$

Agora, A pode ser escrito como $(A \setminus B) \uplus (A \cap B)$ e $B = (B \setminus A) \uplus (A \cap B)$, portanto

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) \\ \mathbb{P}(B) &= \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) \end{aligned}$$

isolando $\mathbb{P}(A \setminus B)$ e $\mathbb{P}(B \setminus A)$ nas equações acima e substituindo na equação (1.4.1) prova a proposição. □

Esse último resultado pode ser generalizado. No caso da união de três eventos a probabilidade pode ser facilmente deduzida da proposição 1.4.9

$$\begin{aligned} \mathbb{P}((A \cup B) \cup C) &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) \\ &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cap C) \cup (B \cap C)) \end{aligned}$$

agora usamos a proposição 1.4.9 nas duas uniões

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

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

que, substituindo na equação anterior, resulta em

$$\mathbb{P}((A \cup B) \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Repetindo essa estratégia podemos estabelecer que

$$\begin{aligned} \mathbb{P}(A \cup B \cup C \cup D) = {} & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(A \cap D) - \mathbb{P}(B \cap C) - \mathbb{P}(B \cap D) \\ & - \mathbb{P}(C \cap D) + \mathbb{P}(A \cap B \cap C) + \mathbb{P}(A \cap B \cap D) + \mathbb{P}(B \cap C \cap D) - \mathbb{P}(A \cap B \cap C \cap D). \end{aligned}$$

Exercício 65. Use a equação acima para responder a pergunta no final da solução do exercício 46.

Exercício 66 (**inclusão–exclusão**). Sejam $A_1, A_2, \dots, A_n$ eventos. Prove que

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n}\mathbb{P}(A_i) - \sum_{1\leqslant i<j\leqslant n}\mathbb{P}(A_i\cap A_j) + \cdots + \\ &+ (-1)^{k+1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}) + \cdots + \\ &+ (-1)^{n+1}\mathbb{P}(A_1\cap A_2\cap\cdots\cap A_n)\end{aligned}$$

A soma

$$(-1)^{k+1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})$$

é feita ao longo dos $\binom{n}{k}$ subconjuntos de k de elementos de $\{1, 2, \dots, n\}$. Escrevemos

$$\boxed{\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})}$$

1.4.10 Proposição (desigualdade de Boole: probabilidade é subaditiva). *Sejam $A_1, A_2, \dots, A_n$ eventos. Então*

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

Demonstração. Se $n = 1$ então $\mathbb{P}(A_1) \leqslant \mathbb{P}(A_1)$, obviamente. Se $n = 2$ então $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$ e como o último termo é positivo $\mathbb{P}(A_1 \cup A_2) \leqslant \mathbb{P}(A_1) + \mathbb{P}(A_2)$.

Sejam $A_1, A_2, \dots, A_n$ eventos e suponhamos, como hipótese do passo indutivo, que

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

Então

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(A_n \cup \bigcup_{i=1}^{n-1} A_i\right) \leqslant \mathbb{P}(A_n) + \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) \leqslant \mathbb{P}(A_n) + \sum_{i=1}^{n-1}\mathbb{P}(A_i) = \sum_{i=1}^{n}\mathbb{P}(A_i)$$

portanto, a desigualdade de Boole segue do Princípio da Indução Finita. □

1.4.11 Proposição (desigualdade de Boole: probabilidade é enumeravelmente subaditiva). *Sejam $A_1, A_2, \dots$ eventos. Então*

$$\boxed{\mathbb{P}\left(\bigcup_{i\geqslant 1} A_i\right) \leqslant \sum_{i\geqslant 1}\mathbb{P}(A_i).}$$

Demonstração. Definimos a seguinte sequência de eventos mutuamente disjuntos: A_1, $A_2 \setminus A_1$, $A_3 \setminus (A_1 \cup A_2), \dots$, de modo geral

$$B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$$

e temos

$$\mathbb{P}\left(\bigcup_{i\geqslant 1} A_i\right) = \mathbb{P}\left(\biguplus_{i\geqslant 1} B_i\right) = \sum_{i\geqslant 1} \mathbb{P}(B_i) \leqslant \sum_{i\geqslant 1} \mathbb{P}(A_i)$$

pois $B_n \subset A_n$ para todo inteiro positivo n. □

1.4.12 A continuidade $\mathbb{P}$ e a σ-aditividade $(*)$: Consideremos uma medida de probabilidade $\mathbb{P}$. Vamos mostrar que essa medida é contínua no sentido dado acima, isto é, para sequências monótonas de eventos, a probabilidade do limite é o limite das probabilidades.

1.4.13 Lema. *Se A_n é uma sequência monótona de eventos então*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right).$$

Se A_n é uma sequência crescente, então o limite pode ser escrito como uma união de eventos disjuntos $\lim_{n\to\infty} A_n = A_1 \uplus (A_2 \setminus A_1) \uplus (A_3 \setminus A_2) \uplus \cdots$ de modo que

$$\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty} A_n\right) &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \cdots \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_{i+1} \setminus A_i) \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n} \left(\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)\right) \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}$$

No caso em que A_n é decrescente tomamos os complementos e temos que $\overline{A_n}$ é crescente, portanto

$$\lim_{n\to\infty} \overline{A_n} = \bigcup_{n\geqslant 1} \overline{A_n} = \overline{\bigcap_{n\geqslant 1} A_n} = \overline{\lim_{n\to\infty} A_n}$$

ademais, do que provamos acima deduzimos

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{A_n}\right) = \lim_{n\to\infty} \mathbb{P}\left(\overline{A_n}\right) = \lim_{n\to\infty} \left(1 - \mathbb{P}\left(\overline{A_n}\right)\right) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n)$$

por outro lado

$$\mathbb{P}\left(\overline{\lim_{n\to\infty} A_n}\right) = 1 - \mathbb{P}\left(\lim_{n\to\infty} A_n\right)$$

portanto $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\lim_{n\to\infty} A_n)$ também vale no caso decrescente. □

De fato, a aditividade enumerável segue da aditividade finita mais a continuidade, ou seja, em termos de axiomas, **não-negatividade, normalização e aditividade enumerável** equivale a **não-negatividade, normalização, aditividade finita e continuidade**.

Suponha que é dada uma função $\mathbb{P}$ definida num espaço de eventos a valores reais não negativos, normalizada, *finitamente* aditiva e contínua no sentido definido acima. Vamos mostrar que $\mathbb{P}\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ para qualquer sequência $(A_i : i \geqslant 1)$ de eventos disjuntos.

Tomemos $B_n := \biguplus_{i=1}^{n} A_i$ para cada $n \geqslant 1$ e temos uma sequência monótona de eventos. Por continuidade

$$\mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \mathbb{P}\left(\bigcup_{n\geqslant 1} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) =$$

mas $\mathbb{P}\left(\bigcup_{n\geqslant 1} B_n\right) = \mathbb{P}\left(\biguplus_{i\geqslant 1} A_i\right)$ e $\mathbb{P}(B_n) = \mathbb{P}\left(\biguplus_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$, com a última igualdade sendo consequência da aditividade finita. De $\mathbb{P}(B_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$, temos

$$\lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i)$$

nessa igualdade, o lado esquerdo vale $\mathbb{P}\left(\biguplus_{i\geqslant 1} A_i\right)$ e o lado direito vale $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

1.4.14 Lemas de Borel–Cantelli (*): Seja A_n, $n \geqslant 1$, uma sequência de eventos. Dado n, o evento "*algum A_k ocorre para $k \geqslant n$*" é

$$\bigcup_{k\geqslant n} A_k$$

e o evento *infinitos A_n ocorrem* (ou, A_n ocorre para infinitos valores de n) é

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k.$$

Por continuidade, já que $B_n := \bigcup_{k\geqslant n} A_k$ é decrescente com n,

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geqslant n} A_k\right)$$

e pela desigualdade de Boole (proposição 1.4.11), supondo que $\sum_k \mathbb{P}(A_k)$ converge

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geqslant n} A_k\right) \leqslant \lim_{n\to\infty} \sum_{k\geqslant n} \mathbb{P}(A_k)$$

mas a convergência assumida acima implica que $\mathbb{P}(A_k) \to 0$ com $k \to \infty$, logo

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

Com isso, provamos o primeiro lema de Borel–Cantelli

1.4.15 Teorema (**Borel–Cantelli**). *Seja A_n, $n \geqslant 1$, uma sequência de eventos aleatórios de um espaço de eventos aleatórios. Se $\sum_n \mathbb{P}(A_n)$ converge, então*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

Usualmente, escrevemos

$$[A_n \text{ infinitas vezes}] := \limsup_{n\to\infty} A_n.$$

Exercício 67. Verifique que

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leqslant \liminf_{n\to\infty} \mathbb{P}(A_n) \leqslant \limsup_{n\to\infty} \mathbb{P}(A_n) \leqslant \mathbb{P}\left(\limsup_{n\to\infty} A_n\right).$$

Para tal, para todo n

1. defina $B_n = \bigcap_{k\geqslant n} A_k$ e prove que
$$\lim_{n\to\infty} B_n = \liminf_{n\to\infty} A_n;$$

2. defina $C_n = \bigcup_{k\geqslant n} A_k$ e prove que
$$\lim_{n\to\infty} C_n = \limsup_{n\to\infty} A_n;$$

3. prove que $B_n \subset A_n \subset C_n$.

Conclua, de 3, a validade das desigualdades.

1.5 Modelo probabilístico.

Um modelo probabilístico para um experimento aleatório consiste de

1. um **espaço amostral** Ω;

2. um **espaço de eventos** $\mathcal{A}$;

3. uma **medida de probabilidade** $\mathbb{P} : \mathcal{A} \to [0, 1]$.

No caso de espaços discretos uma abordagem mais simples é possível para o modelo probabilístico.

1.5.1 Modelo probabilístico discreto: No caso de espaço amostral discreto, todo experimento tem seu modelo probabilístico especificado quando estabelecemos

- o espaço amostral $\Omega = \{\omega_1, \omega_2, \dots\}$
- a função $\mathbb{P}(\omega) = \mathbb{P}(\{\omega\})$ definida para cada evento elementar de modo que:

  (P1') $0 \leqslant \mathbb{P}(\omega) \leqslant 1$, para todo $\omega \in \Omega$, e

  (P2') $\mathbb{P}(\Omega) = \mathbb{P}(\{\omega_1, \omega_2, \dots\}) = \sum_{i=1}^{\infty} \mathbb{P}(\omega_i) = 1$

a qual estendemos para todo $A \subset \Omega$ por

$$\mathbb{P}(A) := \sum_{\omega_i \in A} \mathbb{P}(\omega_i) \tag{1.5.1}$$

Exemplo 37. Uma moeda equilibrada é lançada até sair coroa

- $\Omega = \{\omega_1, \omega_2, \dots, \omega_\infty\}$, onde $\omega_i = (c_1, c_2, \dots, c_i)$ com $c_j = \begin{cases} \text{Co} & \text{se } j = i \\ \text{Ca} & \text{se } 1 \leqslant j < i \end{cases}$ e $\omega_\infty = (\text{Ca}, \text{Ca}, \text{Ca}, \dots)$
- $\mathbb{P}(\omega_i) = \left(\frac{1}{2}\right)^i$ para $i \neq \infty$ e $\mathbb{P}(\omega_\infty) = 0$.

Notemos que

$$\sum_{\omega_i \in \Omega} \mathbb{P}(\omega_i) = \sum_{\omega_i \in \Omega \setminus \{\infty\}} \mathbb{P}(\omega_i) = \sum_{i \geqslant 1} 2^{-i} = 1$$

além disso, $0 \leqslant \mathbb{P}(\omega_i) \leqslant 1$ para todo i. ◊

Exemplo 38. Escolhemos um inteiro positivo ao acaso, a probabilidade de escolher i é $(\frac{1}{2})^i$. Estendemos a probabilidade a qualquer evento A pondo

$$\mathbb{P}(A) = \sum_{a \in A} \mathbb{P}(\{a\}).$$

A probabilidade de escolher um número par é

$$\sum_{\substack{a \in A \\ a \text{ par}}} \mathbb{P}(\{a\}) = \sum_{k \geqslant 1} \left(\frac{1}{2}\right)^{2k} = \sum_{k \geqslant 1} \left(\frac{1}{4}\right)^k = \frac{1}{3}.$$

Portanto, a probabilidade de ocorrer um número ímpar é $2/3$; ◊

Exemplo 39. Um casal é escolhido ao acaso e $(i, j) \in \mathbb{N} \times \mathbb{N}$ representa o número de filhos (i) e o número de filhas (j) do casal. Admitamos que

$$\mathbb{P}(\{(i, j)\}) = \frac{1}{2^{i+j+2}}$$

qual é a probabilidade de um casal não ter filho é a probabilidade de $A = \{(0, j) : j \in \mathbb{N}\}$ e

$$\mathbb{P}(A) = \sum_{j \geqslant 0} \frac{1}{2^{j+2}} = \frac{1}{2}$$

analogamente, a probabilidade de não ter filha é 1/2. A probabilidade de não ter filho nem filha é 1/4. A probabilidade de ter exatamente dois filhos é 3/16. ◊

Exercício 68. Verifique que $\mathbb{P}$ dos exemplos anteriores é uma medida de probabilidade (de acordo com **P1**, **P2** e **P3**). Também, prove que $\mathbb{P}$ como definida em equação (1.5.1), de acordo com **P1'** e **P2'**, é uma medida de probabilidade sobre 2^{Ω} (de acordo com **P1**, **P2** e **P3**).

1.6 Espaço de probabilidade.

probabilidade pode ser estudada do ponto de vista formal/abstrato sem se referir a experimentos e sem que os números associados aos eventos tenham qualquer interpretação. De fato, a probabilidade moderna (axiomática) é uma disciplina matemática que estuda os *espaços de probabilidade* independentemente da natureza dos números que definem a função probabilidade, embora todo modelo probabilístico de um experimento corresponde é um espaço de probabilidades e todo espaço de probabilidades pode ser associado a um modelo probabilístico de algum experimento.

Um *espaço de probabilidade* é uma tripla $(\Omega, \mathcal{E}, \mathbb{P})$ composta por um conjunto Ω, uma σ-álgebra $\mathcal{E}$ de subconjuntos de Ω e uma medida de probabilidade $\mathbb{P} : \mathcal{E} \to \mathbb{R}$, por *medida* entendemos uma função com propriedades especias descritas a seguir.

1.6.1 Espaço de medida (**): um conjunto não vazio M munido de uma σ-álgebra $\mathcal{E}$ é dito *espaço mensurável* e uma medida nesse espaço é uma função $\mu : \mathcal{E} \to [0, +\infty]$ com $\mu(\emptyset) = 0$ e σ-aditiva. A medida μ é *finita* se $\mu(M) < +\infty$. A terna $(M, \mathcal{E}, \mu)$ é um espaço de medida e no caso $\mu(M) = 1$ é um espaço de probabilidade.

Uma dificuldade na construção de espaços de medida é que, em geral, não conseguimos escrever um elemento típico de uma σ-álgebra de modo a explicitar o cômputo da função μ. Por exemplo, um elemento típico da álgebra $\mathcal{A}$ gerada pelo intervalos semiabertos da reta (exemplo 0.6.1, página 31) é uma união finita de intervalos semiabertos disjuntos (proposição 0.6.5, página 33), ou seja, todo $A \in \mathcal{A}$ é da forma

$$A = (a_1, b_1] \uplus (a_2, b_2] \uplus \cdots \uplus (a_n, b_n] \tag{1.6.1}$$

com $-\infty \leqslant a_1 < b_i \leqslant +\infty$ para todo i.

Uma estratégia útil para construir espaços de medida é definirmos uma função (finitamente) aditiva em espaços mais simples, como uma álgebra, e estende-lá para uma medida numa σ-álgebra.

Por exemplo, no caso da semiálgebra $\mathcal{S}$ dos intervalos semiabertos da reta apresentada no exemplo 0.6.1 (pág. 31) tomamos $\ell((a, b]) = b - a$ que é o comprimento (não necessariamente finito) do intervalo semiaberto $(a, b]$. Estendemos aditivamente a função ℓ para a álgebra $\mathcal{A} = \mathfrak{a}(\mathcal{S})$, gerada por $\mathcal{S}$, de modo que para A dado na equação (1.6.1) tenhamos a função aditiva μ_0 dada por

(1.6.2)
$$\begin{aligned}\mu_0(A) &= \ell((a_1, b_1]) + \ell((a_2, b_2]) + \cdots + \ell((a_n, b_n]) \\ &= (b_1 - a_1) + (b_2 - a_2) + \cdots + (b_n - a_n).\end{aligned}$$

Ingenuamente, μ_0 é enumeravelmente aditiva na álgebra $\mathcal{A}$: sejam $I_1, I_2, \ldots$ elementos de $\mathcal{A}$ e suponha que $J = \bigcup_{n\geqslant 1} I_n$ é um elemento de $\mathcal{A}$. Desse último fato, temos que $J = \biguplus_{i=1}^{k} J_i$ com $J_i = (a_i, b_i]$ e como $I_n \in \mathcal{A}$ temos $I_n = \biguplus_{j=1}^{k_n} L_j^{(n)}$, portanto

$$\begin{aligned}\mu_0(J) &= \sum_{i=1}^{k} \ell(J_i) \\ &= \sum_{i=1}^{k} \left(\sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \ell(J_i \cap L_j^{(n)}) \right) \\ &= \sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \sum_{i=1}^{k} \ell(J_i \cap L_j^{(n)}) \\ &= \sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \ell(L_j^{(n)}) \\ &= \sum_{n=1}^{\infty} \mu_0(I_n).\end{aligned}$$

De fato, a função μ_0 definida na álgebra $\mathcal{A}$ pode ser estendida para uma medida $\lambda : \sigma(\mathcal{A}) \to [0, +\infty]$, tal extensão é única e tal medida é chamada de *medida de Lebesgue* na reta.

Extensão de funções: sejam $\mathcal{A}$ e $\mathcal{B}$ duas família de subconjuntos tais que $\mathcal{A} \subset \mathcal{B}$. Dizemos que a função $Q : \mathcal{B} \to \mathbb{R}$ é uma extensão da função $P : \mathcal{A} \to \mathbb{R}$, ou que Q estende P, se

$$P(A) = Q(A) \quad \text{para todo } A \in \mathcal{A}$$

ou seja, as funções coincidem em $\mathcal{A}$.

Dizemos que P é *finita* se $P(\mathcal{A}) < \infty$, que é *aditiva* se

$$P(A \uplus B) = P(A) + P(B) \quad \text{sempre que } A, B \in \mathcal{A}$$

que P é *enumeravelmente aditiva* se

$$P\left(\biguplus_{i\geqslant 1} A_i\right) = \sum_{i\geqslant 1} P(A_i) \quad \text{sempre que } A_i \in \mathcal{A}\ (\forall i) \text{ e } \biguplus_{i\geqslant 1} A_i \in \mathcal{A}.$$

1.6.2 Teorema (**teorema de extensão de Caratheodory**). *Sejam M um conjunto não vazio, $\mathcal{A}$ um álgebra de subconjuntos de M e $P : \mathcal{A} \to [0, +\infty]$ uma função enumeravelmente aditiva com $P(\emptyset) = 0$. Então, existe uma medida $\mu : \sigma(\mathcal{A}) \to [0, +\infty]$ que é uma extensão de P para a σ-álgebra gerada por $\mathcal{A}$. Ademais, se P é finita então a extensão μ é finita e* única. □

Notemos que o teorema, como enunciado acima, não garante que há uma única extensão da equação (1.6.2) para os borelianos da reta. A condição de P ser finita pode ser trocada por uma mais fraca que ainda garante a unicidade da extensão, precisamos apenas exigir que P seja σ-finita: M é coberto por uma família enumerável A_i de elementos da álgebra com $P(A_i) < \infty$. A reta é coberta pelos intervalos $(n, n+1]$, para $n \in \mathbb{Z}$, de comprimento 1 e isso garante a exitência de uma única medida em $\mathbb{R}$ que seja extensão de equação (1.6.2), a medida de Lebesgue.

Garantir a σ-aditividade é a parte difícil na aplicação desse teorema. Lembremos que, em espaço de probablidade, σ-aditividade equivalente a aditividade finita mais continuidade. Geralmente, usamos o teorema de extensão de Caratheodory em conjunto com o que segue.

1.6.3 Proposição. *Sejam M um conjunto não vazio, $\mathcal{A}$ uma álgebra de subconjuntos de M e $P : \mathcal{A} \to [0, +\infty]$ uma função finitamente aditiva com $P(M) = 1$. Se para toda sequência monótona decrescente $(A_n : n \geqslant 1)$ de elementos de $\mathcal{A}$ com $\bigcap_{n \geqslant 1} A_n = \emptyset$ vale*

$$\lim_{n\to\infty} P(A_n) = 0$$

então P é σ-aditiva.

Extensão de função definida em semiálgebra: como fizemos acima no caso da reta, podemos partir de uma semiálgebra, uma estrutura mais fraca na família dos subconjuntos de M, para termos uma medida numa σ-álgebra. Uma função definida na semiálgebra pode ser estendida por aditividade para álgebra gerada (vide proposição 0.6.5), a qual por sua vez, sendo enumeravelmente aditiva, pode ser estendida para uma medida.

1.6.4 Teorema. *Sejam M um conjunto não vazio, $\mathcal{S}$ uma semiálgebra de subconjuntos de M, $P : \mathcal{S} \to [0, 1]$ com $P(M) = 1$ e enumeravelmente aditiva sobre $\mathcal{S}$. Então existe uma σ-álgebra $\mathcal{E} \supset \mathcal{S}$ e uma medida de probabilidade $\mathbb{P}$ sobre $\mathcal{E}$ que estende P.* □

De novo, verificar a aditividade é a parte mais trabalhosa da hipótese e o que precisamos para garanti-lá é: (i) verificar a aditividade finita e (ii) verificar que para toda sequência $(A_n : n \geqslant 1)$ monótona decrescente de elementos de $\mathcal{S}$ com $A_n \to \emptyset$ vale $\mathbb{P}(A_n) \to 0$.

Sobre unicidade da extensão: um resultado bastante útil é o corolário abaixo. Um π-sistema de subconjuntos de S é qualquer família de subconjuntos de S que é fechada para intersecção finita de seus elementos. Por exemplo, $\{(-\infty, x] : x \in \mathbb{R}\}$ é um π-sistema. Notemos que uma semiálgebra é um π-sistema.

1.6.5 Teorema. *Sejam M um conjunto não vazio, $\mathcal{S}$ um π-sistema de subconjuntos de M e $\mathcal{A} := \sigma(\mathcal{S})$. Se $\mathbb{P}$ e $\mathbb{Q}$ são duas medidas finitas com $\mathbb{P}(M) = \mathbb{Q}(M)$ e que coincidem em $\mathcal{S}$ então $\mathbb{P} = \mathbb{Q}$.*

1.6.6 Corolário. *Sejam M um conjunto não vazio, $\mathcal{S}$ uma semiálgebra de subconjuntos de M e $\mathcal{A} := \sigma(\mathcal{S})$. Se $\mathbb{P}$ e $\mathbb{Q}$ são duas medidas de probabilidade que coincidem em $\mathcal{S}$ então $\mathbb{P} = \mathbb{Q}$.*

1.6.7 Probabilidade uniforme no intervalo $[0,1]$ (**): Para definir uma medida de probabilidade em $[0,1]$, "equiprovável" no sentido da probabilidade geométrica, página 76, que estenda equação (1.6.2) usamos o teorema de extensão de Caratheodory enunciado acima, teorema 1.6.2 e a proposição 1.6.3. É mais conveniente definirmos, primeiro, uma medidade de probabilidae em $(0,1]$. Como na equação (1.6.1) e equação (1.6.2), tomamos a álgebra de conjuntos $\mathcal{A}$ consistindo das uniões finitas de intervalos semiabertos disjuntos $A = (a_1, b_1] \uplus \cdots \uplus (a_m, b_m]$ com $0 \leqslant a_1 < b_1 \leqslant a_2 < b_2 \leqslant \ldots \leqslant a_n < b_n \leq 1$ e definimos

$$P(A) := \sum_{i=1}^{n} (b_i - a_i).$$

Temos $P((0,1]) = 1$ e $P(A \uplus B) = P(A) + P(B)$, ou seja, P é finitamente aditiva.

Agora, vamos mostrar que para qualquer sequência monótona decrescente $(A_n : n \geq 1)$ de elementos da álgebra cujo limite é vazio, o limite de $P(A_n)$ é 0 quando $n \to \infty$. Para obternos uma contradição, supomos que $P(A_n) \geqslant 2\epsilon$, para algum $\epsilon > 0$ e todo n numa subsequência dos inteiros positivos, de modo que $P(A_n) \not\to 0$.

É possível provar que $\bigcap_n A_n \neq \emptyset$ para os inteiros n da subsequência declarada acima, o que é uma contradição. A demonstração é bastante técnica e a ideia é como segue. Se

$$A = (a_1, b_1] \cup \cdots \cup (a_m, b_m]$$

é um elemento da álgebra $\mathcal{A}$ então tomamos

$$F := [a_1 + \delta, b_1 - \delta] \cup \cdots \cup [a_m + \delta, b_m - \delta]$$

que está contido em A se escolhemos $\delta > 0$ suficientemente pequeno e, assim, $P(A \setminus F) \leqslant 2\delta m$.

Fazemos isso para cada A_n da subsequência de intereiros tal que $P(A_n) \geqslant 2\epsilon$, ou seja, tomamos $F_n \subset A_n$ com $\delta = \delta(n)$ de modo que $P(A_n \setminus F_n) \leqslant \epsilon 2^{-n}$. Vamos mostrar que $\bigcap_{k \leqslant n} F_k \neq \emptyset$ provando que a medida desse conjunto é positiva:

$$P\left(A_n \setminus \bigcap_{k \leqslant n} F_k\right) \leqslant P\left(\bigcup_{k \leqslant n} A_k \setminus F_k\right) \leqslant \sum_{k \leqslant n} \epsilon 2^{-k} < \epsilon$$

donde temos $P(\bigcap_{k\leqslant n} F_k) > \epsilon$ pois $P(A_n) \geqslant 2\epsilon$, logo $\bigcap_{k\leqslant n} F_k \neq \emptyset$.

Portanto, $\bigcup_{k\leqslant n} \overline{F_k} \neq (0,1]$ donde concluímos que nenhuma união finita de $\overline{F_k}$'s resulta em $(0,1]$.

O ponto final envolve um resultado da Topologia conhecido como teorema de Heine–Borel: *se* $(0,1] = \bigcup_{i\in\Lambda} O_i$ *com* O_i *intervalos abertos então existe* $J \subset \Lambda$ finito *tal que* $(0,1] = \bigcup_{i\in J} O_i$.

Se $\bigcap_n F_n = \emptyset$ para os inteiros n da subsequência declarada acima, então $\overline{\bigcap_n F_n} = \bigcup_n \overline{F_n} = (0,1]$ e o teorema de Heine–Borel implica em $J \subset \mathbb{Z}^+$ finito tal que $\bigcup_{n\in J} \overline{F_n} = (0,1]$, uma contradição.

Finalmente, usando o teorema de extensão de Caratheodory obtemos uma medida de probabilidade $\mathbb{P}$ sobre $((0,1], \mathcal{B}(0,1])$, onde $\mathcal{B}(0,1] = \sigma(\mathcal{A})$ que também é dada pela σ-álgebra gerada pelos abertos de $(0,1]$. Para termos uma medida de probabilidade em $[0,1]$ incluímos o 0 no espaço amostral, tomamos sua probabilidade igual a zero e a σ-álgebra $\mathcal{B}([0,1])$ dos borelianos de $[0,1]$, que também é dada por $\mathcal{B}(0,1] \cup \{A \cup \{0\} : A \in \mathcal{B}(0,1]\}$.

Essa medida de probabilidade uniforme no intervalo $[0,1]$ é a medida de Lebesgue no intervalo $[0,1]$. Existem subconjuntos $A \subset [0,1]$ que não são borelianos e mesmo assim $\mathbb{P}(A)$ está definida. De fato, existe uma σ-álgebra $\mathcal{E}$ que contém $\mathcal{B}([0,1])$ propriamente e uma medida de probabilidade $\mathbb{P}$ tal que

$$\mathbb{P}([a,b]) = \mathbb{P}([a,b)) = \mathbb{P}((a,b]) = \mathbb{P}((a,b)) = b-a, \quad (\forall a,\ 0 \leqslant a \leqslant b \leqslant 1). \tag{1.6.3}$$

Entretanto, não conseguimos estender essa medida para todos os subconjuntos de $[0,1]$.

Prova do fato enunciado na página 79: há casos em que não é possível definir uma medida de probabilidade para todo subconjunto de Ω. No caso em estudo agora, o espaço amostral é o intervalo unitário da reta $[0,1]$, a medida é a uniforme, ou de Lebesgue, como demos acima, em particular, a probabilidade de um intervalo é o comprimento dele. Seja $A \subset [0,1]$ e definamos a r-translação de A

$$A \oplus r := \{a + r : a \in A \text{ e } a + r \leqslant 1\} \cup \{a + r - 1 : a \in A \text{ e } a + r > 1\}$$

para $r \in [0,1]$. Além de que $\mathbb{P}$ satisfazer os axiomas de probabilidade, temos que

$$\mathbb{P}(A \oplus r) = \mathbb{P}(A) \quad (\forall r \in [0,1]) \tag{1.6.4}$$

ou seja, uma translação do conjunto não altera a medida.

1.6.8 Proposição. *Não existe uma medida de probabilidade que esteja definida para todo* $A \subseteq [0,1]$ *e valna equação* (1.6.3) *e equação* (1.6.4).

Demonstração. A prova desse fato é por contradição. Assumamos que exista tal $\mathbb{P}$. Definimos uma relação de equivalência sobre $[0,1]$ pondo $x \sim y$ se, e só se, $x - y \in \mathbb{Q}$. Tomemos H o conjunto dado por um representante de cada classe de equivalência de modo que o representante dos racionais (os quais formam uma classe de equivalência) em não seja o 0.

Para racionais $r_1, r_2 \in \mathbb{Q} \cap [0,1]$ distintos temos $H \oplus r_1 \cap H \oplus r_2 = \emptyset$ pois se x está na interseção então existem $h_1 \in H_1$ e $h_2 \in H_2$ tais que ou $x = h_1 + r_1 = h_2 + r_2$ ou $x = h_1 + r_1 = h_2 + r_2 - 1$ (os outros casos são análogos) portanto $h_1 - h_2 \in Q$, ou seja $h_1 \sim h_2$ e como H tem um representante de cada classe, logo $h_1 = h_2$. Segue que $r_1 = r_2$ e dessa contradição temos $H \oplus r_1 \cap H \oplus r_2 = \emptyset$.

Para todo real $x \in (0,1]$ existe um real $\bar{x} \in H$, que o representa em H, isto é, existe um racional $r_x \in \mathbb{Q} \cap [0,1)$ tal que vale $x - \bar{x} = r_x$ (lembremos que $0 \notin H$), de modo que

$$(0,1] = \biguplus_{r \in \mathbb{Q} \cap [0,1)} H \oplus r$$

e como a união é disjunta e $\mathbb{P}$ probabilidade

$$\mathbb{P}\left((0,1]\right) = \sum_{r \in \mathbb{Q} \cap [0,1)} \mathbb{P}(H \oplus r).$$

Usando 1.6.3 no lado esquerdo e 1.6.4 no lado direito da equação acima

$$1 = \sum_{r \in \mathbb{Q} \cap [0,1)} \mathbb{P}(H)$$

mas o lado direito só pode valer 0, ou ∞ ou $-\infty$, uma contradição. □

Exercícios.

1. Seja $\Omega = \{1,2,3\}$. Descubra condições necessárias e suficientes para os números reais $\alpha = \mathbb{P}(\{1,2\})$, $\beta = \mathbb{P}(\{2,3\})$ e $\gamma = \mathbb{P}(\{1,3\})$ para que $\mathbb{P}$ seja uma medida de probabilidade sobre 2^Ω.

2. Definimos $\mathcal{A}_0$ consistindo do conjunto vazio, dos intervalos $(a,b]$ e de todo conjunto A obtido por união finita de intervalos $A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$ em que $0 \leqslant a_1 < b_1 \leqslant a_2 < b_2 \leqslant \ldots \leqslant a_n < b_n \leq 1$. Prove que $\mathcal{A}_0$ é uma álgebra de subconjuntos de $(0,1]$, isto é, $(0,1] \in \mathcal{A}_0$; $\overline{A} \in \mathcal{A}_0$ para todo $A \in \mathcal{A}_0$; e $A \cup B \in \mathcal{A}_0$ para todos $A \in \mathcal{A}_0$ e $B \in \mathcal{A}_0$.

3. Tome $\Omega = \{1,2,3,4\}$ e descreva a espaço de eventos (σ-álgebra) gerada por $\{\{1\},\{2\}\}$.

4. Sejam $\mathbb{P}$ e $\mathbb{Q}$ medidas de probabilidade sobre o espaço de eventos $\mathcal{A}$ tais que $\mathbb{P}(A) = \mathbb{Q}(A)$ para todo A com $\mathbb{P}(A) \leqslant 1/2$. Prove que $\mathbb{P}(A) = \mathbb{Q}(A)$ para todo $a \in \mathcal{A}$.

5. Sejam $\mathcal{A}$ uma família dos subconjuntos A de $\mathbb{N}$ tal que A é finito ou $\overline{A}$ é finito e defina $\mathbb{P} : \mathcal{A} \to [0,1]$ por $\mathbb{P}(A) = 0$ se A é finito e $\mathbb{P}(A) = 1$ se $\overline{A}$ é finito. É $\mathbb{P}$ finitamente aditiva? É $\mathbb{P}$ σ-aditiva?

6. Um dado equilibrado é lançado até que ocorra um 6. Descreva o modelo probabilístico. Se E_n representa o evento "o dado foi lançado n vezes", descreva os elementos desse evento. O que é o evento $\overline{\bigcup_{i \geqslant 1} E_i}$?

7. Lança-se um par de dados equilibrados até ocorra uma soma 5 ou 7. Descreva o modelo probabilístico. Qual é a probabilidade do resultado 5 ocorrer primeiro?(dica: seja E_n o evento "ocorreu 5 na n-ésima rodada *e* 5 ou 7 não ocorreram nas $n-1$ primeiras rodadas") $0{,}4$

8. Mo jogo de *craps* um jogador lança dois dados. Se a soma for 2, 3, ou 12 o jogador perde e se for 7 ou 11 o jogador ganha; se ocorrer outro resultado o jogador joga novamente até que o primeiro resultado ocorra novamente, caso em que o jogador ganha, ou ocorra um 7, caos em que o jogador perde. Descreva o modelo probabilístico. Determine a probabilidade do jogador ganhar.

9. Numa floresta vivem 20 renas, das quais 5 são capturadas, marcadas e soltas. Algum tempo depois 4 renas são capturadas. Descreva o modelo probabilístico. Qual é a probabilidade de 2 estarem marcadas? $70/323$

10. Há n meias numa gaveta, das quais 3 são vermelhas. Qual é o valor de n se a probabilidade de retirar 2 meias vermelhas da gaveta é 1/2? 4 Descreva o modelo probabilístico.

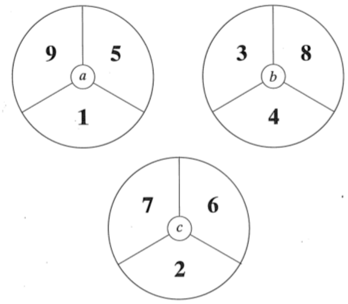
11. Três pessoas fazem registro aleatoriamente num dos 5 hotéis de um cidade, qual é a probabilidade de que se hospedem em hotéis diferentes? $0{,}48$ Descreva o modelo probabilístico.

12. Dois dados equilibrados são lançados n vezes. Descreva o modelo probabilístico. Qual é a probabilidade de que ocorra um duplo 6 pelo menos um vez? Quão grande deve ser n para que a probabilidade seja pelo menos 1/2? $1-(35/36)^n$

13. Se n pessoas, incluindo $Carlos$ e $Felipe$, são dispostas aleatoriamente em linha. Descreva o modelo probabilístico. Qual é a probabilidade de que $Carlos$ esteja ao lado de $Felipe$? E se a formação for um círculo? $2/n;\ 2/(n-1)$

14. Se n bolas são distribuídas aleatoriamente em N caixas, qual é a probabilidade de que m bolas caiam na primeira caixa? $\binom{n}{m}(n-1)^{n-m}/N^n$ Descreva o modelo probabilístico.

15. Dois jogadores A e B jogam o seguinte jogo: A escolhe uma das 3 roletas do desenho abaixo e B escolhe outra das 2 restantes. Então ambos giram a roleta e o qual obtiver o maior resultado é o vencedor. Supondo que as regiões das roletas sejam equiprováveis, qual jogador você preferia ser? Por que?



Jogador B.

Se A escolhe (c), B escolhe (b).

Se A escolhe (b), B escolhe (a).

Se A escolhe (a), B escolhe (c).

Em cada caso, B ganha probabilidade 5/9.

16. Seja $A_1, A_2, \dots$ uma sequência de eventos e defina m função deles uma sequência $B_1, B_2, \dots$ de eventos mutuamente exclusivos tal que

$$\bigcup_{i\geqslant 1} A_i = \bigcup_{i\geqslant 1} B_i.$$

17. Suponha que E seja um evento de um experimento e que o experimento seja executado n vezes. Denote por $n(E)$ o número de vezes que o evento E ocorre nas n repetições e defina $p(E) = \frac{n(E)}{n}$. Mostre que p satisfaz os axiomas **P1**, **P2** e **P3** de medida de probabilidade.

18. Seja $\mathcal{A}$ a família de subconjuntos A de $\mathbb{N}$ tais que o seguinte limite existe

$$\mathbb{P}(A) := \lim_{n\to\infty} \frac{|A \cap \{1, 2, \dots, n\}|}{n}$$

$(\mathbb{N}, \mathcal{A}, \mathbb{P})$ é um modelo probabilístico?

19. Sejam $\Omega \neq \emptyset$, $\mathcal{A} \subset 2^{\Omega}$ e $\omega \in \Omega$. A terna $(\Omega, \mathcal{A}, P)$ com $P(A) = 0$ se $\omega \notin \Omega$ e $P(A) = 1$ se $\omega \in \Omega$ é um modelo probabilístico.

20. Seja $(A_n)_{n\geqslant 1}$ uma sequência de eventos aleatórios de um modelo probabilístico com $\mathbb{P}(A_n) = 1$. Mostre que $\mathbb{P}(\bigcap_{n\geqslant 1} A_n) = 1$.

21. Seja $(A_n)_{n\geqslant 1}$ uma sequência de eventos aleatórios de um modelo probabilístico com $\mathbb{P}(A_n) = 0$. Mostre que $\mathbb{P}(\bigcup_{n\geqslant 1} A_n) = 0$.

22. (**desigualdade de Bonferroni**) Prove que

$$\mathbb{P}\left(\bigcap_{n=1}^{m} A_n\right) \geqslant \sum_{n=1}^{m} \mathbb{P}\left(A_n\right) - (n-1).$$

23. (**desigualdade triangular**) Prove que

$$\mathbb{P}(F \vartriangle G) = \mathbb{P}(F \vartriangle H) + \mathbb{P}(H \vartriangle G), \quad \text{para todo evento } H.$$

24. Uma urna contém bolas k pretas e uma única bola vermelha. Pedro e Paula sorteiam bolas sem reposição alternadamente até que a bola vermelha é extraída. O jogo é ganho pelo jogador que jogar por último. Pedro é um cavalheiro oferece a Paula a opção de escolher se ela quer começar ou não. Paula tem um palpite de que ela tem mais chance se ela começar, afinal, ela pode ter sucesso no primeiro sorteio. Por outro lado, se o seu primeiro sorteio produz uma bola preta, então a chance de Pedro ganhar em seu primeiro sorteio é aumentada, já que a urna atem uma bola preta a menos. Como Paula deve decidir a fim de maximizar sua probabilidade de ganhar?

1.7 Probabilidade condicional.

Retomemos o modelo clássico para motivar a definição de probabilidade condicional, sejam Ω o espaço amostral e $\mathbb{P}$ a medida de probabilidade uniforme. Se A e E são eventos de Ω, com $E \neq \emptyset$, *a probabilidade com que A ocorre dado que E ocorre* é

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

Probabilidade condicional: num modelo probabilístico $(\Omega, \mathcal{A}, \mathbb{P})$ a probabilidade do evento A condicionada ao evento E, em que $\mathbb{P}(E) > 0$, é definida por

$$\boxed{\mathbb{P}(A \mid E) := \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}}$$

e $\mathbb{P}(A \mid E)$ é lido como a probabilidade de A dado E.

Por exemplo, no problema do exercício 43 da página 75 os eventos que interessam são E dado por "a face exibida é laranja" e A dado por "a face não-exibida é laranja". Como metade das faces são laranjas, $\mathbb{P}(E) = 1/2$. Só uma das cartas tem as duas faces laranjas logo $\mathbb{P}(A \cap E) = 1/3$, assim a probabilidade que interessa é

$$\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

1.7.1 Proposição (**teorema da multiplicação**). *Se A e B são eventos de um modelo probabilístico com $\mathbb{P}(A) > 0$ então* $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A)}$. *Analogamente, $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$, se $\mathbb{P}(B) > 0$.* □

Suponhamos que o @alunodaufabc está em dúvida entre matricular-se em Física Quântica ou em Introdução à Probabilidade e à Estatística cujos horários de aula coincidem. Ele estima que a probabilidade de passar com A em Física Quântica é $1/2$ e a de passar com A em Introdução à

Probabilidade e à Estatística é 2/3. Se ele usar uma moeda para escolher a disciplina, qual é a probabilidade de passar com A em IPE? Se A representa "passar com A" e I representa "cursar IPE"

$$\mathbb{P}(I \cap A) = \mathbb{P}(A \mid I)\mathbb{P}(I) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}.$$

Exercício 69. Mostre que para eventos A, B e C vale

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(C \mid A \cap B)\,\mathbb{P}(B \mid A)\,\mathbb{P}(A). \tag{1.7.1}$$

Exercício 70. Sejam $\mathbb{P}$ é uma medida de probabilidade sobre $(\Omega, \mathcal{A})$ e $E \in \mathcal{A}$ um evento com probabilidade positiva. Defina a função $\mathbb{Q} : \mathcal{A} \to \mathbb{R}$ por

$$\mathbb{Q}(A) := \mathbb{P}(A \mid E).$$

Prove que $\mathbb{Q}$ é uma medida de probabilidade para os eventos em $\mathcal{A}$ (i.e, mostre que $\mathbb{Q}$ satisfaz os axiomas de probabilidade a partir do fato de que $\mathbb{P}$ os satisfazem). Verifique, também, que $(E, \{A \cap E : A \in \mathcal{A}\}, \mathbb{Q})$ é um modelo probabilístico.

Como $\mathbb{P}(\cdot \mid E)$ é uma medida de probabilidade todas as propriedades derivadas dos axiomas, como as que são dadas na página 84 valem para $\mathbb{P}(\cdot \mid E)$, como

1. $\mathbb{P}\big(\overline{A} \,\big|\, E\big) = 1 - \mathbb{P}(A \mid E)$
2. $\mathbb{P}(\emptyset \mid E) = 0$
3. $0 \leqslant \mathbb{P}(A \mid E) \leqslant 1$
4. se $A \subset B$ então $\mathbb{P}(A \mid E) \leqslant \mathbb{P}(B \mid E)$
5. $\mathbb{P}(A \cup B \mid E) \leqslant \mathbb{P}(A \mid E) + \mathbb{P}(B \mid E)$
6. $\mathbb{P}(A \cup B \mid E) = \mathbb{P}(A \mid E) + \mathbb{P}(B \mid E) - \mathbb{P}(A \cap B \mid E)$
7. $\mathbb{P}\big(\biguplus_{i=1}^{t} A_i \,\big|\, E\big) = \sum_{i=1}^{t} \mathbb{P}(A_i \mid E)$.
8. $\mathbb{P}\big(\bigcup_{i=1}^{t} A_i \,\big|\, E\big) \leqslant \sum_{i=1}^{t} \mathbb{P}(A_i \mid E)$.
9. $\mathbb{P}\big(\bigcup_{i \geqslant 1} A_i \,\big|\, E\big) \leqslant \sum_{i \geqslant 1} \mathbb{P}(A_i \mid E)$.

1.7.2 Teorema da probabilidade total: observemos que se E e A são eventos, $0 < \mathbb{P}(E) < 1$, então A ocorre se e só se

A e E ocorre

ou

$$A \text{ e } \overline{E} \text{ ocorre}$$

isto é, $A = (A \cap E) \uplus (A \cap \overline{E})$ donde deduzimos

$$\mathbb{P}(A) = \mathbb{P}\Big((A \cap E) \uplus (A \cap \overline{E})\Big) = \mathbb{P}(A \cap E) + \mathbb{P}(A \cap \overline{E}) = \mathbb{P}(A \mid E)\mathbb{P}(E) + \mathbb{P}\Big(A \,\Big|\, \overline{E}\Big)\mathbb{P}(\overline{E}). \tag{1.7.2}$$

Exemplo 40 (**Monty Hall revisitado**). No problema de Monty Hall, se o jogador não troca de porta, ele acerta o carro com probabilidade de acertar uma dentre três portas, i.e., 1/3. Se o jogador troca de porta, então consideramos os eventos A dado por "ganha o carro" e E dado por "acerta na escolha inicial" e, então, $\mathbb{P}(E) = 1/3$, $\mathbb{P}(A \mid E) = 0$ e $\mathbb{P}\Big(A \,\Big|\, \overline{E}\Big) = 1$. Usando a lei da probabilidade total, equação (1.7.2),

$$\mathbb{P}(A) = \mathbb{P}(A \mid E)\mathbb{P}(E) + \mathbb{P}\Big(A \,\Big|\, \overline{E}\Big)\mathbb{P}(\overline{E}) = \frac{2}{3}$$

portanto, é melhor trocar de porta. ◊

1.7.3 Teorema (**teorema da probabilidade total — caso geral**). *Seja* $\{E_1, E_2, \dots, E_n\}$ *uma partição de* Ω *com* $0 < \mathbb{P}(E_i) < 1$ *para todo* i. *Para qualquer evento* A

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(E_i \cap A) = \sum_{i=1}^{n} \mathbb{P}(A \mid E_i)\mathbb{P}(E_i). \tag{1.7.3}$$

Demonstração. Sejam $(\Omega, \mathcal{A}, \mathbb{P})$ um modelo probabilístico e $\{E_1, E_2, \dots, E_n\}$ uma partição de Ω com $0 < \mathbb{P}(E_i) < 1$ para todo i. Se A é um evento aleatório, então para todo $\omega \in A$ existe um único $i \in \{1, 2, \dots, n\}$ tal que $\omega \in E_i$, pois $\omega \in \Omega$ e os E_i's particionam Ω, de modo que $\{A \cap E_1, A \cap E_2, \dots, A \cap E_n\}$ particiona A. Assim, a equação (1.7.3) segue da aditividade e do teorema da multiplicação, proposição 1.7.1. □

Consideremos uma fábrica de sorvete compra galões de leite de 3 fazendas diferentes. Inspeções nas entregas nos fornecem o seguinte:

- A fazenda F_1 entrega todo dia 20% do leite usado e uma inspeção revelou que 20% desse leite estava adulterado.
- A fazenda F_2 entrega 30% do leite usado e a inspeção revelou que 5% estava adulterado.
- A fazenda F_3 entrega 50% do leite usado, 2% adulterado.

Os galões são indistinguíveis. Um é escolhido ao acaso. Qual a probabilidade do evento A definido por "ter leite adulterado"? Dado que a amostra de leite está adulterada, qual é a probabilidade de ter sido escolhida da fazenda F_1?

O espaço amostral Ω representa os galões de leite e pode ser particionado como $\Omega = F_1 \uplus F_2 \uplus F_3$. As inspeções nos fornecem as probabilidades $\mathbb{P}(A \mid F_1) = 0{,}2$, $\mathbb{P}(A \mid F_2) = 0{,}05$ e $\mathbb{P}(A \mid F_3) = 0{,}02$. Usando o teorema da probabilidade total, equação (1.7.3) acima, determinamos

$$\mathbb{P}(A) = \mathbb{P}(A \mid F_1)\mathbb{P}(F_1) + \mathbb{P}(A \mid F_2)\mathbb{P}(F_2) + \mathbb{P}(A \mid F_3)\mathbb{P}(F_3) = 0{,}06504065.$$

Agora, dado que a amostra de leite está adulterada, a probabilidade de ter sido escolhida da fazenda F_1 é

$$\begin{aligned} \mathbb{P}(F_1 \mid A) &= \frac{\mathbb{P}(F_1 \cap A)}{\mathbb{P}(A)} && \text{[por definição]} \\ &= \frac{\mathbb{P}(A \mid F_1)\mathbb{P}(F_1)}{\mathbb{P}(A)} && \text{[pelo teorema do produto]} \\ &= \frac{0{,}2 \cdot 0{,}2}{0{,}06504065} = 0{,}615. \end{aligned} \tag{1.7.4}$$

Exemplo 41 (**urna de Pólya**). Inicialmente, uma urna contém duas bolas, uma branca e uma preta. Em cada instante t, para $t = 1, 2, \dots$, sorteamos uma bola da urna e a devolvemos para a urna junto com uma outra bola da mesma cor dessa sorteada, de modo que o t-ésimo ($t \geqslant 1$) sorteio ocorre com $t+1$ bolas na urna e imediatamente após o t-ésimo sorteio a urna terá $t+2$ bolas.

Seja P_t ($t \geqslant 1$) o evento "a t-ésima bola sorteada é preta"; se não é sorteada uma bola preta então é sorteada uma bola branca, cujo evento é $\overline{P}_t$. Então, pela lei de probabilidade total, como na equação (1.7.2)

$$\mathbb{P}(P_2) = \mathbb{P}(P_2 \mid P_1)\mathbb{P}(P_1) + \mathbb{P}\left(P_2 \,\middle|\, \overline{P}_1\right)\mathbb{P}(\overline{P}_1)$$

e se ocorre P_1, então no instante $t = 1$ (i.e., após o primeiro sorteio e antes do segundo) há 2 bolas pretas dentre 3 bolas, portanto, $\mathbb{P}(P_2 \mid P_1) = 2/3$. Analogamente, $\mathbb{P}\left(P_2 \,\middle|\, \overline{P}_1\right) = 1/3$, de modo que

$$\mathbb{P}(P_2) = \frac{2}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{2} = \mathbb{P}(P_1).$$

Usando a mesma estratégia usada na dedução da equação (1.7.4)

$$\mathbb{P}(P_1 \mid P_2) = \frac{\mathbb{P}(P_2 \mid P_1)\mathbb{P}(P_1)}{\mathbb{P}(P_2)} = \frac{2}{3} = \mathbb{P}(P_2 \mid P_1). \tag{1.7.5}$$

Para computar $\mathbb{P}(P_3)$ precisamos de um pouco mais de esforço. Generalizando o teorema da multiplicação (veja o exercício 69)

$$\mathbb{P}(P_1 \cap P_2 \cap P_3) = \mathbb{P}(P_1)\mathbb{P}(P_2 \mid P_1)\mathbb{P}(P_3 \mid P_1 \cap P_2) = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}.$$

e usando a equação (1.7.3)

$$\begin{aligned} \mathbb{P}(P_3) &= \mathbb{P}(P_1 \cap P_2 \cap P_3) + \mathbb{P}(P_1 \cap \overline{P}_2 \cap P_3) + \mathbb{P}(\overline{P}_1 \cap P_2 \cap P_3) + \mathbb{P}(\overline{P}_1 \cap \overline{P}_2 \cap P_3) \\ &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{2}{4} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{2}{4} + \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

Notemos que $\mathbb{P}(\overline{P}_1 \cap \overline{P}_2 \cap P_3) = \mathbb{P}(P_1 \cap P_2 \cap \overline{P}_3)$ e que $\mathbb{P}(\overline{P}_1 \cap P_2 \cap P_3) = \mathbb{P}(P_1 \cap \overline{P}_2 \cap \overline{P}_3)$ de modo que

$$\mathbb{P}(P_3) = \mathbb{P}(P_1 \cap P_2 \cap P_3) + \mathbb{P}(P_1 \cap \overline{P}_2 \cap P_3) + \mathbb{P}(P_1 \cap P_2 \cap \overline{P}_3) + \mathbb{P}(P_1 \cap \overline{P}_2 \cap \overline{P}_3) = \mathbb{P}(P_1)$$

e tal simetria vale para qualquer n de modo que $\mathbb{P}(P_n) = \mathbb{P}(P_1)$ para todo $n \geqslant 1$. Verificaremos tal fato com mais detalhes a seguir.

No caso do evento P_n, i.e. "a n-ésima bola sorteada é preta", consideremos uma sequência de eventos $X_1, X_2, \ldots, X_{n-1}, P_n$ onde $X_t = P_t$ ou $X_t = \overline{P}_t$, para todo t, $1 \leqslant t < n$. A probabilidade de P_n é a soma das probabilidades

$$\mathbb{P}\left(\left(\bigcap_{t=1}^{n-1} X_t\right) \cap P_n\right) \tag{1.7.6}$$

em que a soma é sobre as 2^{n-1} tais sequências de eventos. Usando o caso geral do teorema da multiplicação, exercício 2 na página 121, a probabilidade na equação (1.7.6) é

$$\mathbb{P}(X_1)\,\mathbb{P}(X_2 \mid X_1)\,\mathbb{P}(X_3 \mid X_1 \cap X_2) \cdots \mathbb{P}(X_{n-1} \mid X_1 \cap \cdots \cap X_{n-2})\,\mathbb{P}(P_n \mid X_1 \cap \cdots \cap X_{n-1}) = \prod_{t=1}^{n} \frac{n_t}{d_t}$$

Os denominadores d_t são fáceis de determinar, $d_t = t+1$, pois em cada sorteio, o número total de bolas aumenta de 1. Sejam $1 \leqslant t_1 < t_2 < \cdots < t_m \leqslant n$ os instantes em que ocorrem sorteio de bolas brancas. Então $n_{t_1} = 1$, $n_{t_2} = 2, \ldots, n_{t_m} = m$. Agora, sejam $1 \leqslant s_1 < s_2 < \cdots < s_{n-m} \leqslant n$ os instantes em que ocorrem sorteio de bolas pretas. Então $n_{s_1} = 1$, $n_{s_2} = 2, \ldots, n_{s_{n-m}} = n-m$. Notemos que o que determina a probabilidade em equação (1.7.6) é quantas ocorrências de P_t há na sequência $X_1, X_2, \ldots, X_{n-1}$, e não em que momento ocorrem, ou seja, a ordem não importa.

Por causa da invariância com respeito a ordem, podemos nos concentrar na probabilidade de ocorrer m sorteios consecutivos de bolas pretas seguidos de $n-m$ sorteios consecutivos de bolas brancas, que é

$$\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{m}{m+1} \cdot \frac{1}{m+2} \cdot \frac{2}{m+3} \cdots \frac{n-m}{n+1} = \frac{m!\,(n-m)!}{(n+1)!} \tag{1.7.7}$$

Agora, usaremos equação (1.7.7) para calcular equação (1.7.6). Para cada natural m, com $m < n$, cada uma das $\binom{n-1}{m}$ sequências $X_1, X_2, \ldots, X_{n-1}, P_n$ com m posições P_t, para $t < n$, têm probabilidade $\mathbb{P}(X_1 \cap X_2 \cap \ldots \cap X_{n-1} \cap P_n)$ dada por equação (1.7.7), portanto

$$\mathbb{P}(P_n) = \sum_{m=0}^{n-1} \binom{n-1}{n} \frac{m!\,(n-m)!}{(n+1)!} = \sum_{m=0}^{n-1} \frac{n-m}{n(n+1)} = \frac{1}{n(n+1)} \sum_{m=1}^{n} m = \frac{1}{2} = \mathbb{P}(P_1)$$

ou seja,

$$\mathbb{P}(P_n) = \mathbb{P}(P_1) \text{ para todo } n \geqslant 1.$$

Decorre daí que $\mathbb{P}(P_j \mid P_k) = \mathbb{P}(P_k \mid P_j)$ pois

$$\mathbb{P}\left(P_j \mid P_k\right) = \frac{\mathbb{P}(P_j \cap P_k)}{\mathbb{P}(P_k)} = \frac{\mathbb{P}\left(P_k \mid P_j\right)\mathbb{P}(P_j)}{\mathbb{P}(P_k)} = \mathbb{P}\left(P_k \mid P_j\right).$$

Ainda, usando equação (1.7.7)

$$\mathbb{P}(\text{há } k \text{ bolas pretas após } n\text{-ésimo sorteio}) = \binom{n}{k}\frac{k!\,(n-k)!}{(n+1)!} = \frac{1}{n+1}.$$

◊

1.7.4 Teorema de Bayes: suponha que *probabilite* é uma doença rara que afeta apenas 0,1% da população e que existe um teste para detectar *probabilite* e que não é perfeito: há 3% de falsos positivos (o teste detecta a doença mas o paciente é são) e 2% de falsos negativos (o teste não detecta a doença no paciente doente). Dado que o teste para um determinado paciente deu positivo, qual é a probabilidade que ele tenha de fato a doença? A lenda diz que a maioria dos médicos respondem 97% sem pestanejar, já que há 3% de falsos positivos. Essa justificativa está errada.

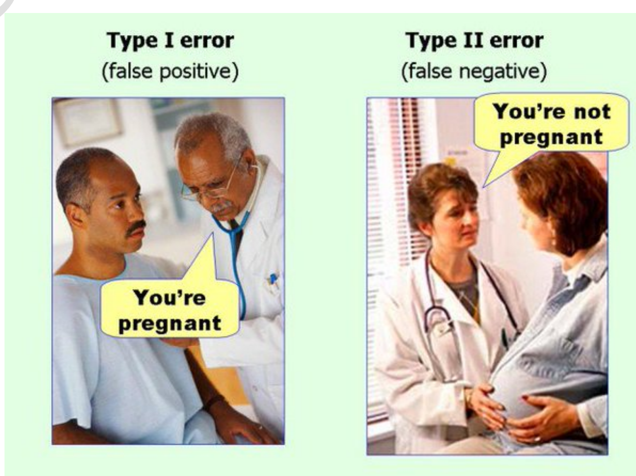
Os eventos de interesse são B dado por "o exame deu positivo" e A dado por "tem *probabilite*". Conhecemos a probabilidade de ter a doença $\mathbb{P}(A) = 0{,}001$, a probabilidade de um falso positivo $\mathbb{P}\left(B \mid \overline{A}\right) = 0{,}03$ e a de um falso negativo $\mathbb{P}\left(\overline{B} \mid A\right) = 0{,}02$. Queremos determinar $\mathbb{P}(A \mid B)$ e da definição de probabilidade condicional, como A e B são eventos com probabilidade positiva

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

Usando o teorema da probabilidade total para o evento B com a partição $\{A, \overline{A}\}$ do espaço amostral e usando a propriedade da probabilidade condicionada do complemento, item 1 em seguida ao exercício 70 na pág. 100, isto é $\mathbb{P}(B \mid A) = 1 - \mathbb{P}\left(\overline{B} \mid A\right)$ e $\mathbb{P}\left(\overline{B} \mid \overline{A}\right) = 1 - \mathbb{P}\left(B \mid \overline{A}\right)$, temos

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}\left(B \mid \overline{A}\right)\mathbb{P}(\overline{A})} = \frac{0{,}98 \cdot 0{,}001}{0{,}98 \cdot 0{,}001 + 0{,}03 \cdot 0{,}999} = 0{,}031664.$$

A chance de ter a doença dado que o teste foi positivo é ligeiramente menor que $3,2\%$.



A dedução feita acima, que é a mesma usada nas equações equação (1.7.4) e equação (1.7.5), pode ser facilmente generalizada para provar o seguinte.

1.7.5 Teorema (Teorema de Bayes). *Sejam $A_1, A_2, \dots, A_m$ eventos que particionam Ω e $B \subset \Omega$ com todos eventos de probabilidade positiva. Para todo $1 \leqslant j \leqslant m$*

$$\boxed{\mathbb{P}\Big(A_j \;\Big|\; B\Big) = \frac{\mathbb{P}\Big(B \;\Big|\; A_j\Big)\mathbb{P}(A_j)}{\sum_{i=1}^{m} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.}$$

Demonstração. A prova é imediata da definição de probabilidade condicional $\mathbb{P}\Big(A_j \;\Big|\; B\Big) = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\mathbb{P}(B)}$ e do teorema da probabilidade total $\mathbb{P}(B) = \sum_{i=1}^{m} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)$. □

Suponhamos que 60% da população adquire a gripe H1N1. Um teste foi desenvolvido para detectar a doença e tem 97% de chance de responder certo. Se uma pessoa tem resultado positivo, então ela pode estar doente ou não. Se o teste foi positivo, com que probabilidade uma pessoa está doente? Denotemos por D o evento das pessoas doentes, por N o seu complemento, o das não-doentes. Denotemos por TP o evento das pessoas que tiverem o teste positivo. Usando o teorema de Bayes

$$\mathbb{P}(D \mid TP) = \frac{\mathbb{P}(TP \mid D)\mathbb{P}(D)}{\mathbb{P}(TP \mid D)\mathbb{P}(D) + \mathbb{P}(TP \mid N)\mathbb{P}(N)} = \frac{0{,}97 \cdot 0{,}6}{0{,}97 \cdot 0{,}6 + 0{,}03 \cdot 0{,}4}$$

logo a probabilidade procurada é $0{,}979798$.

Exemplo 42. Em uma caixa temos três moedas e apenas uma delas é desbalanceada, não sabemos qual. Essa moeda desbalanceada resulta cara com probabilidade 2/3.

De início qualquer uma delas é a moeda desbalanceada com mesma probabilidade. Lançamos uma-a-uma essas moedas e o resultado é (Ca, Ca, Co). O que aprendemos com isso?

Consideremos os eventos E_i definidos por "a i-ésima moeda é a desbalanceada"; logo $\mathbb{P}(E_i) = 1/3$ para todo i. Consideremos o evento B definido por "o resultado dos lançamentos é Ca, Ca, Co".

$$\begin{aligned}
\mathbb{P}(B \mid E_1) &= \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{6} \\
\mathbb{P}(B \mid E_2) &= \frac{1}{2} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{6} \\
\mathbb{P}(B \mid E_3) &= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{12}
\end{aligned}$$

Usando o Teorema de Bayes

$$\mathbb{P}(E_1 \mid B) = \frac{\mathbb{P}(B \mid E_1)\mathbb{P}(E_1)}{\mathbb{P}(B \mid E_1)\mathbb{P}(E_1) + \mathbb{P}(B \mid E_2)\mathbb{P}(E_2) + \mathbb{P}(B \mid E_3)\mathbb{P}(E_3)} = \frac{2}{5}$$

lembremos que $\mathbb{P}(E_1) = 1/3 < \mathbb{P}(E_1 \mid B) = 2/5$, também, $\mathbb{P}(E_2 \mid B) = 2/5$ e $\mathbb{P}(E_3 \mid B) = 1/5$. Logo "aprendemos mais" sobre as moedas depois do lançamento. ◊

1.7.6 Paradoxo de Simpson: é um paradoxo estatístico descrito por Edward Simpson em 1951 e George Udny Yule em 1903, em que o sucesso em grupos parece ser revertida quando os grupos são combinados. Esse resultado é surpreendente e frequentemente encontrado na prática, especialmente nas ciências sociais e as estatísticas médicas. Há vários casos conhecidos e bem estudados, vamos começar com uma simplificação de um caso conhecido.

Em uma universidade P 200 em 1000 homens e 150 em 1000 mulheres escolhem cursar Economia. Na universidade G os números são 30 em 100 homens e 1000 em 4000 mulheres.

Temos as seguintes proporções por sexo

	Homem	Mulher
P	0,2 (200/1000)	0,15 (150/1000)
G	0,3 (30/100)	0,25 (1000/4000)
Ambas	0,21 (230/1100)	0,23 (1150/5000)

e considerando os dois cursos em conjunto a proporção de homens é $\approx 0{,}21$ e a de mulheres 0,23 ou seja, em conjunto a proporção de mulheres é maior, embora, isoladamente, em cada universidade a proporção de homens é maior.

Outro exemplo é de um estudo sobre dois tratamentos H e M para pedras no rim (H e M, não tem relação com sexo). Em função do tamanhos das pedras (pequenas ou grandes) a seguinte tabela mostra o sucesso dos tratamentos

	H	M
Pequenas	0,93 (81/87)	0,87 (234/270)
Grandes	0,73 (192/263)	0,69 (55/80)
Ambos	0,78 (273/350)	0,83 (289/350)

e o mesmo fenômeno ocorre, o tratamento H tem maior taxa de sucesso se consideramos isoladamente as pedras pequenas (0,93) ou as pedras grandes (0,73), no entanto, no quadro geral o tratamento M tem maior taxa de sucesso (0,83).

Em termos gerais, no seguinte esquema genérico para as tabelas acima

	H	$M = \overline{H}$
P	$\mathbb{P}(A \mid P \cap H)$	$\mathbb{P}(A \mid P \cap \overline{H})$
$G = \overline{P}$	$\mathbb{P}\left(A \,\middle|\, \overline{P} \cap H\right)$	$\mathbb{P}\left(A \,\middle|\, \overline{P} \cap \overline{H}\right)$
A	$\mathbb{P}(A \mid H)$	$\mathbb{P}\left(A \,\middle|\, \overline{H}\right)$

o paradoxo ocorre porque as desigualdades

$$\mathbb{P}(A \mid P \cap H) > \mathbb{P}\left(A \,\middle|\, P \cap \overline{H}\right) \text{ e}$$

$$\mathbb{P}\left(A \,\middle|\, \overline{P} \cap H\right) > \mathbb{P}\left(A \,\middle|\, \overline{P} \cap \overline{H}\right) \text{ e}$$

$$\mathbb{P}(A \mid H) < \mathbb{P}(A \mid \overline{H})$$

são compatíveis.

Veja outro exemplo neste link.

1.8 Independência. O conhecimento da ocorrência de um evento B possivelmente afeta a probabilidade de ocorrência de outro evento A, a probabilidade $\mathbb{P}(A \mid B)$ não é, em geral, igual a $\mathbb{P}(A)$. Quando esse não é o caso, isto é $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ dizemos que A é independente de B. Definimos que o evento A é independente do evento B se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{1.8.1}$$

Notemos que se A é independente de B então B é independente de A.

Se os eventos têm probabilidade positiva então decorre da definição acima que a independência de A e B equivale a $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ e $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Por exemplo, se um dado é lançado duas vezes então a probabilidade da soma do dois lançamentos resultar 7 é $6/36 = 1/6$, portanto, a soma ser 7 e o primeiro lançamento resultar 4 tem probabilidade $1/36 = 1/6 \cdot 1/6$, logo são eventos independentes. Por outro lado, a soma ser 5 e o primeiro lançamento resultar 4 tem probabilidade $1/36$, mas a soma ser 5 tem probabilidade $4/36 \neq 1/6 \cdot 1/6$, logo não são eventos independentes.

É imediato da definição o seguinte

1.8.1 Proposição. *Todo evento aleatório A de um modelo probabilístico é independente do evento certo Ω e do evento impossível $\emptyset$.*

1.8.2 Proposição. *Se A é independente de B então A é independente de $\overline{B}$.*

Demonstração. Pelo teorema de probabilidade total $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$ donde deduzimos, usando a independência de A e B, que

$$\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(\overline{B})$$

portanto são eventos independentes. □

Para mais de dois eventos, digamos A, B, C, queremos que A seja independente de B e C se o conhecimento de qualquer informação a respeito da ocorrência de B, de C e de $B \cap C$ não altere a probabilidade de A. Por exemplo, se um dado é lançado duas vezes, sejam A o evento "a soma do dois lançamentos é 7", cuja probabilidade é $1/6$, B o evento "o primeiro lançamento resulta 4" e C o evento "o segundo lançamento resulta 2", ambos B e C têm probabilidade $1/6$, e $B \cap C$ tem probabilidade $1/36$. Então A é independente de B, como já vimos, também, de modo análogo, A é independente de C, entretanto, A não é independente de $B \cap C$, pois $\mathbb{P}(A \mid B \cap C) = 0$. Assim, para A ser independente de B e C devemos ter

$$\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid B) = \mathbb{P}(A \mid C) = \mathbb{P}(A). \tag{1.8.2}$$

Exercício 71. Consideremos a definição acima que A é independente de B se vale a equação equação (1.8.1). Assuma, como definição de "*A é independente de B e C*" se vale a equação equação (1.8.2).

Prove que se A é independente de B e C então A é independente de cada um dos seguintes eventos: $\overline{B}$, $\overline{C}$, $B \cap \overline{C}$, $\overline{B} \cap C$, $\overline{B} \cap \overline{C}$, $\overline{B} \cup \overline{C}$, $B \cup \overline{C}$, $\overline{B} \cup C$, $\emptyset$, Ω.

Em vista do exercício anterior podemos definir que A é independente de B e C se for independente de todo evento do espaço de eventos gerado por $\{B, C\}$ (veja página 34 para a definição de espaço de eventos gerado por uma família de subconjuntos). Ademais, notemos que essa definição é compatível com a definição de "A independente de B" dada anteriormente pois, pela proposição 1.8.1 e proposição 1.8.2, A é independente de todo evento do espaço de eventos gerado por $\{B\}$, o qual é $\{\emptyset, B, \overline{B}, \Omega\}$. Tal definição é equivalente ao enunciado a seguir, que é a que usaremos como definição: *A é independente de B e C se, e somente se, é independente de B, de C e de $B \cap C$*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B \cap C). \tag{1.8.3}$$

Estendemos essa definição para o caso de uma coleção enumerável de eventos.

1.8.3 Independência mútua: uma coleção de eventos $E_1, E_2, \dots, E_n, \dots$ é dita mutuamente independente se para toda subcoleção finita $E_{i_1}, E_{i_2}, \cdots, E_{i_k}$ vale que

$$\mathbb{P}\left(\bigcap_{\ell=1}^{k} E_{i_\ell}\right) = \mathbb{P}(E_{i_1})\mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_\ell}).$$

qualquer que seja $k \in \{2, 3, \dots\}$.

No caso de três eventos, A, B e C são mutuamente independentes se

$$\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A\cap C) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B\cap C) = \mathbb{P}(B)\mathbb{P}(C), \quad \mathbb{P}(A\cap B\cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Por exemplo, consideremos o espaço amostral com 9 elementos dado pelas 6 permutações das letras a, b, c mais as 3 ternas (a, a, a), (b, b, b) e (c, c, c), cada uma com probabilidade $1/9$. Seja E_i o evento "a coordenada i é a". Então $\mathbb{P}(E_i) = 1/3$ e $\mathbb{P}(E_i \cap E_j) = 1/9$ para $i \neq j$, portanto os eventos são 2-a-2 independentes, mas não são independentes pois, por exemplo, a ocorrência de E_1 e E_3 implica na ocorrência de E_2; de fato, $\mathbb{P}(E_1 \cap E_2 \cap E_3) = 1/9$.

Agora, consideremos três lançamentos de uma moeda equilibrada e os eventos E_{12} dado por "o resultado do primeiro e do segundo coincidem"; E_{13} dado por "o resultado do primeira e do terceiro coincidem" e E_{23} dado por "o resultado do segundo e do terceiro coincidem".

$$\mathbb{P}(E_{12}) = \mathbb{P}(E_{13}) = \mathbb{P}(E_{23}) = \frac{1}{2}$$

Os eventos são independentes quando tomados dois-a-dois,

$$\mathbb{P}\left(E_{12} \cap E_{13}\right) = \mathbb{P}\left(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Co}, \mathrm{Co})\}\right) = \frac{1}{4}$$

e, analogamente, para $E_{12} \cap E_{23}$ e $E_{13} \cap E_{23}$. Entretanto, esses eventos não são independentes pois

$$\frac{1}{4} = \mathbb{P}(E_{12} \cap E_{13} \cap E_{23}) \neq \mathbb{P}(E_{12})\mathbb{P}(E_{13})\mathbb{P}(E_{23}) = \frac{1}{8}$$

Exemplo 43. Tomemos $\Omega = \{1,2,3,4,5,6,7,8,9,10\}$ com a probabilidade uniforme (modelo clássico), e consideremos os eventos $A = \{2,3,4,5,6\}$, $B = \{4,5,6,7,8,9\}$, $C = \{2,3,4,5,6,7,8,9,10\}$ e $D = \{1,7,8,9\}$.

Temos as probabilidades $\mathbb{P}(A \mid B) = 1/2 = \mathbb{P}(A)$ e $\mathbb{P}(B \mid A) = 3/5 = \mathbb{P}(B)$. Portanto A e B são eventos independentes. Por outro lado, $\mathbb{P}(A \mid C) = 5/9 > \mathbb{P}(A)$ de modo que A e C não são independentes. Também, $\mathbb{P}(B \mid C) = 2/3 > \mathbb{P}(B)$. Portanto A e C são eventos independentes, assim como B se C. Ainda, $\mathbb{P}(A \mid D) = 0 < \mathbb{P}(A)$ de modo que A e D não são independentes. Também, $\mathbb{P}(B \mid D) = 3/4 > \mathbb{P}(B)$ e $\mathbb{P}(C \mid D) = 3/4 < \mathbb{P}(C)$. ◊

Exemplo 44. Um dado de 4 faces tem uma face de cor Azul, uma face de cor Branca, uma face de cor Cinza, e uma face com as três cores; as faces ocorrem com a mesma probabilidade num lançamento. Representamos a cor que ocorre num lançamento por suas letras iniciais A, B, C. Então $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$ pois cada cor aparece em 2 das 4 faces. Também $\mathbb{P}(A \mid B) = \mathbb{P}(B \mid C) = \mathbb{P}(C \mid A) = \mathbb{P}(B \mid A) = \mathbb{P}(C \mid B) = \mathbb{P}(A \mid C) = 1/2$, portanto, os eventos são 2-a-2 independentes. Entretanto $\mathbb{P}(A \mid B \cap C) = 1$ portanto os eventos não são independentes. ◊

1.8.4 Independência condicional: dizemos que A_1 e A_2 são *condicionalmente independentes* com respeito a B (ou dado B) se

$$\mathbb{P}(A_2 \mid B \cap A_1) = \mathbb{P}(A_2 \mid B) \tag{1.8.4}$$

que equivale a

$$\mathbb{P}(A_2 \cap A_1 \mid B) = \mathbb{P}(A_1 \mid B)\,\mathbb{P}(A_2 \mid B) \tag{1.8.5}$$

Considere uma caixa com duas moedas, uma delas com duas caras e a outra é uma moeda normal. Uma moeda é sorteada e lançada duas vezes. Sejam A e B os eventos "o primeiro lançamento é cara" e "o segundo lançamento é cara", respectivamente. Condicionados a C, dado por "a moeda normal foi a escolhida", os eventos A e B são independentes:

$$\mathbb{P}(A \cap B \mid C) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A \mid C) \cdot \mathbb{P}(B \mid C).$$

Entretanto, A e B não são independentes:

$$\mathbb{P}(A) = \mathbb{P}(A \mid C)\mathbb{P}(C) + \mathbb{P}\left(A \mid \overline{C}\right)\mathbb{P}(\overline{C}) = \frac{1}{2} \cdot \frac{1}{2} + 1\frac{1}{2} = \frac{3}{4};$$

analogamente

$$\mathbb{P}(B) = \mathbb{P}(B \mid C)\mathbb{P}(C) + \mathbb{P}\left(B \mid \overline{C}\right)\mathbb{P}(\overline{C}) = \frac{3}{4}.$$

Agora, $\mathbb{P}(A\cap B) = \mathbb{P}(A \cap B \mid C)\mathbb{P}(C) + \mathbb{P}\left(A \cap B \;\middle|\; \overline{C}\right)\mathbb{P}(\overline{C})$ e podemos usar a independência condicional, usando a equação (1.8.5) para calcular

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C)\mathbb{P}(C) + \mathbb{P}\left(A \;\middle|\; \overline{C}\right)\mathbb{P}\left(B \;\middle|\; \overline{C}\right)\mathbb{P}(\overline{C}) = \frac{5}{8}.$$

Exemplo 45. As seguradoras de automóveis classificam motoristas em *propensos* a acidentes e *não propensos* a acidentes; estimam que os propensos são 30% da população. As estatísticas mostram que os propensos se envolvem em acidente no período de um ano com probabilidade 0,4 e os não propensos com probabilidade 0,2. Seja A o evento definido pelos motoristas propensos. Então, a probabilidade de um novo segurado se envolver em acidente em um ano é, usando o teorema da probabilidade total,

$$\mathbb{P}(A_1) = \mathbb{P}(A_1 \mid A)\mathbb{P}(A) + \mathbb{P}\left(A_1 \;\middle|\; \overline{A}\right)\mathbb{P}(\overline{A}) = 0{,}4 \cdot 0{,}3 + 0{,}2 \cdot 0{,}7 = 0{,}26$$

e se um novo segurado se envolve em acidente nesse prazo, a probabilidade dele ser propenso é, pelo teorema de Bayes,

$$\mathbb{P}(A \mid A_1) = \frac{\mathbb{P}(A_1 \mid A)\mathbb{P}(A)}{\mathbb{P}(A_1)} = \frac{0{,}3 \cdot 0{,}4}{0{,}26} = \frac{6}{13}.$$

Qual a probabilidade de ocorrer um acidente no 2º ano dado que tenha acidentado no 1º ano de contrato? Seja A_2 o evento definido pelos motoristas que se acidentam no 2º ano e assumamos que A_1 e A_2 são *condicionalmente independentes* com respeito a A.

Definimos a medida de probabilidade, sobre os mesmos eventos do modelo probabilístico adotado, $\mathbb{Q}(\cdot) = \mathbb{P}(\cdot \mid A_1)$ (veja o exercício 70). Dessa forma, pelo teorema da probabilidade total

$$\mathbb{Q}(A_2) = \mathbb{Q}(A_2 \mid A)\mathbb{Q}(A) + \mathbb{Q}\left(A_2 \;\middle|\; \overline{A}\right)\mathbb{Q}(\overline{A})$$

onde, por definição,

$$\mathbb{Q}(A_2 \mid A) = \frac{\mathbb{Q}(A_2 \cap A)}{\mathbb{Q}(A)} = \frac{\mathbb{P}(A_2 \cap A \mid A_1)}{\mathbb{P}(A \mid A_1)} = \mathbb{P}(A_2 \mid A \cap A_1).$$

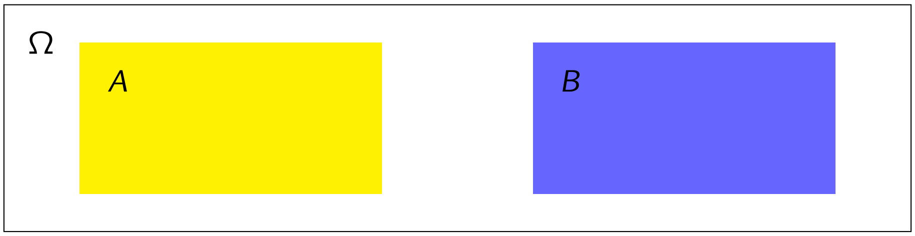
Por hipótese $\mathbb{P}(A_2 \mid A \cap A_1) = \mathbb{P}(A_2 \mid A)$, portanto, $\mathbb{Q}(A_2 \mid A) = 0{,}4$ e temos que

$$\begin{aligned}
&\mathbb{Q}(A_2) = \mathbb{Q}(A_2 \mid A)\mathbb{Q}(A) + \mathbb{Q}\left(A_2 \;\middle|\; \overline{A}\right)\mathbb{Q}(\overline{A}) \Leftrightarrow \\
&\mathbb{P}(A_2 \mid A_1) = \mathbb{P}(A_2 \mid A)\mathbb{P}(A \mid A_1) + \mathbb{P}\left(A_2 \;\middle|\; \overline{A}\right)\mathbb{P}\left(\overline{A} \;\middle|\; A_1\right) \Leftrightarrow \\
&\mathbb{P}(A_2 \mid A_1) = 0{,}4 \cdot \frac{6}{13} + 0{,}2 \cdot \frac{1}{13} = 0{,}29.
\end{aligned}$$

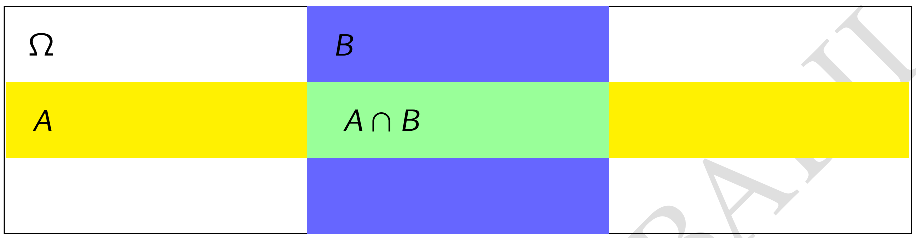
◊

Independência × exclusão: Não confunda esses termos.

Eventos A e B exclusivos mas não independentes:



Eventos A e B independentes mas não exclusivos:



Exemplo 46 (**filtros bayesiano anti-*spam***). A abordagem Bayesiana para filtrar mensagens (uma das ferramentas do spamassassin, por exemplo) foi ideia de Paul Graham que descreve aqui como eles funcionam. Abaixo daremos uma explicação resumida e formal de como usar o teorema de Bayes na classificação de mensagens que são spam. O espaço amostral é dado por $\Omega := \{0{,}1\}^n$ de modo que cada coordenada indica se uma mensagem tem (representado por 1) ou não tem (representado por 0) uma determinada característica e a primeira coordenada de uma n-nupla, especificamente, é 1 se a mensagem é spam e 0 caso, contrário. Assim o evento S dado por $\{1\} \times \{0, 1\}^{n-1}$ representa o evento é spam e $\overline{S} = \{0\} \times \{0, 1\}^{n-1}$ representa o evento não é spam.

Denotamos por C_i, $i \geqslant 2$, o evento "tem a característica i" que corresponde a todos os vetores de Ω cuja i-ésima coordenada é 1.

Vamos assumir que as características $2, 3 \dots, n$ são independentes condicionalmente

$$\mathbb{P}\left(\bigcap_{i \in I} C_i \;\middle|\; S\right) = \prod_{i \in I} \mathbb{P}(C_i \mid S)$$
$$\mathbb{P}\left(\bigcap_{i \in I} C_i \;\middle|\; \overline{S}\right) = \prod_{i \in I} \mathbb{P}\left(C_i \mid \overline{S}\right)$$

para todo $I \subseteq \{2, 3, \dots, n\}$ (pode não ser uma hipótese muito realista. Por exemplo, se a característica 2 é conter a palavra *watch* e a característica 3 é conter a palavra *replica* então um email que contém *replica* tem muito mais chance de conter *watch*; o meu endereço de email recebe várias dessas mensagens que anunciam réplicas de relógios caros e watch só aparece nesses spams; esse pode não ser o caso de um relojoeiro num país de língua inglesa).

Pelo teorema de Bayes

$$\mathbb{P}\left(S \,\middle|\, \bigcap_{i\in I} C_i\right) = \frac{\prod_{i\in I}\mathbb{P}(C_i \mid S)}{\prod_{i\in I}\mathbb{P}(C_i \mid S) + \left(\mathbb{P}(S)/\mathbb{P}(\overline{S})\right)^{|I|-1}\prod_{i\in I}\mathbb{P}\left(C_i \,\middle|\, \overline{S}\right)}. \tag{1.8.6}$$

(use o teorema de Bayes com S, $\overline{S}$ e $\bigcap_i C_i$.)

Agora, suponha que temos 100 mensagens, 50 spams e 50 não spams. O trabalho inicial é definir características das mensagens que são spams das que não são. Por exemplo, nas minhas mensagens muitos dos spams têm a palavra *watch* enquanto que muitos dos não spams têm a palavra *reunião*; a maioria das mensagens têm a palavra *a*, tantos spams quanto não spams, logo *a* não deve ser uma característica levada em consideração.

Para a característica i, seja k_i a quantidade de spams que têm a característica i e seja ℓ_i a quantidade de não spams que têm a característica i; definimos $\mathbb{P}(C_i \mid S) = \frac{k_i}{50}$ e $\mathbb{P}\left(C_i \,\middle|\, \overline{S}\right) = \frac{\ell_i}{50}$ e pelo teorema de Bayes,

$$\mathbb{P}(S \mid C_i) = \frac{\mathbb{P}(C_i \mid S)\mathbb{P}(S)}{\mathbb{P}(C_i \mid S)\mathbb{P}(S) + \mathbb{P}\left(C_i \,\middle|\, \overline{S}\right)\mathbb{P}(\overline{S})} = \frac{k_i\mathbb{P}(S)}{k_i\mathbb{P}(S) + \ell_i\mathbb{P}(\overline{S})}.$$

Recebida uma mensagem, filtramo-lá para determinar suas características, digamos C_i com $i \in I$, e equação (1.8.6) determina a probabilidade dessa mensagem ser spam. Se essa probabilidade estiver abaixo de um limiar pré-fixado, a mensagem não é classificada com *spam*, caso contrário, é classificada como *spam*. ◊

Exemplo 47. Com que probabilidade uma máquina que está funcionando no instante t_0 continuará funcionando até o instante $t_0 + t$ supondo que essa probabilidade dependa somente do comprimento do intervalo $(t_0, t_0 + t)$, supondo que a probabilidade com que a máquina pare durante um intervalo de comprimento Δt proporcional a Δt e supondo que os eventos que consistem da máquina parar durante intervalos disjuntos são independentes?

Se $p(t)$ é a probabilidade procurada. A probabilidade da máquina parar num intervalo de tempo de tamanho Δt é $1 - p(\Delta t) = \alpha\Delta t$, para alguma constante fixa $\alpha > 0$. A probabilidade com que uma máquina que está funcionando no instante t_0 continuará funcionando até o instante $t_0 + t + \Delta t$ é $p(t+\Delta t)$ e pela independência assumida $p(t+\Delta t) = p(t)\cdot p(\Delta t)$ logo $p(t+\Delta t) = p(t)(1-\alpha\Delta t)$ de modo que

$$\frac{p(t+\Delta t) - p(t)}{\Delta t} = -\alpha p(t)$$

que, passando ao limite quando $\Delta t \to 0$, resulta em

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t) = -\alpha p(t)$$

donde temos $p(t) = Ce^{-\alpha t}$ para alguma constante C. De $p(0) = 1$ deduzimos que $C = 1$. ◊

Exemplo 48. Faltando um minuto para meia-noite colocamos numa urna bolas numeradas de 1 a 10 e retiramos a bola 10. Faltando meio minuto para meia-noite colocamos na urna bolas numeradas de 10 a 20 e retiramos a bola 20. E assim por diante. Quando for meia-noite, todas as bolas numeradas com inteiro positivo, exceto as múltiplos de 10, estarão na urna.

Agora, faltando um minuto para meia-noite colocamos numa urna bolas numeradas de 1 a 10 e retiramos a bola 1. Faltando meio minuto para meia-noite colocamos na urna bolas numeradas de 10 a 20 e retiramos a bola 2. E assim por diante. Quando for meia-noite, nenhuma bola estará na urna.

Na variante aleatória, faltando um minuto para meia-noite colocamos numa urna bolas numeradas de 1 a 10 e sorteamos uma bola. Faltando meio minuto para meia-noite colocamos na urna bolas numeradas de 10 a 20 e sorteamos outra bola. E assim por diante. Quando for meia-noite, o que restará na urna? O espaço amostral consiste das sequência de inteiros distintos, com primeiro elemento entre 1 e 10, segundo entre 11 e 20, e assim por diante. Seja E_n o evento "a bola 1 está na urna após n passos"

$$\mathbb{P}(E_n) = \frac{9}{10}\cdot\frac{18}{19}\cdot\frac{27}{28}\cdots\frac{9n}{9n+1} = \prod_{i=1}^{n}\frac{9i}{9i+1}.$$

Os eventos E_n são monótonos decrescentes

$$\mathbb{P}\left(\bigcap_n E_n\right) = \lim_{n\to\infty}\mathbb{P}(E_n) = \lim_{n\to\infty}\left(\left(1+\frac{1}{9}\right)\left(1+\frac{1}{18}\right)\cdots\right)^{-1} \leqslant 0$$

assim com probabilidade 0 a bola 1 (assim como qualquer outra) estará na urna a meia-noite. Se D_n representa o evento "a n-ésima bola está na caixa a meia noite"

$$\mathbb{P}\left(\bigcup_{i\geqslant 1} D_n\right) \leqslant \sum_{i\geqslant 1}\mathbb{P}(D_n) = 0$$

ou seja, a urna estará vazia com probabilidade 1. ◊

1.8.5 Repetições independentes de um experimento: no caso de um experimento aleatório com modelo probabilístico *discreto* $(\Omega, \mathcal{A}, \mathbb{P})$ vamos modelar o experimento composto que corresponde à execução repetida sequencialmente n vezes do experimento original sob condições idênticas, com resultados mutuamente independentes. Tomemos $\Omega = \{\omega_1, \omega_2, \dots\}$ com $p_i := \mathbb{P}(\omega_i)$. Um resultado do experimento composto é uma n-upla $\omega = (\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n}) \in \Omega^n$. Esse espaço amostral é discreto portanto basta-nos definir as probabilidades dos pontos amostrais $\omega \in \Omega^n$. Cada evento elementar $\omega = (\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n})$ é dado pela interseção de n eventos, as saber os eventos "a primeira coordenada é ω_{i_1}", "a segunda coordenada é ω_{i_2}", ..., "a última coordenada é ω_{i_n}", portanto, $\mathbb{Q}(\omega) = p_{i_1}p_{i_2}\cdots p_{i_n}$ é uma medida de probabilidade sobre Ω^n (verifique).

Exercício 72. No contexto descrito acima, sejam $A_1, A_2, \dots, A_n$ eventos tais que, para todo i, existe $B_i \subset \Omega$ tal que temos

$$A_i = \Omega_1 \times \cdots \times \Omega_{i-1} \times B_i \times \Omega_{i+1} \times \cdots \times \Omega_n.$$

Prove que os eventos A_i's são mutuamente independentes.

Exemplo 49. Consideremos uma moeda com probabilidade de p de resultar em cara. Em três lançamentos independentes a probabilidade de só resultar em cara é p^3 e a probabilidade de só resultar em coroa é $(1-p)^3$. Em três lançamentos independentes temos $\binom{3}{1} = 3$ resultados distintos com 1 ocorrência de cara e 2 ocorrências de coroa, cada um tem probabilidade $p(1-p)^2$, portanto, a probabilidade de em três lançamentos independentes termos 1 ocorrência de cara é $\binom{3}{1}p(1-p)^2$. Em três lançamentos independentes temos $\binom{3}{2} = 3$ resultados distintos com 2 ocorrências de cara e 1 ocorrência de coroa, cada um tem probabilidade $p^2(1-p)$, portanto, a probabilidade de em três lançamentos independentes termos 2 ocorrências de cara é $\binom{3}{2}p^2(1-p)$.

De um modo geral, em $n \geqslant 1$ lançamentos independentes da moeda viciada, temos $\binom{n}{k}$ resultados distintos com k ocorrências de cara e $n-k$ ocorrências de coroa, cada um desses resultados tem probabilidade $p^k(1-p)^{n-k}$, portanto, a probabilidade de em n lançamentos independentes termos k ocorrências de cara é $\binom{n}{k}p^k(1-p)^{n-k}$. ◊

A seguir apresentamos um exemplo com espaço de eventos contínuo. Embora não tenhamos construído o modelo probabilístico do exemplo, podemos calcular algumas probabilidades de eventos aleatórios usando a independência.

Exemplo 50. Consideremos uma sequência infinita de lançamentos de uma moeda equilibrada, cada lançamento independente dos outros. Se E_i é o evento, "cara não-ocorre na i-ésima tentativa", então a probabilidade de B_n definido por "cara não ocorre nas n primeiras realizações" é $\mathbb{P}(B_n) = \mathbb{P}(E_1)\mathbb{P}(E_2)\cdots\mathbb{P}(E_n) = (1/2)^n$ por causa da independência das realizações. Portanto, pelo menos uma ocorrência de cara nas n primeiras realizações tem probabilidade $1-(1/2)^n$. Notemos que $B_1 \supset B_2 \supset B_3 \supset \cdots$ logo

$$\mathbb{P}\left(\bigcap_{n\geqslant 1} B_n\right) = \mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} (1/2)^n = 0$$

portanto, cara certamente ocorre.

Qual a probabilidade de ocorrerem k caras nas n primeiras realizações do experimento? Qual a probabilidade de todas as realizações resultarem cara? ◊

1.8.6 Espaço produto ($*$) : se temos dois experimentos com modelos $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ e $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ e que ocorrem de modo independente então podemos definir um modelo probabilístico para a realização conjunta e independente desses experimentos da seguinte forma.

O espaço amostral é dado pelos pares (ω_1, ω_2), i.e., o espaço amostral é $\Omega_1 \times \Omega_2$. Os eventos aleatórios são difíceis de definir formalmente, não é simplesmente $\mathcal{A}_1 \times \mathcal{A}_2$, mas sim o menor espaço de eventos que contém todos os produtos de eventos $A_1 \times A_2$ ou, equivalentemente, a σ-álgebra gerada por todos os produtos de eventos $A_1 \times A_2 \in \mathcal{A}_1 \times \mathcal{A}_2$. No produto definimos a medida de probabilidade $\mathbb{P}(A_1 \times A_2) := \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$. O espaço de probabilidade $(\Omega_1 \times \Omega_2, \sigma(\mathcal{A}_1 \times \mathcal{A}_2), \mathbb{P})$ é chamado de espaço produto.

Isso pode ser estendido para o produto enumerável de espaços. Sejam $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ modelos probabilísticos, para $i = 1, 2, \dots$. Um retângulo mensurável é um conjunto da forma

$$\bigtimes_{i=1}^{\infty} A_i \text{ tal que } \begin{cases} A_i \in \mathcal{A}_i \text{ para uma quantidade finita de índices } i \\ A_i = \Omega_i \text{ para o restante dos índices } i. \end{cases}$$

O espaço de eventos é a σ-álgebra $\mathcal{A}$ gerada pelos retângulos mensuráveis (a qual coincide com a σ-álgebra gerada pela álgebra dos cilindros). Em $\mathcal{A}$ existe uma única medida de probabilidade $\mathbb{P}$ tal que nos retângulos mensuráveis $\bigtimes_{i=1}^{\infty} A_i$ vale

$$\mathbb{P}\left(\bigtimes_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{n} \mathbb{P}_i(A_i)$$

com $n = \max\{i : A_i \neq \Omega_i\}$. Essa medida é a medida-produto.

1.8.7 Eventos cilíndricos (**): para definir um modelo probabilístico para o experimento do exemplo 50, vamos construir um espaço de eventos do experimento. Representemos cara por 1 e coroa por 0 de modo que $\Omega = \{0, 1\}^{\mathbb{N}}$.

Um cilindro de ordem $n \in \mathbb{Z}^+$ de Ω é o *evento* obtido fixando um número finito de coordenadas: para $(a_0, \dots, a_{n-1}) \in \{0, 1\}^n$

$$C_{a_0,\dots,a_{n-1}} := \Big\{(\omega_0, \omega_1, \dots) \in \Omega : \omega_i = a_i \text{ para todo } i,\ 0 \leqslant i \leqslant n-1\Big\} \tag{1.8.7}$$

ao qual queremos atribuir probabilidade $(1/2)^n$. A família

$$\mathcal{S} := \Big\{C_{a_0,\dots,a_{n-1}} : n \in \mathbb{N};\ (a_0, \dots, a_{n-1}) \in \{0, 1\}^n\Big\} \cup \Big\{\emptyset, \Omega\Big\}$$

é uma semiálgebra de subconjuntos de Ω pois a interseção de dois deles $C_{a_0,\dots,a_{n-1}} \cap C_{b_0,\dots,b_{m-1}}$ é um cilindro da forma equação (1.8.7)

$$C_{a_0,\dots,a_{n-1}} \cap C_{b_0,\dots,b_{m-1}} = \begin{cases} C_{a_0,\dots,a_{n-1}}, & \text{se } n \leqslant m \text{ e } a_i = b_i \text{ para todo } i < n \\ C_{b_0,\dots,b_{m-1}}, & \text{se } m \leqslant n \text{ e } a_i = b_i \text{ para todo } i < m \\ \emptyset, & \text{caso contrário.} \end{cases} \tag{1.8.8}$$

Ademais, se $F = \{0,1\}^n \setminus \{(a_0,\dots,a_{n-1})\}$, então

$$\overline{C_{a_0,\dots,a_{n-1}}} = \biguplus_{(b_0,\dots,b_{n-1})\in F} C_{b_0,\dots,b_{n-1}} \tag{1.8.9}$$

As equações equação (1.8.8) e equação (1.8.9) garantem que a família dos cilindros da forma equação (1.8.7) formam uma semiálgebra de subconjuntos de Ω. Em $\mathcal{S}$ definimos $\mathbb{P}_0$ pondo $\mathbb{P}_0(\emptyset) = 0$, $\mathbb{P}_0(\Omega) = 1$) e $\mathbb{P}_0(C_{a_0,\dots,a_{n-1}}) = (1/2)^n$.

Para definir o modelo probabilístico usamos o teorema 1.6.4 para estender $\mathbb{P}_0$ definida em $\mathcal{S}$. Não é difícil verificar que $\mathbb{P}_0$ é aditiva (finita) nos elementos da semiálgebra (verifique). Por fim, é possível provar que $\mathbb{P}_0$ é enumeravelmente aditiva e isso requer um pouco de Topologia: uma família $\{A_n : n \geqslant 1\}$ de subconjuntos tem a *propriedade da interseção finita* (PIF) se a interseção de qualquer quantidade finita de elementos da família tem interseção não vazia. Um espaço topológico é compacto se, e somente se, se a interseção de todos elementos de uma família de subconjuntos fechados com PIF é não vazia.

Seja $\{A_n : n \geqslant 1\}$ uma família monótona decrescente de elementos de $\mathcal{S}$ tal que $\bigcap_{n\geqslant 1} A_n = \emptyset$. Se munimos $\{0,1\}$ da topologia discreta então Ω, com a topologia do produto, é compacto. Cada A_n da sequência é fechado, pois seu complemento é aberto na topologia produto. Portanto, tal família não deve ter a PIF de modo que $A_n = \emptyset$ para todo n suficientemente grande, logo $\lim_{n\to\infty} \mathbb{P}_0(\bigcap_n A_n) = 0$. Disso segue a aditividade enumerável como observamos na página 93.

1.8.8 Lemas de Borel–Cantelli (∗): para registro,

$$[A_n \text{ para } n \text{ grande o suficiente}] := \liminf_{n\to\infty} A_n = \bigcup_{n\geqslant 1}\bigcap_{k\geqslant n} A_k.$$

é o conjunto dos $\omega \in \Omega$ que pertencem a todo A_n para todo n suficientemente grande (e que pode depender de ω).

Seja A_n, $n \geqslant 1$, uma sequência de eventos de um modelo probabilístico. Lembremos que

$$[A_n \text{ infinitas vezes}] = \limsup_{n\to\infty} A_n = \bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k$$

e provamos que se $\sum_n \mathbb{P}(A_n)$ converge, então $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$. Por outro lado,

$$[A_n \text{ para } n \text{ grande o suficiente}] = \liminf_{n\to\infty} A_n = \bigcup_{n\geqslant 1}\bigcap_{k\geqslant n} A_k$$

é o conjunto dos $\omega \in \Omega$ que pertencem a todo A_n para n suficientemente grande; e se $\sum_k \mathbb{P}(A_k) = \infty$

então

$$
\begin{aligned}
1-\mathbb{P}(\limsup_{n\to\infty} A_n) &= \mathbb{P}\left(\overline{\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k}\right)\\
&= \mathbb{P}\left(\bigcup_{n\geqslant 1}\bigcap_{k\geqslant n} \overline{A_k}\right)\\
&= \mathbb{P}\left(\lim_{n\to\infty}\bigcap_{k\geqslant n} \overline{A_k}\right)\\
&= \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k\geqslant n} \overline{A_k}\right)
\end{aligned}
$$

pois $\bigcap_{k\geqslant n}\overline{A_k}$ é uma sequência crescente com n. Ademais,

$$
\bigcap_{k\geqslant n}\overline{A_k}\subset\bigcap_{k=n}^{m}\overline{A_k}\implies \mathbb{P}\left(\bigcap_{k\geqslant n}\overline{A_k}\right)\leqslant\mathbb{P}\left(\bigcap_{k=n}^{m}\overline{A_k}\right)
$$

e se assumimos independência dos eventos A_n, então

$$
\mathbb{P}\left(\bigcap_{k=n}^{m}\overline{A_k}\right)=\prod_{k=n}^{m}\mathbb{P}\left(\overline{A_k}\right)
$$

e como $1-x\leqslant\exp(-x)$ para $x\in[0,1]$

$$
\prod_{k=n}^{m}\mathbb{P}\left(\overline{A_k}\right)=\prod_{k=n}^{m}1-\mathbb{P}\left(A_k\right)\leqslant\exp\left(-\sum_{k=n}^{m}\mathbb{P}(A_k)\right)
$$

e como a série é divergente

$$
\lim_{m\to\infty}\sum_{k=n}^{m}\mathbb{P}(A_k)=\infty
$$

para todo n, logo,

$$
0\leqslant\lim_{m\to\infty}\mathbb{P}\left(\bigcap_{k=n}^{m}\overline{A_k}\right)\leqslant\lim_{m\to\infty}\exp\left(-\sum_{k=n}^{m}\mathbb{P}(A_k)\right)=0
$$

e como vale para todo n

$$
\lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k\geqslant n}\overline{A_k}\right)=0
$$

portanto

$$
\mathbb{P}(\limsup_{n\to\infty} A_n)=1.
$$

Com isso, provamos o segundo lema de Borel–Cantelli

1.8.9 Teorema (Borel–Cantelli). *Seja A_n, $n \geqslant 1$, uma sequência de eventos de um modelo probabilístico. Se $\sum_n \mathbb{P}(A_n)$ não converge e os eventos são independentes, então*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 1.$$

Notemos que a hipótese de independência é necessária pois se fizermos $A_n = A$ para todo n, para algum evento A, então a conclusão do teorema não vale (verifique).

Exemplo 51. Retomemos o exemplo 50 acima. Se A_n denota o eventos "o n-ésimo lançamento é cara" então pelo teorema acima $\mathbb{P}([A_n \text{ infinitas vezes}]) = 1$ de modo que a probabilidade de só ocorrer coroa é 0, embora o evento não seja impossível. ◊

Exemplo 52. Seja E o evento "sequência de 100100 caras" no lançamento de uma moeda equilibrada infinitas vezes. Dividimos a sequência em blocos de tamanho 100100 e denotamos por E_n o evento "sequência de 100100 caras no n-ésimo bloco". Desse modo os eventos E_n são independentes e $\mathbb{P}(E_n) = (1/2)^{100100}$ de modo que $\sum_n \mathbb{P}(E_n) = \infty$. Pelo teorema acima $\mathbb{P}([E_n \text{ infinitas vezes}]) = 1$ portanto $\mathbb{P}(E) = 1$. ◊

Uma aplicação popular do segundo lema de Borel–Cantelli é conhecido como o Teorema do macaco infinito. Ele diz que com probabilidade 1 um macaco digitando aleatoriamente num teclado, eventualmente, digita todos as obras de Shakespeare. Esse exemplo aparece no livro sobre probabilidade de Emile Borel, publicado em 1909. A estratégia é como a do exemplo anterior, considere uma sequência M infinita de letras dos alfabeto (finito) escolhidas aleatória, uniforme e independentemente. Fixemos uma sequência S de comprimento m de letras do alfabeto e consideremos os eventos E_n definidos, como no exemplo acima, "o n-ésimo bloco é a sequência S". Concluímos que S ocorre com probabilidade 1.

1.8.10 Observação (uma estimativa no wikipedia). *Se houvesse tantos macacos quanto átomos no universo observável, todos digitando extremamente rápido durante trilhões de vezes a vida do universo, a probabilidade dos macacos replicarem uma única página de Shakespeare é pequena. Ignorando pontuação, espaçamento e maiúsuculas, um macaco digitando letras uniformemente ao acaso tem uma chance de* 1 *em* 26 *de digitar corretamente a primeira letra de Hamlet. O texto de Hamlet contém aproximadamente 130 mil letras. Assim, há uma probabilidade de um em* $3{,}4 \times 10^{183.946}$ *para obter o texto na primeira tentativa. Mesmo se cada próton no universo observável fosse um macaco com uma máquina de escrever, escrevendo desde o Big Bang até o fim do universo (quando prótons já não existem), eles ainda precisariam de muito (mais de 360 mil ordens de magnitude a mais) mais tempo para ter ainda uma chance* 1 *em* 10.500 *de sucesso. Dito de outra maneira, para termos uma chance de sucesso de um em um trilhão, precisaríamos ter* $10^{360.641}$ *universos feitos de macacos atômicos. Tal como Kittel e Kroemer colocou, "A probabilidade de Hamlet é, portanto, zero no qualquer sentido operacional de um evento e a afirmação de que os macacos devem conseguir dá uma conclusão enganosa sobre números muito, muito grandes".*

1.8.11 Exemplos forenses:

Exemplo 53. Um investigador de polícia está 60% convencido da culpa de um determinado suspeito. No decorrer da investigação uma nova evidência na cena do crime prova que o criminoso tinha uma determinada característica, seu pé-direito tem seis dedos. Sabe-se que 20% da população tem seis dedos no pé direito. Dado que o suspeito tem seis dedos no pé direito, qual é o grau de convencimento do investigador?

Representemos por H o evento "suspeito culpado" e por E o evento "tem seis dedos no pé direito" e temos pelo Teorema de Bayes que

$$\mathbb{P}(H \mid E) = \frac{\mathbb{P}(E \mid H)\mathbb{P}(H)}{\mathbb{P}(E \mid H)\, P(H) + \mathbb{P}\left(E \,\middle|\, \overline{H}\right) P(\overline{H})} = \frac{1 \cdot 0{,}6}{1 \cdot 0{,}6 + 0{,}2 \cdot 0{,}4} = 0{,}882352941$$

ou seja, o grau de convencimento agora é $88,2\%$. ◊

Exemplo 54. No contexto do exemplo anterior, se H é uma hipótese e E uma evidência então

$$\mathbb{P}(H \mid E) = \frac{\mathbb{P}(E \mid H)\, P(H)}{\mathbb{P}(E)} \qquad \text{e} \qquad \mathbb{P}\left(\overline{H} \,\middle|\, E\right) = \frac{\mathbb{P}\left(E \,\middle|\, \overline{H}\right) P(\overline{H})}{\mathbb{P}(E)}$$

portanto

$$\frac{\mathbb{P}(H \mid E))}{\mathbb{P}\left(\overline{H} \,\middle|\, E\right)} = \frac{\mathbb{P}(H)}{\mathbb{P}(\overline{H})} \cdot \frac{\mathbb{P}(E \mid H)}{\mathbb{P}\left(E \,\middle|\, \overline{H}\right)} \tag{1.8.10}$$

onde

$$\frac{\mathbb{P}(H)}{\mathbb{P}(\overline{H})} \quad \text{e} \quad \frac{\mathbb{P}(H \mid E)}{\mathbb{P}\left(\overline{H} \,\middle|\, E\right)}$$

quantificam, respectivamente, quanto provável é a ocorrência de H com relação a não ocorrência de H, e quão mais provável é a ocorrência de H com relação a não ocorrência após a introdução da evidência E.

Portanto, se a hipotese é culpa e a evidencia é uma característica do criminoso, no exemplo anterior temos

$$\frac{\mathbb{P}(E \mid H)}{\mathbb{P}\left(E \,\middle|\, \overline{H}\right)} = \frac{1}{0{,}2} = 5$$

ou seja, após a nova evidência a culpa é 5 vezes mais provável. ◊

Exemplo 55 (**falácia do promotor**). Essa falácia geralmente resulta em assumir que a probabilidade prévia de que uma evidência condenaria um membro escolhido aleatoriamente da população é igual à probabilidade de que ela condenaria o réu. Por exemplo, você ganha na loteria, mas é acusado de fraude. No julgamento o promotor afirma que ganhar na loteria sem trapacear é extremamente improvável portanto é igualmente extremamente improvável que você seja inocente. Isso desconsidera que a probabilidade de qualquer outra pessoa ganhe na loteria é igualmente pequena. A confusão típica é, usando a linguagem dos tribunais, que

Falácia do promotor: se

$$\mathbb{P}(\text{evidência comprovada} \mid \text{inocência do suspeito})$$

é pequena, então

$$\mathbb{P}(\text{inocência do suspeito} \mid \text{evidência comprovada})$$

é pequena.

De fato, essas probabilidade podem ser muito diferentes como mostramos a seguir. No caso de teste de DNA para identificar criminosos, dependendo do número de locus considerado na análise, é possível que a frequência de determinado padrão seja da ordem de 1 em 1 milhão [Technology in Forensic Science]. O erro comum é a confusão resultante de tomar por aproximadamente igual

1. a probabilidade que a amostra de DNA de uma outra pessoa (não o criminoso) case com o padrão do DNA descoberto na cena do crime, dado que a pessoa é inocente;

2. a probabilidade que a pessoa é inocente dado que seu DNA casa com o padrão do DNA descoberto na cena do crime.

A primeira é a probabilidade atestada por um perito em exame de DNA; quando um teste de DNA atesta que "a probabilidade é de 1 em 100 milhões que o DNA da cena do crime case como DNA do suspeito" isso significa que "a chance do casamento de padrões entre o DNA na cena do crime e do DNA do acusado é de 1 em 100 milhões, se o acusado é inocente e não é parente do criminoso", pondo de outro modo $P(\text{evidência} \mid \text{inocência})$ é pequena. A segunda probabilidade, descrita acima, é a que interessa ao promotor para poder acusar sem erro o suspeito e uma vez que o resultado positivo é certo ($\mathbb{P}(\text{inocência} \mid \text{evidência})$), se o acusado deixou a amostra de DNA na cena do crime então

$$\frac{\mathbb{P}(\text{evidência} \mid \text{culpa})}{\mathbb{P}(\text{evidência} \mid \text{inocência})} = 100 \text{ milhões}$$

ou seja, independentemente da chance a priori de culpa, a evidência do DNA torna a probabilidade de culpa 100 milhões de vezes maior que a de inocência, de acordo com a equação equação (1.8.10) acima

$$\frac{\mathbb{P}(\text{culpa} \mid \text{evidência})}{\mathbb{P}(\text{inocência} \mid \text{evidência})} = 100.000.000 \frac{\mathbb{P}(\text{culpa})}{\mathbb{P}(\text{inocência})}.$$

◊

Exemplo 56. O seguinte relato foi extraído de uma reportagem da revista TIME de sexta-feira, 8 de janeiro de 1965. "Por volta do meio-dia, em um dia de Junho último, uma mulher idosa foi assaltada em um beco em San Pedro, na Califórnia. Pouco depois, uma testemunha viu uma garota loira com rabo de cavalo saindo do beco e entrando em um carro amarelo conduzido por um negro com barba e bigode. A polícia prendeu e Janet e Malcolm Collins, um casal que que encaixava na descrição física dada pela testemunha e que possuía um Lincoln amarelo."

Ao apresentar o caso o promotor, que teve acessoria de um professor de matemática local, salientou a baixa probabilidade de haver casais como os Collins em San Pedro, em um carro amarelo pois

- probabilidade de carro amarelo = 1/10;
- probabilidade de homem com bigode = 1/4;
- probabilidade de mulher com rabo de cavalo = 1/10;
- probabilidade mulher tem cabelo loiro = 1/3;
- probabilidade de homem negro com barba = 1/10;
- probabilidade de casal inter-racial = 1/1000;

Multiplicando todos esses valores leva a um valor de 1 em 12 milhões para a probabilidade de haver um casal com as mesmas características de Janet e Malcolm Collins e com uma evidência tão forte o casal acabou sendo condenado. Qual é o problema desse argumento?

Veja este link se quiser mais detalhes sobre as considerações legais do processo e o desdobramento. Há vários casos reais de uso incorreto de probabilidade num julgamento, veja os casos Lucia de Berk, Kevin Sweeney, Sally Clark retratados no *Wikipedia*. ◊

Exercícios.

1. Mostre que para eventos aleatórios A, B, C e D

$$\mathbb{P}(A \cap B \cap C \cap D) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)\,\mathbb{P}(C \mid A \cap B)\,\mathbb{P}(D \mid A \cap B \cap C).$$

   Assuma que todos os eventos envolvidos tenha probabilidade positiva para que as probabilidades condicionais estejam definidas.

2. (**Regra da multiplicação**) Sejam $A_2, A_2, \dots, A_n$ eventos tais que $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$. Então

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\prod_{i=2}^{n-1}\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j=1}^{i-1} A_j\right).$$

3. Use a regra da multiplicação para três eventos (eq. equação (1.7.1)) para determinar qual é a probabilidade de ocorrer três reis ao se retirar três cartas ao acaso de um baralho.

4. No último ano de uma classe com 100 alunos que estão se formando no ensino médio, 42 estudaram matemática, 68 estudaram psicologia, 54 estudaram historia, 22 estudaram matemática e historia, 25 estudaram matemática e psicologia, 7 estudaram historia mas não estudaram matemática nem psicologia, 10 estudaram as três matérias e 8 não estudaram nenhuma. Se um estudante é selecionado ao acaso

   (a) qual é a probabilidade que uma pessoa matriculada em psicologia tenha estudado as três matérias; 5/34

   (b) qual é a probabilidade que uma pessoa não esteja estudando psicologia esteja estudando historia e matemática. 3/8

5. A probabilidade de que um homem casado assista determinado programa de TV é 0,4 e a probabilidade que uma mulher casada é 0,5. A probabilidade que um homem assista dado que sua mulher assista é 0,7. Determine a probabilidade de que

   (a) um casal assista ao programa; 0,35

   (b) uma esposa assista ao programa, dado que o marido o faça; 0,875

   (c) pelo menos uma pessoa do casal assista ao programa. 0,55

6. A probabilidade de que um médico acerte o diagnóstico de uma doença é 0,7. Dado que o médico fez um diagnóstico incorreto, a probabilidade de que o paciente entre com um processo é 0,9. Qual é a probabilidade de que o médico erre o diagnóstico e seja processado pelo paciente.

   0,27

7. Um corretor de imóveis tem oito chaves mestras para abrir diversas casas. Somente uma delas abrirá qualquer uma das casas. Se 40% dessas casas são, normalmente, deixadas destrancadas, qual é a probabilidade de que o corretor entre em uma casa específica se ele selecionou três chaves mestras aleatoriamente antes de sair do escritório? 5/8

8. Uma cidade tem dois carros de bombeiros operando independentemente. Cada carro tem probabilidade 0,96 de estar disponível em caso de necessidade.

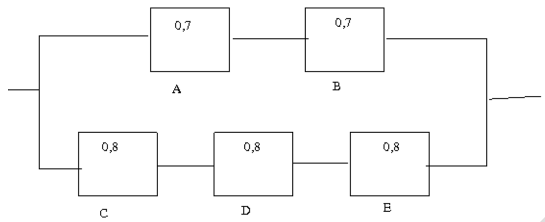
   (a) Qual é a probabilidade de que nenhum carro esteja disponível quando necessário? 0,0026

   (b) Qual é a probabilidade de que algum carro esteja disponível quando necessário? 0,9984

9. Numa variante do jogo de Monty Hall, ao competidor são apresentadas 10 portas. De novo, atrás de uma delas está um carro e atrás das outras há bodes. Depois que o competidor escolhe uma porta, Monty Hall abre outras sete das nove portas restantes que estão fechadas, sem revelar a porta que esconde o carro. Existem agora três portas fechadas: a que o competidor

originalmente escolheu e duas outras. Qual a melhor estratégia, mudar de porta escolhida ou ficar com a escolha original? Além disso, qual é a probabilidade de ganhar, seguindo estas duas estratégias?

10. Um sistema de circuitos é dado na figura abaixo. Assuma que os componentes falham independentemente.



(a) Qual é a probabilidade que o sistema todo funcione? 0,7512

(b) Dado que o sistema funciona, qual é a probabilidade de que o componente A não esteja funcionando? 0,2045

11. Em certa região do país sabe-seque a probabilidade de selecionar um adulto com mais de 40 anos com câncer é 0,05. Se a probabilidade de um médico diagnosticar corretamente uma pessoa com câncer como portadora da doença é 0,78 e a probabilidade de diagnosticar incorretamente uma pessoa sem câncer como sendo portadora da doença é 0,06, qual é a probabilidade de que a pessoa seja diagnosticada com câncer? 0,096 Qual é a probabilidade de que a pessoa diagnosticada com câncer realmente tenha a doença? 0,40625

12. Suponha que quatro inspetores em uma fábrica e filmes tenham que estampar a data de validade em cada pacote de filme ao final da linha de montagem. João, que estampa 20% dos pacotes, não estampa a data de validade em 1 de cada 20 pacotes; Antônio, que estampa 60% dos pacotes, erra uma vez a cada 100 pacotes; Jorge que estampa 15% dos pacotes, erra uma vez a cada 90 pacotes e Patrícia que estampa 5% dos pacotes, erra uma vez a cada 200 pacotes. Se um cliente reclame que sua embalagem de filme não contém a data de validade, qual é a probabilidade de que ela tenha sido inspecionada por João? 0,1124

13. Considere os eventos A dado por "o rio é poluído", B dado por "uma amostra de água testada detecta poluição" e C dado por "a pesca é permitida". Assuma $\mathbb{P}(A) = 0{,}3$, $\mathbb{P}(B \mid A) = 0{,}75$, $\mathbb{P}\left(B \mid \overline{A}\right) = 0{,}2$, $\mathbb{P}(C \mid A \cap B) = 0{,}2$, $\mathbb{P}\left(C \mid \overline{A} \cap B\right) = 0{,}15$, $\mathbb{P}\left(C \mid A \cap \overline{B}\right) = 0{,}8$ e $\mathbb{P}\left(C \mid \overline{A} \cap \overline{B}\right) = 0{,}9$. Determine

a) $\mathbb{P}(A \cap B \cap C)$; 0,045

b) $\mathbb{P}(\overline{B} \cap C)$; 0,564

c) $\mathbb{P}(C)$; 0,63

d) determine a probabilidade de o rio ser poluído, dado que a pesca é permitida e a amostra testada não detectou poluição. 0,1064

14. O teorema da probabilidade total vale para uma partição enumerável por eventos de Ω, digamos $\{E_1, E_2, \dots, E_n, \dots\}$ com $\mathbb{P}(E_i) > 0$ para todo i?

15. No experimento da urna de Pólya, exemplo 41, qual é a probabilidade de que a primeira bola seja preta dado que a n-ésima bola sorteada foi preta? qual é a probabilidade de que a primeira bola seja preta dado que as n próximas bolas sorteadas foram pretas?

16. Generalize o experimento da urna de Pólya, exemplo 41, para o caso em que a urna comece com a bolas brancas, b bolas pretas e a cada sorteio a bola sorteada é reposta na urna junto com mais c bolas da mesma cor.

17. Definimos que o evento A é independente dos eventos B e C se vale a equação equação (1.8.3). Verifique se equação (1.8.3) e equação (1.8.2) são equivalentes. Se sim, dê uma prova, senão dê um contraexemplo.

18. Refaça o exercício 71 para a definição definitiva (eq. equação (1.8.3)) de "A independente de B e C".

19. Mostre que se A e B são eventos independentes, então (a) $\overline{A}$ e $\overline{B}$ são eventos independentes; (b) A e $\overline{B}$ são eventos independentes; (c) $\overline{A}$ e B são eventos independentes.

20. Prove que as equações 1.8.4 e 1.8.5 são equivalentes.

21. Seja $\Omega = \{1, 2, \dots, p\}$ para p primo e $\mathbb{P}(A) = |A|/p$, para todo $A \subset \Omega$. Prove que se A e B são independentes então pelo menos um desses eventos deve ser $\emptyset$ ou Ω.

22. Um pais extremamente democrático mantém três prisioneiros sem qualquer acusação na prisão de Guantánamo. São eles: Said, Osama e Ayman. O carcereiro diz a eles que um será libertado e os outros serão executados, mas não lhe foi permitido dizer o que acontecerá a cada um. Said avalia que ele tem chance de 1/3 de escapar e pensou em perguntar em segredo ao carcereiro qual dos outros prisioneiros seria executado; em seguida avaliou que se o carcereiro respondesse o Osama, por exemplo, então a chance dele escapar passaria a ser 1/2 pois seria ou ele e Osama, ou ele e Ayman. Essa avaliação está correta?

23. (**Lewis Carrol**) Leia a seguinte "prova" de que uma urna não pode ter duas bolas da mesma cor:

Uma urna contém um par de bolas que podem ser da cor branca (B) ou preta (P), de modo que $\mathbb{P}(BP) = \mathbb{P}(PB) = \mathbb{P}(PP) = \mathbb{P}(BB) = \frac{1}{4}$. Adicionamos uma bola preta de modo que $\mathbb{P}(PBP) = \mathbb{P}(PPB) = \mathbb{P}(PPP) = \mathbb{P}(PBB) = \frac{1}{4}$. Agora, sorteamos uma bola e a probabilidade que essa bola é preta é, usando probabilidade total

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

Se a probabilidade de sortear uma bola preta é $\frac{2}{3}$ então o par original de bolas não pode ser PP nem BB, ou seja, não podem ter a mesma cor.

Faça uma análise criteriosa dessa (falsa) dedução.

24. Uma urna contém 6 bolas, três são vermelhas e três azuis. Uma dessas bolas é selecionada ao acaso e removida permanentemente da urna sem que a sua cor seja revelada para um observador, designaremos essa bola por A. Agora, o observador retira bolas um número pré-determinado de vezes, uma de cada vez, de forma aleatória e com reposição. O objetivo desse experimento é formar uma ideia aproximada da proporção de bolas vermelhas e azuis que permaneceram na urna após a retirada da bola A. O observador retirou seis vezes uma bola da urna e todos resultados foram bolas vermelhas. Numa segunda execução, retirou uma bola da urna 600 vezes, 303 delas foram bolas vermelhas e 297 delas azuis. Claramente, ambos tendem a prever que a bola A é, provavelmente, azul. Qual deles, se é que algum, tem a evidência empírica mais forte de sua previsão?

25. Sejam $A_1, A_2, \ldots$ eventos independentes de um modelo probabilístico. Prove que

$$\mathbb{P}\left(\bigcap_{i\geqslant 1} A_i\right) = \prod_{i\geqslant 1} \mathbb{P}(A_i).$$

§2 Variáveis Aleatórias

Se uma moeda é lançada 3 vezes com os resultados independentes. Qual é o número de caras ocorridas? Qual é a probabilidade de termos 2 caras?

resultado 1	resultado 2	resultado 3	Nº de caras
Ca	Ca	Ca	3
Ca	Ca	Co	2
Ca	Co	Ca	2
Ca	Co	Co	1
Co	Ca	Ca	2
Co	Ca	Co	1
Co	Co	Ca	1
Co	Co	Co	0

A probabilidade de ocorrerem exatamente 2 caras é 3/8. Muitas vezes estamos mais interessados numa característica numérica de um evento do que no evento propriamente dito, por exemplo

1. Qual o número de chamadas telefônicas que chegam a uma central em um intervalo de tempo?
2. Qual a distância da origem de um ponto escolhido no círculo unitário?
3. Qual a altura de um cidadão escolhido?
4. Qual o tempo de duração de uma lâmpada escolhida da linha de produção?

Essas são grandezas que dependem do resultado de um experimento, são chamadas de variáveis aleatórias.

Variável aleatória real: uma variável aleatória (v.a.) real de um modelo probabilístico $(\Omega, \mathcal{A}, \mathbb{P})$ é uma função que associa a cada elemento de um espaço amostral Ω um número real

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

de modo que, para todo $t \in \mathbb{R}$,

$$[X \leqslant t] := \{\omega \in \Omega : X(\omega) \leqslant t\} \tag{2.0.1}$$

é um evento aleatório do modelo cuja probabilidade é

$$\mathbb{P}(X \leqslant t) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leqslant t\}).$$

Se uma função definida num espaço amostral a valores reais é ou não uma variável aleatória depende do espaço de eventos $\mathcal{A}$.

Comumente, usamos as letras maiúsculas finais do alfabeto X, Y, Z, W para variáveis aleatórias.

Exemplo 57. Se X é o número de caras em 3 lançamentos de uma moeda então

$$\begin{array}{lll} X((\text{Ca},\text{Ca},\text{Ca})) = 3 & X((\text{Ca},\text{Co},\text{Co})) = 1 & X((\text{Co},\text{Co},\text{Ca})) = 1 \\ X((\text{Ca},\text{Ca},\text{Co})) = 2 & X((\text{Co},\text{Ca},\text{Ca})) = 2 & X((\text{Co},\text{Co},\text{Co})) = 0 \\ X((\text{Ca},\text{Co},\text{Ca})) = 2 & X((\text{Co},\text{Ca},\text{Co})) = 1 & \end{array}$$

e

$$[X \leqslant t] = \begin{cases} \emptyset, & \text{se } t < 0 \\ \big\{(\text{Co},\text{Co},\text{Co})\big\}, & \text{se } 0 \leqslant t < 1 \\ \big\{(\text{Co},\text{Co},\text{Co}), (\text{Co},\text{Co},\text{Ca}), (\text{Co},\text{Ca},\text{Co}), (\text{Ca},\text{Co},\text{Co})\big\}, & \text{se } 1 \leqslant t < 2 \\ \big\{(\text{Co},\text{Co},\text{Co}), (\text{Co},\text{Co},\text{Ca}), (\text{Co},\text{Ca},\text{Co}), (\text{Ca},\text{Co},\text{Co}), & \\ \qquad (\text{Ca},\text{Ca},\text{Co}), (\text{Ca},\text{Co},\text{Ca}), (\text{Co},\text{Ca},\text{Ca})\big\}, & \text{se } 2 \leqslant t < 3 \\ \Omega, & \text{se } t \geqslant 3. \end{cases}$$

Aqui, nesse exemplo, deve ficar claro que X satisfaz equação (2.0.1) pois tem 2^{Ω} como espaço de eventos. ◊

Exemplo 58. Consideremos o modelo clássico para o lançamento de um dado equilibrado. Seja $X : \Omega \to \mathbb{R}$ o resultado do lançamento, isto é, $X(\omega) = \omega$ para todo $\omega \in \Omega$. Para $t < 1$, temos $[X \leqslant t] = \emptyset$. Para $1 \leqslant t < 6$, temos $[X \leqslant t] = \{1, \dots, \lfloor t \rfloor\}$. Para $t \geqslant 6$, temos $[X \leqslant t] = \Omega$. ◊

2.0.1 Proposição. *A condição equação (2.0.1) é trivialmente satisfeita sempre que o espaço amostral é discreto.* □

Exemplo 59 (**variável aleatória constante**). Se para todo $\omega \in \Omega$ há $c \in \mathbb{R}$ tal que $X(\omega) = c$, então X é uma v.a. (verifique) tal que $[X \leqslant t] = \Omega$ caso $t \geqslant c$ e $[X \leqslant t] = \emptyset$ caso contrário. ◊

Exemplo 60. Consideremos o modelo geométrico clássico para o lançamento de um dardo num alvo de raio 3 (exemplo 34). Seja $Y : \Omega \to \mathbb{R}$ a distância euclidiana do ponto atingido ao centro do alvo, de modo que para qualquer $0 \leqslant t \leqslant 3$ o conjunto é $[X \leqslant t]$ é o circulo de raio t, cuja área está bem definida, portanto é um elemento de $\mathcal{A}$. ◊

Se $X, Y : \Omega \to \mathbb{R}$ são variáveis aleatórias então temos, por exemplo, os conjuntos

1. $[X > 3] = \overline{[X \leqslant 3]} = \overline{\{\omega \in \Omega : X(\omega) \leqslant 3\}} = \{\omega \in \Omega : X(\omega) > 3\}$;

2. $[X = 3] = [X \leqslant 3] \cap [X \geqslant 3]$,

3. $[2 \leqslant X < 3] = [X < 3] \cap [X \geqslant 2] = \{\omega \in \Omega : 2 \leqslant X(\omega) < 3\}$;

que, de fato, são eventos aleatórios. De modo análogo ao que foi descrito nos parágrafos acima, definimos os conjuntos $[X = t]$, $[X < t]$, $[X \leqslant t]$ e $[X \geqslant t]$ para uma variável aleatória $X : \Omega \to \mathbb{R}$. Esses conjuntos são eventos aleatórios do modelo pois para quaisquer reais a, b pois podem ser escrito como resultado de operações elementares de conjuntos sobre eventos aleatórios

- $[X > a] = \overline{[X \leqslant a]}$;
- $[a < X \leqslant b] = [X > a] \cap [X \leqslant b]$;
- $[X = a] = \bigcap_{n \geqslant 1} [a - \frac{1}{n} < X \leqslant a]$;
- $[X \geqslant a] = [X > a] \cup [X = a]$;
- $[X < a] = \overline{[x \geqslant a]}$.

O que nos interessa sobre probabilidade com relação à variáveis aleatórias são derivados das probabilidades desses eventos. Por exemplo,

1. os eventos $[Y < 1]$ e $[Y \geqslant 1]$ particionam Ω e, por exemplo, assumindo que ambos têm probabilidade positiva $\mathbb{P}(X = 3) = \mathbb{P}(X = 3 \mid Y < 1)\mathbb{P}(Y < 1) + \mathbb{P}(X = 3 \mid Y \geqslant 1)\mathbb{P}(Y \geqslant 1)$ pelo teorema da probabilidade total.

2. Se X é o número de caras em 3 lançamentos de uma moeda então

$$[X = 2] = \{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca})\}$$
$$[X \geqslant 2] = \{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca})\}$$

e $\mathbb{P}(X = 2) = 3/8$ e $\mathbb{P}(X \geqslant 2) = 1/2$. O evento complementar a $[X \geqslant 2]$ representa o evento $\overline{[X \geqslant 2]} = [X < 2] = \{(\mathrm{Ca}, \mathrm{Co}, \mathrm{Co}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Co}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Co}, \mathrm{Co})\}$ que ocorre com probabilidade $\mathbb{P}(\overline{X \geqslant 2}) = 1 - \mathbb{P}(X \geqslant 2) = \frac{1}{2}$. Se A representa o evento definido por "o primeiro lançamento foi Ca" então

$$\mathbb{P}(X = 2 \mid A) = \frac{\mathbb{P}([X = 2] \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca})\})}{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Co})\})} = \frac{1}{2}.$$

$$\mathbb{P}(X = 2 \mid X \geqslant 2) = \frac{\mathbb{P}([X = 2] \cap [X \geqslant 2])}{\mathbb{P}(X \geqslant 2)} = \frac{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca})\})}{\mathbb{P}(\{(\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}, \mathrm{Ca}), (\mathrm{Co}, \mathrm{Ca}, \mathrm{Ca})\})} = \frac{3}{4}.$$

2.0.2 Função de uma variável aleatória: se $X : \Omega \to \mathbb{R}$ é uma v.a. e $f : \mathbb{R} \to \mathbb{R}$ é uma função real então a função composta pode não ser uma variável aleatória, entretanto será em vários casos úteis.

Exemplo 61. Se X é uma v.a. de $(\Omega, \mathcal{A}, \mathbb{P})$ então podemos definir $Z : \Omega \to \mathbb{R}$ por $Z(\omega) = X(\omega)^2$ para todo $\omega \in \Omega$. A função Z também é uma v.a. pois para todo t não negativo temos $[Z \leqslant t] = [-\sqrt{t} \leqslant X \leqslant \sqrt{t}] \in \mathcal{A}$ para todo $t \in \mathbb{R}$.

Podemos definir $W : \Omega \to \mathbb{R}$ por $W(\omega) = X(\omega)^3$ para todo $\omega \in \Omega$. A função W também é uma v.a. pois $[W \leqslant t] = [X \leqslant \sqrt[3]{t}] \in \mathcal{A}$ para todo $t \in \mathbb{R}$.

Se X é uma v.a. de $(\Omega, \mathcal{A}, \mathbb{P})$ e $a, b \in \mathbb{R}$ $(a \neq 0)$ então $Y : \Omega \to \mathbb{R}$ dada por $Y(\omega) = a \cdot X(\omega) + b$ é uma variável aleatória. De fato, para todo real t temos $[Y \leqslant t] = [X \leqslant (t - b)/a] \in \mathcal{A}$. ◊

Em geral, quando f é "bem comportada", a composta é uma v.a.:

2.0.3 Teorema. *Se X é uma variável aleatória de $(\Omega, \mathcal{A}, \mathbb{P})$ e $f : \mathbb{R} \to \mathbb{R}$ é contínua (ou contínua por partes) então $f(X)$ é uma variável aleatória.*

Esse resultado será provado mais adiante. Uma consequência imediata é:

2.0.4 Corolário. *X^k é uma variável aleatória para todo $k \in \mathbb{N}$.* □

Operações aritméticas com variáveis aleatórias são bem comportadas.

2.0.5 Teorema. *Se X e Y são variáveis aleatórias de um espaço de probabilidades $(\Omega, \mathcal{A}, \mathbb{P})$ então $X + Y$ e $X \cdot Y$ também são variáveis aleatória.*

Demonstração. Sejam X e Y variáveis aleatórias definidas em $(\Omega, \mathcal{A}, \mathbb{P})$.

Para a soma $X + Y(\omega) := X(\omega) + Y(\omega)$ temos que, para qualquer $t \in \mathbb{R}$

$$[X + Y < t] = \bigcup_{r \in \mathbb{Q}} [X < r] \cap [Y < t - r]$$

e o lado direito está em $\mathcal{A}$. Para verificar a igualdade de conjuntos tomemos $\omega \in [X + Y < t]$. Então $X(\omega) + Y(\omega) < t$. Tomemos um racional $r \in \mathbb{Q}$ tal que $X(\omega) < r < t - Y(\omega)$. De $X(\omega) < r$ temos $\omega \in [X < r]$ e de $r < t - Y(\omega)$ temos que $\omega \in [Y < t - r]$. Por outro lado, se $\omega \in [X < r] \cap [Y < t - r]$ então, claramente, $\omega \in [X + Y < t]$.

Para o produto $X \cdot Y(\omega) := X(\omega) \cdot Y(\omega)$ basta notar que

$$X \cdot Y = \frac{1}{2}\Big((X + Y)^2 - X^2 - Y^2\Big)$$

e o lado direito é uma variável aleatória pois, pelo exemplo 61 e pelo parágrafo acima $(X+Y)^2 - X^2 - Y^2$ é variável aleatória e, para concluir, usamos o resultado apresentado no exemplo 61. □

2.0.6 Teorema. *Se $X_1, X_2, \ldots$ são variáveis aleatórias de um espaço de probabilidades $(\Omega, \mathcal{A}, \mathbb{P})$ tais que existe o limite $\lim_{n\to\infty} X_n(\omega)$ para todo $\omega \in \Omega$, então*

$$X = \lim_{n\to\infty} X_n$$

é uma variável aleatória de $(\Omega, \mathcal{A}, \mathbb{P})$.

Demonstração do teorema 2.0.6 (∗): Vamos mostrar que

$$[X \leqslant t] = \bigcap_{m\geqslant 1} \bigcup_{k\geqslant 1} \bigcap_{n\geqslant k} \left[X_n \leqslant t + \frac{1}{m}\right]$$

e como no lado direito temos operações sobre eventos, o resultado é um evento.

Da definição de limite, para cada ω, temos $X_n(\omega) \to X(\omega)$ significa que para todo natural m temos que $\{n \in \mathbb{N} : |X_n(\omega) - X(\omega)| \geqslant 1/m\}$ é finito, ou seja, de acordo com a definição dada na página 31, se $\omega \in [X \leqslant t]$ temos

$$\omega \in \liminf_{n\to\infty} \left[X_n \leqslant t + \frac{1}{m}\right]$$

para todo $m \geqslant 1$, isto é,

$$\omega \in \bigcap_{m\geqslant 1} \bigcup_{k\geqslant 1} \bigcap_{n\geqslant k} \left[X_n \leqslant t + \frac{1}{m}\right].$$

para a recíproca, suponha que

$$\omega \in \bigcup_{k\geqslant 1} \bigcap_{n\geqslant k} \left[X_n \leqslant t + \frac{1}{m}\right] \quad \text{para todo } m \geqslant 1.$$

Então, fixado m, temos que $X_n(\omega) \leqslant t + 1/m$ para todo $n \geqslant k(m)$. Tomando uma subsequência $(n_m : m \geqslant 1 \text{ e } n_m > k(m))$ crescente temos

$$\lim_{m\to\infty} X_{n_m}(\omega) \leqslant \lim_{m\to\infty} t + 1/m \quad \text{e} \quad \lim_{m\to\infty} X_{n_m} = X \quad \text{e} \quad \lim_{m\to\infty} t + \frac{1}{m} = t$$

portanto $\omega \in [X \leqslant t]$. □

Uma definição não usual para $\mathbf{X^{-1}}$: notemos que a função inversa de X pode não estar definida pois X não é, necessariamente, injetora (ela pode ser considerada sobrejetora pois podemos restringir o contradomínio à imagem). Entretanto, definimos X^{-1} para todo $A \subset \mathbb{R}$ por

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}.$$

Uma propriedade importante dessa definição é que ela preserva e comuta com as operações sobre conjuntos (verifique):

$$X^{-1}(\overline{A}) = \overline{X^{-1}(A)}, \quad X^{-1}\left(\bigcup_n A_n\right) = \bigcup_n X^{-1}(A_n) \quad e \quad X^{-1}\left(\bigcap_n A_n\right) = \bigcap_n X^{-1}(A_n).$$

Também usamos a notação

$$[X \in A] := X^{-1}(A)$$

o qual é um evento aleatório sempre que A é boreliano (veja página 35). Daí temos, na definição de variável aleatória real, que $[X \leqslant t] = X^{-1}((-\infty, t]) \in \mathcal{A}$ é equivalente a dizer que $X^{-1}(A) \in \mathcal{A}$ para todo boreliano A, portanto, nesse caso faz sentido falar da probabilidade $\mathbb{P}(X \in A)$.

Demonstração do teorema 2.0.3 (∗): Devemos mostrar que

$$[f(X) \leqslant t] = \{\omega \in \Omega : f(X(\omega)) \leqslant t\} = (f \circ X)^{-1}((-\infty, t])$$

é um evento aleatório de $\mathcal{A}$ para todo real t.

O intervalo (t, ∞) é um aberto da reta portanto $f^{-1}((t, \infty))$ também é um aberto da reta, pois f é contínua e, se é aberto, então é boreliano (veja página 35) de modo que $X^{-1}(f^{-1}((t, \infty)))$ é um evento aleatório, logo seu complemento é um evento aleatório de $\mathcal{A}$, porém

$$\overline{X^{-1}(f^{-1}((t, \infty)))} = X^{-1}(\overline{f^{-1}((t, \infty))}) = X^{-1}(f^{-1}(\overline{(t, \infty)})) = X^{-1}(f^{-1}((-\infty, t]))$$

é um evento aleatório de $\mathcal{A}$.

No caso em que f é contínua por partes, f é contínua em intervalos disjuntos $L_1, L_2, \dots, L_n$ de modo que a função f é dada por $f = f_1 \circ \mathbf{1}_{L_1} + f_2 \circ \mathbf{1}_{L_2} + \cdots + f_n \circ \mathbf{1}_{L_n}$ de modo que $f(X) = f_1 \circ \mathbf{1}_{L_1}(X) + f_2 \circ \mathbf{1}_{L_2}(X) + \cdots + f_n \circ \mathbf{1}_{L_n}(X)$ e a soma de variáveis aleatórias é uma variável aleatória. □

2.0.7 Variável aleatória indicadora: denotamos por $\mathbf{1}_A$ a função indicadora da ocorrência do evento aleatório A, isto é, para todo $\omega \in \Omega$

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A; \\ 0 & \text{se } \omega \notin A. \end{cases}$$

Assim,

$$[\mathbf{1}_A \leqslant t] = \begin{cases} \Omega, & \text{se } t \geqslant 1 \\ \overline{A}, & \text{se } 0 \leqslant t < 1 \\ \emptyset, & \text{caso contrário} \end{cases}$$

de modo que essa função é uma variável aleatória. Notemos que o próprio evento pode ser dado por uma variável aleatória. Por exemplo, se X é o resultado do lançamento de um dado, então $\mathbf{1}_{[X>3]}$ vale 1 se o resultado do lançamento é maior que 3 e 0 se o resultado do lançamento é menor ou igual a 3.

Exercício 73. Considere os eventos A e B de um espaço de probabilidade. Prove que

1. $\mathbf{1}_\Omega = 1$

2. $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}$.

3. $\mathbf{1}_{\overline{A}} = 1 - \mathbf{1}_A$.

4. $A \subset B \Rightarrow \mathbf{1}_A \leqslant \mathbf{1}_B$.

5. $\mathbf{1}_{A\cap B} = \mathbf{1}_A \cdot \mathbf{1}_B$.

2.0.8 Independência de variáveis aleatórias reais: dizemos que X e Y são variáveis aleatórias independentes se os eventos $[X \leqslant a]$ e $[X \leqslant b]$ são independentes para quaisquer $a, b \in \mathbb{R}$.

2.0.9 Variável aleatória, definição geral (∗): se S é um conjunto e $\mathcal{A}_S$ uma σ-álgebra de subconjuntos de S então uma variável aleatória é uma função $X : (\Omega, \mathcal{A}) \to (S, \mathcal{A}_S)$ tal que $X^{-1}(A) \in \mathcal{A}$ para todo $A \in \mathcal{A}_S$. Uma tal função é dita *mensurável*. No caso de v.a. real ($S = \mathbb{R}$) consideramos a σ-álgebra dos borelianos ($\mathcal{A}_S = \mathcal{B}(\mathbb{R})$).

2.0.10 Observação. *Não é fácil encontrar exemplos de funções reais que não são variáveis aleatórias (i.e., que não são mensuráveis). Nos cursos introdutórios não nos preocupamos com a questão do espaço de eventos na definição de variável aleatória posta na equação equação (2.0.1); as funções $X : \Omega \to \mathbb{R}$ encontradas principalmente nas situações práticas, satisfazem o requisito necessário.*

2.1 Função de distribuição acumulada.

Seja $(\Omega, \mathcal{E}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória nesse espaço. Uma maneira simples de descrever as propriedades probabilísticas de X é dada pela sua função de distribuição acumulada.

A função de distribuição acumulada (f.d.a.) de X é a função $F_X : \mathbb{R} \to \mathbb{R}$ dada por

$$F_X(t) := \mathbb{P}(X \leqslant t).$$

É imediato da definição de f.d.a. que

$$0 \leqslant F_X(t) \leqslant 1.$$

Agora, observemos que se $x < y$ então $[X \leqslant x] \subset [X \leqslant y]$ logo $\mathbb{P}(X \leqslant x) \leqslant \mathbb{P}(X \leqslant y)$ pela proposição 1.4.7, ou seja, $F_X(x) \leqslant F_X(y)$, resumindo

$$x < y \Longrightarrow F_X(x) \leqslant F_X(y)$$

o que significa que F_X é não-decrescente. Além dessas duas propriedades, também valem

$$\lim_{t\to+\infty} F_X(t) = 1, \quad \lim_{t\to-\infty} F_X(t) = 0 \quad \text{e} \quad \lim_{t\to a^+} F_X(t) = F_X(a)$$

com as quais temos quatro propriedades que *caracterizam* funções de distribuição, i.e., qualquer função real F que satisfaz

(**F1**) $0 \leqslant F(t) \leqslant 1$ para todo $t \in \mathbb{R}$;

(**F2**) $x < y \Longrightarrow F(x) \leqslant F(y)$ para todos $x, y \in \mathbb{R}$;

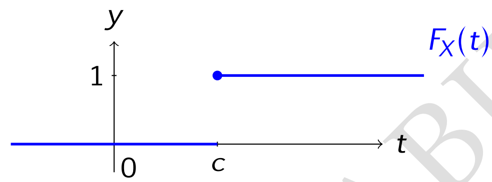
(**F3**) $\lim_{t\to+\infty} F(t) = 1$ e $\lim_{t\to-\infty} F(t) = 0$;

(**F4**) F é contínua à direta: $\lim_{t\to a^+} F(t) = F(a)$, para todo $a \in \mathbb{R}$.

é a f.d.a. de alguma variável aleatória.

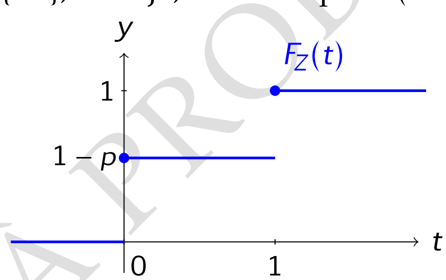
Por exemplo, no caso de variável aleatória constante, $X(\omega) = c$, algum $c \in \mathbb{R}$ e todo $\omega \in \Omega$

$$F_X(t) = \begin{cases} 0, & \text{se } t < c \\ 1, & \text{se } t \geqslant c \end{cases}$$



Exemplo 62. Consideremos uma moeda com probabilidade p de resultar Ca num lançamento. Seja $Z : \Omega \to \mathbb{R}$ a variável indicadora do evento {Ca}, ou seja, Z é dada por $Z(\text{Ca}) = 1$ e $Z(\text{Co}) = 0$.

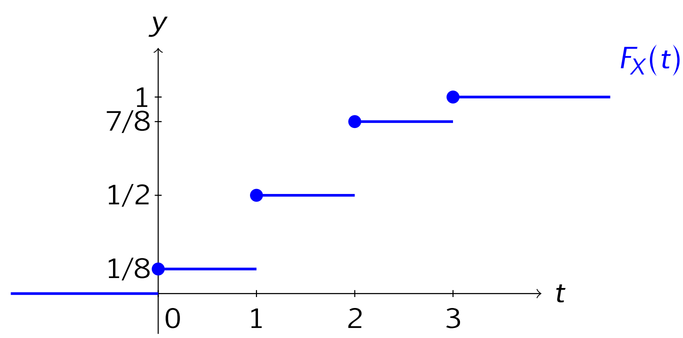
$$F_Z(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1-p & \text{se } 0 \leqslant t < 1 \\ 1, & \text{se } t \geqslant 1 \end{cases}$$



◊

Exemplo 63. No caso dos 3 lançamentos de uma moeda, se X é o número de caras

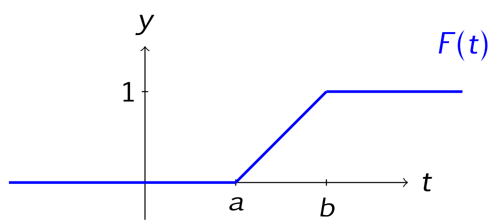
$$F_X(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1/8, & \text{se } 0 \leqslant t < 1 \\ 1/2, & \text{se } 1 \leqslant t < 2 \\ 7/8, & \text{se } 2 \leqslant t < 3 \\ 1, & \text{se } t \geqslant 3 \end{cases}$$



◊

Exemplo 64. Seja F dada por

$$F(t) = \begin{cases} 0, & \text{se } t < a \\ \frac{t-a}{b-a}, & \text{se } a \leqslant t < b \\ 1, & \text{se } t \geqslant b \end{cases}$$



F é contínua em toda a reta, portanto é continua a direita; $0 \leqslant F(t) \leqslant 1$ para todo real t; os limites quando $t \to +\infty$ e quando $t \to -\infty$ são, respectivamente, 1 e 0. Assim, essa função e uma f.d.a. de uma variável aleatória U. $\lozenge$

A variável aleatória U do exemplo acima pode ser vista como a coordenada de um ponto escolhido no intervalo $[a, b]$ no modelo geométrico clássico $([a, b], \mathcal{B}([a, b]), \mathbb{P})$. Para $[c, d] \subset [a, b]$ a probabilidade de sortear um ponto no intervalo $[c, d]$ é calculada usando a f.d.a. do seguinte modo: a probabilidade de $[X < c]$ é $F(c)$, a probabilidade de $[X > d]$ é $1 - F(d)$, um desses eventos ocorre com probabilidade

$$\mathbb{P}([X < c] \uplus [X > d]) = F(c) + 1 - F(d) = 1 - \frac{d - c}{b - a}$$

portanto, $\mathbb{P}(X \in [c, d]) = (d - c)/(b - a)$, que é o que esperávamos. No caso particular de $a = 0$ e $b = 1$ temos que $U(\omega) = \omega$ e no intervalo $[0, 1]$ a f.d.a. vale $F_U(t) = t$.

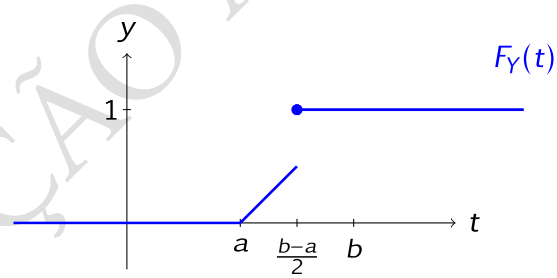
Exemplo 65. Definamos a variável aleatória Y por

$$Y(\omega) = \min\left\{U(\omega), \frac{b - a}{2}\right\}$$

em que U é a variável aleatória do exemplo anterior. Usando a definição de Y é imediato que

$$\mathbb{P}\left(Y \leqslant \frac{b - a}{2}\right) = 1.$$

Para $t < (b - a)/2$, $Y(\omega) \leqslant t$ se e só se $U(\omega) \leqslant t$, logo $F_Y(t) = F_U(t)$ e o gráfico de F_Y é



$\lozenge$

Demonstração das propriedades de F_X (∗): As propriedades nos itens **F3** e **F4** precisam do lema 1.4.13, página 87. Para provar o item **F3** definimos para cada $n \in \mathbb{N}$ a sequência crescente de conjuntos

$$A_n := [X \leqslant n]$$

cuja união para todo natural n é $[X < \infty]$ e temos

$$\lim_{n\to\infty} \mathbb{P}(X \leqslant n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \mathbb{P}(X < \infty) = 1.$$

Agora, para cada real t temos $\lfloor t \rfloor \leqslant t$, logo $\mathbb{P}(X \leqslant \lfloor t \rfloor) \leqslant F_X(t) \leqslant 1$ e como $\lfloor t \rfloor \to \infty$ quando $t \to \infty$ temos $\mathbb{P}(X \leqslant \lfloor t \rfloor) \to 1$, portanto, podemos concluir que $\lim_{t\to+\infty} F(t) = 1$. Para o outro limite definimos a sequência decrescente de conjuntos

$$A_n := [X \leqslant -n]$$

e a dedução é análoga.

Para verificar a continuidade à direita consideremos qualquer sequência t_n tal que $t_n \to a^+$ quando $t \to a^+$. Os eventos $[X \leqslant t_n]$ definem uma sequência decrescente de conjuntos e $\lim_{n\to\infty}[X \leqslant t_n] = [X \leqslant a]$. Pela continuidade de $\mathbb{P}$ vale que

$$\lim_{n\to\infty} \mathbb{P}(X \leqslant t_n) = \mathbb{P}(X \leqslant a)$$

ou $\lim_{n\to\infty} F_X(t_n) = F_X(a)$, portanto a função de distribuição é contínua pela direita. $\square$

Que essas propriedades caracterizam uma f.d.a., pode ser provado de modo relativamente fácil no seguinte caso particular.

2.1.1 Proposição. *Se F é uma função que satisfaz as quatro propriedades enunciadas na página 132, itens **F1**, **F2**, **F3** e **F4** e, além disso, contínua e crescente, então existe uma variável aleatória X sobre $([0,1], \mathcal{B}([0,1]), \mathbb{P})$ (o modelo definido na página 94) com $F_X = F$*

Demonstração. Por ser crescente $\mathfrak{Im}(F) = (0,1)$ e F é invertível. Para cada $\omega \in (0,1)$, definimos $X(\omega) = F^{-1}(\omega)$ e temos $F(X(\omega)) = F(F^{-1}(\omega)) = \omega$. Notemos que $F(X)$ é a variável aleatória denotada por U no exemplo 64 no intervalo $(0,1)$ da reta, cuja f.d.a. é $F_U(t) = t$ para todo $t \in (0,1)$ (leia o parágrafo posterior ao exemplo).

Vale que $X(\omega) \leqslant t$ se, e só se, $F(X(\omega)) \leqslant F(t)$, portanto, $[X \leqslant t]$ é um evento aleatório, ou seja, X é uma variável aleatória real. Então $F_X(t) = \mathbb{P}(X \leqslant t) = \mathbb{P}(F(X) \leqslant F(t)) = F(t)$. $\square$

No caso geral, definimos uma medida de probabilidade sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ da seguinte maneira: dada uma f.d.a. $F : \mathbb{R} \to \mathbb{R}$, consideramos a semiálgebra $\mathcal{S}$ dos semiabertos (exemp. 0.6.1, pág. 31) e definimos

$$P((a,b]) := F(b) - F(a)$$

em que adotamos $F(-\infty) = 0$ e $F(+\infty) = 1$ para simplificar a notação. Pode-se provar que P é enumeravelmente aditiva e, ainda, notemos que $\mathbb{P}(\mathbb{R}) = 1$, portanto, pelo teorema 1.6.4 temos uma medida de probabilidade $\mathbb{P}$ sobre o espaço mensurável $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Resta notar que $X : \mathbb{R} \to \mathbb{R}$ é uma v.a. desse espaço com $F_X = F$.

2.1.2 Mais propriedades de uma f.d.a.: A função de distribuição acumulada tem várias propriedades que ajudam no cálculo de probabilidades, algumas são dadas na proposição abaixo. Usamos a notação

$$F(a-) := \lim_{t \to a^-} F(t).$$

2.1.3 Proposição. *A função de distribuição acumulada F de uma variável aleatória X satisfaz*

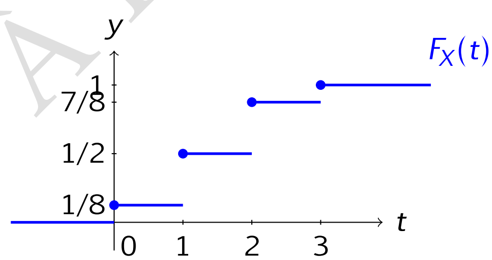
1. $1 - F(t) = \mathbb{P}(X > t)$, *para todo* $t \in \mathbb{R}$;
2. $F(b) - F(a) = \mathbb{P}(a < X \leqslant b)$, *para quaisquer* $a < b$;
3. $F(a) - F(a-) = \mathbb{P}(X = a)$, *para todo real* a.

2.1.4 Corolário. *F é contínua em $a \in \mathbb{R}$ se, e só se,* $\mathbb{P}(X = a) = 0$.

As descontinuidades das funções de distribuição acumulada são do tipo salto. Se F não é contínua em a então o salto em a é de grandeza $\mathbb{P}(X = a) = F(a) - F(a-)$. No exemplo abaixo, o salto em $t = 2$ é $7/8 - 1/2 = 3/8 = \mathbb{P}(X = 2)$. Notemos que a soma dos saltos de tamanho $\geqslant 1/n$ não deve ser maior que 1, portanto, há no máximo n desses saltos; desse fato podemos concluir que há no máximo um número enumerável de pontos de descontinuidade em qualquer função de distribuição acumulada.

Exemplo 66. No caso dos 3 lançamentos de uma moeda, X é o número de caras

$$F_X(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1/8, & \text{se } 0 \leqslant t < 1 \\ 1/2, & \text{se } 1 \leqslant t < 2 \\ 7/8, & \text{se } 2 \leqslant t < 3 \\ 1, & \text{se } t \geqslant 3 \end{cases}$$

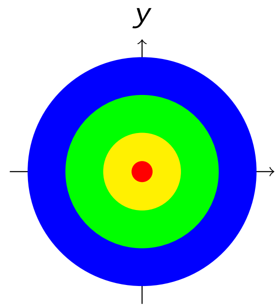


Usando as propriedades de uma função de distribuição temos, por exemplo

- $\mathbb{P}(X > 2) = 1 - F(2) = 1 - 7/8 = 3/8$;
- $\mathbb{P}(X > 3) = 1 - F(3) = 0$;
- $\mathbb{P}(0{,}5 < X \leqslant 2{,}5) = F(2{,}5) - F(0{,}5) = 7/8 - 1/8 = 3/4$;
- $\mathbb{P}(X = 1) = F(1) - F(1-) = 1/2 - 1/8 = 3/8$;
- $\mathbb{P}(X = 1{,}8) = F(1{,}8) - F(1{,}8-) = 1/2 - 1/2 = 0$;
- $\mathbb{P}(X = -1) = F(-1) - F(-1-) = 0 - 0 = 0$;
- $\mathbb{P}(X = 7) = F(7) - F(7-) = 1 - 1 = 0$.

$\lozenge$

Exemplo 67. Consideremos o exemplo 34 em que um dardo acerta aleatoriamente um alvo composto de círculos concêntricos de raios 1/4, 1, 2 e 3 que supomos centrados na origem de um sistema cartesiano.

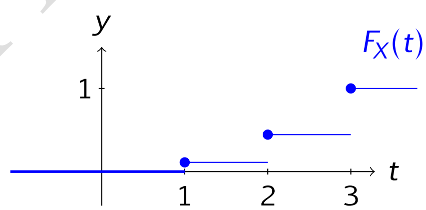


Para cada $k = 1, 2, 3$, considere as regiões A_k da figura acima do seguinte modo:

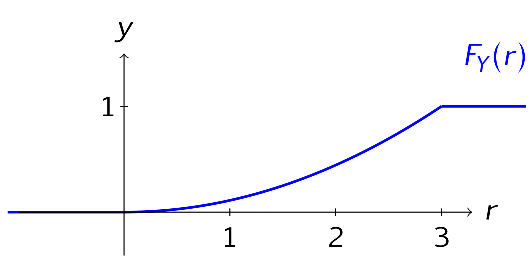
$$\begin{aligned} A_1 &= \left\{(x, y)\colon x^2 + y^2 \leqslant 1\right\} \text{ (Vermelho + Amarelo)} \\ A_2 &= \left\{(x, y)\colon 1 < x^2 + y^2 \leqslant 4\right\} \quad \text{(Verde)} \\ A_3 &= \left\{(x, y)\colon 4 < x^2 + y^2 \leqslant 9\right\} \quad \text{(Azul)} \end{aligned}$$

Se $X(\omega) = k$ quando $\omega \in A_k$ então $\mathbb{P}(X = k) = \mathbb{P}(A_k) = (2k - 1)/9$ e X tem função de distribuição acumulada

$$F_X(t) = \begin{cases} 0, & \text{se } t < 1 \\ \frac{\lfloor t \rfloor^2}{9}, & \text{se } 1 \leqslant t \leqslant 3 \\ 1, & \text{se } t > 3 \end{cases}$$



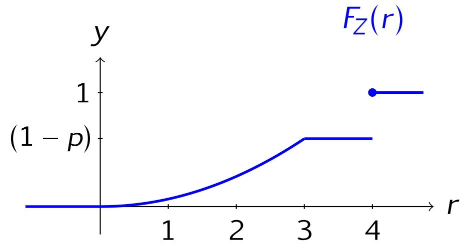
Se $\omega = (x, y) \in \Omega$ e $Y(\omega) = \sqrt{x^2 + y^2}$ é a distância do ponto atingido ao centro, a função de distribuição acumulada de Y é $F_Y(r) = r^2/9$ se $0 \leqslant r \leqslant 3$



Agora, suponha que o jogador erre o alvo com probabilidade p, para algum $p > 0$ fixo, e caso acerte então vale a equação equação (1.3.1). A pontuação Z é, caso acerte, 1 se acertou o verde, 2 o azul e 3 o vermelho ou, caso erre o alvo, 4. Então

$$\mathbb{P}(Z \leqslant r) = \mathbb{P}(Z \leqslant r \mid \text{acertou})\mathbb{P}(\text{acertou}) + \mathbb{P}(Z \leqslant r \mid \text{errou})\mathbb{P}(\text{errou})$$

$$F_Z(r) = \begin{cases} 0 & \text{se } r < 0, \\ (1-p)F_Y(r) & \text{se } 0 \leqslant r < 4, \\ 1 & \text{se } r \geqslant 4. \end{cases}$$



◊

Demonstração da proposição 2.1.3(∗): Os dois primeiros itens são deixados como exercício. O último item precisa do lema 1.4.13, página 87. Temos

$$[X = a] = \bigcup_{n=1}^{\infty} \left[a - \frac{1}{n} < X \leqslant a \right]$$

e por continuidade

$$\begin{aligned} \mathbb{P}(X = a) &= \lim_{n\to\infty} \mathbb{P}\left(a - \frac{1}{n} < X \leqslant a \right) \\ &= \lim_{n\to\infty} \left(F(a) - F(a - \frac{1}{n}) \right) \\ &= F(a) - \lim_{n\to\infty} \left(F(a - \frac{1}{n}) \right). \end{aligned}$$

que é $F(a) - F(a-)$. □

2.1.5 Lei de uma variável aleatória real (∗): se X é uma variável aleatória do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ então podemos definir um espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ na reta em que $\mathcal{B} = \mathcal{B}(\mathbb{R})$ é o espaço de eventos gerado pelos intervalos $(-\infty, x]$ da reta (os borelianos) e $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$, ou seja

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

para todo boreliano A, chamada lei (ou distribuição) da variável aleatória X. A distribuição da variável aleatória transfere a estrutura probabilística sobre um espaço abstrato $(\Omega, \mathcal{A})$ para o espaço mais conhecido $(\mathbb{R}, \mathcal{B})$.

Se F_X é a função de distribuição acumulada de X então vale

$$F_X(t) = \mathbb{P}_X((-\infty, t]) \tag{2.1.1}$$

para todo real t. A lei $\mathbb{P}_X$ é uma função que atribui um número aos borelianos de $\mathbb{R}$ enquanto que a função de distribuição F_X é uma função real, um objeto mais simples de descrever e que incorpora a mesma informação. Por exemplo, a partir da equação (2.1.1) nós podemos calcular

$$\mathbb{P}_X((a, b]) = F_X(b) - F_X(a)$$

para qualquer intervalo semiaberto, portanto, pelo construção dos borelianos, $\mathbb{P}_X$ pode ser definida de modo único em qualquer evento aleatório.

Notemos também que se X e Y são variáveis aleatórias tais que $F_X(t) = F_Y(t)$ para todo t, então $\mathbb{P}_X$ e $\mathbb{P}_Y$ coincidem nos intervalos $(-\infty, t]$, para todo t, portanto, $\mathbb{P}_X = \mathbb{P}_Y$ pelo teorema 1.6.5.

Por outro lado, se $\mathbb{P}$ é uma medida de probabilidade sobre $(\mathbb{R}, \mathcal{B})$ então

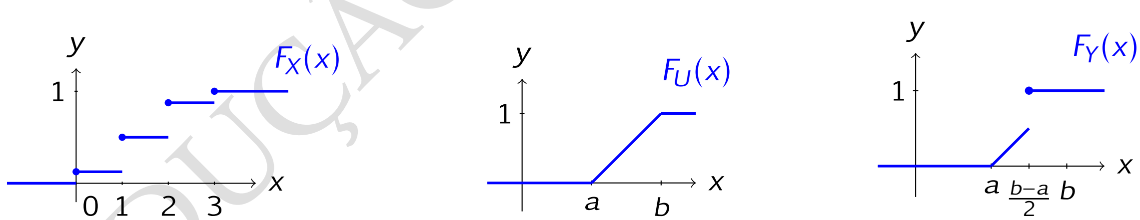
$$F(t) := \mathbb{P}((-\infty, t]) \tag{2.1.2}$$

é uma função de distribuição acumulada. De fato, F definida em 2.1.2 determina $\mathbb{P}$ de modo único. Isso ocorre pois se $\mathbb{Q}$ é uma medida de probabilidade tal que $F(t) = \mathbb{Q}((-\infty, t])$ então o teorema 1.6.5 (pág. 94) garante que $\mathbb{P} = \mathbb{Q}$.

Exercício 74. Seja $\mathbb{P}$ uma medida de probabilidade na reta e F dada por equação (2.1.2). Prove que valem as seguintes identidades

1. $\mathbb{P}((x, y]) = F(y) - F(x)$;
2. $\mathbb{P}([x, y]) = F(y) - F(x-)$;
3. $\mathbb{P}([x, y)) = F(y-) - F(x-)$;
4. $\mathbb{P}((x, y)) = F(y-) - F(x)$;
5. $\mathbb{P}(\{x\}) = F(x) - F(x-)$;

2.1.6 Variáveis aleatórias discretas e contínuas: Lembrando os gráficos das f.d.a.'s dos exemplos 63, 64 e 65



o primeiro corresponde a uma variável aleatória discreta, o segundo a uma variável aleatória contínua e o terceiro corresponde a uma variável aleatória que não é discreta nem contínua.

Variável aleatória discreta: é uma variável aleatória X que assume valores num conjunto enumerável (finito ou infinito) $I = \{x_0, x_1, \dots\} \subset \mathbb{R}$. A função dada por $p_X(x_i) := \mathbb{P}(X = x_i)$ para todo $x_i \in I$ é chamada *função de massa de probabilidade* ou, simplesmente, *função de probabilidade* de X.

Claramente,

$$\sum_{x_i \in I} \mathbb{P}(X = x_i) = 1$$

e qualquer evento que envolve X tem sua probabilidade determinada pelos valores $p_X(x_i)$. A função de distribuição acumulada é dada por

$$F_X(t) = \sum_{i:x_i \leqslant t} p_X(x_i).$$

Ademais, F_X é uma função escada com saltos que ocorrem nos pontos $x_i \in I$.

Variável aleatória (absolutamente) contínua: é uma variável aleatória X para a qual existe uma função $f_X : \mathbb{R} \to [0, \infty)$ tal que

$$F_X(t) = \int_{-\infty}^{t} f_x(x)\,\mathrm{d}x$$

para todo $t \in \mathbb{R}$. A função dada por f_X é chamada *função de densidade de probabilidade* ou, simplesmente, *densidade* de X.

Claramente, a integral de f_X em $\mathbb{R}$ é igual a 1. Reciprocamente, uma função f não negativa cuja integral em $\mathbb{R}$ é igual a 1 é densidade de alguma v.a. pois se definirmos $F : \mathbb{R} \to \mathbb{R}$ por $F(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ valem as propriedades **F1**, **F2**, **F3** e **F4** que caracterizam uma f.d.a.

Sendo F_X a integral indefinida de uma função ela é uma função contínua, de fato F_X é *absolutamente contínua*[1] e, portanto, $F_x' = f_X$ em quase todo ponto[2]. Na prática, podemos verificar que X admite uma densidade se F_X é contínua e se tem derivada no interior de uma quantidade enumerável de intervalos fechados cuja união é $\mathbb{R}$.

Nem discreta, nem contínua: uma mistura de ambas, é o caso da variável aleatória Y dada por $Y(\omega) = \min\{Z(\omega), (b-a)/2\}$ apresentada no exemplo 65 temos

$$\mathbb{P}\left(Y = \frac{b-a}{2}\right) = F_Y\left(\frac{b-a}{2}\right) - F_Y\left(\frac{b-a^-}{2}\right) = \frac{a+b}{2(b-a)}.$$

Se Y fosse uma variável aleatória contínua esse valor deveria ser 0 (por quê?) Essa variável aleatória não é discreta porque sua imagem não é enumerável. Essa variável é dita do tipo *mista*. Ainda, é possível ocorrer de uma v.a. não ser de nenhum dos três tipos; possivelmente, o caso mais conhecido é da *função de Cantor*[3] Uma variável aleatória cuja distribuição é a função de Cantor é dita *singular*:

[1] g é absolutamente contínua em $[a, b]$ se para todo $\varepsilon > 0$, existe $\delta > 0$, para toda coleção finita de intervalos disjuntos $(x_1, y_1), \dots, (x_n, y_n)$

$$\sum_{i=1}^{n} |x_i - y_i| \Rightarrow \sum_{i=1}^{n} |g(x_i) - g(y_i)| < \delta.$$

(1) g é absolutamente contínua se e só se for dada por uma integral indefinida [Teo. 13 do cap. 5 de Royden, *Real Analysys*].

(2) Se g é absolutamente contínua em $[a, b]$ então tem derivada em quase todo ponto de $[a, b]$ [Cor. 11 do cap. 5 de Royden, *Real Analysys*].

[2] Ou seja, o conjuntos dos pontos t tais que $F_x'(t) \neq f_X(t)$ tem medida nula, o que significa que para qualquer $\varepsilon > 0$ (por menor que seja) tal conjunto está contido numa reunião de intervalos de comprimento total menor que ε.

[3] seção 2.2 de Barry James, *Probabilidade: um curso de nível intermediário*.

X é singular se F_X é contínua (não absolutamente) e $F_X(t) = 0$ em quase todo ponto. Informalmente falando, *toda variável aleatória é uma mistura de discreta, absolutamente contínua e singular.*

Na próxima seção apresentamos uma ferramenta que permite um tratamento unificado de variáveis discretas e contínuas.

2.2 Variáveis aleatórias discretas.

Uma variável aleatória X é discreta se assume valores num conjunto $\mathrm{Im}(X) \subset \mathbb{R}$ discreto (enumerável finito ou infinito). A função de distribuição acumulada é dada por

$$F_X(t) = \sum_{x \in \mathrm{Im}(X): x \leqslant t} \mathbb{P}(X = x)$$

em que a soma é sobre todo $x \in \mathrm{Im}(X)$ tal que $x \leqslant t$.

Função de massa de probabilidade: (f.m.p.) ou, simplesmente, função de probabilidade da variável aleatória X é a função $f_X : \mathbb{R} \to [0,1]$ dada por

$$f_X(a) := \mathbb{P}(X = a)$$

de modo que

$$\sum_{a \in \mathrm{Im}(X)} f_X(a) = 1.$$

e para qualquer $S \subset \mathbb{R}$

$$\mathbb{P}(X \in S) = \sum_{a \in S} f_X(a).$$

Por exemplo, no caso do lançamento de 3 moedas, o número de caras tem função de probabilidade

$$f_X(a) = \begin{cases} 0 & \text{se } a \neq 0,1,2,3 \\ 3/8 & \text{se } a = 1 \text{ ou } x = 2 \\ 1/8 & \text{se } a = 0 \text{ ou } x = 3. \end{cases}$$

Exemplo 68. Dado um inteiro positivo n, lançamos uma moeda com probabilidade até que resulte cara ou complete n lançamentos.

A probabilidade de cara é $p \in (0,1)$ e os resultados dos lançamentos são independentes. O número de lançamentos é uma variável aleatória que denotamos por X. Qual a probabilidade de $[X = k]$?

Cada resultado possível com $k < n$ lançamentos tem probabilidade $\mathbb{P}(X = k) = p(1-p)^{k-1}$ e são dois resultado possíveis com n lançamentos, um que termina com cara e outro que termina com coroa, portanto $\mathbb{P}(X = n) = (1-p)^{n-1}$. ◊

Exercício 75. Lançamos uma moeda equilibrada n vezes, independentemente. O número de caras é uma variável aleatória que denotamos por X. Determine $\mathbb{P}(X = k)$.

2.2.1 Principais modelos discretos:

Distribuição uniforme discreta: dado S de cardinalidade k dizemos que X tem distribuição uniforme discreta em S, fato denotado por $X \sim \mathrm{U}(k)$, se

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

Distribuição de Bernoulli:: na prática, ocorrem muitas situações com experimentos para os quais nos interessa apenas dois resultados, por exemplo

1. uma peça é classificada como *boa* ou *defeituosa*;
2. o resultado de um exame médico é *positivo* ou *negativo*;
3. um paciente submetido a um tratamento é *curado* ou *não* da doença;
4. um entrevistado *concorda* ou *não concorda* com a afirmação feita;
5. no lançamento de um dado *ocorre* ou *não ocorre* a face "5".

Nessas situações podemos representar, genericamente, os resultados do experimento com o espaço amostral $\Omega = \{\text{sucesso}, \text{fracasso}\}$ e o modelo probabilístico fica determinado dado $p = \mathbb{P}(\text{sucesso}) \in [0,1]$. Esses experimentos recebem o nome de *Ensaios de Bernoulli* e a variável aleatória indicadora do evento "sucesso" é uma *variável aleatória de Bernoulli* com parâmetro p.

A notação

$$X \sim \text{Bernoulli}(p)$$

indica que X é uma variável aleatória de Bernoulli com parâmetro p; ela assume dois valores: 1 se ocorre sucesso e 0 se ocorre fracasso; sua f.m.p. é dada por

$$\text{be}_p(x) = \begin{cases} p^x(1-p)^{1-x} & \text{se } x \in \{0,1\} \\ 0 & \text{nos outros casos.} \end{cases}$$

Distribuição binomial:: consideremos n repetições *independentes* de um Ensaio de Bernoulli. Seja X o número de sucessos nas repetições.

Por exemplo, um dado equilibrado é lançado 3 vezes. Assumindo independência dos resultados, qual é a probabilidade de se obter a face 5 duas vezes? Se S denota *sucesso*, i.e., "ocorre face 5" e F denota *fracasso*, "não ocorrer face 5" então podemos associar a cada resposta do experimento um elemento de $\Omega = \{SSS, SSF, SFS, FSS, SFF, FSF, FFS, FFF\}$ onde atribuímos $p = \mathbb{P}(S) = 1/6$ e $1 - p = \mathbb{P}(F) = 5/6$. A função de massa de probabilidade para o número X de sucessos é, usando independência das respostas,

ω	x	$f_X(x)$
FFF	0	$(1-p)^3$
SFF, FSF, FFS	1	$3p(1-p)^2$
SSF, SFS, FS	2	$3p^2(1-p)$
SSS	3	p^3

e podemos escrever essa função como $f_X(x) = \binom{3}{x}p^x(1-p)^{3-x}$ para todo $x \in \{0, 1, 2, 3\}$ e $f_X(x) = 0$ nos outros casos. Assim, $\mathbb{P}(X = 2) = \binom{3}{2}p^2(1-p)^{3-2} = 0{,}0694$.

Uma *variável aleatória binomial* de parâmetros $n \in \mathbb{N}$ e $p \in (0, 1)$ é uma variável aleatória com f.m.p.

$$\mathrm{bi}_{n,p}(x) = \begin{cases} \binom{n}{x}p^x(1-p)^{n-x}, & \text{se } x \in \{0, 1, 2, \dots, n\} \\ 0 & \text{nos outros casos} \end{cases}$$

que pode ser vista como o número de sucessos em n ensaios independentes de Bernoulli e com mesma probabilidade p de sucesso. A notação

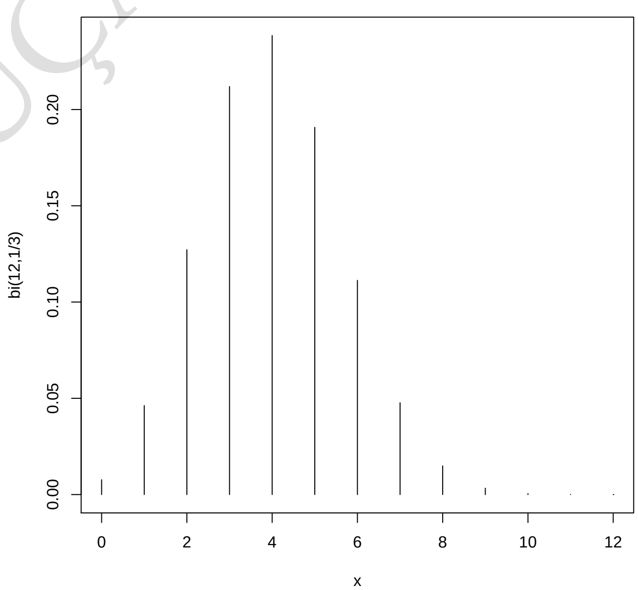
$$X \sim \mathrm{Binomial}(n, p)$$

indica que X é uma variável aleatória com distribuição binomial com parâmetros n e p.

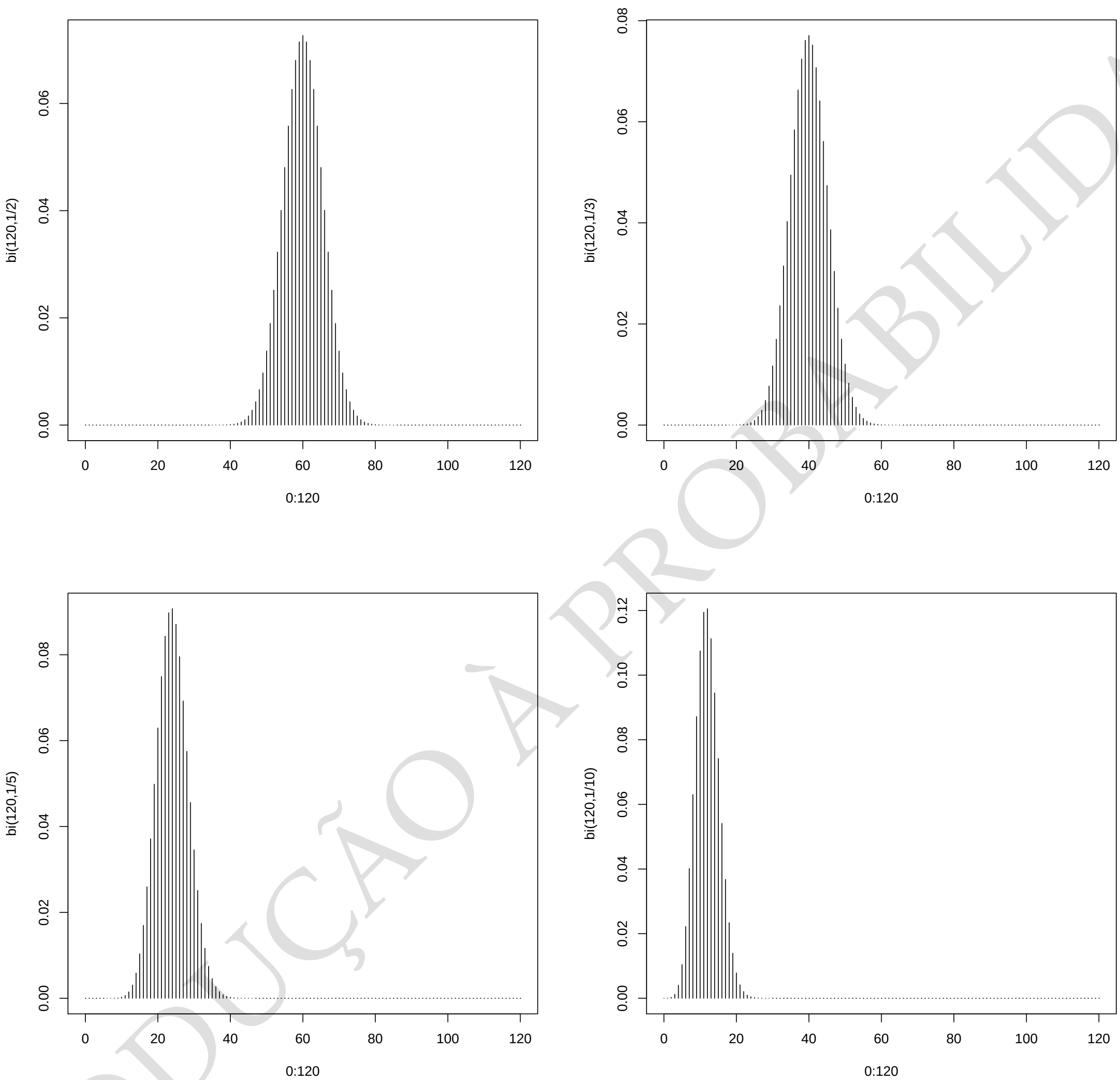
Exemplo 69. Numa prova com 12 questões de múltipla escolha, com 3 alternativas, se todas as respostas forem escolhidas aleatoriamente, então o número de acertos é $X \sim \mathrm{Binomial}(12, \frac{1}{3})$ e a função de massa de probabilidade é

$$\mathrm{bi}_{12,\frac{1}{3}}(x) = \binom{12}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{n-x}$$

tem o seguinte *gráfico de barras*



A seguir, respectivamente, os gráficos mostram a distribuição nos casos de uma prova com 120 questões e 2 alternativas, uma prova com 120 questões e 3 alternativas e uma prova com 120 questões e 5 alternativas e, finalmente, 10 alternativas



◊

Exemplo 70. Três bolas são aleatoriamente retiradas de um saco com 20 bolas numeradas de 1 a 20. Supondo que todas as $\binom{20}{3}$ seleções são equiprováveis, com que probabilidade pelo menos uma tem número maior ou igual a 18?

Se a maior bola selecionada é maior ou igual a 18, então menos uma tem número maior ou igual a 18 e vice-versa. Seja X a maior bola selecionada; X é uma variável aleatória. Queremos determinar com que probabilidade ocorre $[X = 18]$ ou $[X = 19]$ ou $[X = 20]$, ou seja

$$\mathbb{P}(X \geqslant 18) = \mathbb{P}([X = 18] \cup [X = 19] \cup [X = 20]) = \mathbb{P}(X = 18) + \mathbb{P}(X = 19) + \mathbb{P}(X = 20)$$

a igualdade vem do fato dos eventos serem independentes. A quantidade de seleções de três bolas de modo que a maior seja igual a i é $\binom{i-1}{2}$ pois há $i-1$ bolas menores, das quais escolhemos 2. Portanto

$$\mathbb{P}(X = i) = \frac{\binom{i-1}{2}}{\binom{20}{3}}$$

portanto

$$\mathbb{P}(X \geqslant 18) = \frac{\binom{17}{2}}{\binom{20}{3}} + \frac{\binom{18}{2}}{\binom{20}{3}} + \frac{\binom{19}{2}}{\binom{20}{3}} \approx 0{,}403.$$

◊

Exemplo 71. Um equipamento resiste a um teste de choque com probabilidade 3/4. Qual é probabilidade de que em 4 equipamentos testados 2 equipamentos sobrevivam ao choque? Se $X \sim \mathrm{Binomial}(4, \frac{3}{4})$

$$\mathrm{bi}_{4,\frac{3}{4}}(2) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \frac{27}{128} \approx 0{,}21.$$

◊

Exemplo 72. Uma febre atinge 25% dos rebanhos bovinos. Para tratamento foram desenvolvidas três vacinas, que denominaremos $V1$, $V2$ e $V3$ e que foram inoculadas em 10, 17 e 23 animais, respectivamente. Após um período de observação, o número desses animais que ficaram doentes foram, respectivamente, 0, 1 e 2. Há, nessas informações alguma evidência de eficácia das vacinas?

Consideremos três grupos dos mesmos tamanhos compostos por animais não vacinados e vamos calcular a probabilidade dessas populações se saírem tão bem quanto as vacinadas. A probabilidade de um indivíduo ficar doente é 1/4 e se X é a quantidade de animais que ficam doente em um rebanho de tamanho n então $X \sim \mathrm{Binomial}(n, \frac{1}{4})$ e a probabilidade com que x animais fiquem doentes é

$$\mathrm{bi}_{n,\frac{1}{4}}(x) = \binom{n}{x}(1/4)^x(3/4)^{n-x}$$

e para as populações consideradas

	população	proporção de doentes
1.	$n = 10,\ x = 0$	$\mathbb{P}(X = 0) = 0{,}05631351$
2.	$n = 17,\ x = 1$	$\mathbb{P}(X \leqslant 1) = 0{,}05011298$
4.	$n = 23,\ x = 2$	$\mathbb{P}(X \leqslant 2) = 0{,}04920334$

Portanto, a chance de um rebanho não vacinado ser sair melhor que um vacinado é aproximadamente 5%. A vacina $V3$ se sai ligeiramente melhor que as outras. ◊

Exemplo 73. 4 ◊

Exemplo 74. Um fabricante garante que seu produto tem uma taxa de itens defeituosos de 3%. Numa seleção de 20 itens a serem inspecionados, qual é a probabilidade de ao menos um ser defeituoso? Se X é a quantidade de itens defeituosos

$$\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - \binom{20}{0}(0{,}03)^0(0{,}97)^{20} = 0{,}4562.$$

Se 10 carregamentos por mês deixam a fabrica e de cada carregamento 20 itens são inspecionados, com que probabilidade 3 carregamentos tem pelo menos um item defeituoso?

$$\binom{10}{3}(0{,}4562)^3(1 - 0{,}4562)^7 = 0{,}1602.$$

◊

Seja $X \sim \text{Binomial}(n, p)$. Para n e p fixos, quando x varia de 0 a n o valor de $\text{bi}_{n,p}(x)$ cresce monotonicamente e depois decresce monotonicamente. De fato,

$$\frac{\text{bi}_{n,p}(x)}{\text{bi}_{n,p}(x-1)} = \frac{(n-x+1)p}{(1-p)x} \tag{2.2.1}$$

que é crescente se $(n - x + 1)p > (1 - p)x$ ou, equivalentemente, $x < (n + 1)p$.

2.2.2 Proposição. *Seja* $X \sim \text{Binomial}(n, p)$*. Para* n *e* p *fixos, quando* x *varia de* 0 *a* n *o valor de* $\text{bi}_{n,p}(x)$ *cresce monotonicamente e depois decresce monotonicamente atingindo o máximo quando* $x = \lfloor (n+1)p \rfloor$. □

Distribuição de Poisson:: a f.m.p. de uma variável aleatória de Poisson expressa a probabilidade de ocorrência de um determinado número de eventos num intervalo de tempo fixo (ou região do espaço), se estes eventos ocorrem com uma taxa média conhecida e independentemente do tempo desde a última ocorrência. Por exemplo:

1. o número de erros de impressão numa página de livro;
2. o número de chamadas que chega a um *call center*;
3. o número de partículas α descarregadas por um material radioativo em um período fixo de tempo.

Há vários exemplos curiosos de fenômenos aleatórios com essa distribuição na literatura. Começaremos com o seguinte exemplo do livro do Feller: Na segunda guerra mundial a cidade de Londres foi intensamente bombardeada pelos alemães. Para determinar se as bombas tinham um alvo ou foram lançadas aleatoriamente os ingleses dividiram o sul da cidade em 576 pequenas regiões, de tamanho 0,25 km^2. O total de bombas que atingiram a região foi 537, o que dá uma taxa de 0,9323 bombas por região; se n_k é o número de regiões que receberam k bombas, a contagem foi

k	0	1	2	3	4	5 ou mais
n_k	229	211	93	35	7	1

ao qual o modelo de Poisson se ajusta impressionantemente bem, o que levou-os a acreditar que o bombardeio foi aleatório.

William Sealy Gosset, um químico e matemático formado em Oxford, foi contratado, em 1899, pela famosa cervejaria *Arthur Guinness and Son* em Dublin; sua tarefa era para aperfeiçoar o processo de produção de cerveja. Gosset (que publicou artigos sob o pseudônimo de *Student*, porque o seu empregador proibiu publicações por funcionários depois que um colega de trabalho havia divulgado segredos comerciais) trabalhou com o modelo de Poisson para a contagem de células de levedura. No artigo sobre tal trabalho, Gosset discutiu "como a dispersão nas contagens de colonias de levedura foi semelhante ao limite exponencial da distribuição binomial".

Outra aplicação curiosa desta distribuição foi feita por Ladislau Bortkiewicz em 1898, quando lhe foi dada a tarefa de investigar o número de soldados no exército russo morto acidentalmente por coice de cavalo.

A notação

$$X \sim \text{Poisson}(\lambda)$$

indica que X é uma *variável aleatória de Poisson* com parâmetro $\lambda > 0$, ela conta o número de ocorrências de um determinado evento que ocorre a uma taxa λ e cuja f.m.p. é dada por

$$\text{po}_\lambda(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \text{se } x \in \mathbb{N} \\ 0, & \text{nos outros casos.} \end{cases}$$

A distribuição de Poisson pode ser derivada como um caso limite para a distribuição binomial quando o número de ensaios tende ao infinito e a taxa média de ocorrências permanece fixa. Por isso, pode ser usado como uma aproximação da distribuição binomial se n for suficientemente grande e p suficientemente pequena. A taxa de ocorrência de um evento está relacionada com a probabilidade de um evento ocorrer em pequenos subintervalos, por pequeno entendemos o suficiente para que a probabilidade de um evento ocorrer duas vezes nesse intervalo é insignificante. Dividimos o intervalo inteiro em n subintervalos $I_1, \dots, I_n$ de igual tamanho com $n > \lambda$. Isto significa que a probabilidade de ocorrência do evento em um intervalo I_k, para cada k é igual a λ/n. Agora, assume-se que as ocorrências do evento em todo o intervalo pode ser visto como n ensaios de Bernoulli com parâmetro λ/n. Suponha que em n ensaios independentes de Bernoulli com probabilidade de sucesso $p = p(n) = \lambda/n$, para uma constante $\lambda > 0$. A probabilidade de x sucessos é $\text{bi}_{n,\frac{\lambda}{n}}(x)$. Para $x = 0$

$$\text{bi}_{n,\frac{\lambda}{n}}(0) = \left(1 - \frac{\lambda}{n}\right)^n$$

portanto, para n suficientemente grande $\mathrm{bi}_{n,p}(0) \approx \mathrm{e}^{-\lambda}$ (pela definição de e como limite de uma sequência). Agora, usando equação (2.2.1) podemos concluir que

$$\begin{aligned}
\mathrm{bi}_{n,\frac{\lambda}{n}}(1) &\approx \mathrm{bi}_{n,\frac{\lambda}{n}}(0)\lambda &&\approx \lambda\mathrm{e}^{-\lambda}\\
\mathrm{bi}_{n,\frac{\lambda}{n}}(2) &\approx \mathrm{bi}_{n,\frac{\lambda}{n}}(1)\lambda/2 &&\approx (\lambda^2/2)\mathrm{e}^{-\lambda}\\
\mathrm{bi}_{n,\frac{\lambda}{n}}(3) &\approx \mathrm{bi}_{n,\frac{\lambda}{n}}(2)\lambda/6 &&\approx (\lambda^3/3!)\mathrm{e}^{-\lambda}
\end{aligned}$$

que podemos estender usando indução para

$$\mathrm{bi}_{n,\frac{\lambda}{n}}(x) \approx \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!} = \mathrm{po}_\lambda(x)$$

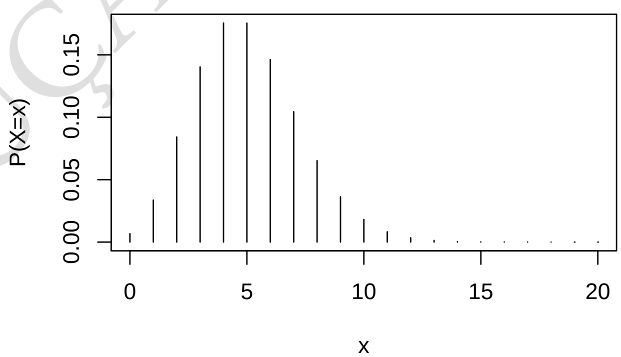
que é conhecido como a *aproximação de Poisson para a distribuição binomial*; em resumo, fixado λ e fixado x, se $Y_n \sim \mathrm{Binomial}(n, \lambda/n)$ e $X \sim \mathrm{Poisson}(\lambda)$, então

$$\mathbb{P}(X = x) = \lim_{n\to\infty} \mathbb{P}(Y_n = x)$$

e uma prova pode ser vista neste link.

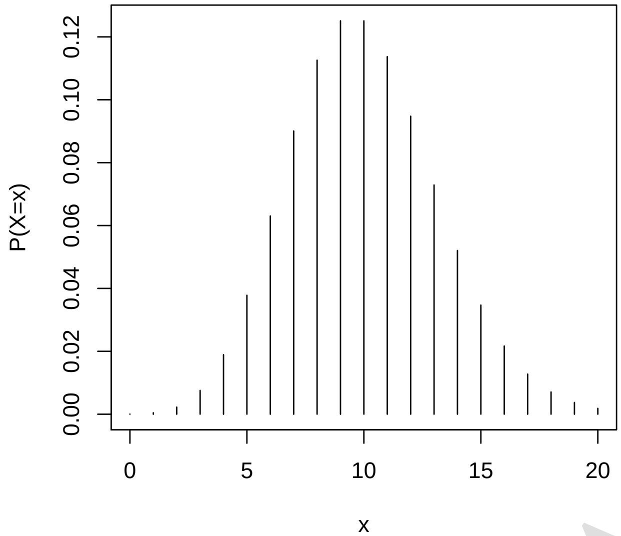
Exemplo 75. Um telefone recebe em média 5 chamadas por minuto. Supondo que a distribuição de Poisson seja um modelo adequado para essa situação, qual a probabilidade com que o telefone não receba chamadas durante um intervalo de 1 minuto?

$$\mathrm{po}_5(0) = \frac{5^0\mathrm{e}^{-5}}{0!} = 0{,}0067$$



◊

Exemplo 76. O número de partículas que contaminam a superfície de um CD no processo de fabricação tem distribuição de Poisson. O número médio de partículas é 0,1 particulas/cm^2 e a área de um CD é 100cm^2. Seja X o número de partículas num CD; $X \sim \mathrm{Poisson}(10)$.



1. a probabilidade de ter 12 partículas é

$$\mathbb{P}(X = 12) = \frac{\mathrm{e}^{-10}10^{12}}{12!} = 0{,}095$$

2. a probabilidade de ter 0 partículas é

$$\mathbb{P}(X = 0) = \frac{\mathrm{e}^{-10}10^{0}}{0!} = 4{,}54 \times 10^{-5}$$

3. a probabilidade de ter $\leqslant 12$ partículas é

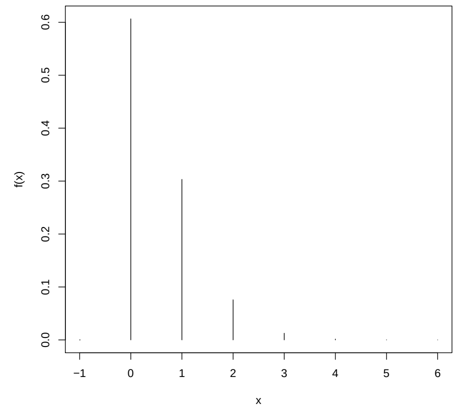
$$\mathbb{P}(X \leqslant 12) = \sum_{x \leqslant 12} \frac{\mathrm{e}^{-10}10^{x}}{x!} = 0{,}792$$

◊

Exemplo 77. Suponha que essas notas tenham erros tipográficos por página que segue uma distribuição de Poisson com $\lambda = 1/2$. Qual é a probabilidade de haver pelo menos um erro nessa página? Se X é o número de erros por página

$$\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - \mathrm{e}^{-1/2} \approx 0{,}393.$$

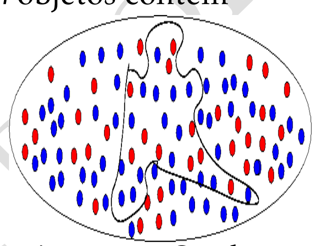
◊



Exercício 76. Prove que a f.m.p. de $X \sim \text{Poisson}(\lambda)$ satisfaz

$$\text{po}_\lambda(x+1) = \frac{\lambda}{x+1}\text{po}_\lambda(x).$$

Distribuição hipergeométrica:: uma coleção de n objetos contém



1. a objetos azuis,
2. $n-a$ objetos vermelhos.

Uma amostra com s elementos é selecionada aleatoriamente. Qual a probabilidade da amostra conter x ($x \leqslant a$) bolas azuis? O número de bolas azuis é uma variável aleatória hipergeométrica. Uma *variável aleatória hipergeométrica* com parâmetros n, a, s tem f.m.p.

$$\text{hg}_{n,a,s}(x) = \begin{cases} \frac{\binom{a}{x}\binom{n-a}{s-x}}{\binom{n}{s}}, & \text{se } x \in \left\{\max\{0, s-(n-a)\}, \dots, \min\{a,n\}\right\} \\ 0, & \text{nos outros casos.} \end{cases}$$

Exemplo 78. Qual a probabilidade de acertar 4 dos 6 números sorteados na mega-sena se todos os resultados são igualmente prováveis? Os parâmetros são $a = s = 6$ e $n = 60$, logo

$$\text{hg}_{60,6,6}(4) = \frac{\binom{6}{4}\binom{54}{2}}{\binom{60}{6}} = 0{,}0004287524.$$

◊

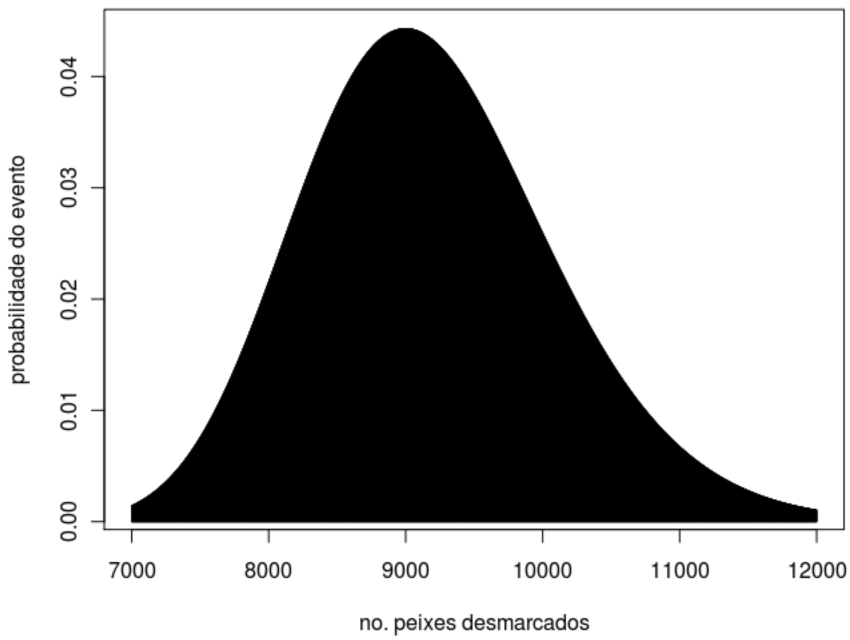
Exemplo 79. Um comprador de componentes elétricos compra os componentes em lote composto de 10 componentes. A política de controle de qualidade é inspecionar 3 componentes escolhidos aleatoriamente e comprar o lote somente se os 3 não apresentarem defeitos. Se 30% dos lotes têm

4 componentes com defeito e 70% apenas 1, qual é a proporção de lotes aceitos? Consideremos os eventos: A definido por "aceita um lote" e B definido por "lote com 4 peças com defeito".

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}\Big(A \;\Big|\; \overline{B}\Big)\mathbb{P}(\overline{B}) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}\frac{3}{10} + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}\frac{7}{10} = \frac{54}{100}.$$

$\Diamond$

Exemplo 80 (**Estimativa de máxima verossimilhança**). Num lago 1000 peixes foram capturados, marcados e devolvidos. Uma nova captura de 1000 peixes é feita e é constatado que 100 deles estão marcados. O que pode ser dito a respeito do tamanho da população de peixes no lago? A probabilidade do evento em função do número (desconhecido) de peixes desmarcados tem *gráfico de barras*



sugere uma população de aproximadamente 9.000 + 1.000 peixes (máxima verossimilhança — estimativa que maximiza a probabilidade do evento ocorrido) e, de fato, essa estimativa por ser feita de modo análogo a equação (2.2.1)). $\Diamond$

Distribuição geométrica:: uma variável aleatória tem *distribuição geométrica* com parâmetro p se tem f.m.p. dada por

$$\mathrm{ge}_p(x) = \begin{cases} (1-p)^{x-1}p, & \text{se } x \in \{1, 2, \dots\} \\ 0, & \text{nos outros casos.} \end{cases}$$

que correspondente ao número de ensaios de Bernoulli independentes com parâmetro p até ocorrer um sucesso. A notação

$$X \sim \text{Geometrica}(p)$$

indica que X é uma variável aleatória geométrica com parâmetro p.

Exemplo 81. Uma urna contém N bolas brancas e M bolas pretas. As bolas são selecionadas aleatoriamente, uma de cada vez, até que saia uma bola preta. Se supormos que cada bola retirada é substituída antes da próxima retirada, qual é a probabilidade com que sejam necessárias exatamente i retiradas?

Se X é o número de retiradas até sair bola preta

$$\mathbb{P}(X = i) = \left(\frac{N}{N+M}\right)^{i-1}\left(\frac{M}{N+M}\right) = \frac{MN^{i-1}}{(N+M)^i}.$$

Com que probabilidade são necessárias pelo menos k retiradas?

$$\begin{aligned}
\mathbb{P}(X \geqslant k) &= \sum_{i=k}^{\infty}\left(\frac{N}{N+M}\right)^{i-1}\left(\frac{M}{N+M}\right)\\
&= \left(\frac{M}{N+M}\right)\sum_{i=k}^{\infty}\left(\frac{N}{N+M}\right)^{i-1}\\
&= \left(\frac{M}{N+M}\right)\frac{\left(\frac{N}{N+M}\right)^{i-1}}{1-\left(\frac{N}{N+M}\right)}\\
&= \left(\frac{N}{N+M}\right)^{k-1}
\end{aligned}$$

Exemplo 82 (**colecionador de cupons**). Considere a seguinte situação: há N tipos de cupons. Um colecionador compra a cada unidade de tempo um cupom aleatório. A probabilidade de obter em cada vez um cupom específico é $1/N$, independentemente das aquisições anteriores.

Uma variável aleatória de interesse é o número de unidades de tempo T até o colecionador reunir pelo menos um cupom de cada tipo. Essa variável aleatória variável toma valores no em $\{N, N+1, \dots\} \cup \{\infty\}$.

Assumindo que o colecionador já esteja de posse de k tipos diferentes de cupons, $0 \leqslant k < N$, a quantidade de tempo até a aquisição de um tipo novo de cupom é uma variável aleatória geométrica com parâmetro $(N-k)/N$.

Distribuição binomial negativa:: Uma *variável aleatória binomial negativa* com parâmetros p e r tem f.m.p. dada por

$$\mathrm{nbi}_{p,r}(x) = \begin{cases} \binom{x-1}{r-1}p^r(1-p)^{x-r}, & \text{se } x \in \{r+1, r+2, \dots\}\\ 0, & \text{nos outros casos.}\end{cases}$$

que correspondente ao número de ensaios de Bernoulli independentes com parâmetro p até ocorrer um total de r sucessos.

Variáveis aleatórias contínuas. Uma variável aleatória X é (absolutamente) *contínua* se

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

para alguma função integrável $f_X : \mathbb{R} \to [0, +\infty)$ tal que $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$.

Função de densidade:: é qualquer função integrável $f : \mathbb{R} \to [0, +\infty)$ satisfazendo $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$. A função f_X é chamada de *função de densidade de probabilidade* (f.d.p) da variável X.

Para nós, sempre valerá que: *uma variável aleatória X admite uma função de densidade se a distribuição F_X é contínua e derivável em todo ponto da reta exceto por um número enumerável deles*. Ademais, $f_X(x)$ é a derivada de $F_X'(x)$ nos pontos x em que a derivada existe, ou seja, exceto para no máximo uma quantidade enumerável de pontos $x \in \mathbb{R}$ temos

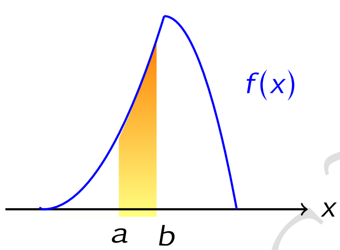
$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x).$$

Assim, basta conhecer a f.d.p. de X para conhecer sua distribuição e vice-versa.

Exercício 77. Mostre que para uma função de densidade f vale que $F(t) := \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ satisfaz as quatro propriedades descritas na pagina 132 e que caracterizam uma função de distribuição.

Para uma variável aleatória contínua X a f.d.a. F_X é contínua em toda reta, portanto, pelo item 3 da proposição 2.1.3 temos que $\mathbb{P}(X = a) = 0$ para todo $a \in \mathbb{R}$, logo $P(a \leqslant X \leqslant b) = P(a \leqslant X < b) = P(a < X < b) = P(a < X \leqslant b)$ e

$$\boxed{P(a < X \leqslant b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x - \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x = \int_{a}^{b} f_X(x)\,\mathrm{d}x.}$$



$\mathbb{P}(a \leqslant X \leqslant b) =$ área da região delimitada pelo gráfico, por $x = a$ e $x = b$ e pelo eixo x no intervalo $[a, b]$.

Daí temos que para qualquer $\varepsilon > 0$ pequeno

$$\mathbb{P}\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_X(x)\,\mathrm{d}x \approx \varepsilon f_X(a) \tag{2.2.2}$$

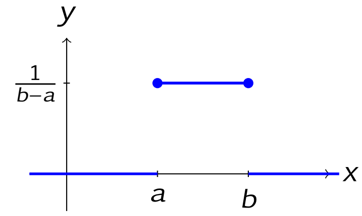
ou seja, X assume valor numa vizinhança de a de diâmetro ϵ muito pequeno com probabilidade aproximadamente $\varepsilon f_X(a)$.

Retomando o exemplo 64, a variável aleatória Z com distribuição dada por

$$F_Z(t) = \begin{cases} 0, & \text{se } t < a \\ \frac{t-a}{b-a}, & \text{se } a \leqslant t < b \\ 1, & \text{se } t \geqslant b \end{cases}$$

é contínua e tem derivada em todo ponto, exceto a e b e

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



2.2.3 Observação. *Os valores $f(a)$ e $f(b)$, no caso acima, podem ser arbitrários pois, para quaisquer que sejam esses valores, a integral $\int_{-\infty}^{t} f(x)\,\mathrm{d}x$ ainda vale $F_Z(t)$.*

Principais modelos contínuos. Retomemos alguns fatos, se X é v.a. contínua valem

1. $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$;
2. $\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1$;
3. $\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(x)\mathrm{d}x = F_X(b) - F_X(a)$ para todo $a \leqslant b$;
4. $\mathbb{P}(X = a) = 0$, para qualquer $a \in \mathbb{R}$.

Distribuição uniforme contínua:: uma variável aleatória contínua X é *uniforme* no intervalo $[a, b]$, para $a < b$, se sua f.d.p. é

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

e denotamos esse fato por $X \sim \text{Uniforme}(a, b)$.

Nesse caso, a probabilidade de X estar num subintervalo de $[a, b]$ é proporcional ao comprimento de tal subintervalo; de fato, para $y < z$ reais

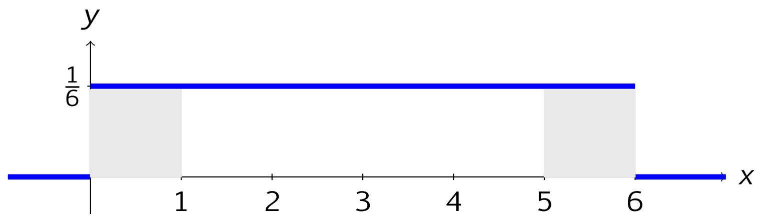
$$\mathbb{P}(y \leqslant X \leqslant z) = \int_y^z \frac{1}{b-a}\,\mathrm{d}x = \frac{z-y}{b-a}.$$

Exemplo 83. Num teste, tubos de PVC de 6 m são submetidos a grande pressão d'água até que o primeiro vazamento ocorra. A distância do início do tubo até o vazamento é uniformemente distribuída. Qual a probabilidade de que o vazamento esteja a no máximo 1 m das extremidades?

Seja $X \sim \text{Uniforme}(0, 6)$ a distância do início do tubo até o vazamento. A probabilidade procurada é

$$\mathbb{P}([0 \leqslant X \leqslant 1] \cup [5 \leqslant X \leqslant 6]) = \mathbb{P}(0 \leqslant X \leqslant 1) + \mathbb{P}(5 \leqslant X \leqslant 6) = \int_0^1 \frac{1}{6}\mathrm{d}x + \int_5^6 \frac{1}{6}\mathrm{d}x = \frac{1}{3}$$

que corresponde à área da região destacada em cinza



O valor médio de $X \sim \text{Uniforme}(a, b)$ é

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a} \mathrm{d}x = \frac{a+b}{2}$$

a variância

$$\mathrm{Var}[X] = \int_a^b x^2 \frac{1}{b-a} \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

de modo que o desvio padrão é $\approx 0{,}29(b-a)$.

Distribuição exponencial:: Uma variável aleatória contínua X é *exponencial* com parâmetro $\lambda > 0$, e denotamos esse fato por

$$X \sim \text{Exponencial}(\lambda)$$

se sua função de densidade é

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

A função de distribuição é

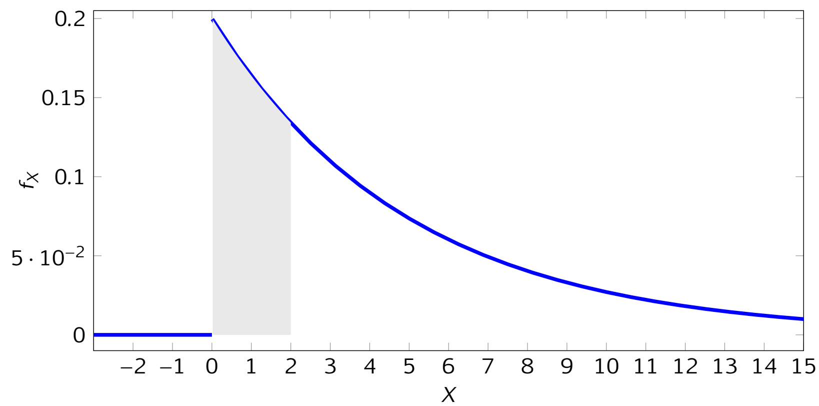
$$F_X(t) = \int_0^t \lambda \mathrm{e}^{-\lambda x} \, \mathrm{d}x = 1 - \mathrm{e}^{-\lambda x}$$

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

Variáveis aleatórias exponenciais são muitas vezes utilizados para modelar a distribuição da quantidade de tempo decorrido até que algum evento particular ocorra. Por exemplo, seja $T \sim \text{Exponencial}(0{,}2)$ o intervalo de tempo (em minutos) entre emissões consecutivas de uma fonte radioativa. A probabilidade de haver uma emissão em até 2 min é $\mathbb{P}(T \leqslant 2) = 1 - \mathbb{P}(T > 2) = 1 - \mathrm{e}^{0{,}2 \cdot 2} \approx 0{,}33$ ou, de outro modo

$$\mathbb{P}(T < 2) = \int_0^2 0{,}2\mathrm{e}^{-0{,}2x} \, \mathrm{d}x = 1 - \mathrm{e}^{0{,}2 \cdot 2}$$

que corresponde a área da região em cinza no gráfico



A probabilidade do intervalo ser maior que 7 dado que foi maior que 5

$$\begin{aligned}\mathbb{P}(T > 7 \mid T > 5) &= \frac{\mathbb{P}([T>7]\cap[T>5])}{\mathbb{P}(T>5)} = \frac{\mathbb{P}(T>7)}{\mathbb{P}(T>5)} \\ &= \frac{\mathrm{e}^{-0,2\cdot 7}}{\mathrm{e}^{-0,2\cdot 5}} = \mathrm{e}^{-0,2\cdot 2} = \mathbb{P}(T>2)\end{aligned}$$

Dizemos que uma variável aleatória T é *sem memória* se

$$\mathbb{P}(T > s+t \mid T > t) = \mathbb{P}(T > s) \tag{2.2.3}$$

para quaisquer $s, t \geqslant 0$. Se X tem distribuição exponencial então é sem memória. Por outro lado, uma variável aleatória contínua sem memória tem distribuição exponencial.

Para ver como isso funciona, imagine que no instante $t_0 = 0$ ligamos um despertador que irá tocar depois de um tempo T que é distribuído exponencialmente com parâmetro λ. Suponha que tivemos que sair e ao voltar, no instante t, descobrimos que o despertador ainda não tocou. Seja S o tempo que decorre partir de então (i.e, observado $[T > t]$) até o despertador tocar.

$$\mathbb{P}(S > s \mid T > t) = \mathbb{P}(T > s+t \mid T > t)$$

que por equação (2.2.3) é $\mathbb{P}(T > t)$. A coisa importante de se notar é que a distribuição do tempo até ocorrer o evento não depende do instante inicial t. A distribuição exponencial é única com essa propriedade.

Exercício 78. Prove que $T \sim \text{Exponencial}(\lambda)$ é sem memória.

Exemplo 84. Suponha que um sistema contenha componentes cujo tempo até falhar é $T \sim \text{Exponencial}(1/5)$ em anos. Se 5 desses componentes são instalados em cada sistema, qual é a probabilidade de que pelo menos 2 componentes ainda estejam funcionando após 8 anos?

Se X é a quantidade de componentes funcionando após 8 anos então $X \sim \text{Binomial}(5, p)$ com $p = \mathbb{P}(T > 8) = \mathrm{e}^{-8/5} \approx 0{,}2$, assim

$$\mathbb{P}(X \geqslant 2) = \sum_{x=2}^{5} \mathrm{bi}_{5,p}(x) = 1 - \sum_{x=0}^{1} \mathrm{bi}_{5,p}(x) \approx 0{,}26.$$

A esperança de $T \sim \text{Exponencial}(\lambda)$ é

$$\mathbb{E}[T] = \int_0^\infty \lambda x e^{-\lambda x} dx = \lambda \left(-\frac{e^{-\lambda x}(\lambda x + 1)}{\lambda^2} \right) \Bigg|_0^\infty = \frac{1}{\lambda},$$

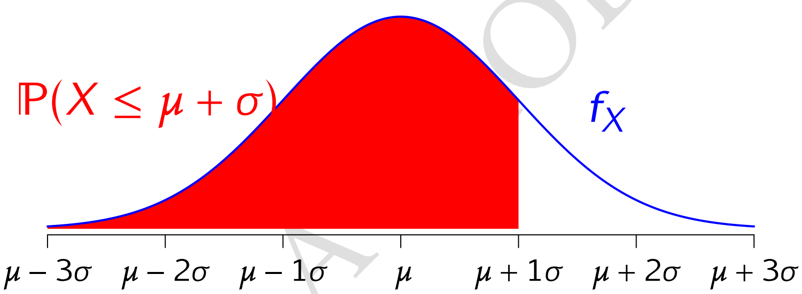
e a variância

$$\sigma^2 = \text{Var}[T] = \int_0^\infty \lambda x^2 e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

2.2.4 A distribuição normal: A variável aleatória X tem distribuição *normal* com parâmetros μ e σ^2, abreviado por $X \sim \mathcal{N}(\mu; \sigma^2)$, se sua função densidade de probabilidade é dada por

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

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



O problema com o qual nos deparamos agora é que

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

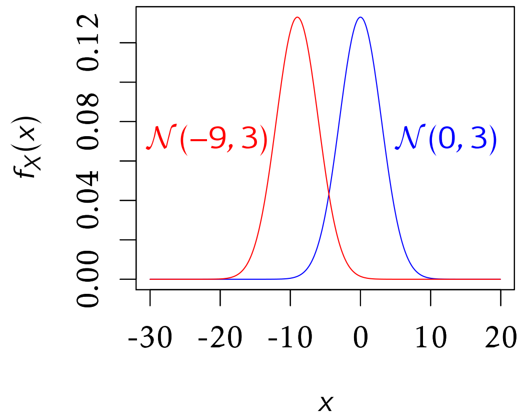
não tem solução analítica.

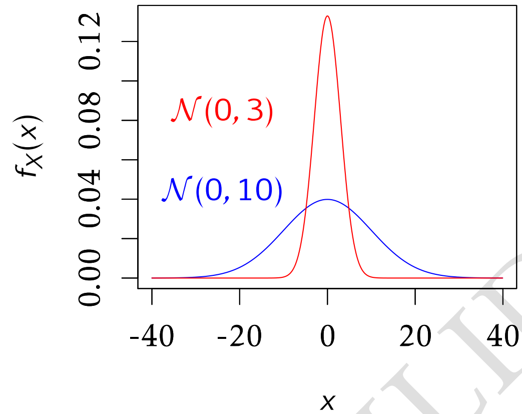
A distribuição normal tem as seguintes propriedades

1. $\mathbb{E}[X] = \mu$ e $\text{Var}[X] = \sigma^2$;
2. $f_X(x) \to 0$ quando $x \to \pm\infty$;
3. μ é ponto de máximo de $f_X(x)$ e $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de $f_X(x)$;
4. o gráfico $(x, f_X(x))$ é simétrico com relação a $x = \mu$;

Exercício 79. Prove as propriedades 2, 3 e 4.

Os gráficos abaixo mostram a função de densidade de normais com parâmetros diferentes





Uma propriedade interessante, e muito útil, é a seguinte

2.2.5 Proposição. *Se* $X \sim \mathcal{N}(\mu;\sigma^2)$ *então* $aX+b \sim \mathcal{N}(a\mu+b;a^2\sigma^2)$, *para quaisquer* $a,b \in \mathbb{R}$.

Demonstração. Escrevemos $Y = aX+b$ e temos

$$F_Y(x) = \mathbb{P}(Y \leqslant x) = \mathbb{P}(aX+b \leqslant x) = \mathbb{P}\left(X \leqslant \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right)$$

se $a > 0$, logo a densidade de Y é

$$F_y'(x) = \frac{1}{a}F_X'\left(\frac{x-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}}e^{-\frac{((x-b)/a-\mu)^2}{2\sigma^2}} = \frac{1}{a\sigma\sqrt{2\pi}}e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}$$

portanto $Y \sim \mathcal{N}(a\mu+b;a^2\sigma^2)$, caso $a < 0$ a mesma conclusão vale e deixamos a verificação como exercício. □

Agora, sabemos pela proposição 2.2.5 acima que para quaisquer $a,b \in \mathbb{R}$

$$\text{se } X \sim \mathcal{N}(\mu;\sigma^2) \text{ então } aX+b \sim \mathcal{N}(a\mu+b;a^2\sigma^2)$$

de modo que se tomarmos $a = 1/\sigma$ e $b = -\mu/\sigma$ temos

$$\text{se } X \sim \mathcal{N}(\mu;\sigma^2) \text{ então } \tfrac{X-\mu}{\sigma} \sim \mathcal{N}\left(\tfrac{1}{\sigma}\mu - \tfrac{\mu}{\sigma}; \left(\tfrac{1}{\sigma}\right)^2\sigma^2\right), \text{ portanto, } \tfrac{X-\mu}{\sigma} \sim \mathcal{N}(0;1)$$

Com isso, temos a seguinte consequência da proposição 2.2.5

2.2.6 Corolário. *Se* $X \sim \mathcal{N}(\mu;\sigma^2)$ *então* $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0;1)$. □

Se $a < x < b$ então $(a-\mu)/\sigma < (x-\mu)/\sigma < (b-\mu)/\sigma$, portanto

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right).$$

Variável aleatória padronizada:: a variável aleatória padronizada de $X \sim \mathcal{N}(\mu;\sigma^2)$ é

$$Z_X := \frac{X-\mu}{\sigma}$$

e temos $Z_X \sim \mathcal{N}(0;1)$. Agora,

$$\boxed{\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z_X < \frac{b-\mu}{\sigma}\right)}$$

de modo que para calcular probabilidades que envolvem uma variável aleatória normal $\mathcal{N}(\mu;\sigma^2)$ basta conhecermos a função de distribuição de $Z \sim \mathcal{N}(0;1)$

$$\Phi(x) := \mathbb{P}(Z \leqslant x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,\mathrm{d}t$$

(é costume usar $\Phi(x)$ para denotar a f.d.a. de uma variável aleatória com distribuição normal).

Exercício 80. Prove que $X \sim \mathcal{N}(0;1)$ tem média 0 e variância 1. Deduze desse fato a média e a variância de $Y \sim \mathcal{N}(\mu;\sigma^2)$.

Tabela da distribuição normal padrão

Para $Z \sim \mathcal{N}(0;1)$, a tabela que usamos é da forma

x	0,00	0,01	0,02	0,03	0,04	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
1,3	$\cdots$	$\cdots$	0,9066	$\cdots$	0,9099	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

e para calcularmos $\mathbb{P}(Z \leqslant 1,32)$ decompomos $1{,}32 = 1{,}3 + 0{,}02$ (parte inteira e primeira casa decimal + segunda casa decimal), em seguida $\mathbb{P}(Z \leqslant 1,32)$ é lido na linha indexada por 1,3 com a coluna indexada por 0,02, no caso $\mathbb{P}(Z \leqslant 1,32) = 0,9066$. Analogamente, $\mathbb{P}(Z \leqslant 1,34) = 0,9099$.

Vejamos mais alguns exemplos de consulta à tabela da distribuição normal

1. quanto é $\mathbb{P}(0 < Z \leqslant 1{,}71)$?

   $\mathbb{P}(0 < Z \leqslant 1{,}71) = \mathbb{P}(Z \leqslant 1{,}71) - \mathbb{P}(Z < 0) = 0{,}9564 - 0{,}5 = 0{,}4564$

2. quanto é $\mathbb{P}(0{,}32 \leqslant Z \leqslant 1{,}71)$?

   $\mathbb{P}(0{,}32 \leqslant Z \leqslant 1{,}71) = \mathbb{P}(Z \leqslant 1{,}71) - \mathbb{P}(Z < 0{,}32) = 0{,}9564 - 0{,}6255 = 0{,}3309$

3. quanto é $\mathbb{P}(Z \leqslant -1{,}71)$?

   $\mathbb{P}(Z \leqslant -1{,}71) = \mathbb{P}(Z \geqslant 1{,}71) = 1 - \mathbb{P}(Z < 1{,}71) = 1 - 0{,}9564 = 0{,}0436$

4. quanto é $\mathbb{P}(-1{,}71 \leqslant Z \leqslant 1{,}71)$?

$$\mathbb{P}(-1{,}71 \leqslant Z \leqslant 1{,}71) = \mathbb{P}(Z \leqslant 1{,}71) - \mathbb{P}(Z < -1{,}71) = 0{,}9564 - 0{,}0436 = 0{,}9128$$

5. ou seja, genericamente, se $Z \sim \mathcal{N}(0;1)$ então para $y \geqslant x \geqslant 0$ reais temos

- $\mathbb{P}(Z \leqslant x) = \Phi(x)$
- $\mathbb{P}(y \leqslant Z \leqslant x) = \Phi(x) - \Phi(y)$
- $\mathbb{P}(Z \leqslant -x) = \Phi(-x) = 1 - \Phi(x)$
- $\mathbb{P}(-x \leqslant Z \leqslant x) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

6. Como encontrar o valor z da distribuição $\mathcal{N}(0;1)$ tal que $\mathbb{P}(Z \leqslant z) = 0,975$?

Consultando a parte relevante da tabela

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

obtemos $z = 1,96$.

7. Qual é o z tal que $\mathbb{P}(0 < Z \leqslant z) = 0,4664$? Como da tabela determinamos probabilidade de forma $\mathbb{P}(Z \leqslant z)$ basta lembrar que, por simetria $\mathbb{P}(Z \leqslant 0) = 0{,}5$ logo, se somarmos $0,5 + 0,4664$ temos $\mathbb{P}(Z \leqslant z) = 0,5 + 0,4664$, portanto $z = 1{,}83$.

O z tal que $\mathbb{P}(Z \geqslant z) = 0,0228$ e determinado por

$$1 - 0,0228 = 0,9772 \Longrightarrow z = 2$$

Distribuição normal padrão $\mathcal{N}(0;1)$ - Tabela f.d.a.

$\mathbb{P}(Z \leq x) = \Phi(x)$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936

Exemplo 85. O tempo gasto no exame vestibular de uma universidade tem distribuição normal, com média 120 min e desvio padrão 15 min.

1. Qual é a probabilidade com que um candidato termine o exame antes de 100 minutos?

   Se X é o tempo gasto no exame vestibular, então $X \sim \mathcal{N}(120; 15^2)$ logo

   $$\mathbb{P}(X < 100) = \mathbb{P}\left(Z_X \leqslant \frac{100 - 120}{15}\right) = \mathbb{P}(Z_X \leqslant -1{,}33) = 1 - \mathbb{P}(Z_X < 1{,}33)$$

   usando a tabela $1 - \mathbb{P}(Z < 1,33) = 1 - 0,9082 = 0{,}0918$.

2. Qual deve ser o tempo de prova de modo a permitir que 95% dos vestibulandos terminem no prazo estipulado?

Para determinar o tempo de prova de modo a permitir que 95% dos vestibulandos terminem no prazo estipulado, devemos encontrar x tal que $\mathbb{P}(X < x) = 0{,}95$, ou seja, tal que

$$\mathbb{P}\left(Z_X \leqslant \frac{x-120}{15}\right) = 0{,}95.$$

Pela tabela $\mathbb{P}(Z \leqslant 1{,}64) = 0{,}95$ portanto

$$\frac{x-120}{15} = 1{,}64$$

ou seja $x = 144{,}6\,\text{min}$.

3. Qual é o intervalo central de tempo, tal que 80% dos estudantes gastam para completar o exame?

   O intervalo central de tempo, tal que 80% dos estudantes gastam para completar o exame é $I = (\mu - \epsilon, \mu + \epsilon)$ tal que $\mathbb{P}(I) = 0{,}8$

$$\mathbb{P}(a \leqslant X \leqslant b) = 0{,}8 \Leftrightarrow \mathbb{P}\left(\frac{a-120}{15} \leqslant Z_X \leqslant \frac{b-120}{15}\right) = 0{,}8$$

   Pela tabela $\mathbb{P}(-1{,}28 \leqslant Z \leqslant 1{,}28) = 0{,}80$, portanto $a = 100{,}8$ e $b = 139{,}2$ minutos.

Exemplo 86. Um sistema considera que um sinal digital será transmitido quando a tensão exceder $0,9\,\text{V}$. Na detecção do sinal o ruído tem distribuição $\mathcal{N}(0; 0{,}45)$. Qual a probabilidade de detectar um sinal quando nada tiver sido enviado?

Se $R \sim \mathcal{N}(0; 0{,}45)$ é a tensão do ruído, então

$$\mathbb{P}(R > 0{,}9) = \mathbb{P}\left(\frac{R}{0{,}45} > \frac{0{,}9}{0{,}45}\right) = \mathbb{P}(Z_X > 2) = 1 - 0{,}97725 = 0{,}02275.$$

O intervalo central que inclui 99% de todas as leituras de ruído é dado por x tal que

$$\mathbb{P}(-x < R < x) = \mathbb{P}\left(\frac{-x}{0{,}45} < \frac{R}{0{,}45} < \frac{x}{0{,}45}\right)\mathbb{P}\left(\frac{-x}{0{,}45} < Z_X < \frac{x}{0{,}45}\right) = 0{,}99.$$

De acordo com a tabela, $x/0{,}45 = 2{,}58$, ou seja, $x = 1{,}16$.

Suponha que quando um sinal é transmitido a média da variável aleatória R mude para $1{,}8\,\text{V}$. Qual a probabilidade do sinal não ser detectado? Seja S a tensão quando um sinal é transmitido.

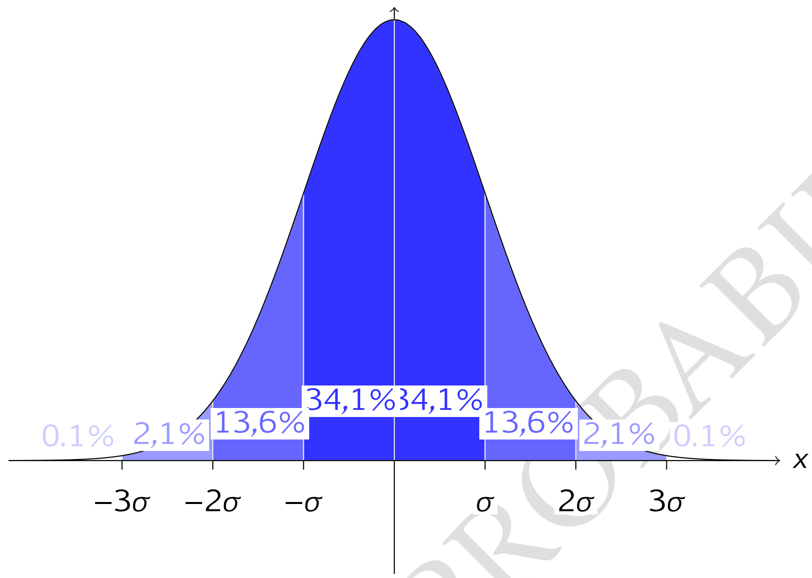
$$\mathbb{P}(S < 0{,}9) = \mathbb{P}\left(\frac{S-1{,}8}{0{,}45} < \frac{0{,}9-1{,}8}{0{,}45}\right) = \mathbb{P}(Z < -2) = 0{,}02275.$$

Essa é a probabilidade com que um sinal é perdido.

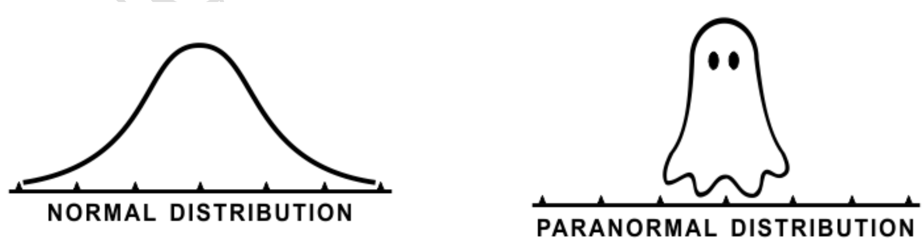
Concentração em torno de μ

Sejam $k \in \mathbb{N}$, $X \sim \mathcal{N}(\mu;\sigma^2)$ e $Z = (X-\mu)/\sigma$ então

$$\mathbb{P}(\mu - k\sigma \leqslant X \leqslant \mu + k\sigma) = \mathbb{P}(-k \leqslant Z \leqslant k) = \mathbb{P}(Z < k) - \mathbb{P}(Z < -k).$$



- Para $k = 1$, $\mathbb{P}(\mu - \sigma \leqslant X \leqslant \mu + \sigma) = \mathbb{P}(-1 \leqslant Z \leqslant 1) = 0{,}682$.
- Para $k = 2$, $\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = \mathbb{P}(-2 \leqslant Z \leqslant 2) = 0{,}954$.
- Para $k = 3$, $\mathbb{P}(\mu - 3\sigma \leqslant X \leqslant \mu + 3\sigma) = \mathbb{P}(-3 \leqslant Z \leqslant 3) = 0{,}997$.



$*$ **Integral de Stieltjes.** Seja f uma função definida no intervalo $[a,b]$ de comprimento positivo e F uma função de distribuição acumulada. Façamos $x_0 = a$, $x_n = b$ e consideremos o intervalo $[a,b]$ dividido pelos pontos $x_1 < x_2 < \cdots < x_{n-1}$ em n partes não necessariamente de mesmo comprimento. Em cada subintervalo escolhemos um ponto arbitrário $\xi_i \in [x_{i-1}, x_i]$, para todo $0 < i \leq n$. Dizemos que que $\mathcal{P} = ((x_0, x_1, \cdots, x_n), (\xi_1, \xi_2, \cdots, \xi_{n-1}))$ é uma *partição rotulada* de $[a,b]$ e sua *norma* é $\Delta\mathcal{P} =$

$\max_{1\leq i\leq n}(x_i - x_{i-1})$. A soma de Riemann–Stieltjes de f com respeito a F e $\mathcal{P}$ é

$$S(f,F,\mathcal{P}) = \sum_{i=1}^{n} f(\xi_i)\,(F(x_i) - F(x_{i-1}))$$

Integral de Stieltjes:: a integral de Riemann–Stieltjes permite um tratamento unificado das variáveis aleatórias discreta e contínua. Também é possível tratar casos como no exemplo 65. A integral de f com respeito a F no intervalo $[a,b]$ é o limite, quando existe,

$$\int_a^b f(x)\,\mathrm{d}F(x) := \lim_{\Delta\mathcal{P}\to 0} S(f,F,\mathcal{P}) \tag{2.2.4}$$

que significa que para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $|\int_a^b f(x)\,\mathrm{d}F(x) - S(f,F,\mathcal{P})| < \varepsilon$ sempre que $\mathcal{P}$ é uma partição rotulada com $\Delta(\mathcal{P}) < \delta$.

$$\int f(x)\,\mathrm{d}F(x) := \lim_{\substack{a\to-\infty\\ b\to+\infty}} \int_a^b f(x)\,\mathrm{d}F(x) \tag{2.2.5}$$

se o limite existe.

É possível provar que se f é contínua e limitada, os limites equação (2.2.4) e equação (2.2.5) existem. Em certos casos, de interesse em Probabilidade, o s limites existem no caso de f não limitada.

2.2.7 Observação. *A integral de Riemann é o caso partilcular $F(x) = x$ em $[a,b]$.*

Se X é uma variável aleatória e tem f.d.a. F então segue da definição que

$$\int_a^b \mathrm{d}F(x) = F(b) - F(a) = \mathbb{P}(a < X \leqslant b)$$

Ademais,

$$\mathbb{P}(X \leqslant b) = \int_{-\infty}^{b} \mathrm{d}F(x) = F(b) - \lim_{t\to-\infty} F(t) = F(b)$$
$$\mathbb{P}(X > a) = \int_{a}^{+\infty} \mathrm{d}F(x) = \lim_{t\to+\infty} F(t) - F(a) = 1 - F(a)$$

A inclusão ou não dos extremos do intervalo faz diferença na integral, se usarmos $b-$ par indicar que b não está incluído.

$$\mathbb{P}(a < X < b) = \int_a^{b-} \mathrm{d}F(x) = F(b-) - F(a).$$

Suponhamos que F seja uma f.d.a. de uma variável aleatória discreta de modo que em $[a, b]$ os saltos de descontinuidade ocorrem em t_i, com $t_0 = a < t_1 < \cdots < t_N = b$ e seja g uma função contínua em $[a, b]$. Façamos

$$d_i := F(t_i) - F(t_i-),\ i = 0, 1, 2, \ldots, N$$

A soma de Riemann–Stieltjes da função g com respeito a função de distribuição acumulada F e a partição rotulada $\mathcal{P} = ((x_0, x_1, \cdots, x_n), (\xi_1, \xi_2, \cdots, \xi_{n-1}))$ é

$$S(g, F, \mathcal{P}) = \sum_{i=1}^{n} g(\xi_i)\left(F(x_i) - F(x_{i-1})\right)$$

e se $x_i - x_{i-1}$ é suficientemente pequeno, então

$$F(x_i) - F(x_{i-1}) = \begin{cases} d_k, & \text{se } t_k \in (x_{i-1}, x_i) \\ 0, & \text{se } t_k \notin (x_{i-1}, x_i) \text{ para qualquer } k \end{cases}$$

ademais, $\xi_i \to t_k$ quando $n \to \infty$ e temos $g(\xi_i) \to g(t_k)$, pois g é contínua, de modo que temos

$$\lim_{n\to\infty} \sum_{i=1}^{n} g(\xi_i)\left(F(x_i) - F(x_{i-1})\right) = \sum_{i=1}^{N} g(t_i) d_i.$$

portanto, se F é uma f.d.a. de uma variável aleatória discreta e em $[a, b]$ os saltos de descontinuidade ocorrem em t_i,

$$\int_a^b g(x)\, \mathrm{d}F(x) = \sum_{i=1}^{N} g(t_i)\mathbb{P}(X = t_i).$$

Agora, se F é diferenciável com derivada $F'(x) = f(x)$ contínua,

$$S(g, F, \mathcal{P}) = \sum_{i=1}^{n} g(\xi_i)\left(F(x_i) - F(x_{i-1})\right) = \sum_{i=1}^{n} g(\xi_i) F'(\eta_i)(x_i - x_{i-1})$$

para algum $\eta_i \in (x_{i-1}, x_i)$ para todo i, pelo Teorema da Valor Médio, logo

$$\lim_{n\to\infty} \sum_{i=1}^{n} g(\xi_i) F'(\eta_i)(x_i - x_{i-1}) = \int_a^b g(x) f(x)\, \mathrm{d}x.$$

portanto, se F é diferenciável com derivada $F'(x) = f(x)$ contínua

$$\int_a^b g(x)\, \mathrm{d}F(x) = \int_a^b g(x) f(x)\, \mathrm{d}x.$$

Propriedades da integral definida:: a integral definida satisfaz as seguintes propriedades, *dado que as integrais existam*. Para $a < c < b$ reais e $\alpha, \beta \in \mathbb{R}$

$$\int_a^b (\alpha \cdot f(x) + \beta \cdot g(x))\,\mathrm{d}F(x) = \alpha \cdot \int_a^b f(x)\,\mathrm{d}F(x) + \beta \cdot \int_a^b g(x)\,\mathrm{d}F(x)$$

$$\int_a^b f(x)\,\mathrm{d}(F(x) + G(x)) = \int_a^b f(x)\,\mathrm{d}F(x) + \int_a^b f(x)\,\mathrm{d}G(x)$$

$$\int_a^b f(x)\,\mathrm{d}F(x) = \int_a^c f(x)\,\mathrm{d}F(x) + \int_c^b f(x)\,\mathrm{d}F(x)$$

ademais de F tem derivada contínua

$$\int_a^b f(x)\,\mathrm{d}F(x) = \int_a^b f(x)f'(x)\,\mathrm{d}x$$

em que a integral da direita é de Riemann.

§3 Esperança matemática

A esperança da variável aleatória X é

$$\mathbb{E}[X] = \int x \, \mathrm{d}F_X(x)$$

se a integral existe. Geometricamente, a esperança é a diferença entre as áreas definidas por (i) o eixo y, a reta $y = 1$ e o gráfico de $y = F(x)$ no intervalo $(0, \infty)$, e por (ii) o eixo x, o gráfico de $y = F(x)$ e o eixo y no intervalo $(-\infty, 0)$.

Se g é uma função real de uma variável real tal que $[g(X) \leq t\}$ é um evento aleatório para todo real t, então

$$\mathbb{E}[g(X)] = \int g(x) \, \mathrm{d}F_X(x)$$

se a integral existe. De fato, consideremos $\{(y_{i-1}, y_i] : i \in \mathbb{Z}\}$ uma partição da reta, rotulada por $\eta_i \in (y_{i-1}, y_i]$ e formemos a soma de Riemann–Stieltjes da função identidade com respeito a f.d.a. F_Y de $Y := g(X)$

$$\begin{aligned}
\sum_i \eta_i (F_Y(y_i) - F_Y(y_{i-1})) &= \sum_i \eta_i \mathbb{P}(Y \in (y_{i-1}, y_i]) \\
&= \sum_i \eta_i \mathbb{P}(g(X) \in (y_{i-1}, y_i]) \\
&= \sum_i \eta_i \mathbb{P}(X \in g^{-1}((y_{i-1}, y_i])) \\
&= \sum_i g(\xi_i) \mathbb{P}(X \in g^{-1}((y_{i-1}, y_i])) \\
&= \sum_i g(\xi_i) \mathbb{P}(X \in (x_{i-1}, x_i])
\end{aligned}$$

com $\xi_i = g^{-1}(\eta_i) \in g^{-1}((y_{i-1}, y_i]) = (x_{i-1}, x_i]$, que é a soma de Riemann–Stieltjes de g com respeito a F_X. Os intervalos $(y_{i-1}, y_i]$ formam uma partição do intervalo todo de modo que $g^{-1}((y_{i-1}, y_i])$ também formam uma partição. Em particular, para $g(x) = ax + b$ temos

$$\mathbb{E}[aX + b] = \int (ax + b) \, \mathrm{d}F(x) = \int ax \, \mathrm{d}F(x) + \int b \, \mathrm{d}F(x) = a \int x \, \mathrm{d}F(x) + b = a\mathbb{E}[X] + b.$$

Exemplo 87. Seja $A \in \mathcal{B}(\mathbb{R})$ e consideremos a variável indicadora $\mathbf{1}_{X \in A}$ de ocorrência do evento $[X \in A]$. Então

$$\mathbb{E}[\mathbf{1}_{X \in A}] = \int \mathbf{1}_{X \in A} \, \mathrm{d}F(x) = \int_A \mathrm{d}F(x)$$

e, por outro lado,

$$\mathbb{E}\left[\mathbf{1}_{X\in A}\right] = 0\mathbb{P}(\mathbf{1}_{X\in A} = 0) + 1\mathbb{P}(\mathbf{1}_{X\in A} = 1) = \mathbb{P}(\mathbf{1}_{X\in A} = 1) = \mathbb{P}(X \in A).$$

Donde concluímos que $\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A \mathrm{d}F(x)$.

Uma variável aleatória é dita integrável se $\mathbb{E}\left[|X|\right] < \infty$.

Uma relação fundamental entre probabilidade e esperança é dada pela variável aleatória indicadora, para todo evento aleatório A

$$\mathbb{P}(A) = \mathbb{E}\left[\mathbf{1}_A\right]. \tag{3.0.1}$$

Da equação equação (3.0.1) acima temos, por exemplo, usando as propriedades dadas no exercício 73, que

$$\mathbb{P}(A \cup B) = \mathbb{E}\left[\mathbf{1}_{A\cup B}\right] = \mathbb{E}\left[\mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}\right] = \mathbb{E}\left[\mathbf{1}_A\right] + \mathbb{E}\left[\mathbf{1}_B\right] - \mathbb{E}\left[\mathbf{1}_{A\cap B}\right] = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Se X é uma variável aleatória com função de (massa/densidade) de probabilidade f então a esperança é

- $\mathbb{E}\left[X\right] = \sum_i x_i f(x_i)$, se X é *discreta*
- $\mathbb{E}\left[X\right] = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x$, se X é *contínua*

desde que a série/integral esteja definida.

No caso de uma variável aleatória discreta

$$\mathbb{E}\left[X\right] = \lim_{n\to\infty}\left(\sum_{i=1}^{n} x_i f(x_i)\right) \tag{3.0.2}$$

de modo que se X assume somente valores não negativos então ou a soma (série) é um número real ou é $+\infty$. No caso geral de uma variável aleatória discreta, definimos as variáveis aleatórias não-negativas

$$X^+(\omega) := \max\{X(\omega), 0\} \quad \text{e} \quad X^-(\omega) := -\min\{X(\omega), 0\}$$

de modo que $X = X^+ - X^-$ donde (cf. teo. 3.0.14, página 193)

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]$$

portanto, se $\mathbb{E}\left[X^+\right] = \mathbb{E}\left[X^-\right] = \infty$ então $\mathbb{E}\left[X\right]$ não está definida; senão $\mathbb{E}\left[X\right]$ está definida e

1. se $\mathbb{E}\left[X^+\right] = z$ e $\mathbb{E}\left[X^-\right] = \infty$ então $\mathbb{E}\left[X\right] = z - \infty = -\infty$;
2. se $\mathbb{E}\left[X^+\right] = \infty$ e $\mathbb{E}\left[X^-\right] = z$ então $\mathbb{E}\left[X\right] = \infty - z = \infty$;

3. se $\mathbb{E}[X^+] = y$ e $\mathbb{E}[X^-] = z$ então $\mathbb{E}[X] \in \mathbb{R}$ e dizemos que X é **integrável**, isto é, quando

$$\sum_i |x_i| f(x_i) < \infty. \tag{3.0.3}$$

Na soma, a convergência absoluta na equação (3.0.3) significa que se alterarmos o ordem dos fatores o limite equação (3.0.2) não muda, fato que usamos anteriormente na prova do teorema 3.0.3 e que, por exemplo, nos permite escrever (verifique a igualdade)

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega).$$

No caso contínuo o tratamento é análogo e X é **integrável** se, e só se, $\mathbb{E}[X^+], \mathbb{E}[X^-] < \infty$, ou seja, quando $\int_{-\infty}^{+\infty} |x| f(x)\,\mathrm{d}x < \infty$. No caso em que exatamente um de $\mathbb{E}[X^+]$ e $\mathbb{E}[X^-]$ é infinito $\mathbb{E}[X]$ está definida mas não é integrável.

Uma variável aleatória é dita integrável se $\mathbb{E}[|X|] < \infty$.

Propriedades da esperança: A esperança de variáveis aleatórias tem as seguintes propriedades.

Exercício 81. Escrevemos $X \leqslant Y$ se $X(\omega) \leqslant Y(\omega)$, $\forall \omega \in \Omega$. Prove que a esperança é monótona

$$X \leqslant Y \Rightarrow \mathbb{E}[X] \leqslant \mathbb{E}[Y]. \tag{3.0.4}$$

Prove que se $\mathbb{P}(X \leqslant Y) = 1$ então $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$.

Exercício 82. Para $c \in \mathbb{R}$ constante, mostre que se $\mathbb{P}(X = c) = 1$ então $\mathbb{E}[X] = c$.

Exercício 83. Prove que se $\mathbb{P}(a \leqslant X \leqslant b) = 1$ então $a \leqslant \mathbb{E}[X] \leqslant b$.

Exercício 84. Se $Z \geqslant 0$ e $\mathbb{E}[Z] = 0$ então $\mathbb{P}(Z = 0) = 1$.

Exercício 85. Se $\mathbb{P}(Z \geqslant 0) = 1$ então Z é integrável.

Esperança e variância de uma variável aleatória discreta. Se X é uma variável aleatória discreta com função de massa de probabilidade f_X então a esperança (também chamada de valor médio ou valor esperado) da variável aleatória X é, simplesmente, a média ponderada dos valores da função de probabilidade

$$\boxed{\mathbb{E}[X] := \sum_x x \cdot f_X(x)} \tag{3.0.5}$$

onde a soma é sobre os valores de x tais que $f_X(x) > 0$, ou seja, é a média do valores x na imagem de X ponderada pela probabilidade com que X assume esse valor, a saber $f_X(x) = \mathbb{P}(X = x)$.

O suporte de uma função é o conjunto de pontos do domínio em que a função é diferente de 0. Quando o suporte de f_X é finito equação (3.0.5) fica bem definida. Por exemplo, no caso de uma variável aleatória constante, digamos $X(\omega) = c$ para algum $c \in \mathbb{R}$, temos

$$\mathbb{E}[X] = \sum_x x \cdot f_X(x) = cf_X(c) = c.$$

Se $\mathbf{1}_A$ é a *variável aleatória indicadora* da ocorrência do evento A (definição na pág. 131) então $\mathbb{E}[\mathbf{1}_A] = 1\mathbb{P}(A) + 0\mathbb{P}(\overline{A}) = \mathbb{P}(A)$. Notemos que $\mathbf{1}_A \sim \text{Bernoulli}(p)$ para $p = \mathbb{P}(A)$ e, de fato, vale

3.0.1 Proposição. *Se* $X \sim \text{Bernoulli}(p)$ *então* $\mathbb{E}[X] = p$.

Exemplo 88. Seja X o resultado de um lançamento de um dado,

$$\mathbb{E}[X] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}.$$

Qual a probabilidade $\mathbb{P}(X = 7/2)$?

Esse exemplo chama a atenção para o fato de que o *valor esperado* para o resultado do lançamento de dum dado é um valor que não ocorre como imagem de X. Além disso, como os resultados têm a mesma probabilidade o valor esperado é a média no sentido usual, a soma de seis valores dividida por seis.

Exemplo 89. Num jogo de azar em cada aposta você ganha \$1.000.000 com probabilidade p e \$10 com probabilidade $1 - p$. Se Y é o ganho numa aposta, então a esperança de ganho numa aposta é

$$\mathbb{E}[Y] = 10^6 p + 10(1 - p).$$

No caso de $p = 1/2$, temos $\mathbb{E}[Y] = 500.005$, qual é a probabilidade de você ganhar \$500.005 numa aposta?

No exemplo anterior, se $p = 1/100$ então a probabilidade de ganhar um valor alto é muito pequeno, comparado com a probabilidade ganhar 10 reais. Apesar de que, com grande probabilidade, o ganho numa aposta seja de \$10 ainda assim o valor esperado de ganho numa única aposta é grande, $\mathbb{E}[Y] = 10.009{,}90$.

Exemplo 90. Num jogo com 3 moedas, você ganha \$5 se ocorrerem três caras ou três coroas, você perde \$3 se ocorrer uma ou duas caras, se Y é o ganho numa rodada então a esperança do ganho é

$$\mathbb{E}[Y] = 5\frac{1}{4} - 3\frac{3}{4} = -1.$$

Com que probabilidade um jogador ganha \$7 em três rodadas consecutivas e independentes?

Exemplo 91. Qual é o valor médio da soma dos pontos no lançamento de dois dados? O espaço amostral é composto por 36 eventos elementares igualmente prováveis, se X é o resultado da soma dos lançamentos, então sua função de massa de probabilidade é

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

O valor esperado da soma é

$$\mathbb{E}[X] = 2\frac{1}{36} + 3\frac{2}{36} + \cdots + 11\frac{2}{36} + 12\frac{1}{36} = \frac{252}{36} = 7.$$

Exemplo 92 ($\mathbb{E}[X] = \infty$). Numa urna estão 1 bola branca e 1 bola preta; uma bola é escolhida ao acaso, se for preta ela é devolvida e mais uma bola preta é colocada na urna e o sorteio é repetido, se sair bola branca o experimento termina. Se X é o número de rodadas até terminar então

$$\mathbb{P}(X = k) = \frac{1}{k(k+1)}$$

e a média é

$$\sum_{k=1}^{\infty} k\frac{1}{k(k+1)} = \sum_{k=2}^{\infty} \frac{1}{k} = \infty$$

(essa é a série harmônica).

3.0.2 Observação. *No caso do último exemplo dizemos que a variável aleatória não tem esperança finita.*

3.0.3 Teorema. *Se X é uma variável aleatória com esperança finita e g uma função real a valores reais, então $g(X)$ é uma variável aleatória e*

$$\boxed{\mathbb{E}[g(X)] = \sum_{x} g(x) f_X(x)}$$

com a soma sobre todo x tal que $f_X(x) > 0$.

Demonstração. Definamos $Y := g(X)$ e temos

$$\mathbb{E}[Y] = \sum_{y} y\,\mathbb{P}(Y = y) = \sum_{y} y\,\mathbb{P}(g(X) = y).$$

Se $\omega \in [g(X) = y]$ então existe um $x \in \mathbb{R}$ tal que $X(\omega) = x$ e $g(x) = y$ e, se $\omega \in [X = x]$ para algum x tal que $g(x) = y$ então $\omega \in [g(X) = y]$, portanto

$$[g(X) = y] = \bigcup_{\substack{x \\ g(x)=y}} [X = x]$$

em que a união é sobre todo x tal que $g(x) = y$ e é uma união de eventos mutuamente exclusivos, de modo que

$$\mathbb{E}[Y] = \sum_{y} y\,\mathbb{P}(g(X) = y) = \sum_{y} y \left(\sum_{\substack{x \\ g(x)=y}} \mathbb{P}(X = x) \right) = \sum_{y} \sum_{\substack{x \\ g(x)=y}} g(x)\mathbb{P}(X = x) = \sum_{x} g(x)\mathbb{P}(X = x)$$

que é o resultado que queríamos obter. □

3.0.4 Corolário (linearidade da esperança). *Para quaisquer* $a, b \in \mathbb{R}$, *vale* $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

Demonstração. Basta tomar $g(x) = ax + b$ no teorema. □

3.0.5 Proposição. *Se* $X \sim \mathrm{Binomial}(n, p)$ *e* k *é um inteiro positivo então*

$$\mathbb{E}\left[X^k\right] = np\mathbb{E}\left[(Y+1)^{k-1}\right]$$

em que $Y \sim \mathrm{Binomial}(n-1, p)$.

Demonstração. Pelo teorema 3.0.3 acima

$$\mathbb{E}\left[X^k\right] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

e usando o exercício 47

$$i\binom{n}{i} = n\binom{n-1}{i-1}$$

de modo que

$$\mathbb{E}\left[X^k\right] = \sum_{i=1}^{n} i^{k-1} n \binom{n-1}{i-1} p^i (1-p)^{n-i} = np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i}$$

fazendo a mudança de variável $j = i - 1$

$$\mathbb{E}\left[X^k\right] = np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

e resta observar que a soma acima $\sum_j (j+1)^{k-1}\binom{n-1}{j} p^j (1-p)^{n-1-j}$ é exatamente $\mathbb{E}\left[(Y+1)^{k-1}\right]$ para $Y \sim \mathrm{Binomial}(n-1, p)$, também pelo teorema 3.0.3 acima. □

3.0.6 Corolário. *Se* $X \sim \mathrm{Binomial}(n, p)$ *então* $\mathbb{E}[X] = np$.

Demonstração. Basta tomar $k = 1$ na proposição 3.0.5. □

Exemplo 93. Seja X a variável aleatória que descreve o número de carros lavados num lava-rápido em 1 hora, cuja função de probabilidade é

x	4	5	6	7	8	9
$f_X(x)$	1/12	1/12	1/4	1/4	1/6	1/6

Se para x carros lavados o atendente recebe $2x - 1$ reais do gerente, então o ganho médio, por hora, é

$$\mathbb{E}[2X - 1] = \sum_{x=4}^{9} (2x-1) f_X(x) = 12{,}67.$$

Variância: A variância da variável aleatória X é uma medida de quão dispersos estão os valores que a variável assume com relação ao valor médio. Por exemplo, se $\mathbb{P}(X = 100) = \mathbb{P}(X = -100) = 1/2$ e $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$ então X e Y têm valor esperado 0 mas Y assume valores mais próximos da média que X. Essa característica da distribuição dos valores em torno da média não é capturada pela esperança. Uma opção é computar o valor médio da distância entre um valor de X e o valor médio $\mathbb{E}[X]$, isto é, a esperança de $|X - \mathbb{E}[X]|$, entretanto, uma solução mais conveniente do ponto de vista matemático é calcular o valor médio do quadrado desses desvios $|X - \mathbb{E}[X]|^2$.

Variância:: a variância de uma variável aleatória X de esperança finita é o valor esperado da variável aleatória $g(X) = (X - \mathbb{E}[X])^2$.

$$\boxed{\mathrm{Var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}$$

Da definição temos

$$\mathrm{Var}[X] = \sum_x (x - \mathbb{E}[X])^2 f_X(x)$$

e o seguinte exercício fornece um modo, em geral, mais fácil para computar a variância.

Exercício 86. Prove que

$$\boxed{\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.}$$

Desvio padrão:: é definido como a raiz quadrada positiva da variância

$$\boxed{\sigma_X = \sqrt{\mathrm{Var}(X)}.}$$

Suponhamos que X é, por exemplo, a quantidade de refrigerante engarrafada por uma máquina de uma fábrica em ml (mililitros). Então $\mathrm{Var}[X]$ é a dispersão dos valores de X com respeito a média em ml^2, o desvio padrão é uma grandeza em ml.

Exemplo 94. Se X é o resultado do lançamento de um dado

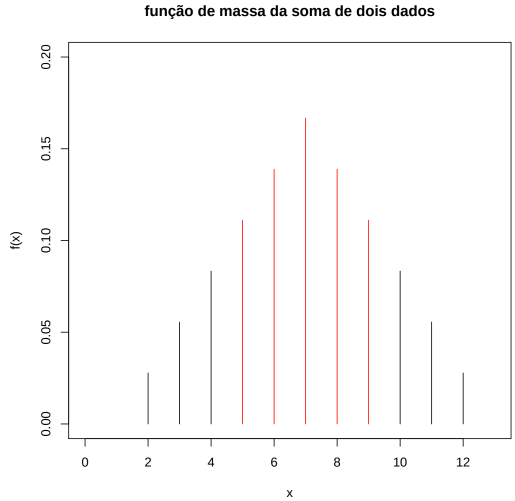
$$\mathbb{E}\left[X^2\right] = \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6}$$

e a variância no valor resultante do lançamento de dado é $91/6 - (7/2)^2 = 35/12 \approx 2,91$ (veja o exemplo 88). O desvio padrão é $\approx 1,7$.

Exemplo 95. Retomando o exemplo 91, se X é o resultado da soma do lançamento de dois dados, então o valor esperado da soma é 7 e o da variância é

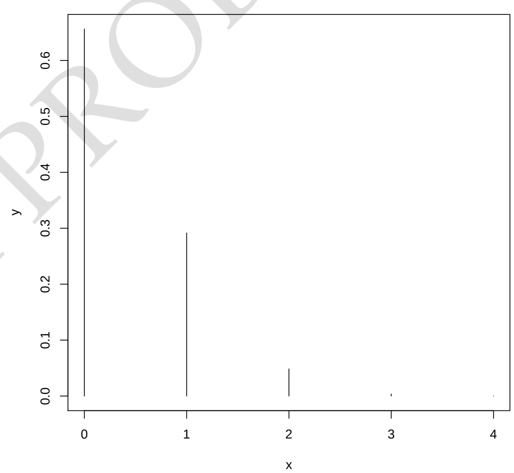
$$\mathrm{Var}(X) = (2-7)^2(1/36) + (3-7)^2(2/36) + \cdots + (11-7)^2(2/36) + (12-7)^2(1/36) = \frac{210}{36} = 5,83.$$

O desvio padrão vale $\sigma_X \approx 2,41$. Notemos que no intervalo $(\mathbb{E}[X] - \sigma_X, \mathbb{E}[X] + \sigma_X) = \{5, 6, 7, 8, 9\}$ está concentrado 2/3 da massa de probabilidade.



Exemplo 96. Um canal digital transmite informação em pacotes de 4 bits. Os bit podem ser recebidos com erro e X denota o número de bits errados num pacote, com função de massa de probabilidade

x	$f(x)$
0	$0,6561$
1	$0,2916$
2	$0,0486$
3	$0,0036$
4	$0,0001$



e o valor médio do número de bits errados é

$$\begin{aligned}\mathbb{E}[X] &= 0\cdot f(0) + 1\cdot f(1) + 2\cdot f(2) + 3\cdot f(3) + 4\cdot f(4)\\ &= 0\cdot 0,6561 + 1\cdot 0,2916 + 2\cdot 0,0486 + 3\cdot 0,0036 + 4\cdot 0,0001\\ &= 0,4\end{aligned}$$

e a variância

$$\begin{aligned}\mathrm{Var}[X] &= \sum_{x=0}^{4}(x-\mathbb{E}[X])^2 f(x)\\ &= 0,16\cdot 0,6561 + 0,36\cdot 0,2916 + 2,56\cdot 0,0486 + 6,76\cdot 0,0036 + 12,96\cdot 0,001296\\ &= 0,36\end{aligned}$$

portanto o desvio padrão é $0,6$.

Exemplo 97 (**variância de uma variável Bernoulli**). Se $X \sim$ Bernoulli(p) então

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

pois $\mathbb{E}\left[X^2\right] = 1^2 p + 0^2(1-p) = p$.

Exemplo 98 (**variância de uma variável binomial**). Se $X \sim$ Binomial(n, p) então, pela proposição 3.0.5 e seu corolário

$$\text{Var}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = np\mathbb{E}[Y+1] - (np)^2$$

com $Y \sim$ Binomial(n, p). Pela linearidade da esperança $\mathbb{E}[Y+1] = \mathbb{E}[Y] + 1$, logo $\mathbb{E}[Y+1] = (n-1)p+1$. Assim,

$$\text{Var}[X] = np(n-1)p + np - (np)^2 = np - np^2 = np(1-p).$$

Exercício 87 (**esperança e variância de uma variável Poisson**). A esperança e a variância de $X \sim$ Poisson(λ) é

$$\mathbb{E}[X] = \text{Var}[X] = \lambda.$$

Use que $\sum_{j\geqslant 0} \frac{\lambda^j}{j!} = e^\lambda$ para provar que $\mathbb{E}[X] = \lambda$. Prove que $\mathbb{E}\left[X^2\right] = \lambda(\lambda+1)$. Conclua que $\text{Var}[X] = \lambda$.

Exercício 88 (**esperança e variância de uma variável hipergeométrica**). A média e a variância de uma variável aleatória hipergeométrica são dadas por

$$\mathbb{E}[X] = sp \qquad \text{e} \qquad \text{Var}[X] = sp(1-p)\frac{n-s}{n-1}$$

em que $p = \frac{a}{n}$. A prova dessas fórmulas é similar à prova para variável aleatória binomial, segue da identidade $\mathbb{E}\left[X^k\right] = \frac{sa}{n}\mathbb{E}\left[(Y+1)^k\right]$ em que Y é uma variável aleatória hipergeométrica com parâmetros $n-1, a-1, s-1$. Use as identidades

$$i\binom{a}{i} = a\binom{a-1}{i-1} \qquad \text{e} \qquad s\binom{n}{s} = n\binom{n-1}{a-1}$$

e obtenha delas $\mathbb{E}\left[X^k\right] = \frac{as}{n}\mathbb{E}\left[(Y+1)^k\right]$. Em seguida, deduza a esperança e variância.

3.0.7 Proposição. *Se X é uma variável aleatória discreta com esperança finita e $a, b \in R$, então*

$$\text{Var}[aX+b] = a^2\text{Var}[X].$$

Demonstração. Sejam X, a e b como no enunciado. Abaixo usamos várias vezes a linearidade da esperança, corolário 3.0.4.

Da definição $\text{Var}[aX+b] = \mathbb{E}\left[(aX+b)^2\right] - (\mathbb{E}\left[aX+b\right])^2$. Do primeiro termo deduzimos

$$\mathbb{E}\left[(aX+b)^2\right] = \mathbb{E}\left[a^2X^2+2abX+b^2\right] = \mathbb{E}\left[a^2X^2\right] + \mathbb{E}\left[2abX\right] + \mathbb{E}\left[b^2\right] = a^2\mathbb{E}\left[X^2\right] + 2ab\mathbb{E}\left[X\right] + b^2$$

e do segunda termo $(\mathbb{E}\left[aX+b\right])^2 = (a\mathbb{E}\left[X\right]+b)^2 = a^2(\mathbb{E}\left[X\right])^2 + 2ab\mathbb{E}\left[X\right] + b^2$ portanto

$$\text{Var}(aX+b) = a^2\mathbb{E}\left[X^2\right] - a^2(\mathbb{E}\left[X\right])^2 = a^2\left(\mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2\right) = a^2\text{Var}[X].$$

□

Esperança e variância de variáveis aleatórias contínuas: Se X é uma variável aleatória com função de densidade de probabilidade f_X então o valor médio (ou valor esperado, ou esperança) da variável aleatória X é

$$\mathbb{E}\left[X\right] := \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x \tag{3.0.6}$$

3.0.8 Observação (uma justificativa informal para equação (3.0.6)). *A definição de valor médio no caso discreto é intuitiva. No caso contínuo podemos justificar, ingenuamente, da seguinte maneira: Sejam $I_n = (y_n, y_{n+1}]$, para $n \in \mathbb{Z}$, uma coleção de intervalos centrados em x_n que* particiona *a reta e que, por simplicidade, supomos todos do mesmo comprimento ε. Definimos a variável aleatória discreta Y sobre o mesmo espaço amostral dada por*

$$Y = \sum_n x_n \mathbf{1}_{[X\in I_n]}$$

que assume os valores x_n ($n \in \mathbb{Z}$). Assim $[Y = x_n] = [X \in I_n]$ e a esperança de Y é

$$\mathbb{E}\left[Y\right] = \sum_n x_n \mathbb{P}(Y = x_n) = \sum_n x_n \mathbb{P}(X \in I_n).$$

Notemos que se $\omega \in [X \in I_n]$, então $X(\omega) \in I_n$ e $Y(\omega) = x_n$, logo

$$|Y(\omega) - X(\omega)| \leqslant \frac{|y_{n+1} - y_n|}{2} = \frac{\varepsilon}{2} \quad (\forall \omega \in \Omega).$$

Disso, a definição de esperança para a variável X deve satisfazer $|\mathbb{E}\left[X\right] - \mathbb{E}\left[Y\right]| \leqslant \frac{\varepsilon}{2}$, logo $EX = \lim_{\varepsilon\to 0}\mathbb{E}\left[Y\right]$

$$\mathbb{E}\left[X\right] = \lim_{\varepsilon\to 0}\sum_n x_n\mathbb{P}(y_n < X \leqslant y_{n+1}) = \lim_{\varepsilon\to 0}\sum_n x_n(F_X(y_{n+1}) - F_X(y_n)) = \int_{-\infty}^{+\infty} x\,\mathrm{d}F_X = \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x$$

(lembremos que $F_X' = f_X$).

Por exemplo, seja T o tempo de vida útil de um equipamento eletrônico em horas. T tem f.d.p.

$$f(t) = \begin{cases} \frac{20.000}{t^3} & \text{se } t > 100 \\ 0 & \text{caso contrário.} \end{cases}$$

O tempo médio de vida é

$$\mathbb{E}[T] = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_{100}^{+\infty} \frac{20.000}{x^2}\,\mathrm{d}x = 200 \text{ horas.}$$

Exemplo 99 ($\mathbb{E}[X] = \infty$). Seja X uma variável com f.d.p.

$$f(x) = \begin{cases} \frac{10}{x^2} & \text{se } x > 10 \\ 0 & \text{caso contrário.} \end{cases}$$

X tem esperança

$$\mathbb{E}[X] = \int_{10}^{\infty} x\frac{10}{x^2}\,\mathrm{d}x = \int_{10}^{\infty} \frac{10}{x}\,\mathrm{d}x = 10\ln(x)\Big|_{10}^{\infty} = \infty$$

3.0.9 Observação. *No caso acima dizemos que a variável aleatória não tem esperança finita.*

3.0.10 Teorema. *Se X é uma variável aleatória com esperança finita e função de densidade de probabilidade f_X e g uma função real a valores reais contínua, então $g(X)$ é uma variável aleatória contínua e*

$$\boxed{\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x}$$

Demonstração. Começamos deduzindo que, para X variável aleatória contínua

$$\begin{aligned} \mathbb{P}(X > a) &= 1 - F_X(a) = \int_{-\infty}^{+\infty} f_X(u)\,\mathrm{d}u - \int_{-\infty}^{a} f_X(u)\,\mathrm{d}u \\ &= \int_{-\infty}^{a} f_X(u)\,\mathrm{d}u + \int_{a}^{+\infty} f_X(u)\,\mathrm{d}u - \int_{-\infty}^{a} f_X(u)\,\mathrm{d}u \\ &= \int_{a}^{+\infty} f_X(u)\,\mathrm{d}u \end{aligned}$$

agora, se $X \geqslant 0$ então

$$\begin{aligned} \int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x &= \int_0^{+\infty}\int_x^{+\infty} f_X(u)\,\mathrm{d}u\,\mathrm{d}x \\ &= \int_0^{+\infty}\int_0^{u} f_X(u)\,\mathrm{d}x\,\mathrm{d}u \\ &= \int_0^{+\infty} u f_X(u)\,\mathrm{d}u = \mathbb{E}[X]. \end{aligned}$$

Para provar o teorema, primeiro assumiremos que $g(x) \geqslant 0$ para todo x. Então,

$$\mathbb{E}[g(X)] = \int_0^{+\infty} \mathbb{P}(g(X) > x)\,\mathrm{d}x = \int_0^{+\infty}\int_B f(u)\,\mathrm{d}u\,\mathrm{d}x$$

em que $B = \{u \in \mathbb{R} : g(u) > x\}$

$$\mathbb{E}[g(X)] = \int_0^{+\infty} \int_B f(u)\, du\, dx = \int_0^{+\infty} \int_0^{g(u)} f(u)\, dx\, d(u) = \int_0^{+\infty} g(u)f(u)\, du$$

que prova o enunciado pelo teorema para g não negativa. Pra finalizar, se g assume valores reais então definimos as variáveis aleatórias não negativas

$$g^+(x) = \max\{g(x), 0\} \quad \text{e} \quad g^-(x) = \max\{-g(x), 0\}$$

e temos que $g(x) = g^+(x) - g^-(x)$, portanto,

$$\begin{aligned}
\mathbb{E}[g(X)] &= \mathbb{E}[g^+(X)] - \mathbb{E}[g^-(X)] \\
&= \int_0^{+\infty} g^+(u)f(u)\, du - \int_0^{+\infty} g^-(u)f(u)\, du \\
&= \int_0^{+\infty} g(u)f(u)\, du.
\end{aligned}$$

□

3.0.11 Corolário (linearidade da esperança). *Para quaisquer* $a, b \in \mathbb{R}$, *vale* $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

Demonstração. Basta tomar $g(x) = ax + b$ no teorema. □

Por exemplo, seja Z uma variável aleatória com f.d.p

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

e $g : \mathbb{R} \to \mathbb{R}$ é dada por $g(x) = 4x + 3$ então

$$\mathbb{E}[g(Z)] = \mathbb{E}[4Z + 3] = \int_{-1}^{2} (4x + 3)\frac{x^2}{3}\, dx = 8.$$

Variância

A **variância** da variável aleatória contínua X é dada pelo valor esperado de $g(X) = (X - \mathbb{E}[X])^2$, sempre que $\mathbb{E}[X] < \infty$

$$\boxed{\mathrm{Var}[X] := \mathbb{E}[X - \mathbb{E}[X]]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2}$$

donde temos

$$\mathrm{Var}[X] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f(x)\, dx = \int_{-\infty}^{+\infty} x^2 f(x)\, dx - \left(\int_{-\infty}^{+\infty} x f(x)\, dx\right)^2$$

O **desvio padrão** é definido como a raiz quadrada positiva da variância, caso seja finita,

$$\sigma_X = \sqrt{\mathrm{Var}[X]}.$$

A variância tem a mesma propriedade do caso discreto.

3.0.12 Proposição. *Se X, Y são variáveis aleatórias contínuas com esperança finita e $a, b \in R$, então*

$$\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X].$$

Exercício 89. Num jogo de apostas, se o ganho é x e a perda é y em cada rodada, então o ganho médio é

$$x \cdot \mathbb{P}(\text{ocorrencias favoraveis}) + y \cdot \mathbb{P}(\text{ocorrencias desfavoraveis}).$$

Uma variável aleatória U não-negativa em função de distribuição acumulada F e densidade $f = F'$. Um jogo lhe é oferecido da seguinte forma: você pode escolher um número não negativo c, se $U > c$ então você ganha a quantidade c, caso contrário, você não ganha nada. Como exemplo, suponha que U é a altura (medida em cm) da próxima pessoa entrando em uma estação ferroviária pública específica. Se você escolher $c = 100$, então você quase certamente ganha essa quantia. Um valor de $c = 200$ dobraria a quantia se você ganhar, mas reduz drasticamente a sua probabilidade de ganhar. Encontre uma equação para caracterizar o valor de c que maximiza o ganho médio.

Exercícios.

1. Prove que $1 - F_X(a) = \mathbb{P}(X > a)$, para todo $a \in \mathbb{R}$.

2. Prove que $F_X(b) - F_X(a) = \mathbb{P}(a < X \leqslant b)$ para quaisquer $a < b$.

3. Seja X uma variável aleatória e $a, b \in \mathbb{R}$. Defina a função Y por $Y(\omega) = aX(\omega) + b$, para todo $\omega \in \Omega$. (a) Verifique que Y é uma variável aleatória, isto é, $[Y \leqslant t]$ é um evento aleatório, para todo $t \in \mathbb{R}$. (b) Se F é função de distribuição acumulada de X, determine a função de distribuição acumulada de Y.

4. Sejam X e Y variáveis aleatórias de um mesmo modelo probabilístico. Verifique que $X + Y$, dada por $X + Y(\omega) := X(\omega) + Y(\omega)$, é variável aleatória e $X \cdot Y$, dada por $X \cdot Y(\omega) := X(\omega) \cdot Y(\omega)$, é uma variável aleatória.

5. Duas bolas são escolhidas aleatoriamente de uma urna que contém 8 bolas brancas, 4 pretas e 2 laranjas. Suponha que ganhemos \$2,00 para cada bola preta selecionada e percamos \$1,00 para cada bola branca selecionada. Suponha que X represente nosso ganho. Qual são os valores possíveis de X e quais são as probabilidades associadas a cada valor?

6. Suponha que um dado equilibrado seja lançado duas vezes. Determine os possíveis valores que as seguintes variáveis aleatórias podem assumir:

   (a) o valor máximo dentre os resultados dos dois lances;

   (b) o valor mínimo dentre os resultados dos dois lances;

   (c) a soma dos valores dos dois lançamentos;

   (d) o valor da primeira jogada menos o valor da segunda jogada.

   Calcule as probabilidades associadas as variáveis aleatórias nas letras (a) a (d).

7. Verifique que a seguinte função $F : \mathbb{R} \to \mathbb{R}$ é de distribuição acumulada

$$F(t) = \begin{cases} 0, & \text{se } x < 0 \\ \frac{2x+1}{2}, & \text{se } 0 \leqslant x < 1/2 \\ 1, & \text{caso contrário.} \end{cases}$$

   Calcule as seguintes probabilidades supondo que X é uma variável aleatória com função de distribuição F

   a) $\mathbb{P}(X < 1/4)$; b) $\mathbb{P}(1/4 < X < 3/4)$; c) $\mathbb{P}(X < 3/4 \mid X > 1/4)$.

8. Prove que se F é função de distribuição acumulada de X então

   a) $\mathbb{P}(a < X < b) = F(b^-) - F(a)$;
   b) $\mathbb{P}(a \leqslant X < b) = F(b^-) - F(a^-)$;
   c) $\mathbb{P}(a \leqslant X < b) = F(b) - F(a^-)$;

9. Prove que se X é variável aleatória discreta com função de massa de probabilidade f então

$$\mathbb{P}(a \leqslant X \leqslant b) = \sum_{x=a}^{b} f(x)$$

   onde a soma é sobre todo x tal que $a \leqslant x \leqslant b$ e $f(x) > 0$.

10. Prove que $X \sim$ Geometrica(p) é sem memória, isto é

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

11. Seja X o número de lançamentos (independentes) de uma moeda até sair cara. Suponha que $\mathbb{P}(\text{Ca}) = p$ para algum $p \in (0, 1)$ fixo. Determine a função de massa de probabilidade de X e a função de distribuição de probabilidade de X. $f(x) = (1-p)^{x-1}p$

12. Se ensaios de Bernoulli com probabilidade de sucesso p são realizados de modo independente até que hajam r sucessos. Qual é a probabilidade de serem necessários n ensaios ($n \geqslant r$)? $\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$

13. Se ensaios de Bernoulli com probabilidade de sucesso p são realizados de modo independentemente, qual é a probabilidade de que r sucessos ocorram antes que m fracassos? $\sum_{n=r}^{r+m-1}\binom{n-1}{r-1}p^{r}(1-p)^{n-r}$

14. Em um teste de múltiplas escolhas com 3 respostas possíveis para cada uma das 5 questões, com que probabilidade um estudante acerte pelo menos 4 questões apenas chutando aleatoriamente? 11/243

15. Quatro jogadas independentes de uma moeda honesta são feitas. Seja X o número de caras. Determine a f.m.p. de $X - 2$.

16. Um sujeito se alega paranormal. Num teste, uma moeda honesta é jogada 10 vezes e pede-se ao homem a previsão dos resultados. Ele acerta 7 respostas. Qual é a probabilidade desse evento supondo que o homem chuta aleatoriamente cada previsão?

17. Suponha que se saiba que o número de carros que chegam a um cruzamento específico durante um período de tempo de 20s é caracterizado pela função de massa

$$f(x) = \frac{e^{-6}6^x}{x!}$$

para todo $x \in \mathbb{N}$. Determine a probabilidade com que num período específico de 20s mais de oito carros cheguem ao cruzamento. 0,1528 Determine com que probabilidade apenas 2 carros cheguem. 0,446.

18. Um livro de jogos de azar recomenda o seguinte para ganhar na roleta: aposte 1 real no vermelho, se der vermelho (probabilidade 18/38) pegue o lucro de 1 real e desista. Se perder a aposta (probabilidade 20/38) faça apostas adicionais de 1 real no vermelho nos próximos dois giros e depois desista. Seja X o lucro obtido com essa estratégia.

    a) determine $\mathbb{P}(X > 0)$; 0,5918
    b) essa é uma estratégia de vitória? não

19. Um motor de avião tem probabilidade $1 - p$ de falhar durante o voo, de forma independente dos outros motores. Se o avião precisa da maioria dos motores para terminar o voo de forma segura, para quais valores de p é preferível um avião de 5 motores a um de 3? $p \geqslant 1/2$

20. Uma moeda com probabilidade p para cara é lançada 10 vezes. Dado que o número de caras é 6 qual é a probabilidade com que os 3 primeiros resultados são

a) CARA,COROA,COROA 1/10

b) COROA,CARA,COROA 1/10

21. **Paradigma de Poisson**: Consideremos n eventos sendo que o evento i ocorre com probabilidade p_i. Se os valores de p_i são pequenos e o eventos independentes (ou "fracamente dependentes") então os eventos que ocorrem têm aproximadamente distribuição de Poisson com parâmetro $\sum_{i=1}^{n} p_i$.

    Em resposta a um ataque de 10 mísseis, um contra-ataque de 500 mísseis antiaéreos é efetuado de modo que o alvo de um míssil antiaéreo é qualquer um dos 10 mísseis, independentemente, com probabilidade 0,1. Use o paradigma de Poisson para determinar uma aproximação para a probabilidade de todos os 10 mísseis serem atingidos. $(1-e^{-5})^{80}$

22. Os times A e B jogam uma série de partidas. O vencedor é aquele que ganhar primeiro 3 partidas. A vence cada partida com probabilidade p independentemente das outras partidas. Qual é a probabilidade condicional de que A vença

    a) a série dado que tenha ganho a primeira partida; $\sum_{i=2}^{4}\binom{4}{i}p^i(1-p)^{4-i}$

    b) a primeira partida dado que ganhou a série. $(\sum_{i=2}^{4}\binom{4}{i}p^{i+1}(1-p)^{4-i})/(\sum_{i=3}^{5}\binom{5}{i}p^i(1-p)^{5-i})$

23. Cada um dos membros de um corpo de 7 jurados toma a decisão correta com probabilidade 0,7, de forma independente um dos outros. Se a decisão do corpo é feita pela regra da maioria, com que probabilidade a decisão correta é tomada? $\sum_{i=4}^{7}\binom{7}{i}(0,7)^i(0,3)^{7-i}$ Dado que 4 juízes tenham a mesma opinião dos jurados, qual é a probabilidade dos jurados tomarem a decisão correta? 0,7

24. Numa roleta com números de 0 a 36 e um 00 João sempre aposta nos números de 1 a 12. Com que probabilidade

    a) João perde suas 5 primeira apostas; 0,15

    b) sua primeira vitória ocorra na quarta aposta. 0,1012

25. Um comprador de transistores os compra em lote de 20 só se a inspeção de 4 deles não apresentar defeito. Um transistor é defeituoso com probabilidade $0,1$ de modo independente dos demais. Qual é a proporção de lotes rejeitados? 0,3439

26. Há 3 auto-estradas num país. O número de acidentes que ocorrem diariamente nelas é uma variável aleatória de Poisson com parâmetros $0,3$, $0,5$ e $0,7$. Determine o número esperado de acidentes para hoje. 1,5

27. Sejam X_1 e X_2 variáveis aleatórias independentes e identicamente distribuídas, com $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. Considere $X_3 = X_1 X_2$. As variáveis aleatórias X_1 , X_2 e X_3 são independentes? São independentes duas a duas?

28. Um aluno estuda 12 exercícios, dos quais o professor vai escolher 6 aleatoriamente para uma prova. O estudante sabe resolver 9 dos 12 problemas. Seja X o número de exercícios resolvidos por ele na prova. (a) Qual a função de distribuição de X? Hipergeométrica(6, 9, 12) (b) Calcule a probabilidade de que o aluno resolva ao menos 5 exercícios da prova. 1/2

29. Uma urna contém cinco bolas numeradas de 1 a 5. Duas bolas são retiradas simultaneamente. Obtenha a função de probabilidade e faça o gráfico da função de distribuição das seguintes variáveis aleatórias: (a) o maior número sorteado. (b) a soma dos números retirados.

30. Um carregamento com 7 televisores contém dois aparelhos defeituosos. Um hotel compra 3 desses 7 aparelhos ao acaso. Se X é o número de aparelhos comprados,

    a) determine a função de probabilidade de X; 0-2/7;1-4/7;2-1/7
    b) determine a função de distribuição acumulada de X;
    c) usando a f.d.a, compute $\mathbb{P}(X = 1)$ e $P(0 < X \leqslant 2)$.

31. Determine o valor de c para que cada uma das seguintes funções seja uma função de probabilidade

    a) $f(x) = c(x^2 + 4)$ para $x = 0, 1, 2, 3$, caso contrário $f(x) = 0$; 1/30
    b) $f(x) = c\binom{2}{x}\binom{3}{3-x}$, para $x = 0, 1, 2$, caso contrário $f(x) = 0$; 1/10
    c) $f(x) = c\sqrt{x}$, para $0 < x < 1$, caso contrário $f(x) = 0$; 3/2

32. O número de horas, em unidades de 100 horas, que uma família usa o aspirador de pó é uma variável aleatória com f.d.p.

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

Determine a probabilidade com que uma família use o aspirador

    a) menos de 120 horas; 0,68
    b) entre 50 e 100 horas. 0,375

33. Um fator importante no combustível sólido de um míssil é a distribuição do tamanho de partículas, partículas grandes podem acarretar sérios problemas. Dos dados de produção foi determinado que o tamanho (em μm) das partículas é caracterizada por

$$f(x) = \begin{cases} 3x^{-4} & \text{se } x > 1; \\ 0, & \text{caso contrário.} \end{cases}$$

Verifique se essa é uma f.d.p. válida e, caso seja, determine a f.d.a. F e determine qual é a probabilidade de que uma partícula aleatória exceda $4\,\mu m$? 0,0156

34. Se Y tem distribuição uniforme no intervalo $(0, 5)$, qual é a probabilidade de que as raízes da equação $4x^2 + 4xY + Y + 2 = 0$ sejam ambas reais?

35. Uma loja de comércio eletrônico envia *emails* com ofertas especiais a seus clientes cadastrados. Suponha que, após o recebimento de uma mensagem, a proporção de clientes que efetivam uma compra é uma variável aleatória com densidade dada por $f(x) = cx(1-x)^5$, para $0 \leqslant x \leqslant 1$ e $f(x) = 0$ para os outros valores de x. (a) Encontre o valor de c. (b) Calcule a probabilidade de que um *email* resulte em alguma compra para mais de 50% dos seus destinatários

36. Um casal combina de se encontrar em certo local perto das 12h30. Suponha que o homem chega em uma hora uniformemente distribuída entre 12h15 e 12h45 e a mulher, independentemente, chega em uma hora uniformemente distribuída entre 12h00 e 13h00. Encontre as probabilidades de que (a) o primeiro a chegar não espere mais que 5 minutos pelo segundo. (b) a mulher chegue primeiro.

37. Para $X \sim \mathcal{N}(10, 35)$ calcule

   a) $\mathbb{P}(X > 5)$ 0,7977
   b) $\mathbb{P}(4 < X < 16)$ 0,6827
   c) $\mathbb{P}(X < 8)$ 0,3695
   d) $\mathbb{P}(X < 20)$ 0,9522
   e) $\mathbb{P}(X > 16)$ 0,1587

38. O tempo em horas para a manutenção de uma máquina é uma variável aleatória com distribuição Exp(1/2).

   a) Qual é a probabilidade com que um reparo dure mais de 2 horas? e^{-1}
   b) Qual é a probabilidade com que um reparo dure mais de 10 horas dado que sua duração seja superior a 9 horas? $e^{-1/2}$

39. Determine a f.d.a. de $X \sim$ Uniforme($[a, b]$) e esboce o seu gráfico.

40. Seja X uma variável aleatória contínua que interpretamos como a vida útil (tempo) de um item, sejam F e f a f.d.a. e f.d.p. de X, respectivamente. A *taxa de falhas* é a função definida por

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

De um argumento que mostra que λ pode ser interpretada como a probabilidade (condicionada a idade) de que um item com idade t apresente falha (i.e., $\mathbb{P}(t < X < t + \Delta t \mid X > t)$. Veja equação (2.2.2)).

Determine $\lambda(t)$ no caso $X \sim \text{Exp}(\alpha)$.

41. Sejam X_1 e X_2 variável aleatória independentes exponencialmente distribuídas com parâmetros λ_1 e λ_2, respectivamente. Mostre que $min\{X_1, X_2\}$ é exponencialmente distribuída com parâmetro $\lambda_1 + \lambda_2$.

42. Em um julgamento de paternidade, um perito atesta que a extensão em dias da gestação humana segue a distribuição $\mathcal{N}(279; 100)$. O réu é capaz de provar que estava fora do país durante um período que começou 290 dias antes do nascimento da criança e terminou 240 dias depois do nascimento. Se o réu é de fato o pais da criança, qual é a probabilidade com que mão possa ter tido uma gestação muito longa ou muito curta indicada pela testemunha? 0,0241

43. Defina uma família de eventos E_a para cada $a \in (0,1)$, com a propriedade $\mathbb{P}(E_a) = 1$ para todo a mas $\mathbb{P}(\bigcap_a E_a) = 0$.(dica: defina E_a em termos de $X \sim$ Uniforme($[0,1]$).)

44. Seja X uma variável aleatória discreta que assume valores em $\mathbb{N}$. Prove que

$$\mathbb{E}[X] = \sum_{i \geqslant 1} \mathbb{P}(X \geqslant i).$$

45. Prove que se $X \sim$ Geometrica(p) então

$$\mathbb{E}[X] = \frac{1}{p}.$$

(Dica: use o exercício 1)

46. Você tem R\$1.000,00 para comprar uma mercadoria que é vendida a R\$2,00 o quilo. Na semana que vem a mesma mercadoria será vendida a R\$1,00 ou R\$4,00, com igual probabilidade.

    a) Se você quer maximizar a quantidade esperada de dinheiro que terá no fim de semana, qual estratégia empregar?

    b) Se você quer maximizar a quantidade esperada de mercadoria que terá no fim de semana, qual estratégia empregar?

47. Uma companhia de seguros vende uma apólice dizendo que uma quantidade A de dinheiro deve ser paga se um determinado evento E ocorra em uma ano. Estima-se que E ocorra em um ano com probabilidade p. Qual é o preço da apólice se o lucro esperado da companhia é 10%? $A(p + 1/10)$

48. Um piloto deseja segurar seu avião no valor de \$200.000. A empresa de seguro estima que uma perda total pode ocorrer com probabilidade $0,002$, uma perda de 50% com probabilidade $0,01$ e uma perda de 25% com probabilidade $0,1$. Ignorando as outras perdas parciais, que prêmio a empresa de seguros deveria cobrar a cada ano para ter um lucro de \$500? 6.900

49. Se o lucro de um vendedor de automóveis, em unidades de \$5.000 é, por automóvel vendido, uma variável aleatória X que tem f.d.p

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

determine o lucro médio por automóvel.

Qual é o lucro médio se o lucro por automóvel vendido é $g(X) = X^2$? 833,33 Determine $\mathrm{Var}[g(X)]$

50. O tempo, em minutos, para a reação humana ao gás lacrimejante tem f.d.p.

$$f(x) = \begin{cases} \frac{1}{4}e^{-x/4} & \text{se } x \geqslant 0 \\ 0 & \text{caso contrário.} \end{cases}$$

Determine o tempo médio para reação. 4 Determine $\mathbb{E}\left[X^2\right]$ 32 e $\mathrm{Var}[X]$ 16.

51. Se X tem distribuição binomial com média 6 e variância 2,4 determine $\mathbb{P}(X = 5)$.

52. O número de ovos que um inseto deposita numa árvore é uma variável aleatória de Poisson com parâmetro λ. Entretanto, tal variável só pode ser observada se for positiva, pois se for 0 não se sabe dizer se o inseto esteve na árvore. Se Y é o número de ovos depositados então

$$\mathbb{P}(Y = i) = \mathbb{P}(X = i \mid X > 0)$$

para $X \sim \text{Poisson}(\lambda)$. Determine $\mathbb{E}[Y]$. $\lambda/(1 - e^{-\lambda})$

53. Cada partida que você joga resulta em vitória com probabilidade p. Você planeja jogar 5 partidas mas se você vencer a quinta partida então terá que jogar até perder.

   a) Qual é o número esperado de partidas jogadas? $4 + 1/(1 - p)$
   b) Qual é o número esperado de partidas perdidas? $4(1 - p) + 1$

54. Se X tem distribuição normal com média $0,5$ e $P(X > 9) = 0,2$, qual é o valor de $\mathrm{Var}[X]$ aproximadamente? 22,66

55. Em um processo industrial, o diâmetro de um rolamento é uma parte importante. O comprador determina que as especificações do diâmetro sejam $3 \pm 0,01$ cm e nenhuma peça fora da especificação será aceita. Sabe-se que o diâmetro do rolamento tem distribuição $\mathcal{N}(3,0; 0,005^2)$. Em média, quantos rolamentos serão inutilizados? 4,56%

56. Calibradores são usados para rejeitar componentes de certa dimensão fora da especificação $1,5 \pm d$. Essa medição é distribuída com média $1,5$ e desvio-padrão $0,2$. Determine o valor par d de modo que as medições cubram 95% das medições. 0,392

57. Certa máquina fabrica resistores elétricos com uma resistência média de 40 ohms e desvio-padrão de 2 ohms. Supondo que a resistência siga uma distribuição normal e que pode ser medida em qualquer grau de acuidade, qual é a porcentagem de resistores que terão uma resistência maior que 43 ohms. 6,68%

58. A nota média de um exame é 74 e o desvio-padrão 7. Se 12% da classe recebe A e as notas são ajustadas para seguir um distribuição normal qual é a menor nota a receber um A e a mais alta a receber um B? 83,82

59. Uma prova é considerada boa se a distribuição das notas daqueles que participaram segue, aproximadamente, uma normal. O professor usa as notas para estimar μ e σ e atribui A para aqueles com nota superior a $\mu + \sigma$, B para aqueles com nota entre μ e $\mu + \sigma$ C para aqueles com nota entre $\mu - \sigma$ e μ, D para aqueles com nota entre $\mu - 2\sigma$ e $\mu - \sigma$ e E para aqueles com nota abaixo de $\mu - 2\sigma$. Qual é a porcentagem da classe que recebe cada um dos conceitos? 16%,34%,34%,14%,2%

60. Seja Y uma variável aleatória não-negativa. Prove que se Y assume valores inteiros (portanto é discreta) então $\mathbb{E}[Y] = \sum_x \mathbb{P}(Y \geqslant x)$. Prove que se Y é contínua então $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > x)\,\mathrm{d}x$.

61. A mediana de uma variável aleatória com f.d.a. F é o valor m tal que $F(m) = \frac{1}{2}$. Determine a mediana de X quando

    a) $X \sim \mathrm{Uniform}e([a, b])$

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

    c) $X \sim \mathrm{Exp}(\lambda)$.

Vetores aleatórios.

Vetor aleatório:: um vetor aleatório X com valores em $\mathbb{R}^n$ é formado por n variáveis aleatórias reais $\mathsf{X} = (X_1, X_2, \dots, X_n)$. Sua distribuição é caracterizada pela função de distribuição n-dimensional

$$F_{\mathsf{X}}(x_1, \dots, x_n) := P(X \leqslant x_1,\ X \leqslant x_2, \dots,\ X \leqslant x_n)$$

em que $X \leqslant x_1,\ X \leqslant x_2, \dots,\ X \leqslant x_n$ é uma simplificação para

$$[X \leqslant x_1] \cap [X \leqslant x_2] \cap \cdots \cap [X \leqslant x_n].$$

A seguir nos limitaremos ao caso bidimensional, a generalização para mais que duas variáveis é fácil.

Função de distribuição acumulada conjunta:: sejam $X,\ Y$ são variáveis aleatórias sobre o mesmo espaço amostral. A f.d.a. conjunta é a função que associa ao par (a, b) de números reais o valor

$$F_{X,Y}(a, b) := \mathbb{P}(X \leqslant a,\ Y \leqslant b).$$

As funções de distribuição de X e Y podem ser inferidas da distribuição conjunta

$$F_X(a) = \mathbb{P}(X \leq a) = \mathbb{P}(X \leq a, Y < \infty).$$

Ademais, se $b_1 < b_2$ então $[Y \leqslant b_1] \subset [Y \leqslant b_2]$, portanto

$$[X \leqslant a] \cap [Y \leqslant b_1] \subset [X \leqslant a] \cap [Y \leqslant b_2]$$

e fazendo $E_n = [X \leqslant a] \cap [Y \leqslant n]$ temos uma sequência crescente (em n) de eventos que pela continuidade de $\mathbb{P}$

$$\lim_{n\to\infty} \mathbb{P}(E_n) = \mathbb{P}\left(\bigcup_{n\geqslant 0} E_n\right)$$

ou seja

$$F_X(a) = \lim_{n\to\infty} F_{X,Y}(a, n)$$

chamada distribuição marginal de X. Analogamente, a distribuição marginal de Y é

$$F_Y(b) = \lim_{a\to\infty} F_{X,Y}(a, b).$$

Exercício 90. Mostre que

$$\mathbb{P}(X > a,\ Y > b) = 1 - F_X(a) - F_Y(b) + F_{X,Y}(a, b).$$

* **Fubini e Tonelli::** Generalizaremos a noção de v.a. para função com valores em $\mathbb{R}^n$. O espaço de eventos é gerado por $\prod_{i=1}^n(-\infty, x_i]$.

f é integrável se

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |f(x_1, \ldots, x_n)| \,\mathrm{d}x_1 \ldots \mathrm{d}x_n < \infty$$

3.0.13 Teorema. *Se f definida no $\mathbb{R}^2$ é mensurável*

1. *se f é positiva*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \,\mathrm{d}x \right) \mathrm{d}y \tag{3.0.7}$$

2. *se $\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x, y)| \,\mathrm{d}x \right) \mathrm{d}y < \infty$ ou $\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x, y)| \,\mathrm{d}y \right) \mathrm{d}x < \infty$ então equação (3.0.7) vale e podemos definir*

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) \,\mathrm{d}x\mathrm{d}y := \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \,\mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \,\mathrm{d}y \right) \mathrm{d}x$$

Função de massa de probabilidade conjunta:: se X e Y são variável aleatória discretas então

$$f_{X,Y}(a, b) = \mathbb{P}(X = a, \ Y = b)$$

é a função de probabilidade conjunta, e para tal valem

- $f_{X,Y}(a, b) \geqslant 0$;
- $\displaystyle\sum_a \sum_b f_{X,Y}(a, b) = 1$;
- $\mathbb{P}((X, Y) \in A) = \displaystyle\sum_{(a,b) \in A} f_{X,Y}(a, b)$, para qualquer conjunto enumerável A de pares de valores reais;

Funções de probabilidade marginais:: de X e de Y são, respectivamente

$$f_X(a) := \sum_b f_{X,Y}(a, b) \quad \text{e} \quad f_Y(b) := \sum_a f_{X,Y}(a, b).$$

Exemplo 100. Dois refis de caneta são selecionados ao acaso dentre 3 azuis, 2 vermelhos e 3 verdes. Sejam X o número de refis azuis e Y o número de refis vermelhos selecionados. A distribuição conjunta é, para $x, y \in \{0, 1, 2\}$ sujeito a condição $x + y \leqslant 2$,

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

cujos valores está descrito na tabela abaixo

$x\backslash y$	0	1	2	$f_X(x)$
0	3/28	3/14	1/28	10/28
1	9/28	3/14	0	15/28
2	3/28	0	0	3/28
$f_Y(y)$	15/28	12/28	1/28	1

A probabilidade $\mathbb{P}((X,Y) \in \{(x,y) \mid x+y \leqslant 1\})$ é $f(0,0)+f(0,1)+f(1,0) = 9/14$. A função de massa marginal de X em i é dada pela soma da linha i da tabela, por exemplo, $f_X(0) = 3/28+3/14+1/28$. Analogamente, a função de massa marginal de Y é dada pela soma da coluna correspondente, por exemplo $f_Y(1) = 3/14+3/14+0$.

Variáveis aleatórias conjuntamente contínuas:: X e Y são variáveis aleatórias conjuntamente contínuas se existe uma função de densidade de probabilidade conjunta $f_{X,Y}$ tal que para quaisquer $a,\ b \in \mathbb{R}$

$$F_{X,Y}(a,b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

e para tal $f_{X,Y}$ valem

- $f_{X,Y}(x,y) \geqslant 0$;
- $\displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$;
- $\mathbb{P}((X,Y) \in A) = \displaystyle\iint_A f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$;

e cada uma das variáveis é, individualmente, contínua com respectivas **funções de densidades marginais**

$$f_X(x) := \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}y \quad \text{e} \quad f_Y(y) := \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}x.$$

Exercício 91. Mostre que

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

quando as derivadas parciais existem.

Exemplo 101. Um banco resolveu apostar num serviço de *drive-thru*, além do atendimento convencional. Em um dia, X é a proporção de tempo que o *drive-thru* está em uso e Y a proporção de tempo que o caixa convencional está em uso, assim $(X,Y) \in \{(x,y) \mid 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1\}$. Estudos indicam que a função de distribuição conjunta é

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & \text{se } 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1 \\ 0 & \text{caso contrário.} \end{cases}$$

A probabilidade de nenhuma das alternativas estar ocupada em mais de um quarto do tempo é

$$\begin{aligned}
\mathbb{P}\left(0 \leqslant X \leqslant \frac{1}{4},\ 0 \leqslant Y \leqslant \frac{1}{4}\right) &= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x+y^2)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}x\,\mathrm{d}x\,\mathrm{d}y + \int_0^{1/4}\int_0^{1/4} \frac{6}{5}y^2\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{7}{640}.
\end{aligned}$$

Ademais

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{6}{5}(x+y)^2\,\mathrm{d}y = \int_0^1 \frac{6}{5}(x+y)^2\,\mathrm{d}y = \frac{6}{5}x + \frac{2}{5}$$

é a função de densidade de probabilidade do tempo que o *drive-thru* está ocupado para $x \in [0,1]$, e $f_X(x) = 0$ para os outros valores de x. Também,

$$f_Y(y) = \int_{-\infty}^{+\infty} \frac{6}{5}(x+y)^2\,\mathrm{d}x = \int_0^1 \frac{6}{5}(x+y)^2\,\mathrm{d}x = \frac{6}{5}y^2 + \frac{3}{5}$$

é a função de densidade de probabilidade do tempo que o caixa convencional está ocupado para $y \in [0,1]$ e $f_Y(y) = 0$ para os outros valores de y.

Exemplo 102. Sejam X e Y as coordenadas de um ponto escolhido no círculo de raio 1 e centro na origem de um sistema de coordenadas com densidade conjunta

$$f(x,y) = \begin{cases} \frac{1}{\pi} & \text{se } x^2+y^2 \leqslant 1 \\ 0 & \text{se } x^2+y^2 > 1 \end{cases}$$

A densidade marginal de X é

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{1}{\pi}\,\mathrm{d}y = \frac{2}{\pi}\sqrt{1-x^2}$$

para todo x tal que $x^2 \leqslant 1$ e $f_X(x) = 0$ nos outros casos. Analogamente,

$$f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$$

para todo x tal que $y^2 \leqslant 1$ e $f_Y(y) = 0$ nos outros casos. A distância do ponto escolhido para a origem é a variável aleatória $D = D(X, Y) = \sqrt{X^2 + Y^2}$ cuja distribuição é, para $a \in [0, 1]$, dada por

$$
\begin{aligned}
F_D(a) &= \mathbb{P}\left(\sqrt{X^2 + Y^2} \leqslant a\right) \\
&= \iint_{\{(x,y): x^2+y^2 \leqslant a^2\}} f(x, y)\, \mathrm{d}x\, \mathrm{d}y \\
&= \iint_{\{(x,y): x^2+y^2 \leqslant a^2\}} \frac{1}{\pi}\, \mathrm{d}x\, \mathrm{d}y \\
&= \frac{1}{\pi} \iint_{\{(x,y): x^2+y^2 \leqslant a^2\}} \mathrm{d}x\, \mathrm{d}y \\
&= \frac{1}{\pi} \pi a^2 \\
&= a^2.
\end{aligned}
$$

Se X e Y têm distribuição conjunta $f_{X,Y}(x, y)$ e h é uma função a duas variáveis reais a valores reais, então o valor médio de $h(X, Y)$ é

$$
\mathbb{E}\left[h(X, Y)\right] := \begin{cases} \displaystyle\sum_x \sum_y h(x, y) f_{X,Y}(x, y) \\ \displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y \end{cases}
$$

nos casos discreto e contínuo, respectivamente.

Exercício 92. O que é $\mathbb{E}\left[h(X, Y)\right]$ no caso $h(X, Y) = X$?

Soma de variáveis aleatórias: Sejam X e Y variáveis aleatórias de um modelo probabilístico. A soma delas é a variável aleatória $X + Y$ dada por

$$(X + Y)(\omega) := X(\omega) + Y(\omega).$$

Se $h(x, y) = ax + by$, para reais a, b, então

$$
\begin{aligned}
\mathbb{E}\left[h(X, Y)\right] &= \sum_x \sum_y (ax + by) f_{X,Y}(x, y) \\
&= \sum_x ax \sum_y f_{X,Y}(x, y) + \sum_y by \sum_y f_{X,Y}(x, y) \\
&= \sum_x ax f_X(x) + \sum_y by f_Y(y) \\
&= a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right].
\end{aligned}
$$

assim como

$$
\begin{aligned}
\mathbb{E}\left[h(X,Y)\right] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(ax+by)f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y\\
&= \int_{-\infty}^{+\infty} ax\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}\right)\mathrm{d}x + \int_{-\infty}^{+\infty} by\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}x\right)\mathrm{d}y\\
&= \int_{-\infty}^{+\infty} axf_X(x)\,\mathrm{d}x + \int_{-\infty}^{+\infty} byf_Y(y)\,\mathrm{d}y\\
&= a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right].
\end{aligned}
$$

3.0.14 Teorema (**linearidade da esperança**). *Se X e Y são variáveis aleatórias de um modelo probabilístico e a e b são reais então*

$$\boxed{\mathbb{E}\left[aX+bY\right] = a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right]}\,.$$

Retomando o exemplo 91, se dois dados honestos são lançados e X é o resultado do primeiro dado e Y o resultado do segundo dado, então o valor esperado para a soma é $E(X+Y) = \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right] = \frac{7}{2} + \frac{7}{2} = 7$.

Exercício 93. Se $X_1,\dots,X_n$ são variáveis aleatórias discretas com esperança finita do mesmo modelo probabilístico então $\mathbb{E}\left[X_1+\cdots+X_n\right] = \mathbb{E}\left[X_1\right]+\cdots+\mathbb{E}\left[X_n\right]$.

Por exemplo, se n dados honestos são lançados, então o valor esperado para a soma dos resultados é $\mathbb{E}\left[X_1+\cdots+X_n\right] = \mathbb{E}\left[X_1\right]+\cdots+\mathbb{E}\left[X_n\right] = n\frac{7}{2}$, em que X_i é o valor do resultado de um lançamento.

Exemplo 103 (**Esperança de uma variável aleatória** Binomial(n,p)). Sejam $Y_i, Y_2, \dots, Y_n$ variáveis aleatórias com distribuição de Bernoulli com parâmetro p. Então $X = Y_i + Y_2 + \cdots + Y_n$ conta o número de sucesso em n ensaios de Bernoulli. Para todo i temos $\mathbb{E}\left[Y_i\right] = p$, portanto $\mathbb{E}\left[X\right] = np$.

A proposição 2.2.2, página 146, sugere que os maiores valores de $\mathrm{bi}_{n,p}$ (os valores mais prováveis de uma variável aleatória binomial) estão em torno do valor médio (reveja os gráficos do exemplo 69, página 143).

Sejam X e Y variáveis aleatórias, com esperança finita e definidas sobre o mesmo espaço amostral. O produto delas é $X\cdot Y(\omega) := X(\omega)\cdot Y(\omega)$. Em geral *não vale* $\mathbb{E}\left[X\cdot Y\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$.

Exemplo 104 ($\mathbb{E}\left[X\cdot X\right] \neq \mathbb{E}\left[X\right]\mathbb{E}\left[X\right]$). Seja X o resultado do lançamento de um dado, como no exemplo 88.

$$\mathbb{E}\left[X^2\right] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}.$$

Notemos que, com o resultado do exemplo 88 podemos concluir que $\mathbb{E}\left[X\cdot X\right] \neq \mathbb{E}\left[X\right]\mathbb{E}\left[X\right]$.

3.0.15 Proposição. *Se X, Y são variáveis aleatórias com esperança finita e $a,b\in R$, então* $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2(\mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right])$.

Demonstração. Sejam X, Y, a, b como no enunciado e vamos usar a notação $\mu_X = \mathbb{E}[X]$ e $\mu_Y = \mathbb{E}[Y]$. Abaixo usamos várias vezes a linearidade da esperança.

Usando a definição $\mathrm{Var}(X+Y) = \mathbb{E}\left[(X+Y-\mathbb{E}[X+Y])^2\right]$

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}\left[(X+Y-\mathbb{E}[X+Y])^2\right] \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])+(Y-E(Y))\right]^2 \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2+(Y-\mathbb{E}[Y])^2+2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right]+\mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right]+\mathbb{E}\left[2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\
&= \mathrm{Var}[X]+\mathrm{Var}(Y)+\mathbb{E}\left[2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\
&= \mathrm{Var}[X]+\mathrm{Var}(Y)+2(\mathbb{E}[XY]-\mathbb{E}[X]\,\mathbb{E}[Y])
\end{aligned}$$

pois $\mathbb{E}[2((X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = 2\mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y]]$ e como $\mathbb{E}[X\mathbb{E}[Y]] = \mathbb{E}[Y\mathbb{E}[X]] = \mathbb{E}[X]\,\mathbb{E}[Y]$ os termos do meio se cancelam. □

O termo $\mathbb{E}[XY]-\mathbb{E}[X]\,\mathbb{E}[Y]$ no item 2 do resultado acima é chamado de **covariância** de X e Y e denotado $\mathrm{cov}(X, Y)$.

Exercício 94. Verifique que se $\mathbb{E}[X\cdot Y] = \mathbb{E}[X]\,\mathbb{E}[Y]$ então $\mathrm{Var}(X+Y) = \mathrm{Var}[X]+\mathrm{Var}(Y)$.

Variáveis aleatórias independentes. Dizemos que as variável aleatória X e Y são independentes se para quaisquer $a, b \in \mathbb{R}$

$$F_{X,Y}(a,b) = F_X(a)F_Y(b)$$

Equivalentemente, no caso específico de variáveis aleatórias ambas *discretas* ou ambas *contínuas* X e Y, elas são independentes se, e só se,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

Uma coleção de variáveis é dita independente se para toda subcoleção finita $X_1, \dots, X_k$ vale

$$F(x_1, x_2, \dots, x_k) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_k}(x_k) \text{ para quaisquer } x_1, x_2, \dots, x_k \in \mathbb{R}.$$

3.0.16 Proposição. *Se X, Y são conjuntamente contínuas com $f_{X,Y}(x,y) = g(x)h(y)$ de modo que $g(x), h(x) \geqslant 0$ para todo $x \in \mathbb{R}$ e $\int_{-\infty}^{\infty} g(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} h(x)\,\mathrm{d}x = 1$ então X e Y são independentes e com as respectivas densidades g e h.*

Demonstração. Exercício. □

Exemplo 105. Se

$$f(x,y) = \begin{cases} 24xy, & \text{se } x \in [0,1],\ y \in [0,1],\ x+y \leqslant 1 \\ 0, & \text{caso contrário.} \end{cases}$$

é a distribuição conjunta de X e Y então $f_X(3/4) = f_Y(3/4) = 9/16$ e $f(3/4, 3/4) = 0$ portanto $f_X(x)f_Y(y) \neq f(x,y)$ ou seja, embora $f(x,y) = g(x)h(y)$ as funções g e h não podem ser feitas distribuições marginais.

Exemplo 106 (**O problema da agulha de Buffon revisitado**). De volta ao exemplo 1.3, temos um piso com linhas paralelas que distam $2t$ cm onde deixamos cair uma agulha de comprimento 2ℓ cm, $\ell < t$. Qual é a probabilidade de que a agulha irá cruzar uma linha?

Sejam R e Θ variáveis aleatórias independentes e uniformemente distribuídas em $[0,t)$ e $[0,\pi)$, portanto $f_R(r) = 1/r$ e $f_\Theta(\theta) = 1/\theta$ nesses intervalos e $f_R(r) = f_\Theta(\theta) = 0$ fora deles, respectivamente. A densidade conjunta é

$$f(r,\theta) = \begin{cases} \dfrac{1}{t\pi} & \text{em } A = [0,t) \times [0,\pi) \\ 0 & \text{caso contrário.} \end{cases}$$

Se $r < \ell \sin(\theta)$ então a agulha cruzou uma das divisórias e a probabilidade desse evento é

$$\begin{aligned} \mathbb{P}\left(R < \ell \sin(\Theta)\right) &= \iint_{\{(r,\theta)\in A : r < \frac{\ell}{2}\sin(\theta)\}} \frac{1}{\pi t} \,\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^\pi \int_0^{\ell \sin(\theta)} \frac{1}{\pi t} \,\mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{1}{\pi t} \int_0^\pi \int_0^{\ell \sin(\theta)} \mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{1}{\pi t} \int_0^\pi \ell \sin(\theta)\,\mathrm{d}\theta \\ &= \frac{2\ell}{\pi t}. \end{aligned}$$

Notemos que se X e Y são independentes então $\mathbb{E}\left[g(X)h(Y)\right] = \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$, pra quaisquer funções reais g, h. De fato

$$\begin{aligned} \mathbb{E}\left[(g(X)h(Y)\right] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y) f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} g(x) f_X(x)\,\mathrm{d}x \int_{-\infty}^{+\infty} h(y) f_Y(y)\,\mathrm{d}y \\ &= \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]. \end{aligned}$$

Dedução análoga vale para distribuição conjunta discreta. Em particular, se X e Y são independentes então

$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]. \tag{3.0.8}$$

Exercício 95. Ache exemplos de variável aleatória que satisfazem equação (3.0.8) mas que não são independentes.

Soma de variáveis aleatórias independentes:: Sejam X e Y variáveis aleatórias discretas, independentes e com função de massa f_X e f_Y, função de distribuição F_X e F_Y, respectivamente. Se $Z := X + Y$, então

$$\begin{aligned}\mathbb{P}(Z \leqslant z) &= \mathbb{P}(X+Y \leqslant z) = \mathbb{P}(X \leqslant z - Y)\\ &= \sum_y \mathbb{P}(X \leqslant z-y \mid Y = y)\mathbb{P}(Y=y)\end{aligned}$$

como as variável aleatória são independentes $\mathbb{P}(X \leqslant z-y \mid Y=y) = \mathbb{P}(X \leqslant z-y)$, logo

$$F_{X+Y}(z) = \sum_y F_X(z-y)F_Y(y).$$

Exercício 96. Prove que, com as mesmas hipóteses,

$$f_{X+Y}(z) = \sum_y f_X(z-y)f_Y(y).$$

Agora, se X e Y são variável aleatória conjuntamente contínuas e independentes então

$$F_{X+Y}(z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{+\infty} f_X(x)\left(\int_{-\infty}^{z-x} f_Y(y)\,\mathrm{d}y\right)\mathrm{d}x$$

portanto

$$F_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)F_Y(z-x)\,\mathrm{d}x$$

ademais, $f_{X+Y}(z) = (\mathrm{d}/\mathrm{d}z)F_{X+Y}(z)$

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)\frac{\mathrm{d}}{\mathrm{d}z}F_Y(z-x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\,\mathrm{d}x.$$

Soma de variáveis aleatórias uniforme

$X, Y \sim$ Uniforme$([0,1])$

$$f_{X+Y}(z) = \int_0^1 1 \cdot f_Y(z-x)\,\mathrm{d}x$$

$0 \leqslant z \leqslant 1$, $f_{X+Y}(z) = \int_0^z dx = z$
$1 < z \leqslant 2$, $f_{X+Y}(z) = \int_{z-1}^1 dx = 2 - z$

Soma de variáveis aleatórias Poisson

$X \sim$ Poisson(λ_X)
$Y \sim$ Poisson(λ_Y)

$$f_{X+Y}(z) = \sum_x f_X(x)f_Y(z-x) = \frac{\mathrm{e}^{-(\lambda_X+\lambda_Y)}}{z!}(\lambda_X+\lambda_Y)^z$$

Soma de variáveis aleatórias Binomial

$X \sim \text{Binomial}(n, p)$
$Y \sim \text{Binomial}(m, p)$
$X + Y \sim \text{Binomial}(n + m, p)$

Soma de variáveis aleatórias normais

Vimos que se $X \sim \mathcal{N}(\mu; \sigma^2)$ então $aX + b \sim \mathcal{N}(a\mu + b; a^2\sigma^2)$, para quaisquer $a, b \in \mathbb{R}$. De fato, vale uma afirmação mais geral:

3.0.17 Teorema. *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, então*

$$\sum_{i=1}^{n} c_i X_i n \sim \mathcal{N}\left(\sum_{i=1}^{n} c_i \mu_i; \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

para quaisquer constantes $c_1, c_2, \ldots, c_n \in \mathbb{R}$.

3.0.18 Corolário. *Se $M_n = \frac{\sum_i X_i}{n}$ então*

$$M_n \sim \mathcal{N}\left(\frac{\sum_i \mu_i}{n}; \frac{\sum_i {\sigma_i}^2}{n^2}\right).$$

Caso as variáveis tenham a mesma distribuição

$$M_n \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right).$$

Exemplo 107. Na fabricação de placas retangulares há pequenas perturbações de modo que o comprimento e a largura em centímetros de uma placa escolhida ao acaso são variáveis aleatórias $C \sim \mathcal{N}(2; 0{,}1^2)$ e $L \sim \mathcal{N}(5; 0{,}2^2)$, respectivamente. Qual a probabilidade do perímetro exceder 15 cm?

Se Y é a variável aleatória pra o perímetro de uma placa escolhida ao acaso, então $Y = 2C + 2L$ e pelo teorema acima $Y \sim \mathcal{N}(14; 0{,}2)$ logo

$$\mathbb{P}(Y > 15) = \mathbb{P}\left(\frac{Y - 14}{\sqrt{0,2}} > \frac{15 - 14}{\sqrt{0{,}2}}\right) = \mathbb{P}(Z > 2{,}236) = 0{,}0129$$

Exemplo 108. O engarrafamento de um refrigerante de 300ml tem variações de modo que o volume do líquido numa garrafa é uma variável aleatória com distribuição $\mathcal{N}(300; 25^2)$. Numa inspeção, 10 garrafas são selecionadas e o volume de cada garrafa, V_1, V_2, $\ldots$, V_{10} é medido, de modo que se a **média amostral**

$$M_n = \frac{V_1 + V_2 + \cdots + V_{10}}{10}$$

for menor que 290 (ml) então a engarrafadora é multada. Qual é a probabilidade de multa?

$M_n \sim \mathcal{N}(300; \frac{25^2}{10})$ de modo que

$$\mathbb{P}(M_n < 290) = \mathbb{P}(Z < -1{,}26) = 0,1038$$

Definimos $XY(\omega) := X(\omega)Y(\omega)$. Se $\mathbf{1}_{[X=a]\cap[Y=b]}(\omega)$ é a variável aleatória indicadora de $\omega \in [X = a] \cap [Y = b]$, então, variando a, b nos reais tais que $f_X(a), f_Y(b) > 0$, temos

$$XY(\omega) = \sum_a \sum_b a \cdot b \cdot \mathbf{1}_{[X=a]\cap[Y=b]}(\omega) \qquad (\forall \omega \in \Omega)$$

portanto

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_a \sum_b ab\mathbb{E}\left[\mathbf{1}_{[X=a]\cap[Y=b]}\right] \\
&= \sum_a \sum_b ab\mathbb{P}([X=a]\cap[Y=b])
\end{aligned}$$

agora, se assumimos X e Y independentes, então

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_a \sum_b ab\mathbb{P}([X=a]\cap[Y=b]) \\
&= \sum_a \sum_b ab\mathbb{P}(X=a)\mathbb{P}(Y=b) \\
&= \sum_a aP(X=a) \sum_b b\mathbb{P}(Y=b) \\
&= \sum_a af_X(a) \sum_b bf_X(b) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y].
\end{aligned}$$

ou seja, provamos que

3.0.19 Proposição. *Se X e Y são variáveis aleatórias independentes então* $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

Como corolário segue da proposição 3.0.7 que

3.0.20 Corolário. *Se X e Y são variáveis aleatórias independentes então* $\mathrm{Var}(X+Y) = \mathrm{Var}[X] + \mathrm{Var}(Y)$.

3.0.21 Observação. *Esses conceitos desenvolvidos até aqui podem ser estendidos, de modo natural, ao caso de 3 ou mais variável aleatória.*

Exemplo 109. Sejam X, Y, Z variável aleatória independentes e uniformemente distribuídas sobre $[0,1]$. Então

$$\begin{aligned}
\mathbb{P}(X \geqslant YZ) &= \iiint\limits_{\{(x,y,z):x\geqslant yz\}} f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\
&= \int_0^1 \int_0^1 \int_{yz}^1 \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\
&= \int_0^1 \int_0^1 (1-yz)\,\mathrm{d}y\,\mathrm{d}z \\
&= \int_0^1 \left(1 - \frac{z}{2}\right) \mathrm{d}z \\
&= \frac{3}{4}.
\end{aligned}$$

§3 Aproximações, Desigualdades e teoremas limite

Momento de uma váriavel aleatória:: No que segue as variáveis têm esperança (momentos) finita.

Desigualdades de Markov e Bienaymé–Chebyshev: Seja λ um real positivo e Z uma variável aleatória que assume valores não negativos. Definamos a variável aleatória Y por

$$Y(\omega) := \begin{cases} \lambda, & \text{se } Z(\omega) \geqslant \lambda \\ 0, & \text{nos outros casos.} \end{cases}$$

Da definição temos $Y \leqslant Z$ donde, por equação (3.0.4), deduzimos que $\mathbb{E}[Y] \leqslant \mathbb{E}[Z]$. Também, o valor esperado de Y é

$$\mathbb{E}[Y] = \lambda\,\mathbb{P}(Y = \lambda) + 0\,\mathbb{P}(Y \neq \lambda) = \lambda\,\mathbb{P}(Z \geqslant \lambda)$$

logo $\mathbb{E}[Z] \leqslant \lambda \mathbb{P}(Z \geqslant \lambda)$. Disso concluímos

3.0.22 Teorema (**Desigualdade de Markov**). *Se λ é real positivo e Z uma variável aleatória que assume valores não negativos, então*

$$\mathbb{P}(Z \geqslant \lambda) \leqslant \frac{\mathbb{E}[Z]}{\lambda}.$$

Se X é uma variável aleatória, podemos tomar $Z = (X - \mathbb{E}[X])^2$ que é uma variável aleatória não negativa. Lembremos que $\mathbb{E}[Z] = \mathrm{Var}[X]$. Da desigualdade de Markov com a constante λ^2 no lugar de λ deduzimos

$$\mathbb{P}\left((X - \mathbb{E}[X])^2 \geqslant \lambda^2\right) \leqslant \frac{\mathrm{Var}[X]}{\lambda^2}.$$

Disso concluímos

3.0.23 Teorema (**Desigualdade de Bienaymé–Chebyshev**). *Se λ é real positivo e X uma variável aleatória então*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant \lambda) \leqslant \frac{\mathrm{Var}[X]}{\lambda^2}. \tag{3.0.9}$$

Se fizermos $\lambda = k\sigma$ na equação (3.0.9) (σ é o desvio padrão de X) obtemos a probabilidade de X desviar de $\mathbb{E}[X]$ por pelo menos k desvios padrão

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant k\sigma) \leqslant \frac{1}{k^2}. \tag{3.0.10}$$

Exemplo 110. Numa doença rara, um paciente se recupera com probabilidade 0,4. Numa escolha aleatória de 15 pacientes doentes, seja X o número de sobreviventes. Então, de 3 a 9 pacientes sobrevivem com probabilidade

$$\begin{aligned}\mathbb{P}(3 \leqslant X \leqslant 9) &= \sum_{x=3}^{9} \binom{15}{x} (0,4)^x (0,6)^{15-x} \\ &= 0{,}94\end{aligned}$$

Agora, usando a desigualdade de Bienaymé–Chebyshev, equação equação (3.0.10), para determinar $X \in (\mu - 2\sigma, \mu + 2\sigma)$, em que $\mu = \mathbb{E}[X]$, temos

$$\mathbb{P}(|X - \mu| \geqslant 2\sigma) \leqslant \frac{1}{4}$$

portanto $X \in (\mu - 2\sigma, \mu + 2\sigma)$ com probabilidade $3/4 = 0{,}75$. A variável aleatória X tem média e variância dadas, respectivamente, por $\mu = 6$ e $\sigma^2 = 3{,}6$, portanto

$$\mathbb{P}(2{,}20 \leqslant X \leqslant 9{,}89) = 0{,}75$$

e como X é inteiro $[3 \leqslant X \leqslant 9] = [2{,}20 \leqslant X \leqslant 9{,}89]$.

A desigualdade de Bienaymé–Chebyshev vale para qualquer distribuição e, por isso, a estimativa não é muito precisa. Compare os dois valores obtidos no exemplo acima. No caso de $X \sim \mathcal{N}(\mu; \sigma^2)$, por exemplo, sabemos que $\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \approx 0{,}95$ enquanto que equação (3.0.10) garante apenas $\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = 0{,}75$.

Exemplo 111. Um produto é distribuído em lotes de 40 unidades. Um lote é inaceitável se três ou mais itens apresentam defeito. O departamento de controle de qualidade de um comprador adotou o plano de selecionar aleatoriamente 5 unidades de cada lote comprado e rejeitar o lote se um item inspecionado for defeituoso. Num lote com 3 itens defeituoso, a probabilidade de haver um defeituoso numa amostra de 5 é

$$\frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0{,}3017$$

ou seja, a estratégia detecta um lote ruim em apenas 30% dos casos. Como já vimos, se X é o número de itens defeituosos na amostra, então $X \in (\mu - 2\sigma, \mu + 2\sigma)$ com probabilidade pelo menos $3/4$, pela desigualdade de Bienaymé–Chebyshev. Nesse caso $\mu = 0{,}375$ e $\sigma = 0{,}558$ e temos então que 1 item em 5 é defeituoso com probabilidade $\geqslant 3/4$, ou seja, em 75% dos casos.

Exemplo 112. Consideremos n lançamentos de uma moeda com os resultados independentes. Seja X_i a variável indicadora (definida na pág. 131) do evento "ocorre cara no i-ésimo lançamento". A variável aleatória

$$S_n = X_1 + \cdots + X_n$$

é a quantidade de ocorrência de cara e $\frac{S_n}{n}$ é a fração de caras. A esperança de S_n é, usando o exercício 93

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \left(\sum_{i=1}^{n} \mathbb{E}[X_i]\right) = \frac{n}{2}.$$

e a variância é $\mathrm{Var}(S_n) = \mathbb{E}\left[S_n^2\right] - \mathbb{E}\left[S_n\right]^2$, assim precisamos calcular $\mathbb{E}\left[S_n^2\right]$.

$$\mathbb{E}\left[S_n^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[X_i X_j\right],$$

se $i \neq j$ então

$$\mathbb{E}\left[X_i \cdot X_j\right] = \mathbb{P}(X_i \cdot X_j = 1) = \mathbb{P}([X_i = 1] \cap [X_j = 1]) = \frac{1}{2} \cdot \frac{1}{2}$$

e se $i = j$ então

$$\mathbb{E}\left[X_i \cdot X_j\right] = E(X_i^2) = P(X_i = 1) = \frac{1}{2}$$

portanto

$$\mathbb{E}\left[S_n^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[X_i X_j\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\left[X_i X_j\right] = \frac{n}{2} + \frac{n(n-1)}{4}.$$

De volta à variância de S_n temos

$$\mathrm{Var}\left(S_n\right) = \mathbb{E}\left[S_n^2\right] - \mathbb{E}\left[S_n\right]^2 = \frac{n}{2} + \frac{n(n-1)}{4} - \left(\frac{n}{2}\right)^2 = \frac{n}{4}.$$

A esperança de S_n/n é, usando linearidade da esperança (corolário 3.0.4)

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n}\mathbb{E}\left[S_n\right] = \frac{1}{2}.$$

e a variância é (proposição 3.0.7),

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\mathrm{Var}[S_n] = \frac{1}{n^2}\frac{n}{4} = \frac{1}{4n}.$$

Usando a desigualdade de Bienaymé–Chebyshev, eq. equação (3.0.9), para qualquer $\lambda > 0$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geqslant \lambda\right) \leqslant \frac{1}{4n\lambda^2}$$

portanto

$$\mathbb{P}\left(\frac{1}{2} - \lambda \leqslant \frac{S_n}{n} \leqslant \frac{1}{2} + \lambda\right) > 1 - \frac{1}{4n\lambda^2}.$$

Por exemplo, fazendo $\lambda = n^{-1/2}$ temos que o número médio de caras está no intervalo $(\frac{1}{2} - \frac{1}{\sqrt{n}}, \frac{1}{2} + \frac{1}{\sqrt{n}})$ com probabilidade maior que $3/4$.

Mais que isso, a probabilidade tende a 1 quando $n \to \infty$. Esse resultado foi provado pela primeira vez em 1713, por Jacob Bernoulli sem usar a desigualdade de Bienaymé–Chebyshev, desconhecida na época.

Exemplo 113. Numa eleição, seja p a fração (desconhecida) da população que vota no candidato D. Para simplificar, assumimos que um voto em voto em D é ensaio de Bernoulli com parâmetro p.

Suponha que são realizadas n entrevistas como o objetivo de estimar o valor de p. Se $V_i \sim \text{Bernoulli}(p)$ é a variável aleatória indicadora do i-ésimo voto ser para D, para $1 \leqslant i \leqslant n$, então

$$S_n = \sum_{i=1}^{n} V_i \sim \text{Binomial}(n, p)$$

é o total de entrevistados a favor de D.

Queremos determinar n para obtermos uma estimativa para p com erro de 4 pontos percentuais com pelo menos 95% de certeza, i.e.,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant 0{,}04\right) < 0{,}05. \tag{3.0.11}$$

Usando a desigualdade de Chebyshev equação (3.0.9),

$$\mathbb{P}\left(|S_n - np| \geqslant 0{,}04n\right) \leqslant \frac{np(1-p)}{0{,}0016n^2} \leqslant \frac{1}{4 \cdot 0{,}0016 \cdot n}$$

e fazendo o lado direito igual a $0{,}05$ e resolvendo para n obtemos $n = 3125$ entrevistas para que S_n/n estime p dentro dos parâmetros de exigência.

Esse valor de n está superestimado, muito por causa da generalidade da desigualdade de Bienaymé–Chebyshev, ela não usa nenhuma particularidade da distribuição binomial que é bem concentrada em torno da média. A equação equação (3.0.11) equivale a

$$\mathbb{P}\left(S_n \leqslant (p-0{,}04)n\right) + \mathbb{P}\left(S_n \geqslant (p+0{,}04)n\right) < 0{,}05$$

e a soma a esquerda da desigualdade é $F_{S_n}((p-0{,}04)n) + F_{S_n}(((1-p)-0{,}04)n)$ que é máxima para $p = 0{,}5$ para todo n fixo, portanto

$$\begin{aligned} \mathbb{P}\left(\frac{S_n}{n} \leqslant p - 0{,}04\right) + \mathbb{P}\left(\frac{S_n}{n} \geqslant p + 0{,}04\right) &\leqslant F_{S_n}((0{,}5-0{,}04)n) + F_{S_n}((0{,}5-0{,}04)n) \\ &= 2F_{S_n}(0{,}46n) \end{aligned}$$

que é menor que $0,05$ para $n \geqslant 624$. Essa estimativa é bem próxima a de dados reais. Informações obtidas no sítio do IBOPE referentes ao 1º turno das eleições municipais de 2008 exibem a seguinte quantidade de entrevistados de acordo como o erro e o grau de confiança:

cidade	no. de eleitores	no. de entrevistas	proporção eleitores entrevistados	margem de erro	grau de confiança
São Paulo	8.198.282	1.204	0,01468%	3%	95%
Rio de Janeiro	4.579.282	1.204	0,02629%	3%	95%
Belo Horizonte	1.772.227	1.204	0,06793%	3%	95%
Santo André	533.428	504	0,09448%	4%	95%
Diadema	301.229	504	0,1673%	4%	95%
São Carlos	154.572	504	0,326%	4%	95%
Cubatão	91.693	504	0,5496%	4%	95%
Registro	41.001	504	1,2292%	4%	95%
Campinas	724.143	2.000	0,2761%	3%	99%
São José dos Campos	414.353	2.000	0,4826%	3%	99%
Ribeirão Preto	388.690	2.000	0,5145%	3%	99%

Adiante, quando estudarmos o Teorema Central do Limite, retomaremos esse exemplo e mostraremos uma regra bastante prática para estimar o tamanho n da amostra de eleitores.

Leis dos Grandes Números:

Convergência em probabilidade:: Quando X e uma sequência X_n, $n \geqslant 1$ são variáveis aleatórias de um mesmo modelo probabilístico, dizemos que X_n converge para X em probabilidade, denotado por

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} X$$

se para todo $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1.$$

3.0.24 Teorema (**Lei Fraca dos Grandes Números**). *Sejam $X_1, X_2, \dots, X_n$ variável aleatória's independentes e identicamente distribuídas, cada uma com média $\mathbb{E}[X_i] = \mu$ finita e variância $\mathrm{Var}[X] = \sigma^2$ finita. Então*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\mathrm{P}} \mu$$

Demonstração. Sejam $X_1, X_2, \dots$ uma sequência de variáveis aleatórias independentes, identicamente distribuídas e todas com valor esperado μ e variância σ^2 finitos. Pondo

$$M_n := \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

temos

$$\mathbb{E}[M_n] = \mu \ \text{ e } \ \mathrm{Var}[M_n] = \frac{\sigma^2}{n}.$$

Usando a desigualdade de Chebyshev equação (3.0.9), concluímos

$$\lim_{n\to\infty} \mathbb{P}(|M_n - \mu| \geqslant \lambda) \leqslant \lim_{n\to\infty} \frac{\sigma^2}{n\lambda^2} = 0.$$

para todo $\lambda > 0$. $\square$

Exemplo 114. Sejam $g : [0,1] \to [0,1]$ uma função contínua e $X_i \sim \text{Uniforme}([0,1])$ variáveis aleatórias independentes, $i \geqslant 1$. As variáveis aleatórias $g(X_i)$, para todo $i \geqslant 1$, são independentes de mesma distribuição, portanto, pela lei dos grandes números

$$\frac{g(X_1)+g(X_2)+\cdots+g(X_n)}{n} \xrightarrow[n\to\infty]{\mathrm{P}} \mu$$

em que $\mu = \mathbb{E}\,[g(X_i)]$ para todo i. Se $X \sim \text{Uniforme}([0,1])$ então

$$\mu = \mathbb{E}\,[g(X)] = \int_0^1 g(x) f_X(x)\,\mathrm{d}x = \int_0^1 g(x)\mathrm{d}x$$

e a variância é

$$\sigma^2 = \mathbb{E}\,[X-\mu]^2 = \int_0^1 (g(x)-\mu)^2\,\mathrm{d}x \leqslant 1$$

pois $|g(x)-\mu| \leqslant 1$, portanto,

$$\int_0^1 g(x)\,\mathrm{d}x \approx \frac{g(X_1)+g(X_2)+\cdots+g(X_n)}{n} = M_n$$

e a Bienaymé–Chebyshev nos diz quão boa é essa estimativa

$$\mathbb{P}\,(|M_n-\mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon^2} < \frac{1}{n\varepsilon^2}$$

para qualquer erro $\varepsilon > 0$. Por exemplo, para $\varepsilon = 0{,}001$, se $n = 100.000.000$ então

$$|M_n-\mu| = \left| M_n - \int_0^1 g(x)\mathrm{d}x \right| < 0{,}001$$

com probabilidade $> 0{,}99$.

Exercício 97. Uma aposta de 1 real tem ganho esperado de $-0{,}0141$. O que a Lei dos Grandes Números nos diz sobre os seus ganhos se você fizer um grande número apostas de 1 real? Será que ela lhe assegura que suas perdas serão pequenas? Será que ela lhe assegura que, se n for muito grande você vai perder?

Exercício 98. Pedro e Paula ambos querem cortar um pedaço de papel retangular. Como ambos são probabilistas eles determinam a forma exata do retângulo utilizando realizações de uma variável aleatória positiva, digamos U, como se segue. Pedro é preguiçoso e gera apenas uma única realização dessa variável aleatória; então ele corta um quadrado que tem comprimento e largura igual a esse valor. Paula gosta de diversidade e gera duas realizações independentes de U. Ela, então, corta um retângulo com largura igual a primeira realização e comprimento igual ao da segunda realização. (a) Serão as áreas cortadas por Pedro e Paula diferentes em média? (b) se forem, Pedro ou Paula deverá ter um retângulo com área maior?

Convergência quase-certa:: Quando X e uma sequência x_n, $n \geqslant 1$ são variáveis aleatórias de um mesmo modelo probabilístico, dizemos que X_n converge para X quase certamente, denotado por

$$X_n \xrightarrow[n\to\infty]{\text{q.c.}} X$$

se existe $A \subset \Omega$ tal que $\mathbb{P}(A) = 1$ e

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \text{ para todo } \omega \in A$$

ou,

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

É possível provar que

convergência quase certa $\Rightarrow$ convergência em probabilidade

e, também, exibir contra-exemplo para a recíproca.

Lei Forte dos Grandes Números:: estabelece que, sob as mesmas hipóteses da lei fraca,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{q.c.}} \mu$$

A Lei dos Grandes Números mostra que o modelo probabilístico, e os axiomas de probabilidade em particular, é consistente com a interpretação frequentista de probabilidade.

Teorema Central do Limite. Consideremos uma sequência $X_1, X_2, \dots$ de variáveis aleatórias *independentes* com a *mesma distribuição*, de esperança μ finita e variância $\sigma^2 > 0$ finita. A soma das n primeiras variáveis aleatórias

$$S_n := X_1 + \cdots + X_n$$

tem esperança $\mathbb{E}[S_n] = n\mu$ e variância $\text{Var}[S_n] = n\sigma^2$ e a média

$$M_n := \frac{X_1 + \cdots + X_n}{n}$$

tem esperança $\mathbb{E}[M_n] = \mu$ e variância $\text{Var}[M_n] = \sigma^2/n$.

Vimos, pela lei de grandes números, que $M_n \approx \mu$. O Teorema Central do Limite diz que, *para n grande*, a variável aleatória normalizada de S_n e M_n

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

têm distribuição aproximadamente $\mathcal{N}(0;1)$. Portanto, $M_n = \mu + (\sigma/\sqrt{n})Z_n$, ou seja, o que veremos agora é que, de fato, $M_n \approx \mu + (\sigma/\sqrt{n})Z$ com $Z \sim \mathcal{N}(0;1)$.

Por ora, ignoremos o real significado de *para n grande* e vejamos alguns exemplos.

Exemplo 115. Lâmpadas produzidas numa fábrica têm vida útil em horas regida pela distribuição normal $\mathcal{N}(800; 40^2)$. Uma seleção aleatória simples de tamanho 16 tem vida útil média M_{16} com distribuição aproximada $\mathcal{N}(800; 40^2/16)$. A probabilidade da vida útil média ser menor que 775 horas é aproximadamente

$$\mathbb{P}(M_{16} < 775) = \mathbb{P}\left(\frac{M_{16} - 800}{40/4} < \frac{775 - 800}{40/4}\right) = \mathbb{P}(Z_{16} < -2,5) \approx 0{,}0062.$$

De fato, a aproximação aqui é uma igualdade pois a combinação linear de variáveis com distribuição normal tem distribuição normal.

Exemplo 116. As chamadas telefônicas numa empresa têm duração em minutos que segue a distribuição exponencial com parâmetro 1/3, portanto tem média 3 e variância 9. Uma amostra aleatória simples com 50 chamadas tem probabilidade da média amostral não ultrapassar 4 min dada por

$$\mathbb{P}(M_{50} \leqslant 4) = \mathbb{P}(Z_{50} \leqslant 2,36) \approx 0{,}991.$$

Convergência em distribuição:: Quando X e uma sequência X_n, $n \geqslant 1$, são variáveis aleatórias do mesmo espaço de probabilidade, dizemos que X_n converge para X em distribuição, denotado por

$$X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$$

se $F_{X_n} \to F_X$ quando $n \to \infty$, isto é

$$\lim_{n\to\infty} \mathbb{P}(X_n \leqslant x) = \mathbb{P}(X \leqslant x)$$

de modo que podemos usar X como modelo probabilístico aproximado para X_n e quanto maior n melhor é a aproximação.

Formalmente, o teorema central do limite é enunciado como

3.0.25 Teorema (**Teorema Central do Limite (TCL)**). *Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias* mutuamente independentes *com a* mesma distribuição, *de esperança μ e variância $\sigma^2 > 0$ finitas. Então*

$$\frac{M_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{D}} Z$$

para $Z \sim \mathcal{N}(0; 1)$.

Por exemplo, se uma moeda equilibrada é lançada n vezes e S_n é a quantidade de caras. Para $n = 100$ temos $\mathbb{E}[S_{100}] = 50$ e $\mathrm{Var}[S_{100}] = 25$. A probabilidade de termos mais que 55 caras é, aproximadamente,

$$\mathbb{P}(S_{100} > 55) = \mathbb{P}\left(\frac{S_{100} - 50}{5} > \frac{55 - 50}{5}\right) = \mathbb{P}(Z_{100}) > 1) \approx 0{,}16.$$

Agora, para $n = 400$, qual a probabilidade com que $S_{400} > 220$ (note-se que $220/400 = 55/10$)?

$$\mathbb{P}(S_{400} > 220) = \mathbb{P}\left(\frac{S_{400} - 200}{10} > \frac{220 - 200}{10}\right) = \mathbb{P}(Z_{400}) > 2) \approx 0{,}025.$$

Com que probabilidade $40 \leqslant S_{100} \leqslant 60$?

$$\mathbb{P}(40 \leqslant S_{100} \leqslant 60) = \mathbb{P}\left(\frac{40 - 50}{5} \leqslant \frac{S_{100} - 50}{5} \leqslant \frac{60 - 50}{5}\right) = \mathbb{P}(-2 \leqslant Z_{100}) \leqslant 2) \approx 0{,}954.$$

Aproximação para a Binomial: No caso que $X_1, \dots, X_n$ são variáveis aleatórias independentes com distribuição Bernoulli(p) temos $S_n = X_1 + \cdots + X_n \sim$ Binomial(n, p). Pondo $q = 1 - p$ e $Z_n = (M_n - \mu)/(\sigma/\sqrt{n})$ o teorema central do limite nos dá

$$\mathbb{P}(k \leqslant S_n \leqslant l) \approx \mathbb{P}\left(\frac{k - np}{\sqrt{npq}} \leqslant Z_n \leqslant \frac{l - np}{\sqrt{npq}}\right) \tag{3.0.12}$$

Por exemplo, se $X \sim$ Binomial(225; 0,2) então pela aproximação dada na equação (3.0.12)

$$\mathbb{P}(39 \leqslant X \leqslant 48) \approx \mathbb{P}\left(-1 \leqslant Z \leqslant 0{,}5\right) = 0{,}5328072 \tag{3.0.13}$$

ainda, o valor exato é

$$\mathbb{P}(39 \leqslant X \leqslant 48) = \sum_{j=39}^{48} \binom{225}{j} (0,2)^j (0{,}8)^{225-j} = 0{,}5852713.$$

Entretanto $0{,}0417 = \mathbb{P}(X = 39) \approx \mathbb{P}(-1 \leqslant Z \leqslant -1) = 0$ e aproximação seria melhor se fizéssemos

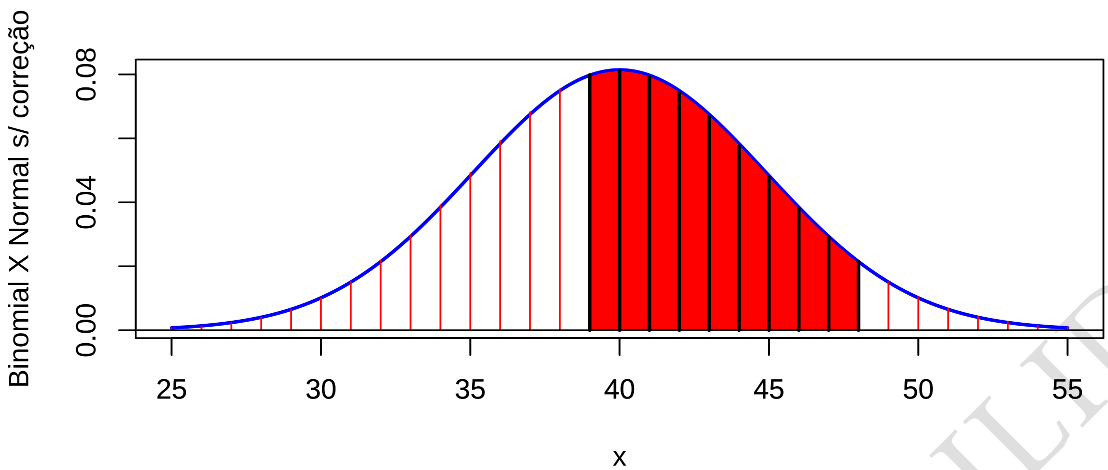
$$\mathbb{P}(X = 39) = \mathbb{P}(38,5 \leqslant X \leqslant 39,5) \approx \mathbb{P}\left(-1,083 \leqslant Z \leqslant -0,916\right) = 0{,}0403$$

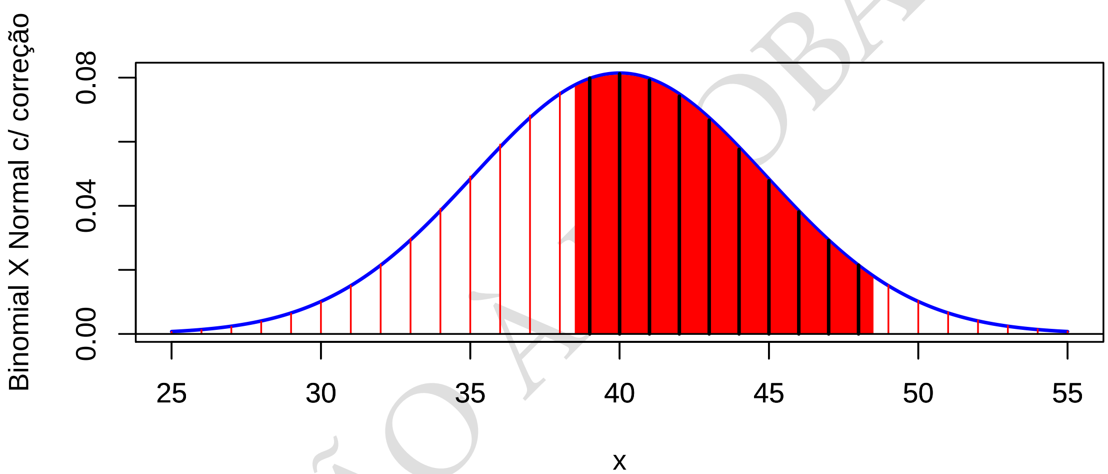
que chamamos de *correção de continuidade*, o que melhora a aproximação

$$\boxed{\mathbb{P}(k \leqslant X \leqslant l) = \mathbb{P}\left(k - \frac{1}{2} \leqslant X \leqslant l + \frac{1}{2}\right) \approx \mathbb{P}\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leqslant Z \leqslant \frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)}$$

agora, com a mesma técnica, temos uma aproximação melhor equação (3.0.13)

$$\mathbb{P}(39 \leqslant X \leqslant 48) \approx \mathbb{P}\left(-1,083 \leqslant Z \leqslant 0{,}583\right) = 0{,}5806491.$$





Exemplo 117. Um teste tem 200 perguntas com 4 alternativas cada, das quais apenas uma é correta. Qual a probabilidade aproximada que o estudante acerte por chute entre 25 e 30 questões para 80 das 200 questões. Seja $X \sim \text{Binomial}(80, 1/4)$ o número de respostas certas. Usando a correção por continuidade

$$\mathbb{P}(25 \leqslant X \leqslant 30) \approx \mathbb{P}(1,16 \leqslant Z \leqslant 2,71) = 0{,}1196602.$$

(o valor correto é 0,1192705)

```
 Aproximacao de P(X<x) da Normal para Binomial(n,p)=( 20 , 0.5 )

     Bin(x,n,p)   Aprox.      Erro        c/ correcao  Erro         Razao erros
0    0.00000095   0.00000387 -0.00000292  0.00001076 -0.00000981  3.35976699
1    0.00002003   0.00002850 -0.00000847  0.00007196 -0.00005194  6.13204207
```

2	0.00020123	0.00017331	0.00002792	0.00039812	-0.00019689	-7.05303706
3	0.00128841	0.00087256	0.00041585	0.00182522	-0.00053680	-1.29084332
4	0.00590897	0.00364518	0.00226379	0.00695315	-0.00104418	-0.46125469
5	0.02069473	0.01267366	0.00802107	0.02208567	-0.00139094	-0.17341068
6	0.05765915	0.03681914	0.02084001	0.05876243	-0.00110328	-0.05294070
7	0.13158798	0.08985625	0.04173173	0.13177624	-0.00018826	-0.00451111
8	0.25172234	0.18554668	0.06617565	0.25116748	0.00055486	0.00838463
9	0.41190147	0.32736042	0.08454105	0.41153164	0.00036984	0.00437465
10	0.58809853	0.50000000	0.08809853	0.58846836	-0.00036984	-0.00419799
11	0.74827766	0.67263958	0.07563809	0.74883252	-0.00055486	-0.00733570
12	0.86841202	0.81445332	0.05395870	0.86822376	0.00018826	0.00348890
13	0.94234085	0.91014375	0.03219710	0.94123757	0.00110328	0.03426659
14	0.97930527	0.96318086	0.01612440	0.97791433	0.00139094	0.08626303
15	0.99409103	0.98732634	0.00676469	0.99304685	0.00104418	0.15435768
16	0.99871159	0.99635482	0.00235677	0.99817478	0.00053680	0.22777120
17	0.99979877	0.99912744	0.00067133	0.99960188	0.00019689	0.29328140
18	0.99997997	0.99982669	0.00015328	0.99992804	0.00005194	0.33883687
19	0.99999905	0.99997150	0.00002754	0.99998924	0.00000981	0.35599322
20	1.00000000	0.99999613	0.00000387	0.99999867	0.00000133	0.34301686

Voltemos ao exemplo 113: Numa eleição com dois candidatos, seja p a fração (desconhecida) da população que vota no candidato D. Para simplificar, assumimos que só há 2 respostas possíveis e um voto em D é ensaio de Bernoulli com parâmetro p. São realizadas n entrevistas: modelamos o i-ésimo entrevistado como $V_i \sim \text{Bernoulli}(p)$ qu é a variável aleatória indicadora do i-ésimo voto ser para D, para todo $1 \leqslant i \leqslant n$. Temos então

$$\begin{aligned} S_n &= V_1 + \cdots + V_n \sim \text{Binomial}(n,p) \text{ é o número de votos em } D;\\ \hat{p} &= M_n = \frac{S_n}{n} \text{ é a proporção da amostra de votos em } D;\\ p &= \text{ é a proporção desconhecida da população de votos em } D \end{aligned}$$

Exemplo 118. Para uma primeira aproximação grosseira, usamos que $\hat{p} \approx p + (\sigma/\sqrt{n})Z \sim \mathcal{N}\left(p; \frac{\sigma^2}{n}\right)$, que $\sigma^2 = p(1-p) \leqslant 1/4$ e que uma variável aleatória normalmente distribuída tem probabilidade 0,95 de estar entre 2 desvios padrão da esperança, portanto

$$\mathbb{P}\left(p - 2\frac{\sigma}{\sqrt{n}} \leqslant \hat{p} \leqslant p + 2\frac{\sigma}{\sqrt{n}}\right) \geqslant \mathbb{P}\left(p - \frac{1}{\sqrt{n}} \leqslant \hat{p} \leqslant p + \frac{1}{\sqrt{n}}\right) \approx 0{,}95$$

de modo que uma estimativa com margem de erro de 4 pontos percentuais e com 95% de certeza temos $1/\sqrt{n} = 0{,}04$, logo $n = 625$. Para uma estimativa com margem de erro de 3 pontos percentuais e com 95% de certeza temos $1/\sqrt{n} = 0{,}03$ logo $n = 1112$. Para determinar o valor de n para uma estimativa com margem de erro de 3 pontos percentuais e com 99% de certeza precisamos de 3 desvios padrão, i.e.,

$$\mathbb{P}\left(p - 3\frac{\sigma}{\sqrt{n}} \leqslant \hat{p} \leqslant p + 3\frac{\sigma}{\sqrt{n}}\right) \geqslant \mathbb{P}\left(p - \frac{1{,}5}{\sqrt{n}} \leqslant \hat{p} \leqslant p + \frac{1{,}5}{\sqrt{n}}\right) \approx 0{,}99$$

portanto, temos $1{,}5/\sqrt{n} = 0{,}03$ logo $n = 1667$.

Vejamos agora esses cáculos com um pouco apurados. Sejam ε a margem de erro tolerada e γ o grau de confiança da estimativa

$$\begin{aligned}
\gamma = \mathbb{P}(-\varepsilon \leqslant p - \hat{p} \leqslant \varepsilon) &= \mathbb{P}\left(\frac{-\varepsilon}{\sqrt{p(1-p)/n}} \leqslant \frac{p-\hat{p}}{\sqrt{p(1-p)/n}} \leqslant \frac{\varepsilon}{\sqrt{p(1-p)/n}}\right) \\
&\approx \mathbb{P}\left(\frac{-\varepsilon}{\sqrt{p(1-p)/n}} \leqslant Z \leqslant \frac{\varepsilon}{\sqrt{p(1-p)/n}}\right) \\
&= \mathbb{P}\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \leqslant Z \leqslant \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= 2\mathbb{P}\left(Z \leqslant \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1
\end{aligned}$$

Queremos z_γ (obtido da tabela da distribuição da normal padrão) tal que $2\mathbb{P}\left(Z \leqslant z_\gamma\right) - 1 = \gamma$. Por exemplo, para $\gamma = 0{,}95$, de $\mathbb{P}\left(Z \leqslant z_\gamma\right) = 1 + \gamma/2 = \frac{1{,}95}{2}$ tiramos $z_\gamma = 1{,}96$. Descoberto tal z_γ precisamos escolher n de modo que

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} = z_\gamma$$

e se usarmos a estimativa mais conservadora $p(1-p) \leqslant 1/4$

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \geqslant \frac{\varepsilon\sqrt{n}}{\sqrt{4}}$$

portanto é suficente termos n tal que

$$\frac{\varepsilon\sqrt{n}}{\sqrt{4}} = z_\gamma$$

ou seja

$$n = \frac{z_\gamma^2}{4\varepsilon^2}.$$

Exemplo 119. Para $\varepsilon = 0{,}04$ e $\gamma = 0{,}95$

$$n = \frac{1,96^2}{4 \cdot 0{,}04^2} = 600{,}25$$

Para uma estimativa com erro de 3 pontos percentuais e 95% de grau de confiança

$$n = \frac{1,96^2}{4 \cdot 0{,}03^2} = 1068.$$

Analogamente, n para uma estimativa com erro de 3 pontos percentuais com 99% de grau de confiança então $z_{0,99} = 2,57$ e

$$n = \frac{2,57^2}{4 \cdot 0{,}03^2} = 1835.$$

Informações obtidas no sítio do IBOPE referentes ao 1º turno das eleições municipais de 2008:

cidade	no. de eleitores	no. de entrevistas	proporção eleitores entrevistados	margem de erro	grau de confiança
São Paulo	8.198.282	1.204	0,01468%	3%	95%
Rio de Janeiro	4.579.282	1.204	0,02629%	3%	95%
Belo Horizonte	1.772.227	1.204	0,06793%	3%	95%
Santo André	533.428	504	0,09448%	4%	95%
Diadema	301.229	504	0,1673%	4%	95%
São Carlos	154.572	504	0,326%	4%	95%
Cubatão	91.693	504	0,5496%	4%	95%
Registro	41.001	504	1,2292%	4%	95%
Campinas	724.143	2.000	0,2761%	3%	99%
São José dos Campos	414.353	2.000	0,4826%	3%	99%
Ribeirão Preto	388.690	2.000	0,5145%	3%	99%

3.0.26 Observação. *O valor de n não depende do tamanho da população.*

No exemplo acima provamos que

$$\mathbb{P}\left(\hat{p} - 1,96\sqrt{4/n} \leqslant p \leqslant \hat{p} + 1,96\sqrt{4/n}\right) \geqslant 0{,}95$$

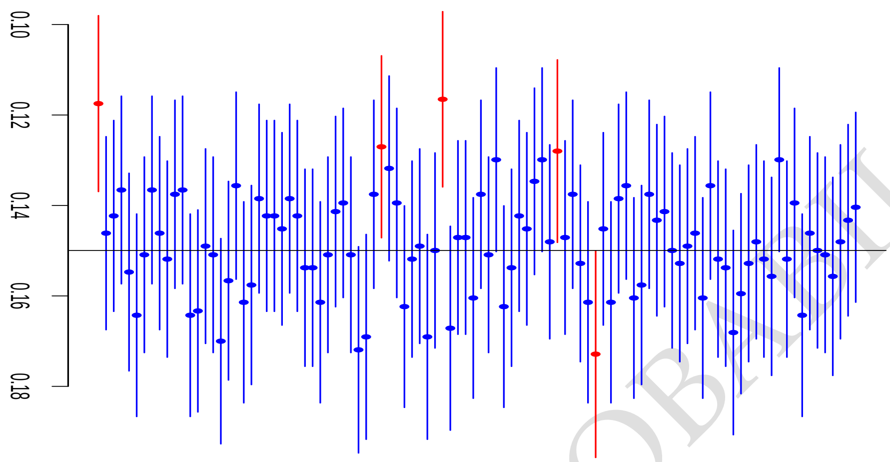
e dizemos que $(\hat{p} - 1,96\sqrt{4/n}\,,\ \hat{p} + 1,96\sqrt{4/n})$ é um **intervalo de confiança** para p com grau de confiança 95%. Notemos que p é um valor médio desconhecido e $\hat{p}$ é uma variável aleatória, portanto o intervalo é aleatório.

Intervalos de confiança: Vamos usar o TCL para resolver o seguinte problema. Queremos estimar a média μ (desconhecida) de alguma característica de uma população pela média M_n de uma amostra aleatória de n indivíduos da população como, por exemplo, fizemos no exemplo 113, com erro controlado: Dados $\epsilon > 0$ e $\gamma \in (0,1)$, para cada amostra aleatória $X_1, \dots, X_n$ queremos uma estimativa intervalar $(M_n - \varepsilon\,,\ M_n + \varepsilon)$ para a média μ da população com grau de confiança γ, isto é,

$$\mathbb{P}\left(\mu \in (M_n - \varepsilon\,,\ M_n + \varepsilon)\right) \geqslant \gamma \tag{3.0.14}$$

que pode ser interpretado assim: num número grande de amostras do mesmo tamanho, se obtivermos um intervalo com grau de confiança, por exemplo 0,95, para cada uma delas, então 95% desses intervalos contém o parâmetro μ.

Por exemplo, o seguinte gráfico apresenta cem intervalos para amostras aleatórias simples de acordo com Bernoulli(0,15) de tamanho $n = 1047$, para cada amostra foi computado um intervalo de confiança de grau 0,95. Em vermelho estão os intervalos que mão contêm a média populacional 0,15:



Tamanho da amostra:: O tamanho n para uma amostra aleatória $X_1, \dots, X_n$ satisfazer equação (3.0.14) pode ser determinada como no caso binomial explicado acima.

$$\begin{aligned}
\gamma &= \mathbb{P}(M_n - \varepsilon \leqslant \mu \leqslant M_n + \varepsilon) \\
&= \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leqslant \frac{M_n - \mu}{\sigma/\sqrt{n}} \leqslant \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\
&\approx \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leqslant Z_n \leqslant \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\
&= \mathbb{P}\left(-\frac{\varepsilon\sqrt{n}}{\sigma^2}\sigma \leqslant Z_n \leqslant \frac{\varepsilon\sqrt{n}}{\sigma^2}\sigma\right)) \\
&= 2\mathbb{P}\left(Z_n \leqslant \frac{\varepsilon\sqrt{n}}{\sigma^2}\sigma\right) - 1
\end{aligned}$$

Tomemos o valor z_γ tal que

$$\mathbb{P}(Z \leqslant z_\gamma) = \frac{1+\gamma}{2}$$

e de $\epsilon\sqrt{n}/\sigma = z_\gamma$ deduzimos

$$n = \left(\frac{z_\gamma \sigma}{\varepsilon}\right)^2$$

e a estimativa intervalar para μ com grau de confiança γ é

$$\mathrm{IC}(\epsilon, \gamma) := \left(M_n - z_\gamma \frac{\sigma}{\sqrt{n}}\,,\, M_n + z_\gamma \frac{\sigma}{\sqrt{n}}\right)$$

Exemplo 120. A renda per-capita domiciliar numa certa região tem desvio padrão 250 reais e média desconhecida. Se desejamos estimar a renda média da população com erro 50 reais e confiabilidade $\gamma = 0{,}95$ quantos domicílios deveremos consultar? Já sabemos que $z_\gamma = 1{,}96$, então

$$n = \left(\frac{z_\gamma}{\epsilon}\right)^2 \sigma^2 = \left(\frac{1{,}96}{50}\right)^2 250^2 = 96{,}04.$$

Exemplo 121. Um provedor de internet monitora o a duração da conexão dos clientes a fim de dimensionar os seus servidores. A média e a distribuição desse tempo são desconhecidos mas o desvio padrão é $\sqrt{50}$ minutos. Numa amostra de 500 conexões o valor médio foi 25 minutos; o que podemos disser a respeito da média com grau de confiança 92%? Como o tamanho da amostra é razoavelmente grande, podemos usar o TCL e aproximar a distribuição por uma normal. Um intervalo de confiança para o tempo de conexão é

$$\left(M - z_\gamma \frac{\sigma}{\sqrt{n}}, M + z_\gamma \frac{\sigma}{\sqrt{n}}\right) = (24.45, 25.55).$$

Em virtude do uso do TCL, o intervalo acima é com grau de confiança *aproximadamente* 0,92.

Na prática não conhecemos σ^2 e devemos substituí-lo por uma estimativa amostral, que pode ser

$$S^2 = \frac{1}{n-1}\sum (x_i - M_n)^2$$

Exemplo 122. O tempo de reação de um remédio pode ser considerado como tendo distribuição normal. Num teste, 20 pacientes foram sorteados e os tempo anotados:

$$\begin{matrix} 2,9 & 3,4 & 3,5 & 4,1 & 4,6 & 4,7 & 4,5 & 3,8 & 5,3 & 4,9 \\ 4,8 & 5,7 & 5,8 & 5,0 & 3,4 & 5,9 & 6,3 & 4,6 & 5,5 & 6,2 \end{matrix}$$

então, a variância amostral é $S^2 = 0{,}992079$ e o intervalo a 95% é

$$\left(M - z_{0{,}95}\sqrt{\frac{S^2}{n}}, M + z_{0{,}95}\sqrt{\frac{S^2}{n}}\right) = (4,278843\,,\ 5,211157).$$

3.0.27 Aproximação de Stirling: Seja $S_n = X_1 + \cdots + X_n$ com $X_i \sim$ Poisson(1) variável aleatória's independentes. Então

$$\mathbb{P}(n-1 < S_n \leqslant n) = \mathbb{P}\left(\frac{-1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leqslant 0\right) \approx \int_{-1/\sqrt{n}}^{0} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}\,\mathrm{d}x \approx \frac{1}{\sqrt{2\pi n}}$$

portanto

$$\mathbb{P}(S_n = n) = \frac{\mathrm{e}^{-n} n^n}{n!} \approx \frac{1}{\sqrt{2\pi n}}$$

donde

$$n! \approx n^{1/2+n}\mathrm{e}^{-n}\sqrt{2\pi}.$$

Funções geradoras. Se $a_0, a_1, a_2, \dots$ é uma sequência de números reais e

$$A(s) = \sum_{i \geqslant 0} a_i s^i$$

converge em algum intervalo então $A(s)$ é uma **função geradora** da sequência $(a_i)_{i \in \mathbb{N}}$. Por exemplo,

$$\frac{1}{1-s} = 1 + s + s^2 + s^3 + \cdots$$

é função geradora da sequência $(1)_{i \in \mathbb{N}} = 1, 1, 1, \dots$

$$e^s = 1 + \frac{s}{1} + \frac{s^2}{2!} + \frac{s^3}{3!} + \cdots$$

é função geradora da sequência $(1/j!)_{j \in \mathbb{N}}$ e

$$\frac{1}{\sqrt{1-4s}} = \sum_{k \geqslant 0} \binom{2k}{k} s^k$$

é função geradora da sequência $\left(\binom{2k}{k}\right)_{k \in \mathbb{N}}$.

Exemplo 123. Se X é o resultado do lançamento de um dado, então a função geradora para a função de probabilidade de X é

$$D(s) = \frac{1}{6}(s + s^2 + s^3 + s^4 + s^5 + s^6)$$

e notemos que a função de probabilidade para a soma do lançamento de dois dados

$$\begin{aligned} D(s)D(s) &= \frac{1}{36}(s + s^2 + s^3 + s^4 + s^5 + s^6)(s + s^2 + s^3 + s^4 + s^5 + s^6) \\ &= \frac{1}{36}(s^2 + 2s^3 + 3s^4 + 4s^5 + 5s^6 + 5s^7 + 5s^8 + 5s^9 + 3s^{10} + 2s^{11} + s^{12}) \end{aligned}$$

Se X assume valores naturais, a função geradora de probabilidade de X é a função geradora da função de massa de probabilidade de X

$$P_X(s) := \mathbb{E}\left[s^X\right] = \sum_{i \geqslant 0} \mathbb{P}(X = i) s^i.$$

A série converge absolutamente para, pelo menos, todo $s \in (-1, 1)$. Derivando P_X com relação a s

$$P_X'(s) = \sum_{i \geqslant 1} i \mathbb{P}(X = i) s^{i-1}$$

portanto

$$P_X'(1) = \mathbb{E}[X]. \tag{3.0.15}$$

Exercício 99. Mostre que

$$\mathrm{Var}[X] = P_X''(1) + P_X'(1) - P_X'(1)^2. \tag{3.0.16}$$

Se X e Y são variáveis aleatórias independentes então s^X e s^Y são independentes, portanto, $\mathbb{E}\left[s^{X+Y}\right] = \mathbb{E}\left[s^X s^Y\right] = \mathbb{E}\left[s^X\right]\mathbb{E}\left[s^Y\right]$ ou seja,

$$P_{X+Y}(s) = P_X(s)P_Y(s)$$

e, usando indução podemos provar que

$$P_{X_1+\cdots+X_n}(s) = P_{X_1}(s)\cdots P_{X_n}(s)$$

agora

3.0.28 Teorema. *Se $X_1, X_2, \ldots$ é uma sequência de variáveis aleatórias independentes de mesma distribuição e com função geradora P_X e se N é uma variável aleatória independente das demais variáveis aleatórias e tem função geradora P_N, então a soma $X_1 + \cdots + X_N$ tem função geradora*

$$P_{X_1+\cdots+X_N}(s) = P_N(P_X(s)). \tag{3.0.17}$$

De fato,

$$P_{X_1+\cdots+X_N}(s) = \mathbb{E}\left[s^{X_1+\cdots+X_N}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{X_1+\cdots+X_N} \,\middle|\, N\right]\right]$$

e o lado direito vale

$$\sum_n \mathbb{E}\left[s^{X_1+\cdots+X_N} \,\middle|\, N = n\right]\mathbb{P}(N = n) = \sum_n \mathbb{E}\left[s^{X_1+\cdots+X_n}\right]\mathbb{P}(N = n)$$

e pela independência ficamos com

$$\sum_n \mathbb{E}\left[s^{X_1}\right]\cdots\mathbb{E}\left[s^{X_n}\right]\mathbb{P}(N = n) = \sum_n P_X(s)^n\mathbb{P}(N = n) = P_N(P_X(s)).$$

$*$ **Sobre convergência**

Acima, não nos preocupamos com a convergência das séries mas podemos diferenciar, ou integrar, como fizemos acima para $\mathbb{E}[X] = P_X'(S)$? Embora não seja óbvio, é sempre seguro assumir convergência de $P_X(s)$ quando $|s| < 1$.

A função geradora de probabilidade de uma variável aleatória que assume valores nos naturais

$$\mathbb{E}\left[s^X\right] = \sum_{i\geqslant 0} f_X(i)s^i$$

converge (absolutamente) para todo s, $|s| < R$ e

$$R = \frac{1}{\limsup_{i\to\infty} |f_X(i)|^{1/i}}$$

e diverge se $|s| > R$. O número R sempre existe, no nosso caso $R \geqslant 1$, e é chamado de raio de convergência. $P_X(s)$ pode ser derivada e integrada, qualquer número de vezes, dentro do raio de convergência.

Um teorema devido a Abel garante que $\lim_{s\to 1^-} P_X(s) = P_X(1) = \sum_{i\geqslant 0} f_X(i)$, ou seja a função geradora é contínua a esquerda em $s = 1$. Assim, escrevemos $P_X(1)$ como abreviação de $\lim_{s\to 1^-} P_X(s)$.

$\ast$ Demonstrações.

Exercícios.

1. Se λ é real positivo e X uma variável aleatória então

$$\mathbb{P}(|X|\geqslant \lambda) \leqslant \frac{\mathbb{E}\left[|X|^t\right]}{\lambda^t} \tag{3.0.18}$$

   para todo $t > 0$.

2. Mostre que se $\varepsilon > 0$ e X assume os valores $-\varepsilon$ e ε com probabilidade $1/2$, cada, então vale a igualdade na desigualdade de Chebyshev.

3. Uma moeda honesta é lançada 100 vezes. O número esperado de caras é 50, e o desvio padrão para o número de caras é 5. O que a desigualdade de Chebyshev nos diz sobre a probabilidade de que o número de caras que ocorrem desvia do número esperado por três ou mais desvios-padrão (ou seja, em pelo menos 15)?

4. A variável aleatória X tem média 10 e variância 4. Use a desigualdade de Chebyshev para determinar (a) $\mathbb{P}(|X-10|\geqslant 3)$ $\leqslant 4/9$; (b) $\mathbb{P}(|X-10|< 3)$ $\geqslant 5/9$; (c) $\mathbb{P}(5 < X < 15)$ $\geqslant 21/24$; (d) o valor de c para que $\mathbb{P}(|X-10|\geqslant c) \leqslant 0{,}04$ 10.

5. Quantas vezes uma moeda deve ser lançada no mínimo para que a proporção número de caras esteja no intervalo $(\frac{1}{4}, \frac{3}{4})$ com probabilidade pelo menos $0{,}95$? 80

6. Uma fábrica de refrigerantes precisa calibrar a máquina de engarrafamento quando ela passa a variar a medida de refrigerante engarrafado fora de determinado limite. Pra controle de qualidade, a fábrica escolhe uma amostra de n de cada lote e verifica se cada garrafa passa ou não passa no teste de volume. Quanto deve valer n, pelo menos, para que a proporção de garrafas boas examinadas não difira da proporção real de garrafas boas por mais que $0{,}04$ com probabilidade pelo menos $0{,}95$? 3125

7. Considere n lançamentos independentes de um dado. Sejam X_j o resultado do n-ésimo lançamento e $S_n = X_1 + X_2 + \cdots + X_n$. Prove que

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| < \varepsilon\right) \to 1 \text{ quando } n \to \infty$$

   para qualquer $\varepsilon > 0$.

8. Evidências a respeito de culpa ou inocência de um réu podem ser resumidas pelo valor de uma variável aleatória X cuja média μ dependem da culpa do réu. Se ele é inocente, $\mu = 1$; se ele é culpado, $\mu = 2$. O juiz considerará o réu culpado se $X > c$ para algum valor de c adequadamente escolhido.

a) Se o juiz quer estar 95% certo de que um inocente não será condenado, qual deve ser o valor de c? 2,996

b) Usando o valor obtido acima, qual é a probabilidade com que um réu culpado é condenado? 0,2236

9. A renda per-capita domiciliar numa certa região tem desvio padrão 250 reais e média desconhecida. Se desejamos estimar a renda média da população com erro 50 reais e confiabilidade $\gamma = 0{,}95$ quantos domicílios deveremos consultar? 97

10. O comprimento de jacarés adultos segue a distribuição normal com parâmetros μ e $\sigma^2 = 0{,}01$. Uma amostra de 10 animais foi sorteada e o comprimento médio dessa amostra é $1{,}69\,\mathrm{m}$. Forneça um intervalo de confiança para μ com grau de confiança 0,95. (1.63,1.75)

11. Baseado em estudos preliminares, é possível admitir que a vida média das baterias automotivas segue a distribuição normal com desvio padrão de 4,5 meses. Qual deve ser o tamanho da amostra para que a amplitude do intervalo de 90% de confiança para a vida média seja 3 meses.? 25

12. Queremos estimar a proporção p de cura através do uso de um medicamento. Um experimento aplicou o medicamento em 200 pacientes escolhidos ao acaso, dos quais 160 se curaram, O que podemos dizer a respeito de p? $p \in (0{,}732,\ 0{,}869)$ com probabilidade 0,95

13. *Monsieur* Leclerc, o conde de Buffon, empolgou-se com seus experimetos (veja exemplos 1.3 e 106) e engoliu uma das agulhas. Assumindo que a agulha não tem nenhuma direção preferida dentro do estômago, qual é a distribuição do comprimento da agulha que aparece na imagem do raio-X do estômago?

14. Prove a **desigualde de Jensen**: Se f é uma função real, contínua e convexa então

$$\mathbb{E}\left[f(X)\right] \geqslant f(\mathbb{E}\left[X\right]).$$

15. Prove a **desigualdade de Cauchy–Schwarz**

$$\mathbb{E}\left[XY\right] \leqslant \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]}.$$

§4 Distribuição condicional

Distribuição condicional.

Distribuição condicional:: se X e Y são variáveis aleatórias e $x \in \mathbb{R}$ é tal que e $f_X(x) > 0$ então a função de distribuição condicionada de Y dado $[X = x]$ é a função em $y \in \mathbb{R}$ dada por

$$\boxed{F_{Y|X}(y|x) := \mathbb{P}(Y \leqslant y \mid X = x)}$$

Se X e Y são variável aleatória conjuntamente contínuas com densidade $f_{X,Y}(x, y)$

$$F_{Y|X}(y|x) := \int_{-\infty}^{y} \frac{f_{X,Y}(x, v)}{f_X(x)} \,\mathrm{d}v$$

e escrevemos $\mathbb{P}(Y \leqslant y \mid X = x)$, apesar de $\mathbb{P}(X = x) = 0$.

Função de probabilidade condicional:: a função de probabilidade de Y dado $[X = x]$ é

$$\boxed{f_{Y|X}(y|x) := \frac{f_{X,Y}(x, y)}{f_X(x)}.}$$

Exemplo 124. Sejam X e Y variáveis aleatórias independentes com $f_X(a) = f_Y(a) = \binom{n}{a}p^a(1-p)^{n-a}$, todo $a \in \{1, 2, \dots, n\}$, algum $p \in (0, 1)$. Então

$$\begin{aligned}\mathbb{P}(X + Y = m) &= \sum_{i=0}^{n} \mathbb{P}(X = i,\ Y = m - i) = \sum_{i=0}^{n} \mathbb{P}(X = i)\mathbb{P}(Y = m - i)\\ &= \sum_{i=0}^{n} \binom{n}{i}\binom{n}{m-i} p^m(1-p)^{n-m+n} = \binom{2n}{m} p^m(1-p)^{2n-m}\end{aligned}$$

pois $\binom{a+b}{k} = \sum_{i=0}^{n}\binom{a}{i}\binom{b}{k-i}$. Desse fato temos

$$\begin{aligned}\mathbb{P}(X = k \mid X + Y = m) &= \frac{\mathbb{P}(X = k\,,\ X + Y = m)}{\mathbb{P}(X + Y = m)}\\ &= \frac{\mathbb{P}(X = k\,,\ Y = m - k)}{\mathbb{P}(X + Y = m)}\\ &= \frac{\mathbb{P}(X = k)\mathbb{P}(Y = m - k)}{\mathbb{P}(X + Y = m)}\\ &= \frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-m+k}}{\binom{2n}{m}p^m(1-p)^{2n-m}}\\ &= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}\end{aligned}$$

ou seja, X condicionada a $X + Y = m$ tem distribuição hipergeométrica com parâmetros $2n, n, m$.

Notemos que, caso as variável aleatória sejam independentes vale $f_{Y|X}(y|x) = f_Y(y)$ e vice-versa, i.e., a igualdade implica independência.

Exemplo 125. De volta ao exemplo 101, a distribuição de Y condicionada a $X = 8$ é

$$f(y \mid 0{,}8) = \frac{f(0{,}8, y)}{f_X(0{,}8)} = \frac{1{,}2(0{,}8 + y^2)}{1{,}2 \cdot 0{,}8 + 0{,}4}, \quad \forall y \in (0, 1)$$

A probabilidade do caixa convencional estar ocupado metade do tempo, dado que $X = 8$, é

$$\mathbb{P}(Y \leqslant 0{,}5 \mid X = 8) = \int_{-\infty}^{0{,}5} f(y \mid 0{,}8)\,\mathrm{d}y = 0{,}39.$$

Usando a distribuição marginal temos que $\mathbb{P}(Y \leqslant 0{,}5) = 0{,}35$.

Esperança condicional. Se X e Y são variável aleatória e $x \in \mathbb{R}$ é tal que e $f_X(x) > 0$ então

$$\mathbb{E}[Y \mid X = x] := \begin{cases} \displaystyle\sum_y y f_{Y|X}(y|x) & \text{no caso discreto} \\ \displaystyle\int_{-\infty}^{+\infty} y f_{Y|X}(y|x)\,\mathrm{d}y & \text{no caso contínuo.} \end{cases}$$

Notemos que quando as variável aleatória são independentes $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y]$ (verifique).

Se $\mu : \mathbb{R} \to \mathbb{R}$ é dada por

$$\mu(x) := \mathbb{E}[Y \mid X = x]$$

então $\mu(X)$ é a **esperança condicional** de Y dado X, denotada

$$\mathbb{E}[Y \mid X]$$

que embora sugira uma média é, de fato, uma variável aleatória, a qual tem a seguinte propriedade

Exercício 100. Verifique

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

Logo

$$\mathbb{E}[Y] = \begin{cases} \displaystyle\sum_x \mathbb{E}[Y \mid X = x])\mathbb{P}(X = x) & \text{no caso discreto} \\ \displaystyle\int_{-\infty}^{+\infty} \mathbb{E}[Y \mid X = x]\, f_X(x)\,\mathrm{d}x & \text{no caso contínuo.} \end{cases} \tag{3.0.19}$$

Se $h(X, Y)$ tem esperança finita então

$$\mathbb{E}[h(X, Y) \mid Y = y] = \mathbb{E}[h(X, y) \mid Y = y].$$

Exemplo 126. De volta ao exemplo 101, a distribuição de Y condicionada a $[X = 8]$ é

$$f_{Y|X}(y \mid 0{,}8) = \frac{f_{X,Y}(0{,}8, y)}{f_X(0{,}8)} = \frac{1{,}2(0{,}8 + y^2)}{1{,}2 \cdot 0{,}8 + 0{,}4}, \quad \forall y \in (0, 1)$$

A probabilidade do caixa convencional estar ocupado metade do tempo, dado que $X = 8$, é

$$\mathbb{P}(Y \leqslant 0{,}5 \mid X = 8) = \int_{-\infty}^{0{,}5} f_{Y|X}(y \mid 0{,}8)\,\mathrm{d}y = 0{,}39.$$

Usando a distribuição marginal temos que $\mathbb{P}(Y \leqslant 0{,}5) = 0{,}35$. Ademais $\mathbb{E}\left[Y\right] = 6$ e

$$\mathbb{E}\left[Y \mid X = 8\right] = \int_{-\infty}^{+\infty} y f_{Y|X}(y \mid 0{,}8)\,\mathrm{d}y = 0{,}574.$$

Exemplo 127. Sejam X, Y como no exemplo 124. X condicionada a $X + Y = m$ tem distribuição hipergeométrica com parâmetros $2n, n, m$, logo

$$\mathbb{E}\left[X \mid X + Y = m\right] = \frac{m}{2}.$$

Exemplo 128 (**Esperança de variável aleatória geométrica**). Seja X o número de lançamentos de uma moeda até sair cara, o que ocorre com probabilidade p, seja Y a variável aleatória indicadora de “cara no primeiro lançamento”. Assim,

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}\left[X \mid Y = 0\right](1 - p) + \mathbb{E}\left[X \mid Y = 1\right]p$$

mas $\mathbb{E}\left[X \mid Y = 1\right] = 1$ e $\mathbb{E}\left[X \mid Y = 0\right] = \mathbb{E}\left[(\right] X + 1)$, portanto

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X\right](1 - p) + (1 - p) + p$$

que resolvendo para $\mathbb{E}\left[X\right]$ resulta em $\mathbb{E}\left[X\right] = 1/p$, como já sabíamos. Do mesmo modo, temos

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[X^2 \,\middle|\, Y = 0\right](1 - p) + \mathbb{E}\left[X^2 \,\middle|\, Y = 1\right]p = \mathbb{E}\left[(X + 1)^2\right](1 - p) + p$$

que resolvendo para $\mathbb{E}\left[X^2\right]$ resulta em

$$\mathbb{E}\left[X^2\right] = \frac{2 - p^2}{p}$$

e com isso

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 = \frac{2 - p^2}{p} - \frac{1}{p}^2 = \frac{1 - p^2}{p}.$$

Exemplo 129 (**Soma de variável aleatória com número aleatório de termos**). Num determinado dia do cotidiano de uma loja N pessoas entram na loja, com $\mathbb{E}\left[N\right] = 50$, e o gasto dessa pessoas são dados pelas variável aleatória independentes $X_1, X_2, \ldots, X_N$ com $\mathbb{E}\left[X\right]_i = \mu = 80$ reais, para todo i, e independentes de N.

A quantia gasta num determinado dia é $X = \sum_{i=1}^{N} X_i$ cuja média é

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid N\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right]$$

Agora,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i \,\middle|\, N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu$$

por causa da independência, por fim

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right] = \mathbb{E}\left[N\mu\right] = \mathbb{E}\left[N\right]\mu = 4.000$$

reais.

Exercício 101. Considere o espaço amostral das sequências definidas pelas permutações de $\{1, 2, 3\}$

$$\Omega = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1), (1, 1, 1), (2, 2, 2), (3, 3, 3)\}.$$

Seja $X_i(\omega)$ a i-ésima coordenada de ω, para $i = 1, 2, 3$. Defina N a variável aleatória igual a X_2. Prove que

- para todo $i, j \in \{1, 2, 3\}$, vale $\mathbb{P}(X_i = j) = 1/3$;
- as variáveis X_i não são independentes;
- $\displaystyle\sum_{i=1}^{\mathbb{E}[N]} \mathbb{E}\left[X_i\right] = 4$ e conclua que

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] \neq \sum_{i=1}^{\mathbb{E}[N]} \mathbb{E}\left[X\right]_i.$$

Exemplo 130. Sejam $U_1, U_2, \ldots$ variável aleatória com distribuição Uniforme($(0, 1)$) independentes e $m(x)$ o valor esperado para $N(x)$ que é o menor n para o qual $U_1 + \cdots + U_n > x$. Certamente,

$$\mathbb{E}\left[N(x)\right] U_1 = y = \begin{cases} 1, & \text{se } y > x \\ 1 + m(x - y), & \text{se } y \leqslant x. \end{cases}$$

Portanto,

$$m(x) = \int_0^1 \mathbb{E}\left[N(x) \mid U_1 = y\right] \mathrm{d}y = 1 + \int_0^x m(x - y)\,\mathrm{d}y = 1 + \int_0^x m(u)\,\mathrm{d}u.$$

Derivando os extremos dessa cadeia de igualdades obtemos $m'(x) = m(x)$, ou seja, $m(x) = a\mathrm{e}^x$ para alguma constante real a e como $m(0) = 1$ temos

$$m(x) = \mathrm{e}^x$$

é o número esperado de termos para que a soma de variável aleatória uniformes em $(0, 1)$ ultrapasse o valor x.

Exercício 102. Mostre que para constantes $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY \mid Z] = a\mathbb{E}[X \mid Z] + b\mathbb{E}[Y \mid Z].$$

Exercício 103. Mostre que $\mu(X) = \mathbb{E}[Y \mid X]$ satisfaz

$$\mathbb{E}[\mu(X)g(X)] = \mathbb{E}[Yg(X)]$$

para qualquer função g pra qual as esperanças acima existam.

Exercício 104. Uma galinha bota N ovos, em que $N \sim \text{Poisson}(\lambda)$. Cada ovo vinga com probabilidade p e independente dos outros ovos. Calcule $\mathbb{E}[N \mid K]$, $\mathbb{E}[K]$ e $\mathbb{E}[K \mid N]$, em que K é o número de pintinhos.

Exercícios.

1. Se X, Y são variável aleatória discretas com função de probabilidade conjunta

$$f(x, y) = \begin{cases} cxy, & \text{se } x, y \in \{1, 2, 3, 4\} \text{ e } x + y \geqslant 5 \\ 0, & \text{caso contrário} \end{cases}$$

determine o valor de c. Essas variáveis aleatórias são independentes? Com que probabilidade X é ímpar dado que $X + Y$ é par?

2. Sejam X o mínimo e Y o máximo de três números sorteados sem reposição de $\{0, 1, \ldots, 9\}$. Determine a função de probabilidade conjunta de X e Y bem como as distribuições marginais. Obtenha a função de probabilidade de $Y - X$.

3. Um número aleatório N de dados são lançados e para $i \geqslant 1$ inteiro $\mathbb{P}(N = i) = (1/2)^i$. Denotemos por S a soma dos resultados dos dados. Determine as probabilidade de

   a) $N = 2$ dado que $S = 3$,

   b) $S = 3$ dado que N é par.

4. Sejam X e Y variáveis aleatórias discretas cuja função de probabilidade conjunta é dada pela seguinte tabela:

$X \backslash Y$	1	2
1	1/20	1/5
2	0	1/10
3	1/10	1/10
4	1/20	2/5

Obtenha a função de probabilidade de X dado que $Y = 2$ e calcule $\mathbb{E}\left[X^2 \mid Y\right] = 2)$.

5. Sejam X e Y variáveis aleatórias com função densidade conjunta

$$f(x, y) = \begin{cases} 8xy & \text{se } 0 < y < x < 1, \\ 0 & \text{caso contrário.} \end{cases}$$

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

6. Escolhe-se ao acaso um ponto (X, Y) no triângulo $\{(x, y) \in \mathbb{R}^2 : 0 < y < x < 1\}$. Calcule $\mathbb{E}[X \mid Y]$. Obtenha $\mathbb{E}[Y \mid X]$ e $\mathbb{E}\left[Y^2 \mid X\right]$. Usando o passo anterior determine $\mathbb{E}\left[(X - Y)^2 \mid X\right]$.

7. Sejam X e Y variáveis aleatórias tais que $\mathbb{E}[X \mid Y] = Y$ e $\mathbb{E}[Y \mid X] = X$. Prove que $\mathbb{P}(X = Y) = 1$.

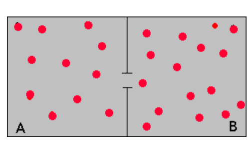
8. Escolhe-se ao acaso um número X entre os inteiros de 1 a n. A seguir, um número Y é escolhido aleatoriamente entre os mesmos números, excluídos os menores que X. (a) Quanto vale $\mathbb{E}[Y \mid X]$? (b) Mostre que $\mathrm{Cov}(X, Y) = \mathrm{Var}[X]/2$.

9. Sejam X o mínimo e Y o máximo de três números sorteados ao acaso, sem reposição, do conjunto $\{1, 2, \ldots, 6\}$. (a) Determine a função de probabilidade conjunta de X e Y, bem como as marginais. (b) Obtenha a função de probabilidade de X dado que $Y = 5$. (c) Calcule $\mathbb{E}[X \mid Y - X = 3]$.

§7 Processos estocásticos a tempo discreto.

Processos estocásticos a tempo discreto:: é uma sequência de $(X_i)_{i \in \mathbb{N}}$ de variáveis aleatórias definidas sobre o mesmo espaço amostral que assumem valor num **conjunto de estados** S enumerável. O índice i é interpretado como tempo e o valor de X_i *é o estado do processo no instante* i.

O modelo de Ehrenfest: Um sistema é formado por dois compartimentos A e B ligados por uma pequena abertura. No início, N moléculas de um gás estão no compartimento A e o compartimento B está vazio.



Em cada instante uma úncia molécula troca de compartimento e todas as moléculas de um mesmo compartimento tem a mesma probabilidade de passar pela abertura.

Parece claro que o sistema deve, eventualmente, chegar a um equilíbrio com as moléculas distribuídas igualmente entre A e B. No entanto, as moléculas podem se reorganizar desproporcionadamente em A ou em B. À primeira vista, não há sentido óbvio de fluxo de tempo neste experimento, que é uma realização discreta de dinâmica molecular, onde as equações de de trajetórias são reversíveis no tempo. Como leis reversíveis de movimento podem, intrinsecamente, conter uma direção preferencial de fluxo de tempo? Se começar com todas as moléculas de um lado, o que evita que o reagrupamento de todas as moléculas de volta para o lado do estado inicial? Quando Boltzmann foi desafiado com essa objeção, apocrifamente ele respondeu, impaciente, "Versuch es doch!" ("Experimente!")[1]

Seja X_t a quantidade de moléculas no instante t que estão no compartimento B, $X_0 = 0$. Conhecido X_t, no próximo momento, X_{t+1}, temos

$$X_{t+1} = \begin{cases} X_t - 1 & \text{caso uma molécula tenha passado de } B \text{ para } A \\ X_t + 1 & \text{caso uma molécula tenha passado de } A \text{ para } B \end{cases}$$

e essa variáveis aleatórias definem um processo estocástico sobre o conjunto de estados $S = S(N) := \{1, 2, \dots, N\}$. Ademais

$$\begin{aligned} \mathbb{P}(X_1 = 1 \mid {}_0 = 0) &= 1 \\ \mathbb{P}(X_{t+1} = N-1 \mid X_t = N) &= 1 \qquad (t > 0) \\ \mathbb{P}(X_{t+1} = k+1 \mid X_t = k) &= \frac{N-k}{N} \quad (t > 0) \\ \mathbb{P}(X_{t+1} = k-1 \mid X_t = k) &= \frac{k}{N} \quad (t > 0) \end{aligned}$$

em qualquer outro caso $\mathbb{P}(X_{t+1} = j \mid X_t = i) = 0$.

Matriz das transições:: é a matriz quadrada $P = (p_{i,j})$ obtida fazendo $p_{i,j} = \mathbb{P}(X_{t+1} = j \mid X_t = i) = 0$. Por exemplo, para $N = 3$ temos

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

e se denotarmos a função de massa de probabilidade $f^{(t)}$ da variável aleatória X_t por um vetor, fazendo $f_k^{(t)} = f^{(t)}(k) = \mathbb{P}(X_t = k)$ para todo $k \in S$

$$\begin{aligned} f^{(0)} &= (1, 0, 0, 0) \\ f^{(1)} &= (0, 1, 0, 0) \\ f^{(2)} &= (1/3, 0, 2/3, 0) \end{aligned}$$

[1] Nonconvergence of the Ehrenfest thought experiment, C. R. MacCluer, Am. J. Phys. 77, 695 (2009); View online: http://dx.doi.org/10.1119/1.3130022

e, se $t > 2$ temos $\mathbb{P}(X_{t+1} = k) = \sum_{s=0}^{3} \mathbb{P}(X_{t+1} = k \mid X_t = s)\mathbb{P}(X_t = s)$ pela Lei de Probabilidade Total, i.e.,

$$f^{(t+1)}(k) = f^{(t)}(k-1)p_{k,k-1} + f^{(t)}(k+1)p_{k,k+1}$$

para todo k.

De um modo geral, conhecida uma **distribuição inicial** de probabilidades $f^{(0)}$ sobre $S(N)$ e a matriz de transições P então para todo $t > 0$

$$\begin{aligned} f^{(t+1)}(k) &= \textstyle\sum_{s=0}^{N} \mathbb{P}(X_{t+1} = k \mid X_t = s)\mathbb{P}(X_t = s) \\ &= \mathbb{P}(X_{t+1} = k \mid X_t = k-1)\mathbb{P}(X_t = k-1) + \mathbb{P}(X_{t+1} = k \mid X_t = k+1)\mathbb{P}(X_t = k+1) \end{aligned}$$

ou seja

$$\boxed{f^{(t+1)} = f^{(t)}P = f^{(0)}P^t.}$$

No nosso exemplo

$$f^{(3)} = f^{(2)}P = (1/3, 0, 2/3, 0)\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (0, 7/9, 2/9, 0).$$

Distribuição invariante:: é um vetor $\pi = (\pi_i)_{i \in S}$ tal que $\pi_k = \mathbb{P}(X_t = k)$ tal que $\boxed{\pi = \pi P}$ A matriz P do modelo de Ehrenfest admite uma distribuição invariante e isso nos permite algumas conclusões interessantes a respeito desse processo. Essa distribuição é, para $k \in S(N)$,

$$\pi_k := \frac{1}{2^N}\binom{N}{k}.$$

Não é difícil verificar que essa distribuição é simétrica em torno de $k = N/2$, logo o sistema tem maior probabilidade de estar num estado com aproximadamente metade das moléculas em cada compartimento.

No caso $N = 3$, por exemplo, $\pi = (1/8, 3/8, 3/8, 1/8)$.

Uma das consequências é que pelo Teorema Ergódico para Cadeias de Markov (Teorema 3.0.38, página 258 — esse processo satisfaz as hipóteses do teorema) para todo $k \in S$, com probabilidade 1 vale que

$$\lim_{n \to \infty} \frac{|t : 0 < t \leqslant n,\ X_t = k|}{n} = \pi_k$$

isto é, π_k é a fração limite do número de visitas ao estado k em n passos. Assim, a fração do tempo que o processo passa no estado $N/2$ é

$$\pi_{N/2} = \frac{1}{2^N}\binom{N}{N/2} \approx \frac{1}{2^N}\frac{2^N}{\sqrt{\pi N/2}} = \frac{1}{\sqrt{\pi N/2}}$$

onde usamos no coeficiente binomial uma aproximação via fórmula de Stirling, e probabilidade de todas as moléculas voltarem para o compartimento A é muito menor

$$\pi_0 = \frac{1}{2^N}.$$

Outra consequência (equação 3.0.36 do Teorema teorema 3.0.36) é a respeito do tempo de retorno. Definamos

$$T := \inf\{t \in \mathbb{N} : t > 0,\ X_t = 0\}$$

que é o tempo do primeiro retorno a origem. Então,

$$\mathbb{E}\left[T\right] = \frac{1}{\pi_0} = 2^N.$$

Assim, se $N = 1000$ então $\mathbb{E}\,[(]\,T) = 2^{1000}$ e se medimos o tempo em segundos então 2^{1000} segundos $\approx 10^{993}$ anos (juntas, todas as pessoas do mundo já viveram $\approx 10^{11}$ anos).

Uma falha desse modelo é que não temos $f^{(t)} \to \pi$ quando $t \to \infty$. Uma ligeira modificação do experimento, para um modelo mais atraente fisicamente, é apresentada no exercício 108 abaixo. No modelo modificado há convergência para distribuição estacionária.

Exercício 105. Com a distribuição invariante sobre o conjunto de estados, qual a probabilidade do compartimento ter 0 partículas e de ter $N/2$ partículas.

Exercício 106. Descreva P^2.

Exercício 107. Se P é a matriz das transições e P^2 tem elementos $(p_{i,j}^{(2)})$ então $p_{i,j}^{(2)}$ é a probabilidade de que evento? Analogamente para $n > 2$, $P^n = (p_{i,j}^{(n)})$ e $p_{i,j}^{(n)}$ é a probabilidade de que evento?

Exercício 108. Considere o experimento de Ehrenfest com a seguinte modificação: em cada momento, independentemente do passado, um compartimento é escolhido aleatoriamente em seguida e uma molécula é escolhida aleatoriamente; a molécula escolhida vai para o compartimento escolhido. Descreva o processo (distribuição inicial e as transições de estado). Compute a função de massa de X_3. Esse processo é irredutível? Qual o período dos estados?

Exercício 109. Defina Q, a matriz das transições no processo modificado dado acima. Mostre que $q_{i,j}^{(n)} \to \binom{N}{j}\frac{1}{2}^n$, quando $n \to \infty$.

Processos de ramificação: Originalmente, processos de ramificações foram considerados por Galton e Watson nos anos de 1870 quando estes procuravam uma explicação quantitativa para o fenômeno do desaparecimento de sobrenomes, mesmo em uma população crescente. Sob a suposição de que cada homem em uma dada família tem a probabilidade p_k de ter k filhos e que o sobrenome de família é passado aos filhos, então desejamos determinar a probabilidade que após n gerações um indivíduo não tenha descendentes homens.

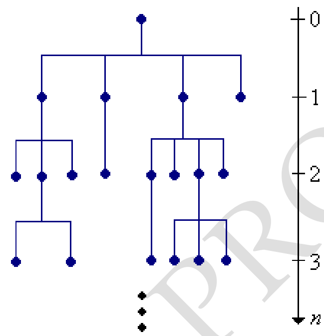
Suponhamos que no tempo $n = 0$ há um indivíduo que no tempo $n = 1$ gera um número aleatório de descendentes segundo a variável aleatória N com massa p_k em $k \in \mathbb{N}$. Suponhamos agora que esses também geram no tempo $n = 2$, cada um, independentemente, por um número aleatório de novos descendentes que possuem a mesma distribuição de N, e assim por diante.

Podemos construir o processo tomando para cada $n \in \mathbb{N}$ uma sequência de variáveis aleatórias $\{N_m^{(n)}\}_m$, o número de filhos do m-ésimo indivíduo da n-ésima geração, cada uma com a mesma distribuição de N, definindo $Z_0 = 1$ e definindo, indutivamente, para $n \geq 1$

$$Z_n = N_1^{(n-1)} + \cdots + N_{Z_{n-1}}^{(n-1)}$$

Z_n nos dá o tamanho da população na n-ésima geração. Em particular, $Z_1 = N_1^{(0)}$. Reforçamos que $\mathbb{P}(N_j^{(n)} = k) = p_k$ para todo $k \in \mathbb{N}$, com $p_0 \in (0, 1)$.

Por exemplo, numa realização representada pela figura



$Z_0 = 1$, $Z_1 = 4$, $Z_2 = 8$ e $Z_3 = 6$.

O número médio de filhos por indivíduo é $\mathbb{E}[Z_1] = \mathbb{E}[N] = \sum_{k\geqslant 0} kp_k$ e, intuitivamente, se esse número for pequeno, digamos menor que 1, então a população encolhe e a chance de sobreviver é pequena; se for grande, digamos maior que 1, então a chance de sobreviver é grande.

O processo estocástico $\{Z_n : n \geq 0\}$ é chamado de processo de ramificação. O conjunto de estados é $\mathbb{N}$ e as transições desse processo são

$$p_{i,j} = \begin{cases} 0 & \text{se } i = 0 \text{ e } j > 0; \\ 1 & \text{se } i = 0 \text{ e } j = 0; \\ \mathbb{P}(Z_{n+1} = j \mid Z_n = i) = \mathbb{P}\left(\sum_{r=1}^{i} N_r^{(n)} = j\right) & \text{se } i \geq 0 \text{ e } j \geq 0. \end{cases}$$

Notemos que se numa geração há 0 indivíduos, então não há indivíduos nas gerações seguintes

$$[Z_n = 0] \subset [Z_{n+1} = 0]. \tag{3.0.20}$$

Dizemos que 0 é um estado **absorvente** e representa a extinção do processo

$$\mathbb{P}(\text{extinção}) = \mathbb{P}\left(\bigcup_{n\geq 1}[Z_n = 0]\right).$$

O evento complementar é chamado de sobrevivência

$$\mathbb{P}(\text{sobrevivência}) = 1 - \mathbb{P}(\text{extinção}).$$

Notemos que $p_0 > 0$ implica em $\mathbb{P}(\text{extinção}) > 0$.

Seja

$$q_n := \mathbb{P}(Z_n = 0)$$

para todo $n \in \mathbb{N}$ e por equação (3.0.20) $q_n \leqslant q_{n+1}$, isto é, $(q_n)_{n\geqslant 0}$ é uma sequência não-decrescente e limitada

$$0 \leqslant q_1 \leqslant q_2 \leqslant \cdots \leqslant 1$$

portanto existe o limite $q = \lim_{n\to\infty} q_n$ e pela continuidade de $\mathbb{P}$

$$\mathbb{P}\left(\bigcup_{n\geq 1}[Z_n = 0]\right) = \mathbb{P}\left(\lim_{n\to\infty}[Z_n = 0]\right) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = q,$$

ou seja, $q = \mathbb{P}(\text{extinção})$ e satisfaz $q = \mathbb{P}_N(q)$: No nosso contexto equação (3.0.17) fica

$$P_{Z_{n+1}}(s) = P_{N_1^{(n)}+\cdots+N_{Z_n}^{(n)}}(s) = P_{Z_n}(P_N(s)).$$

para $n \geqslant 1$ e $s \in [0, 1)$.

Iterando a expressão acima, a função geradora de probabilidade de Z_n é obtida em função da função geradora de probabilidade de N, pois $P_{Z_0}(s) = s$ e

$$\begin{aligned} P_{Z_n}(P_N(s)) = \mathbb{P}_{Z_{n-1}}(P_N(P_N(s))) = \cdots &= P_N \circ P_N \circ \cdots \circ P_N(s) \\ &= P_N(P_{Z_n}(s)) \end{aligned}$$

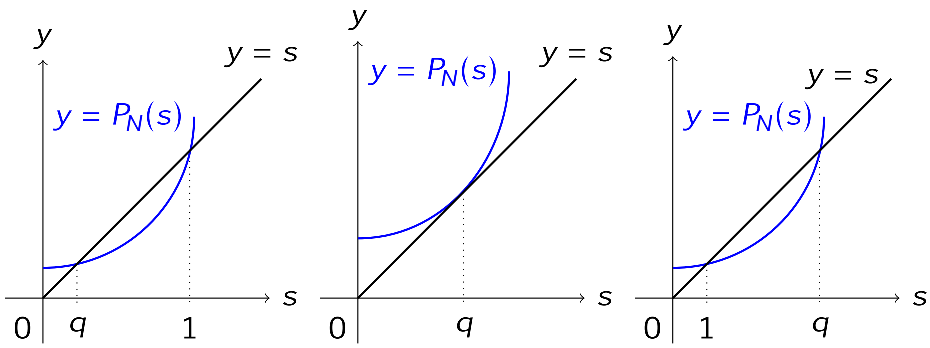
em que a composição na equação (3.0.21) tem n termos de P_N. Como $P_{Z_n}(0) = \mathbb{P}(Z_n = 0)$ da equação (3.0.22) temos

$$q_{n+1} = P_N(q_n)$$

e como P_N é contínua resulta que $q = P_N(q)$.

Analisemos as soluções da equação $s = P_N(s)$ para $s \in [0, 1]$. Primeiro, observamos que para $s \geqslant 0$ vale que as derivadas são positivas, $P_N'(s)$, $P_N''(s) \geqslant 0$, portanto a função é crescente e côncava pra cima, portanto, a reta $y = s$ encontra a curva $y = P_N(s)$ em no máximo dois pontos do plano cartesiano, digamos que de coordenadas iguais a s_0 e a s_1. Como $1 = P_N(1)$ as possibilidades para as raízes da equação $s = P_N(s)$ são

$$(a)\ s_0 < 1 \text{ e } s_1 = 1,\ (b)\ s_0 = s_1 = 1 \text{ e } (c)\ s_0 = 1 \text{ e } s_1 > 1.$$



Como probabilidade não pode ser maior que 1, nos casos (b) e (c) podemos concluir que $q = 1$, ou seja, a extinção é certa. Lembremos que $q_{n+1} = P_N(q_n)$, $q_0 = 0$ e que $q_n \uparrow q$. Então,

$$q_{n+1} = P_N(q_n) \leqslant \mathbb{P}_N(q) = q \text{ para } q = s_0 \text{ e } q = s_1$$

pois P_N é crescente, portanto $q_n \leqslant \min\{s_0, s_1\}$ para todo $n \geqslant 0$. Portanto, $(q_n)_{n\geqslant 0}$ converge para a menor solução positiva de $s = P_N(s)$.

Ademais, por equação (3.0.15) a inclinação da reta tangente em $(1,1)$ na curva $y = P_N(s)$ é $\mu = \mathbb{E}\left[()\, N\right)$, de modo que no caso (a) $\mu > 1$, no caso (b) $\mu = 1$ e no caso (c) $\mu < 1$. Em resumo,

3.0.29 Teorema. *Considere um processo de ramificação em que cada indivíduo dá origem a k outros indivíduos com probabilidade p_k, a função geradora $P(s) = \sum_{k\geqslant 0} p_k s^k$ e q a menor raiz positiva da equação $s = P(s)$. Se $0 < p_0 < 1$ e cada indivíduo gera em média μ outros, então $\mathbb{P}(\text{extinção}) = q$ e*

$$\begin{aligned} \mu > 1 &\implies q < 1 \\ \mu \leqslant 1 &\implies q = 1. \end{aligned}$$

□

Alternativamente, como consequência da equação (3.0.22) acima

$$P'_{Z_{n+1}}(s) = P'_N(P_{Z_n}(s))P'_{Z_n}(s)$$

pela regra da cadeia. Com isso e equação (3.0.15)

$$\mathbb{E}\left[Z_{n+1}\right] = P'_N(P_{Z_n}(1))P'_{Z_n}(1) = P'_N(1)P'_{Z_n}(1) = \mu\mathbb{E}\left[Z_n\right]$$

se $\mu := \mathbb{E}\left[Z_1\right]$. Portanto, para todo $n \in \mathbb{N}$

$$\mathbb{E}\left[Z_{n+1}\right] = \mu^{n+1}.$$

Assim, se $\mu < 1$ então, usando a desigualdade de Markov que diz que $\mathbb{P}(Z_n \geqslant 1) \leqslant \mathbb{E}\left[Z_n\right]$,

$$\lim_{n\to\infty} \mathbb{P}(Z_n = 0) = 1.$$

Notemos que $[Z_n = 0] \subset [Z_{n+1} = 0]$ e por continuidade de probabilidade

$$\mathbb{P}(\text{extinção}) = \mathbb{P}\left(\bigcup_{n\geq 1}[Z_n = 0]\right) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0)$$

donde concluimos que $\mathbb{P}(\text{extinção}) = 1$.

Exercício 110. Mostre que no processo de ramificação com $p_o = p_1 = 1/4$ e $p_2 = 1/2$ a probabilidade de extinção é $1/2$.

Exercício 111. Defina $p_k := bp^{k-1}$, com $b > 0$ e $0 < p < 1$, para $k \geqslant 1$ e $p_0 = 1 - \sum_{k\geqslant 1} p_k$.

1. Determine $P_N(s)$.

2. Determine $\mathbb{E}[N]$ em função de p e b.

3. Determine a probabilidade de extinção q em função de p e b.

O modelo básico de processos de ramificação tem muitas aplicações em problemas de crescimento populacional, e também no estudo de reações em cadeia na química, fissão nuclear e filas de espera.

Suponha que os indivíduos representam nêutrons que estão sujeitos a colisões e que em cada colisão nêutron se divide em k nêutrons. Se p é a probabilidade de colisão então o número de nêutrons é modelado como um processo de ramificação.

Uma aplicação interessante é a seguinte. Suponha que queremos analisar o comportamento de uma fila formada por clientes que chegam a um determinado caixa. Suponha que em cada instante de tempo existem duas possibilidades ou (i) chega um cliente, com probabilidade p, ou (ii) não chega cliente, com probabilidade $1 - p$. Ademais, se o caixa está livre, quando cliente chega, então esse cliente é atendido, por outro lado, se o caixa está ocupado então o cliente se junta no fim da fila. Os tempos de serviço são independentes e com a mesma distribuição. A geração 0 é do primeiro a chegar e a geração n $(n > 0)$ consiste dos clientes que chegam durante tempo de atendimento aos clientes da geração $n - 1$.

Ruína do jogador: Em um jogo, um jogador A começa com i reais e outro, B, com j reais, com $i + j = n$. Em cada rodada A vence com probabilidade p, de modo independente a cada rodada, e B com probabilidade $q = 1 - p$. Em cada rodada o vencedor leva 1 real (e o perdedor perde um real). O jogo continua até que a fortuna de A seja 0 ou n, nesse momento ele para de jogar.

Denotamos a quantia do jogador A no instante t pela variável X_t e temos $X_0 = i$ e as probabilidades de transição

$$\begin{aligned} p(0,0) &= p(n,n) = 1 \\ p(i,i+1) &= p \ \text{ e } \ p(i,i-1) = q, \end{aligned}$$

para $1 \leqslant i \leqslant n$ definem um processo estocásticos sobre o conjunto de estados $S = \{0, 1, 2, \dots, n\}$. As demis transições têm probabilidade 0.

Vamos determinar a probabilidade de A chegar ao estado n antes do estado 0, esse é um problema clássico da probabilidade conhecido como o *problema da ruína do jogador*.

Seja P_i a probabilidade de o jogador chegar antes ao estado n dado que $X_0 = i$. Então

$$P_i = pP_{i+1} + qP_{i-1}$$

para todo $i \in \{1, \dots, n\}$, que equivale a

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

para todo $i \in \{1, \dots, n\}$. Para facilitar usaremos $q = 1 - p$. Segue da igualdade acima que

$$\begin{aligned} P_{i+1} - P_i &= \frac{q}{p}(P_i - P_{i-1}) \\ &= \left(\frac{q}{p}\right)^2 (P_{i-1} - P_{i-2}) \\ &= \cdots \\ &= \left(\frac{q}{p}\right)^i (P_1 - P_0). \end{aligned}$$

donde concluímos, usando que $P_0 = 0$, que

$$P_{i+1} = P_i + \left(\frac{q}{p}\right)^i P_1 \tag{3.0.23}$$

para todo $i \in \{1, \dots, n\}$. Iterando

$$\begin{aligned} P_{i+1} &= P_i + \left(\frac{q}{p}\right)^i P_1 \\ &= P_{i-1} + \left(\left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^i\right) P_1 \\ &= P_{i-2} + \left(\left(\frac{q}{p}\right)^{i-2} + \left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^i\right) P_1 \\ &= \cdots \\ &= \left(1 + \frac{q}{p} + \cdots + \left(\frac{q}{p}\right)^{i-2} + \left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^i\right) P_1 \end{aligned}$$

e de $P_n = 1$ temos

$$1 = \left(1 + \frac{q}{p} + \cdots + \left(\frac{q}{p}\right)^{n-2} + \left(\frac{q}{p}\right)^{n-1}\right) P_1.$$

Se $p \neq q$ então a soma acima é de uma progressão geométrica cuja valor é $1 - (q/p)/1 - (q/p)^n$ logo

$$P_1 = \begin{cases} \dfrac{1}{n} & \text{se } p = q \\[2ex] \dfrac{1 - (q/p)}{1 - (q/p)^n} & \text{se } p \neq q. \end{cases}$$

e substituindo na equação (3.0.23)

$$P_i = \begin{cases} \dfrac{i}{n} & \text{se } p = q \\ \\ \dfrac{1-(q/p)^i}{1-(q/p)^n} & \text{se } p \neq q. \end{cases}$$

Notemos que se j é muito grande ($j \to \infty$) a probabilidade do jogador A ficar "infinitamente rico" é 0 se $p \leqslant 1/2$ pois $\frac{i}{n}$, e $\frac{1-(q/p)^i}{1-(q/p)^n}$ tendem a 0 ($n \to \infty$), ou seja, a ruína de B é praticamente impossível.

Por outro lado, se A joga melhor que B, ou seja, $p > 1/2$ então $(q/p)^n \to 0$ (quando $n \to \infty$) logo $P_i \sim 1-(q/p)^i > 0$ é a probabilidade de ruína de B (a de A é $\sim (q/p)^i$), ou seja, mesmo com capital menor, a habilidade melhor garante lhe garante menor chance de ruína.

Passeio aleatório: Sejam Y_n, para $n \in \mathbb{N}$, variável aleatória independentes com mesma distribuição $\mathbb{P}(Y_n = +1) = \mathbb{P}(Y_n = -1) = 1/2$ e $X_n := Y_1 + \cdots + Y_n$ que assumem valores no conjunto de estados $S = \mathbb{Z}$. No instante inicial $X_0 = 0$ e as probabilidades das transições não-nulas são $p_{i,i-1} = p_{i,i+1} = 1/2$ (*passeio aleatório simétrico*). O processo estocástico $(X_n)_{n\in\mathbb{N}}$ assim definido é chamado de *passeio aleatório em uma dimensão*.

O passeio aleatório é um processo **homogêneo** com respeito ao tempo, que significa que a probabilidade de ir do estado i ao estado j em n passos não depende do instante particular que o processo se encontra em i, ou seja,

$$\mathbb{P}(X_{m+n} = j \mid X_m = i) = \mathbb{P}(X_{k+n} = j \mid X_k = i)$$

para quaisquer i, j, k, m, n.

Um *retorno à origem* ocorre no instante n se $X_n = 0$. O evento $X_n = 0$ só ocorre para n par, digamos $n = 2m$ para $m \geqslant 1$. A quantidade de caminhos de tamanho $2m$ é 2^{2m}, pois cada passo tem duas possibilidades, esquerda ou direita. Desses, os que retornam para a origem são os que metade dos passos são pra direita e metade pra esquerda, portanto se escolhemos quais instantes são passos para a direita, nos restantes os passos são para a esquerda, ou seja, a quantidade de tais caminhos são $\binom{2m}{m}$ e

$$\mathbb{P}(X_{2m} = 0) = \binom{2m}{m} 2^{-2m}$$

e nos outros casos $\mathbb{P}(X_n = 0) = 0$.

Podemos escrever a função geradora de probabilidade do retorno $R(S)$ tomando $r_n := \mathbb{P}(X_n = 0)$

$$\begin{aligned} R(s) = \sum_{n\geqslant 0} r_n s^n &= \sum_{m\geqslant 0} \binom{2m}{m} 2^{-2m} s^{2m} \\ &= \sum_{m\geqslant 0} \binom{2m}{m} \left(\frac{s^2}{4}\right)^m \\ &= \frac{1}{\sqrt{1-s^2}} \end{aligned}$$

para todo $s \in (-1, 1)$.

Dentre os retornos à origem, destacamos especial atenção ao primeiro retorno, isto é, estamos interessados na probabilidade do passeio retornar a origem em algum momento futuro

$$T_0 := \min\{n > 0 : X_n = 0\}.$$

Denotemos por p_k a probabilidade de primeiro retorno a origem em k passos, $p_1 = \mathbb{P}(X_1 = 0)$ e para $k > 1$

$$p_k = \mathbb{P}(X_1 \neq 0,\ X_2 \neq 0, \ldots, X_{k-1} \neq 0, X_k = 0) = \mathbb{P}(T_0 = k).$$

As probabilidades de retorno r_n e p_k estão relacionadas da seguinte maneira, para todo $n \geqslant 1$

$$r_n = \sum_{k=1}^{n} p_k r_{n-k}.$$

De fato,

$$\begin{aligned} r_n &= \sum_{k=1}^{n} \mathbb{P}(X_n = 0 \mid X_1 \neq 0,\ X_2 \neq 0, \ldots, X_{k-1} \neq 0, X_k = 0) p_k \\ &= \sum_{k=1}^{n} \mathbb{P}(X_n = 0 \mid X_k = 0) p_k \end{aligned}$$

usando a homogeneidade $\mathbb{P}(X_n = 0 \mid X_k = 0) = \mathbb{P}(X_{n-k} = 0) = r_{n-k}$.

Vamos usar $R(s)$ para determinar a função geradora de probabilidade do *primeiro* retorno $P(S)$

$$P(s) = \sum_{k \geqslant 1} p_k s^k.$$

Agora, a função geradora da probabilidade de retorno pode ser reescrita como

$$\begin{aligned} R(s) &= 1 + \sum_{n \geqslant 1} r_n s^n \\ &= 1 + \sum_{n \geqslant 1} \left(\sum_{k=1}^{n} p_k r_{n-k} \right) s^n \\ &= 1 + \sum_{n \geqslant 2} \left(\sum_{k=1}^{n-1} p_k r_{n-k} \right) s^n + \sum_{n \geqslant 1} p_n r_0 s^n \\ &= 1 + \sum_{n \geqslant 2} \left(\sum_{k=1}^{n-1} p_k r_{n-k} \right) s^n + \sum_{n \geqslant 1} p_n s^n \\ &= 1 + P(s)(R(s) - 1) + P(s) \\ &= 1 + P(s)R(s) \end{aligned}$$

portanto,

$$P(s) = 1 - \sqrt{1 - s^2}.$$

O processo retorna a origem com probabilidade

$$\sum_{n \geqslant 0} p_n = P(1) = 1$$

além disso o tempo esperado para o retorno é

$$\sum_{n \geqslant 0} n p_n = P'(1) = \infty.$$

Se $\mathbb{P}(T_0 < \infty) = \sum_{n \geqslant 0} p_n = 1$ então dizemos que o passeio é **recorrente**. Se $\mathbb{P}(T_0 < \infty) < 1$ então dizemos que o passeio é **transiente**.

Exercício 112. Verifique que $P'(s) = 2sR(s)$ e então conclua que

$$p_{2m+2} = \frac{1}{m+1} r_{2m} = \frac{1}{m+1} r_{2m} = \frac{1}{m+1} \binom{2m}{m} 2^{-2m}.$$

Exercício 113. Deduza da aproximação de Stirling que

$$r_{2m} \approx \frac{1}{\sqrt{\pi m}}.$$

Exercício 114. Considere transições

$$p_{i,i-1} = p \text{ e } p_{i,i+1} = q$$

com $p + q = 1$. Mostre que também vale $P(s) = 1 + P(s)R(s)$ e ainda $R(s) = (1 - 4pqs^2)^{-1/2}$ e $P(s) = 1 - (1 - 4pqs^2)^{1/2}$. Conclua que se $p \neq 1/2$ então o passeio é transiente.

Exercícios.

1. Argumente que no modelo de Ehrenfest vale que para quaisquer $i, j \in S$, existe $t \in \mathbb{N}$ tal que $p_{i,j}^{(t)} > 0$. Nesse caso dizemos que o processo é **irredutível**.

2. Mostre que no modelo de Ehrenfest temos $p_{i,i}^{(t)} = 0$ se t é ímpar. O **período** de um estado é $d = \text{mdc}\{t : p_{i,i}^{(t)} = 0\}$. Mostre que os estados de S têm período 2.

3. Mostre que no modelo de Ehrenfest vale $\pi_i p_{i,j} = \pi_j p_{j,i}$ para a distribuição invariante π. Nesse caso chamamos o processo de **reversível**.

4. Mostre que no processo de ramificação temos

$$P''_{Z_{n+1}}(s) = P''_N(P_{Z_n}(s))(P'_{Z_n}(s))^2 + P'_N(P_{Z_n}(s))P''_{Z_n}(s)$$

   e use equação (3.0.16) para concluir que

$$\text{Var}[Z_{n+1}] = \text{Var}[Z_n]^2 \mu^{n-1}(1 + \mu + \mu^2 + \cdots + \mu^{n-1}).$$

5. Mostre que no processo de ramificação temos $\mathbb{P}(Z_n = 0 \mid Z_{n-1} = i) = \mathbb{P}(N = 0)^i$.

6. Um exemplo de processo de ramificação bem conhecido é devido a Lotka e estuda a evolução da descendência de uma família (homens americanos com dados baseados num censo de 1920). Lotka mostrou que $p_0 = 0{,}4825$ e $p_k = (0{,}2126)(0{,}5893)^{k-1}$ ($k \geqslant 1$) descreve o processo. Determine a probabilidade de extinção.

7. Descreva a matriz de transições do processod e ramificação para cada um dos seguintes distribuições para p_k:

   a) a distribuição de Poisson com parâmetro $m > 0$;

   b) a distribuição binomial com parâmetros $k \in \mathbb{N}$ e $p \in (0, 1)$;

   c) a distribuição geométrica em com parâmetro $1 - p \in (0, 1)$.

8. Um ingresso de teatro custa 5 reais e todas as $2n$ pessoas na fila pra compra o ingresso tem ou uma nota de 5 ou uma nota de 10 reais, metade das pessoas em cada caso. O vendedor de ingressos começa com o caixa vazio, i.e., não tem troco no primeiro momento. Com que probabilidade, numa fila aleatória, nenhum comprador fique sem troco? $1/(n+1)$

9. Considere

$$p_k(r) = \mathbb{P}(X_1 \neq r,\ X_2 \neq r, \ldots, X_{k-1} \neq r, X_k = r)$$

   a probabilidade da primeira visita ao estado r em k passo, e sua função geradora

$$P_r(s) = \sum_{k \geqslant 0} p_k(r) s^k.$$

   Prove que $P_1(s) = (1/s)\left(1 - \sqrt{1 - s^2}\right)$ e que $P_r(s) = (P_1(s))^r$.

10. Sejam $e_1 = (1, 0)$, $-e_1 = (-1, 0)$, $e_2 = (0, 1)$ e $-e_2 = (0, -1)$. Consideremos o passeio aleatório em duas dimensões, em que o conjunto de estados é $S = \mathbb{Z} \times \mathbb{Z}$, o estado inicial é $X_0 = (0, 0)$ e $X_{i+1} = X_i \pm e_j$ sendo que cada um dos quatro possíveis casos para $\pm e_j$ têm a mesma probabilidade, a saber 1/4. Essa descrição corresponde ao passeio aleatório em duas dimensões. Prove que o retorno a origem é certo, ou seja, o passeio aleatório é recorrente.

11. Formalize o passeio aleatório em três dimensões e prove que é transiente.

§8 Cadeias de Markov a tempo discreto. No que segue fixamos $T = \mathbb{N}$ e a menos que seja dito o contrário $S \subset \mathbb{N}$. Nessas condições um processo estocástico é uma cadeia de Markov se vale a condição de Markov de que o estado futuro depende somente do estado atual e é independente dos estados passados

$$\mathbb{P}(X_{n+1} = j \mid X_n = i,\ X_{n-1} = s_{n-1}, \cdots,\ X_0 = s_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \tag{3.0.24}$$

e ela é dita *homogênea* se essa probabilidade não depende do tempo n, que é o caso que estudaremos a seguir. Como todas as cadeias desse capítulo são homogêneas nós omitiremos esse adjetivo daqui em diante.

Elementos básicos: Uma sequência $(\rho_i)_{i\in S}$ com $\rho_i \geqslant 0$, de modo que $\sum_{i\in S} \rho_i = 1$ é chamada *vetor de probabilidades* ou *distribuição da variável aleatória* X no caso em que $\rho_i = \mathbb{P}(X = i)$. Chamamos X_0 de *estado inicial* e sua função de massa de *distribuição inicial*.

Uma cadeia de Markov é caracterizada por

1. conjuntos de estados S,
2. distribuição inicial $(\rho_i)_{i\in S}$,
3. probabilidades de transição do estado i para o estado j, para todo $(i,j) \in S^2$,

$$p(i,j) \geqslant 0 \quad \text{com} \quad \sum_{j\in S} p(i,j) = 1$$

de modo que um o processo estocástico $\{X_t\}_{t\in\mathbb{N}}$ com valores em S é uma **cadeia de Markov homogênea** com distribuição inicial ρ e probabilidade transição p se a distribuição conjunta de $X_0, \ldots, X_n$, para todo $n \in \mathbb{N}$, é

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, \cdots, X_n = s_n) = \rho_{s_0} \prod_{i=0}^{n-1} p(s_i, s_{i+1}) \tag{3.0.25}$$

para toda escolha de $s_0, s_1, \ldots, s_n \in S$.

Da distribuição conjunta temos

$$\begin{aligned}
\mathbb{P}\left(X_{n+1} = s_{n+1} \,\middle|\, \bigcap_{\ell=0}^{n} [X_\ell = s_\ell] \right) &= \frac{\mathbb{P}\left(\bigcap_{\ell=0}^{n+1} [X_\ell = s_\ell]\right)}{\mathbb{P}\left(\bigcap_{\ell=0}^{n} [X_\ell = s_\ell]\right)} \\
&= \frac{\rho_{s_0} \prod_{i=0}^{n+1} p(s_i, s_{i+1})}{\rho_{s_0} \prod_{i=0}^{n} p(s_i, s_{i+1})} \\
&= p(s_n, s_{n+1})
\end{aligned}$$

e

$$
\begin{aligned}
\mathbb{P}\left(X_{n+1} = s_{n+1} \,\middle|\, X_n = s_n\right) &= \frac{\mathbb{P}\left(X_{n+1} = s_{n+1},\ X_n = s_n\right)}{\mathbb{P}(X_n = s_n)} \\
&= \frac{\sum_{s_0,\dots,s_{n-1}} \rho_{s_0} \prod_{i=0}^{n} p(s_i, s_{i+1})}{\sum_{s_0,\dots,s_{n-1}} \rho_{s_0} \prod_{i=0}^{n-1} p(s_i, s_{i+1})} \\
&= p(s_n, s_{n+1})
\end{aligned}
$$

que é a condição de Markov equação (3.0.24). Da condição de Markov, eq. equação (3.0.24), deduzimos a distribuição conjunta, eq. equação (3.0.25), pelo exercício 2 (pág. 121). Ademais, fixado n, temos a partir dos itens 2 e 3 acima que

$$
\sum_{s_0\in S}\sum_{s_1\in S} \cdots \sum_{s_{n-1}\in S}\sum_{s_n\in S} \rho_{s_0} \prod_{i=0}^{n-1} p(s_i, s_{i+1}) =
$$
$$
\sum_{s_0\in S}\sum_{s_1\in S} \cdots \sum_{s_{n-1}\in S} \rho_{s_0} \prod_{i=0}^{n-2} p(s_i, s_{i+1}) \sum_{s_n\in S} p(s_{n-1}, s_n) =
$$
$$
\sum_{s_0\in S}\sum_{s_1\in S} \cdots \sum_{s_{n-1}\in S} \rho_{s_0} \prod_{i=0}^{n-2} p(s_i, s_{i+1}) = \cdots = \sum_{s_0\in S}\sum_{s_1\in S} \rho_{s_0} p(s_0, s_1) = \sum_{s_0\in S} \rho_{s_0} = 1
$$

portanto 3.0.25 é uma atribuição de probabilidades válida no espaço das realizações das n variáveis. A mesma dedução vale para qualquer valor de n portanto a definição da distribuição é consistente (o Teorema de Kolmogorov garante a existência das variáveis).

Exercício 115 (**Propriedade de Markov**). Prove que se $\{X_t\}_{t\in\mathbb{N}}$ é uma cadeia de Markov com respeito ao conjunto de estados S, distribuição inicial $(\rho_i)_{i\in S}$ e probabilidades de transição $\{p(i,j)\}_{i,j\in S}$, então condicionado a $[X_m = i]$ o processo $\{X_{t+m}\}_{t\in\mathbb{N}}$ é uma cadeia de Markov com respeito a $(\delta_i(j))_{i\in S}$ e $\{p(i,j)\}_{i,j\in S}$ e independente de $X_0,\dots,X_m$. Acima, $\delta_i(j) = 1$ se e só se $i = j$.

Identidade de Chapman–Kolmogorov

Convencionamos que $p_{i,j}^{(0)} = \delta_i(j)$ e para $n > 0$

$$
p_{i,j}^{(n)} := \mathbb{P}(X_n = j \mid X_0 = i)
$$

é a transição do estado i para o estado j em n passos.

Notemos que $\mathbb{P}(A,\ B \mid C) = \mathbb{P}(A \mid B,\ C)\mathbb{P}(B \mid C)$ e usando esse fato podemos deduzir

$$
\begin{aligned}
p_{i,j}^{(k+t)} &= \mathbb{P}(X_{k+t} = j \mid X_0 = i) \\
&= \sum_{s\in S} \mathbb{P}(X_{k+t} = j,\ X_k = s \mid X_0 = i) \\
&= \sum_{s\in S} \mathbb{P}(X_{k+t} = j \mid X_k = s,\ X_0 = i)\mathbb{P}(X_k = s \mid X_0 = i) \\
&= \sum_{s\in S} p_{i,s}^{(k)} p_{s,j}^{(t)}
\end{aligned}
$$

portanto, vale a seguinte identidade que é bastante útil.

3.0.30 Lema (Identidade de Chapman–Kolmogorov). *Se $p_{i,j}$ são as probabilidades de transição de uma cadeia de Markov então*

$$p_{i,j}^{(k+t)} = \sum_{s\in S} p_{i,s}^{(k)} p_{s,j}^{(t)} \tag{3.0.26}$$

para quaisquer $i,\ j \in S$.

Exemplos: Convencionamos que quando é dada uma definição de uma cadeia de Markov, as probabilidades das transições que não são dadas explicitamente são iguais a zero. Ademais, usualmente não damos a distribuição inicial ρ_s porque ou esse parâmetro não é relevante no interesse do momento ou é trivial (i.e., $\mathbb{P}(X_0 = s) = 1$ para algum s).

Exemplo 131. Numa pilha com n cartas de baralho distintas numeradas de 1 a n, definimos uma cadeia de Markov tomando um estado para cada permutação π, em que $\pi(i)$ é a posição da i-ésima carta; X_0 é a permutação identidade com probabilidade 1. Uma transição entre estados é obtida retirando a carta topo e colocando-á numa posição arbitrária escolhida uniformemente entre as cartas restantes na pilha, logo

$$p(\pi, \sigma) = \mathbb{P}(X_{t+1} = \sigma \mid X_t = \pi) = \frac{1}{n}$$

para todo $t \in \mathbb{N}$ e com a permutação σ obtida a partir da permutação π pelo processo descrito acima (nos casos em que σ não pode ser obtida pelo processo descrito vale $p(\pi, \sigma) = 0$, como convencionamos). Um problema interessante é estimar o valor n, caso exista, para o qual a distribuição de X_n seja aproximadamente uniforme.

Exemplo 132. Sejam $Y_i \sim \text{Bernoulli}(p)$ variáveis aleatórias independentes, $X_0 = 0$ e para $i > 0$, tomemos $X_i = Y_i + Y_{i-1}$. Então

$$\mathbb{P}(X_{i+1} = 2 \mid X_i = 1,\ X_{i-1} = 2) = 0,$$

enquanto que

$$\mathbb{P}(X_{i+1} = 2 \mid X_i = 1,\ X_{i-1} = 0) = p,$$

logo $\{X_i\}$ não é uma cadeia de Markov por não satisfazer equação (3.0.24).

Exemplo 133. Definimos uma cadeia de Markov se tomamos o conjunto de estados como sendo os inteiros positivos, $S = \mathbb{N} \setminus \{0\}$, o estado inicial $X_0 = 1$ com probabilidade 1 e as transições probabilidades

$$p(i,1) = \frac{1}{i+1}$$
$$p(i,i+1) = \frac{i}{i+1}$$

para todo inteiro $i \geqslant 1$.

Exemplo 134 (passeio aleatório em uma dimensão). Tomemos o conjunto de estados como os inteiros, $S = \mathbb{Z}$, o estado inicial $X_0 = 0$ com probabilidade 1 e as transições de estados têm as probabilidades

$$p(i,i-1) = 1-p$$
$$p(i,i+1) = p$$

para todo inteiro i e algum $p \in (0,1)$. Essa cadeia é um *passeio aleatório* pelos inteiros.

Exemplo 135. Seja G um grafo finito. Definimos uma cadeia de Markov tomando S como o conjunto de vértices do grafo e as transições são definidas de modo que se $X_t = v$ então X_{t+1} é qualquer vértice adjacente a v com probabilidade uniforme. Esse tipo de cadeia de Markov é conhecida como passeio aleatório num grafo.

Exemplo 136 (**Cadeia de Markov não-homogênea**). Suponha $Y_1, Y_3, \dots, Y_{2i+1}, \dots$ variável aleatória independentes e identicamente distribuídas com

$$\mathbb{P}(Y_{2k+1} = 1) = \mathbb{P}(Y_{2k+1} = -1) = \frac{1}{2}$$

para todo $k \geqslant 0$. Agora, tomemos $Y_{2k} = Y_{2k+1}Y_{2k-1}$ para todo $k > 0$. A sequência de variável aleatória $\{Y_i\}_{i \geqslant 1}$ não define uma cadeia de Markov pois

$$\mathbb{P}(Y_{2k+1} = 1 \mid Y_{2k} = -1) = \frac{1}{2}$$

enquanto que

$$\mathbb{P}(Y_{2k+1} = 1 \mid Y_{2k} = -1,\ Y_{2k-1} = -1) = 1.$$

No entanto, $Z_n = (Y_n, Y_{n+1})$ é uma cadeia de Markov não homogênea pois, por exemplo

$$\mathbb{P}(Z_{n+1} = (1,1) \mid Z_n = (1,1)) = \begin{cases} 1/2 & \text{se } n \text{ par} \\ 1 & \text{se } n \text{ ímpar.} \end{cases}$$

Exemplo 137 (**Cadeia de Markov não-homogênea**). Consideremos uma caixa com v bolas vermelhas e a bolas azuis e o processo de retirar aleatoriamente uma bola da caixa sem reposição. Seja X_i o número de bolas vermelhas na caixa na i-ésima rodada. A probabliade de transição não depende só do número de bolas vermelhas (estado), mas também depende do momento i.

Exemplo 138 (**ruína do jogador**). Em um jogo, o jogador ganha 1 real com probabilidade p, ou perde 1 real com probabilidade $1-p$, de modo independente a cada rodada de até que sua fortuna seja 0 ou n, nesse momento ele pára de jogar. Denotamos a quantia do jogador no instante t pela variável X_t e temos que

$$
\begin{aligned}
&p(0,0) = p(n,n) = 1\\
&p(i,i+1) = p \ \text{ e } \ p(i,i-1) = 1-p,
\end{aligned}
$$

para $1 \leqslant i \leqslant n$, são as transições de estado de uma cadeia de Markov.

Exemplo 139. Consideremos a cadeia de Markov sobre $S = \{0,1,\dots,n\}$ com estado inicial $X_0 = 0$ e as transições de estados têm as probabilidades

$$
\begin{aligned}
&p(0,1) = 1 = p(n,n)\\
&p(j,j+1) = 1/2\\
&p(j,j-1) = 1/2
\end{aligned}
$$

para todo inteiro $j \geqslant 1$. Vamos estimar o número esperado de passos até a cadeia chegar ao estado n. Seja Y_j o número de passos para atingir n a partir de j. Se o estado atual é $j > 0$ e o próximo $j-1$ então $Y_j = 1 + Y_{j-1}$, senão $Y_j = 1 + Y_{j+1}$, portanto, por equação (3.0.19)

$$
\mathbb{E}\left[Y_j\right] = \mathbb{E}\left[Y_{j-1}+1\right]\frac{1}{2} + \mathbb{E}\left[Y_{j+1}+1\right]\frac{1}{2} = \frac{1}{2}\left(\mathbb{E}\left[Y_{j-1}\right] + \mathbb{E}\left[Y_{j+1}\right]\right) + 1
$$

e temos

$$
\begin{cases}
2\mathbb{E}\left[Y_j\right] = \mathbb{E}\left[Y_{j-1}\right] + \mathbb{E}\left[Y_{j+1}\right] + 2, & \text{se } 0 < j < n\\
\mathbb{E}\left[Y_0\right] = \mathbb{E}\left[Y_1\right] + 1,\\
\mathbb{E}\left[Y_n\right] = 0.
\end{cases}
$$

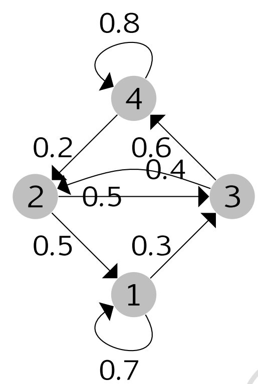
cuja solução é $\mathbb{E}\left[Y_j\right] = n^2 - j^2$ (verifique). Finalmente, $\mathbb{E}\left[Y_0\right] = n^2$. Pela desigualdade de Markov, equação equação (3.0.18), a probabilidade da cadeia não chegar no estado n em $2n^2$ passos é

$$
\mathbb{P}(Y_0 \geqslant 2n^2) \leqslant \frac{\mathbb{E}\left[Y_0\right]}{2n^2} = \frac{1}{2} \tag{3.0.27}
$$

e, analogamente, tende a zero a probabilidade da cadeia não chegar no estado n em $n^2\omega(n)$ passos qualquer que seja $\omega(n)$ que tenda ao infinito com n, por exemplo, $\omega(n) = \log\log n$.

Exemplo 140. Suponha que a ocorrência de chuva num dia é determinada pela condição do clima nos dois dias anteriores do seguinte modo: amanhã chove com probabilidade $0,7$ se chove nos dois dias anteriores; amanhã chove com probabilidade 0,5 se chove hoje mas não choveu ontem; amanhã chove com probabilidade 0,4 se não chove hoje mas choveu ontem; amanhã chove com probabilidade

0,2 se não chove nos dois dias anteriores. Identificamos cada uma das quatro situações acima com seguintes estados **1** se choveu hoje e ontem; **2** se choveu hoje mas não choveu ontem; **3** se não choveu hoje mas choveu ontem; e **4** se não choveu nem hoje e nem ontem. O estado no dia X_{n+1} depende da condição nos dias anteriores; assim a transição de $X_n = 1$ para $X_{n+1} = 3$, por exemplo, ocorre quando não chove amanhã $(n+1)$ mas choveu hoje, dado que choveu hoje (n) e ontem $(n-1)$, nessa configuração a probabilidade de não-chuva amanhã se choveu hoje e ontem é $1 - 0{,}7 = 0{,}3$. O seguinte diagrama ilustra as transições de estados com suas probabilidades.



Exercício 116. Um dado é lançado repetidamente. Quais das seguintes sequências de variáveis formam um cadeia de Markov?

1. X_n é o maior resultado até a n-ésima rodada;
2. Y_n é a quantidade de 6 em n rodadas;
3. no instante r, Z_r é o tempo desde o 6 mais recente;
4. no instante r, W_r é o tempo até o próximo 6.

Exercício 117 (**Equação de Wald e ruína do jogador**). Sejam $Z_1, Z_2, \ldots$ variável aleatória independentes, identicamente distribuídas, e com esperança $\mathbb{E}[Z]$ finita. Chamamos a variável aleatória N de *tempo de parada* para a sequência $\{Z_i\}_{i\geqslant 1}$ se o evento $[N = n]$ é independente de Z_t para $t > n$, para todo n. Prove que se N tem esperança $\mathbb{E}[N]$ finita, então

$$\mathbb{E}\left[\sum_{i=1}^{N} Z_i\right] = \mathbb{E}[N]\,\mathbb{E}[Z]. \tag{3.0.28}$$

No exemplo 138, seja i o estado inicial D_i o tempo de duração do jogo, ou seja, o tempo esperado até a cadeia atingir o estado 0 ou o estado n. Assuma que as esperanças sejam finitas e use a equação equação (3.0.28) para determinar D_i.

Classificação de estados. Os estados de uma cadeia de Markov são classificados em dois tipos fundamentais. Um estado é *recorrente* (ou, persistente) se, uma vez tendo estado nele, é certo que a cadeia eventualmente volta a ele, caso contrário o estado é dito *transiente* (ou, transitório). Quando o estado é recorrente, a cadeia visita-o infinitas vezes pois uma vez que a cadeia esteja em tal estado é como um reinicio, ela voltará a ele com probabilidade 1, e no caso transiente o estado é visitado um número finito de vezes. Notemos que isso leva a conclusão de que *se* $|S|$ *é finito então pelo menos um estado é recorrente*.

Transiência e recorrência: Para $n \in \mathbb{N}$, seja $f_{i,j}^{(n)}$ a probabilidade do evento "primeira passagem pelo estado j a partir do estado i em n passos"

$$X_0 = i,\ X_1 \neq j,\ X_2 \neq j,\ \cdots,\ X_{n-1} \neq j,\ X_n = j$$

com $f_{i,j}^{(0)} = 0$, assim $f_{i,i}^{(n)}$ é a *probabilidade do primeiro retorno a* i *em* n *passos* e definimos

$$f_{i,i} := \sum_{n>0} f_{i,i}^{(n)}$$

que é a probabilidade de eventualmente retornar ao estado i. Com essa definição, um estado i é **recorrente** se e somente se $f_{i,i} = 1$, caso contrário $f_{i,i} < 1$ e chamamos i de **transiente**.

Distribuição do primeiro retorno.: No caso em que i é recorrente $f_{i,i} = 1$ e temos uma função de massa probabilidade $\{f_{i,i}^{(n)}\}_{n>0}$ para o tempo de retorno ao estado i e média dessa variável aleatória, chamado de *tempo médio de recorrência* (ou, tempo médio de retorno) para i, é

$$\mu_i := \sum_{n>0} n f_{i,i}^{(n)}.$$

No caso de i transiente convencionamos $\mu_i = \infty$.

Definimos a variável aleatória

$$T_i := \min\{n \geqslant 1 : X_n = i\}$$

e fica convencionado que $T_i = \infty$ se a cadeia nunca passa por i. Claramente, $\mathbb{P}(T_i = \infty \mid X_0 = i) > 0$ se, e só se, i é transiente, ademais

$$\mu_i = \mathbb{E}\,[T_i \mid X_0] = i).$$

Ressalvamos que $\mu_i = \infty$ pode ocorrer para i recorrente. Dado que o retorno ao estado inicial i é certo, é natural perguntar se o tempo médio de retorno é finito. No caso $\mu_i = \infty$ o estado i é dito **recorrente nulo**, no caso de média finita, $\mu_i < \infty$, o estado i é dito **recorrente positivo** (ou **não-nulo**).

Vimos que no passeio aleatório simétrico em uma dimensão o estado 0 é recorrente nulo.

Exemplo 141. No exemplo 133, $S = \mathbb{N} \setminus \{0\}$, $X_0 = 1$, $p(i,1) = \frac{1}{i+1}$ e $p(i, i+1) = \frac{i}{i+1}$ para todo inteiro $i \geqslant 1$. A probabilidade de não retornar ao estado 1 nos $n-1$ primeiros passos é

$$\prod_{j=1}^{n-1} \frac{j}{j+1} = \frac{1}{n}$$

portanto, por continuidade de $\mathbb{P}$, a probabilidade de nunca voltar ao estado 1 é 0 e a probabilidade de voltar ao 1 no n-ésimo passo é

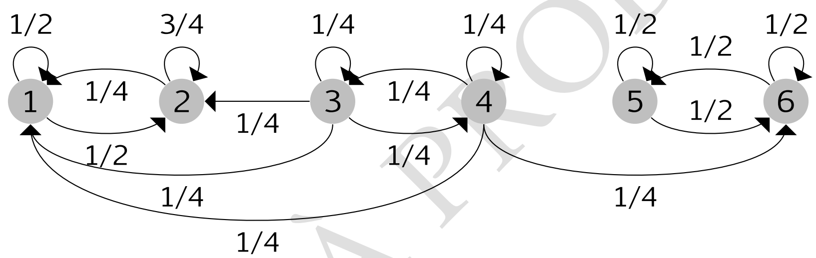
$$f_{1,1}^{(n)} = \frac{1}{n}\frac{1}{n+1}$$

e, a partir disso, o tempo médio de recorrência do estado 1 é

$$\mu_1 = \sum_{n>0} n f_{1,1}^{(n)} = \sum_{n>0} \frac{1}{n+1} = \infty.$$

Em resumo, o estado 1 é recorrente nulo.

Exemplo 142. No caso da cadeia representada no diagrama abaixo



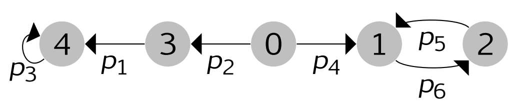
$$f_{1,1}^{(1)} = \frac{1}{2} \qquad \text{e} \qquad f_{1,1}^{(n)} = \frac{1}{2}\left(\frac{3}{4}\right)^{n-2}\frac{1}{4}\ (n \geqslant 2)$$

portanto

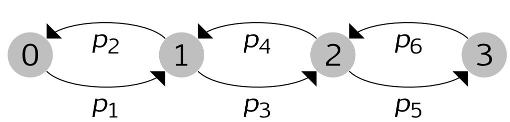
$$\mu_1 = \frac{1}{2} + \frac{1}{8}\sum_{n\geqslant 2} n\left(\frac{3}{4}\right)^{n-2} = \frac{1}{2} + \frac{1}{6}\sum_{n\geqslant 2} n\left(\frac{3}{4}\right)^{n-1} = \frac{1}{2} + \frac{1}{6}15 = 3$$

Os estados 0 e n da cadeia do exemplo 138, quando são atingidos a cadeia não muda mais de estado; nesses casos chamamos tais estados de **absorvente**, que é um estado recorrente positivo.

Exemplo 143. Na cadeia representada pelo esquema abaixo o estado 4 é absorvente, os estados 0 e 3 são transiente e os estados 1 e 2 são recorrentes



Período: Numa cadeia de Markov que pode ser representada como



se $X_0 = 0$ então $p_{0,0}^{(n)} > 0$ só pode ocorrer na cadeia em instantes n da forma $n = 2k$, $n = 4k$, $n = 6k$. Para o estado 1, as probabilidades de retorno positivas só ocorrem para $n = 2k$, $n = 4k$. Para o estado 2 ocorre o mesmo fenômeno que no estado 1 e o caso do estado 3 é semelhante ao caso do estado 0.

O período de um estado i numa cadeia de Markov é

$$\tau(i) = \mathrm{mdc}\left\{n : p_{i,i}^{(n)} > 0\right\}$$

Caso $\tau(i) > 1$ dizemos que i é **periódico**, nesse caso $p_{i,i}^{(n)} = 0$ a menos que n seja múltiplo de τ_i, e caso $\tau(i) = 1$ dizemos que i é **aperiódico**. No exemplo acima todos os estados são periódicos de período $t = 2$.

3.0.31 Observação. *Suponha que, como no exemplo acima, numa cadeia de Markov todos os estados tenham o mesmo período $t > 1$ e que para cada par de estados (i, j) existe um instante n para o qual $p_{i,j}^{(n)} > 0$ (de fato, essa 2 hipóteses são, em certo sentido, redundantes, veja o lema 3.0.33).*

Para todo estado k da cadeia devem existir instantes m e n tais que

$$p_{0,k}^{(n)},\ p_{k,0}^{(m)} > 0$$

e como $p_{0,0}^{(n+m)} \geqslant p_{0,k}^{(n)} p_{k,0}^{(m)} > 0$ devemos ter que t divide $n + m$. Fixando m concluímos que todo n para o qual $p_{0,k}^{(n)} > 0$ é da forma $a + vt$ para $0 \leqslant a < t$ inteiro. Assim, podemos particionar S em classes $S_0, S_1, \dots, S_{t-1}$ (no exemplo acima $S_0 = \{0, 2\}$ e $S_1 = \{1, 3\}$) para os valores de a acima de modo que se $k \in S_a$ então $p_{0,k}^{(n)} = 0$ a menos que $n = a + vt$. *Agora consideramos os estados na ordem $S_0, \dots, S_{t-1}, S_0$, ciclicamente, e um passo da cadeia sai de um estado para um estado na classe a direita, e a cada t passos a cadeia está de volta à mesma classe.*

Um estado ao qual a cadeia não retorna e considerado aperiódico. Um estado recorrente não-nulo e aperiódico é dito **ergódico**.

Exercício 118. Prove que

$$p_{j,k}^{(n)} = \sum_{t=1}^{n} f_{j,k}^{(t)} p_{k,k}^{(n-t)}.$$

Para tal, deduza da propriedade de Markov que

$$\mathbb{P}(X_0 = j,\ X_1 \neq k,\ X_2 \neq k,\ \cdots,\ X_{t-1} \neq k,\ X_t = k,\ X_n = k) =$$

$$\mathbb{P}(X_0 = j,\ X_1 \neq k,\ X_2 \neq k,\ \cdots,\ X_{n-1} \neq k,\ X_t = k \mid X_0 = j)\mathbb{P}(X_n = k \mid X_t = k).$$

Essa classificação de estados pode ser caracterizada pelas probabilidades de transições da maneira que deduziremos a seguir.

Seja $I_{[X_n=i]}$ a variável aleatória indicadora do evento $[X_n = i]$. Então a soma para todo $n \geqslant 0$ das variáveis $I_{[X_n=i]}$ é o número de vezes que a cadeia visita o estado i,

$$V_i := \sum_{n\geqslant 0} I_{[X_n=i]}$$

que afirmamos acima ser finito se e só se n é transiente. De fato, pelo exercício 60,

$$\mathbb{E}\left[V_i \mid X_0 = i\right] = \sum_{k\geqslant 1} \mathbb{P}(V_i \geqslant k \mid X_0 = i)$$

e, saindo de i, visitar j pelo menos k vezes equivale a visitar j e em seguida visitar j pelo menos $k-1$ vezes, o que ocorre com probabilidade $f_{i,j}(f_{j,j})^{k-1}$ logo, fazendo $j = i$ temos

$$\mathbb{E}\left[V_i \mid X_0 = i\right] = \sum_{k\geqslant 1} (f_{i,i})^k = \begin{cases} \frac{f_{i,i}}{1-f_{i,i}} & \text{se } f_{i,i} < 1 \\ \infty & \text{se } f_{i,i} = 1 \end{cases}$$

Pela linearidade da esperança,

$$\mathbb{E}\left[V_i \mid X_0 = i\right] = \sum_{n\geqslant 0} \mathbb{E}\left[I_{[X_n=i]} \,\Big|\, X_0 = i\right] = \sum_{n\geqslant 0} \mathbb{P}\left(X_n = i \mid X_0 = i\right) = \sum_{n\geqslant 0} p_{i,i}^{(n)}$$

e as caracterizações seguem das considerações acima a respeito da esperança condicionada: i é transiente se e só se $\sum_{n\geqslant 0} p_{i,i}^{(n)} < \infty$; i é recorrente se e só se $\sum_{n\geqslant 0} p_{i,i}^{(n)} = \infty$. Ademais, o exercício 118 acima implica que no caso i transiente $\sum_{n\geqslant 0} p_{j,i}^{(n)} < \infty$ para todo j, portanto $\lim_{n\to\infty} p_{j,i}^{(n)} = 0$ para todo j. No caso i recorrente, um resultado que veremos (**xxx**) implica em $\sum_{n\geqslant 0} p_{i,i}^{(n)} = \infty$ e $\lim_{n\to\infty} p_{i,i}^{(n)} = 0$ no caso recorrente nulo e aperiódico, e também pelo exercício 118, $\lim_{n\to\infty} p_{j,i}^{(n)} = 0$, para todo j.

3.0.32 Lema. *O estado i é*

- *transiente se e só se* $\displaystyle\sum_{n\geqslant 0} p_{i,i}^{(n)} < \infty$. *Nesse caso,* $\displaystyle\sum_{n\geqslant 0} p_{j,i}^{(n)} < \infty$ *para todo j, portanto,* $p_{j,i}^{(n)} \to 0$ *quando $n \to \infty$, para todo j.*
- *recorrente se e só se* $\displaystyle\sum_{n\geqslant 0} p_{i,i}^{(n)} = \infty$. *Nesse caso* $\displaystyle\sum_{n\geqslant 0} p_{j,i}^{(n)} = \infty$ *para todo j tal que $f_{j,i} > 0$. Ainda se i é recorrente nulo, então* $p_{j,i}^{(n)} \to 0$, *quando $n \to \infty$, para todo j.*

Classificação das cadeias. A condição $p_{i,j}^{(n)} > 0$ para algum n equivale a $f_{i,j} > 0$ e significa que a partir do estado i a cadeia eventualmente atinge o estado j; nesse caso dizemos que j é *acessível* a partir de i e escrevemos $i \to j$. Quando j não é acessível a partir de i a probabilidade da cadeia chegar em j saindo de i é

$$\mathbb{P}\left(\bigcup_{n\geqslant 0} [X_n = j] \mid X_0 = i\right) \leqslant \sum_{n\geqslant 0} \mathbb{P}\left(X_n = j \mid X_0 = i\right) = 0.$$

Escrevemos $i \leftrightarrow j$ se $i \to j$ e $j \to i$ e dizemos que os estados i e j se *comunicam*. Deixamos para o leitor a verificação de que $\leftrightarrow$ define uma relação de equivalência sobre S, i.e.,

1. $i \leftrightarrow i$,
2. se $i \leftrightarrow j$, então $j \leftrightarrow i$,
3. se $i \leftrightarrow k$ e $k \leftrightarrow j$ então $i \leftrightarrow j$,

portanto, ela particiona S em classes de equivalência, que chamaremos *classe de comunicação*. A importância dessa partição é pelo fato enunciado no exercício a seguir.

3.0.33 Lema. *Recorrência, transiência e período são propriedades de classe de equivalência de $\leftrightarrow$, isto é, dois estados que se comunicam tem a mesma classificação.*

Demonstração. Sejam i e j estados que se comunicam e sejam n e m inteiros positivos tais que $p_{i,j}^{(n)} > 0$ e $p_{j,i}^{(m)} > 0$. Para todo t

$$p_{i,i}^{(t+n+m)} \geqslant p_{i,j}^{(n)} p_{j,j}^{(t)} p_{j,i}^{(m)} \tag{3.0.29}$$

portanto, $p_{i,i}^{(t+n+m)} > 0$ sempre que $p_{j,j}^{(t)} > 0$.

Se fizermos $t = 0$ na equação equação (3.0.29), resulta que $p_{i,i}^{(n+m)} > 0$ portanto $\tau(i)$ divide $n + m$. O lado esquerdo da equação equação (3.0.29) é nulo a menos nos períodos múltiplos de $\tau(i)$, portanto $p_{j,j}^{(t)} > 0$ só quando t é múltiplo de $\tau(i)$, portanto $\tau(j)$ divide $\tau(i)$. Trocando os papéis de i e j concluiremos que $\tau(i)$ divide $\tau(j)$. Portanto, os períodos são iguais.

Se somamos os dois lados equação (3.0.29) sobre todo natural t temos que se $\sum_t p_{j,j}^{(t)}$ não converge, então também não converge o lado esquerdo, portanto se j é recorrente então também será o estado i. Trocando os papéis de i e j concluiremos que se i é recorrente então j é recorrente. Ademais, se j é transiente então também será i e vice-versa. □

Um conjunto $C \subset S$ é **irredutível** se $i \leftrightarrow j$ para todos $i,\ j \in C$, portanto, cada classe de comunicação é um subconjunto irredutível maximal. Quando um conjunto não é irredutível, dizemos *redutível*. Quando há uma única classe de comunicação dizemos que a cadeia é uma *cadeia de Markov irredutível*. Dizemos que C é **fechado** se $p_{i,j} = 0$ para $i \in C$ e $j \notin C$.

Exercício 119. Prove a seguinte afirmação. Se $i \in S$ é recorrente então existe um único $C \subset S$ fechado e irredutível ao qual i pertence. Ademais, para quaisquer $j, k \in C$ vale que $f_{j,k} = f_{k,j} = 1$.

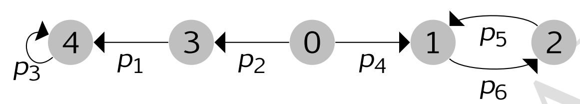
No exemplo 140 é fácil ver a partir do diagrama que a cadeia é irredutível. A cadeia do exemplo 138, claramente, não é irredutível por causa dos estados absorventes 0 e n. Toda cadeia com pelo menos dois estados e com pelo menos um deles absorvente é uma cadeia redutível.

Num conjunto irredutível de estados todos os estados são do mesmo tipo, logo podemos classificar um conjunto de estados, ou a própria cadeia de Markov, irredutível como

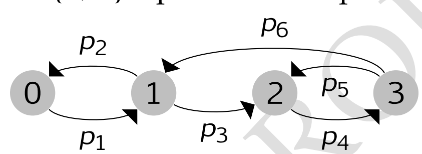
- **aperiódica** se todos os seus estados o forem;
- **periódica** se todos os seus estados o forem;
- **transiente** se todos os seus estados o forem;
- **recorrente** se todos os seus estados o forem.

3.0.34 Observação. *Notemos que a Observação 3.0.31 se aplica a uma classe de comunicação, ou a uma cadeia irredutível, pois os estados dela têm o mesmo período.*

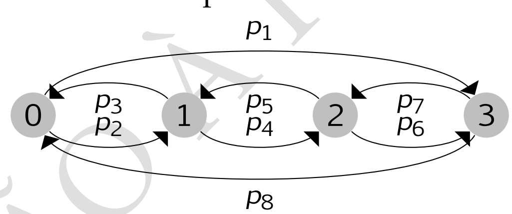
Exemplo 144.



é uma cadeia redutível, a classe $\{1,2\}$ é periódica de período 2, o estado 4 é absorvente.

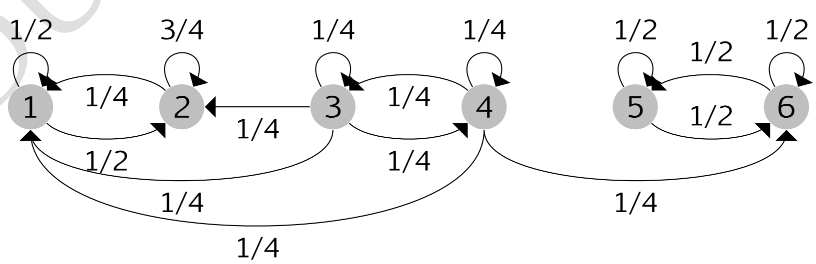


é uma cadeia aperiódica e irredutível.



é uma cadeia irredutível, recorrente e periódica.

Exemplo 145. No caso da cadeia representada abaixo



A cadeia é redutível e o conjunto dos estados têm três classes de equivalência de comunicação $S = \{\{1,2\},\{3,4\},\{5,6\}\}$. Os conjuntos $\{1,2\}$ e $\{5,6\}$ são irredutíveis e recorrentes; $\{3,4\}$ é transiente pois uma vez que a cadeia deixa o conjunto ela não volta mais a ele. Todos os estados têm período 1, pois $p_{i,i} > 0$, portanto a cadeia é aperiódica.

Exemplo 146 (passeio aleatório, exemplo 134). Vejamos o caso do passeio aleatório não-simétrico em uma dimensão, onde

$$p_{i,i+1} = p = 1 - p_{i,i-1}.$$

Claramente, a cadeia é irredutível logo todos os estados são ou recorrentes ou transientes. Fixemo-nos no estado 0 e vamos determinar $\sum_{n\geqslant 0} p_{0,0}^{(n)}$.

É impossível estar num estado par com um número ímpar de movimentos

$$p_{0,0}^{(2n-1)} = 0, \text{ para } n \geqslant 1.$$

Após $2n$ passos a cadeia estará de volta em 0 se e só se metade dos passo foi num sentido (e.g., do estado atual para o maior, o que ocorre com probabilidade p) e metade no sentido oposto, portanto

$$p_{0,0}^{(2n)} = \binom{2n}{n} p^n(1-p)^n = \frac{(2n)!}{n!}n!p^n(1-p)^n \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

onde $a_n \sim b_n$ significa que $\lim_{n\to\infty}(a_n/b_n) = 1$ e $n! \sim n^n e^{-n}\sqrt{2\pi n}$. Agora, se $a_n \sim b_n$ então $\sum_n a_n$ converge se, e só se, $\sum_n b_n$ converge, portanto precisamos estudar a convergência de

$$\sum_{n\geqslant 1} \frac{(4p(1-p))^n}{\sqrt{\pi n}}.$$

No numerador $4p(1-p) \leqslant 1$ e vale a igualdade se e somente se $p = 1/2$. Portanto a série é infinita se e só se $p = 1/2$ e como a cadeia é irredutível e o 0 recorrente, a cadeia é recorrente. Se $p \neq 1/2$ então a cadeia é transiente.

Exercício 120. Prove que toda classe de comunicação com estados recorrentes é fechada.

Exercício 121. Prove que toda classe de comunicação fechada e finita é recorrente.

Exercício 122 (Teorema de decomposição). O conjunto de estados S pode ser particionado como

$$S = T \cup C_1 \cup C_2 \cup \cdots$$

de modo que T é o subconjunto dos estados transientes da cadeia e C_i, para todo i, são conjuntos fechados e irredutíveis de estados recorrentes. Prove essa afirmação.

Essa classificação diz-nos que quando estamos interessado no comprtamento de longo prazo basta estudarmos as cadeias irredutíveis. Se X_0 está em alguma classe de equivalência de comunicação C_ℓ que é recorrente, então a evolução da cadeia se restringe a essa classe pois, assumindo que $i \in C_\ell$ e $j \notin C_\ell$ de modo que $i \to j$, por conseguinte $j \not\to i$, temos de $[X_1 = j] \subset \bigcup_{n\geqslant 1}[X_n \neq i]$

$$\mathbb{P}\left(\bigcup_{n\geqslant 1}[X_n \neq i] \,\middle|\, X_0 = i\right) \geqslant p_{i,j} > 0$$

contrariando o fato de i ser recorrente. Se, por outro lado, X_0 no conjunto T dos estados transientes implica em ou a cadeia ficar em T (consulte seção XV.8 do Feller para a probabilidade desse evento) ou se mover para algum C_ℓ e não sair mais dessa classe.

Exemplo 147. Uma cadeia de Markov com estados $\{1,2\}$ e matriz de transições

$$P = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}$$

não corresponde a uma cadeia irredutível pois para todo inteiro $k > 0$

$$P^k = \begin{pmatrix} 1 & 0 \\ \frac{2^k-1}{2^k} & \frac{1}{2^k} \end{pmatrix}$$

portanto $1 \not\to 2$.

Exemplo 148. Uma cadeia com estados $\{1,2,3\}$ e matriz de transições

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

é redutível pois P^k é da forma

$$\begin{pmatrix} \frac{x}{3^k} & \frac{3^k-x}{3^k} & 0 \\ \frac{y}{3^{k-1}} & \frac{3^{k-1}-y}{3^k} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

com $0 < x < 3^k$ e com $0 < y < 3^{k-1}$ e para todo inteiro $k > 0$.

Exemplo 149. Uma cadeia com 4 estados e matriz de transições

$$P = \begin{pmatrix} 0 & 4/5 & 0 & 1/5 \\ 1 & 0 & 0 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

é *periódica* pois P^k na posição $(1,1)$ é diferente de zero somente para k par, em outras palavras, a partir do estado 1 a cadeia somente volta ao estado 1 em número par de passos.

Exemplo 150. Considere uma cadeia de Markov com matriz de transição

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

cujas classes de comunicação são $\{0,1\}$ que é recorrente, $\{3\}$ que é absorvente e $\{2\}$ que é transiente. Se o inicio é no estado 2, então a probabilidade p com que a cadeia entra na classe recorrente $\{0,1\}$ é, condicionando em X_1, dada por

$$p = \frac{1}{4}1 + \frac{1}{4}1 + \frac{1}{4}p + \frac{1}{4}0 = \frac{1}{2} + \frac{1}{4}p$$

logo $p = \frac{2}{3}$; e com probabilidade $\frac{1}{3}$ a cadeia é absorvida em 3.

No caso da matriz de transição

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 0 & 1/3 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

as classes de comunicação são $\{0,1\}$, $\{2,3\}$ e $\{4,5\}$, respectivamente, recorrente aperiódica, transiente e recorrente periódica. Se p_i é a probabilidade de entrar na primeira classe descrita acima a partir do estado i, $i = 2,3$, então, condicionando em X_1,

$$\begin{aligned} p_2 &= \frac{1}{2}1 + 0\cdot 1 + 0p_2 + \frac{1}{3}p_3 + \frac{1}{6}0 + \frac{1}{6}0 \\ p_3 &= \frac{1}{6}1 + \frac{1}{6}1 + \frac{1}{6}p_2 + 0p_3 + \frac{1}{3}0 + \frac{1}{6}0 \end{aligned}$$

cuja solução é $p_2 = \frac{8}{17}$ e $p_3 = \frac{7}{17}$.

Representação matricial e Distribuição invariante. A matriz $P = (p_{i,j})$ definida pelas probabilidades de transição $p_{i,j} = p(i,j)$ é dita **matriz de transições** da cadeia de Markov. Ainda, para $k \in \mathbb{N}$, se

$$p_{i,j}^{(k)} := \mathbb{P}(X_{t+k} = j \mid X_t = i)$$

é a transição do estado i para o estado j em k passos, que não depende de t por homogeneidade temporal, então

$$p_{i,j}^{(k)} = \sum_{s\in S} p_{i,s}p_{s,j}^{(k-1)}$$

para todo inteiro $k > 1$ e com $p_{i,j}^{(1)} = p_{i,j}$ e *convencionamos* $p_{i,j}^{(0)} = \delta_i(j)$. Logo a k-ésima potência da matriz P é a matriz de transição em k passos

$$P^k = (p_{i,j}^{(k)}).$$

Notemos que $\sum_j p_{i,j}^{(k)} = 1$, para qualquer $k \in \mathbb{N}$. Uma matriz cujas entradas são não-negativas e as linhas somam 1 é uma *matriz estocástica*.

A evolução da cadeia é determinada pela matriz P e boa parte dos estudos das cadeias de Markov são reduzidas, principalmente no caso finito, ao estudo das propriedades algébricas desses elementos.

Exemplo 151. No caso do exemplo 140 dado que choveu na segunda-feira e na terça-feira, qual é a probabilidade de chover na quinta? A matriz toda está descrita abaixo

$$P = \begin{pmatrix} 0{,}7 & 0 & 0{,}3 & 0 \\ 0{,}5 & 0 & 0{,}5 & 0 \\ 0 & 0{,}4 & 0 & 0{,}6 \\ 0 & 0{,}2 & 0 & 0{,}8 \end{pmatrix}$$

A matriz de transição em 2 passos é a matriz

$$P^2 = \begin{pmatrix} 0{,}49 & 0{,}12 & 0{,}21 & 0{,}18 \\ 0{,}35 & 0{,}20 & 0{,}15 & 0{,}30 \\ 0{,}20 & 0{,}12 & 0{,}20 & 0{,}48 \\ 0{,}10 & 0{,}16 & 0{,}10 & 0{,}64 \end{pmatrix}$$

e chover na quinta equivale ao processo estar no estado 0 ou no estado 1, logo a probabilidade é $p_{0,0}^{(2)} + p_{0,1}^{(2)} = 0{,}61$.

No exemplo acima, se fizermos $\pi^{(0)} = \rho$, a distribuição inicial, então $\pi^{(1)} = \pi^{(0)}P$ é a função de massa de probabilidade de X_1, na posição i do vetor lemos $\pi_i^{(1)} = \mathbb{P}(X_i = i)$. Por exemplo, para

$$\pi^{(0)} = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

a distribuição uniforme em S, então temos

$$\pi^{(1)} = \begin{pmatrix} 0{,}285 & 0{,}15 & 0{,}165 & 0{,}4 \end{pmatrix}$$

Analogamente, $\pi^{(2)}$ é a função de massa de X_2 e, em geral, para $n \in \mathbb{N}$

$$\pi^{(n+1)} = \pi^{(n)}P$$

é a função de massa de X_{n+1}. Notemos que

$$\pi^{(n+1)} = \pi^{(n)}P = (\pi^{(n-1)}P)P = \cdots = \pi^{(0)}P^n.$$

Distribuição invariante.: Um vetor de probabilidades $\pi = (\pi_j)_{j \in S}$ que satisfaz

$$\pi_j = \sum_{k \in S} \pi_k p_{k,j}$$

ou, na forma matricial,

$$\pi = \pi P$$

é chamado de **distribuição invariante** ou **distribuição estacionária** da cadeia de Markov com matriz de transições P.

No último exemplo,

$$\pi = \begin{pmatrix} 0.25 & 0.15 & 0.15 & 0.45 \end{pmatrix}$$

é invariante, como o leitor pode verificar. No caso do modelo de Ehrenfest vimos que

$$\pi(k) = \frac{1}{2^N}\binom{N}{k} \qquad 0 \leqslant k \leqslant N$$

é invariante.

Quando S é finito temos que se

$$p_{i,j}^{(n)} \to \pi_j \text{ quando } n \to \infty \text{ para todo estado } j$$

então $\pi = (\pi_j)_j$ é invariante. De fato

$$\pi_j = \lim_{n\to\infty} p_{i,j}^{(n)} = \lim_{n\to\infty} \sum_{k\in S} p_{i,k}^{(n)} p_{k,j} = \sum_{k\in S} \lim_{n\to\infty} p_{i,k}^{(n)} p_{k,j} = \sum_{k\in S} \pi_k p_{k,j}$$

e

$$\sum_{i\in S} \pi_i = \sum_{i\in S} \lim_{n\to\infty} p_{i,j}^{(n)} = \lim_{n\to\infty} \sum_{i\in S} p_{i,j}^{(n)} = 1.$$

Quando S não é finito isso não vale necessariamente, por exemplo, no caso assimétrico ($p \neq 1/2$) do passeio aleatório em $\mathbb{Z}$, exemplo 146, temos $p_{i,j}^{(n)} \to 0$ pois a cadeia é transiente e então o vetor limite π não é um vetor de probabilidades. Também, no caso S finito a existência de um vetor invariante é garantido para qualquer matriz estocástica. Esse fato decorre do Teorema do Ponto Fixo de Brower: *Para toda função contínua $f : T \to T$, onde $T \subset \mathbb{R}^n$ é compacto e convexo, existe $x \in T$ tal que $f(x) = x$.* Agora, consideremos para todo $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$ a norma L^1

$$\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i|$$

e o conjunto (compacto e convexo)

$$T = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \geqslant 0 \text{ e } \|\mathbf{v}\|_1 = 1\}.$$

Seja P uma matriz estocástica e considere a função linear de T em T dada por $\mathsf{x} \mapsto \mathsf{x}P$. Por ser linear a função é contínua, portanto, pelo Teorema da Ponto Fixo de Brower deduzimos a existência de um vetor invariante.

3.0.35 Teorema. *Uma cadeia de Markov irredutível tem distribuição invariante se, e só se, a cadeia for recorrente não-nula. Nesse caso,*

$$\pi_j = \frac{1}{\mu_j}$$

para cada estado j é a única distribuição estacionária

Provaremos esse teorema com o seguinte roteiro:

1. uma cadeia irredutível e recorrente admite para cada estado k um vetor invariante positivo $\gamma = (\gamma_i)_{i\in S}$ tal que $\gamma_k = 1$;
2. para qualquer vetor invariante positivo $\lambda = (\lambda_i)_{i\in S}$ tal que $\lambda_k = 1$ vale $\lambda = \gamma$, ou seja, o vetor dado no item anterior é único;
3. se k é um estado reccorrente não-nulo então existe uma única distribuição invariante $\pi = (\pi_i)_{i\in S}$ tal que $\pi_k = \frac{1}{\mu_k}$, em que μ_k é o tempo médio de retorno para k;
4. a existência de π afirmada no item anterior implica que todo estado da cadeia é recorrente não-nulo.

Demonstração. Consideremos uma cadeia de Markov irredutível e recorrente $\{X_t\}_{t\in\mathbb{N}}$ com distribuição inicial ρ e matriz de transição $P = (p_{i,j})_{i,j\in S}$ e vamos construir uma distribuição invariante $\pi = (\pi_j)_{j\in S}$.

Fixemos um estado k e seja $\gamma_i = \gamma_i(k)$ o número de visitas ao estado i entre duas visitas a k, em média, ou seja

$$\gamma_i(k) := \mathbb{E}\left[\sum_{t=1}^{T_k} 1_{[X_t=i]} \,\middle|\, X_0 = k\right] = \sum_{t\geqslant 1} \mathbb{P}(X_t = i, t \leqslant T_k \mid X_0 = k))$$

donde $\gamma_k(k) = 1$. Vamos mostrar que o vetor $\gamma = \gamma(k)$ dado por

$$\gamma = (\gamma_i)_{i\in S}$$

é um *vetor* invariante.

Primeiro, notemos que pela propriedade de Markov, condicionado ao evento $[X_{t-1} = j]$, o processo estocástico $X_{t-1}, X_t, \dots$ é uma cadeia de Markov com respeito ao estado inicial j e matriz de transição P independente das variáveis aleatórias $X_0, X_1, X_2, \dots, X_{t-1}$, também o evento $[t \leqslant T_k]$ depende somente de $X_0, X_1, X_2, \dots, X_{t-1}$ portanto, como $T_k < \infty$

$$\mathbb{P}(X_{t-1} = j, X_t = i, t \leqslant T_k \mid X_0 = k) = \mathbb{P}(X_{t-1} = j, t \leqslant T_k \mid X_0 = k)p_{j,i}$$

e segue da definição de γ_i que

$$\gamma_i = \sum_{j\in S}\sum_{t\geqslant 1}\mathbb{P}(X_{t-1}=j, X_t=i, t\leqslant T_k \mid X_0=k)$$

logo

$$\begin{aligned}
\gamma_i &= \sum_{j\in S}\sum_{t\geqslant 1}\mathbb{P}(X_{t-1}=j, t\leqslant T_k \mid X_0=k)p_{j,i}\\
&= \sum_{j\in S}p_{j,i}\sum_{t\geqslant 1}\mathbb{P}(X_{t-1}=j, t\leqslant T_k \mid X_0=k)\\
&= \sum_{j\in S}p_{j,i}\sum_{t\geqslant 0}\mathbb{P}(X_{t}=j, t\leqslant T_k-1 \mid X_0=k)\\
&= \sum_{j\in S}p_{j,i}\gamma_j
\end{aligned}$$

o que dá a invariância do vetor $\gamma=\gamma(k)$. Ademais, para o estado k fixado, pela irredutibilidade da cadeia para todo estado j devem existir $n,m>0$ tais que $p_{j,k}^{(n)}>0$ e $p_{k,j}^{(m)}>0$. Logo

$$\begin{aligned}
1=\gamma_k &= \textstyle\sum_{i\in S}p_{i,k}^{(n)}\gamma_i\\
&\geqslant p_{j,k}^{(n)}\gamma_j
\end{aligned}$$

portanto $\gamma_j=\gamma_j(k)$ é finito. Também,

$$\begin{aligned}
\gamma_j &= \sum_{i\in S}p_{i,j}^{(m)}\gamma_i\\
&\geqslant p_{k,j}^{(m)}\gamma_k
\end{aligned}$$

portanto $\gamma_j=\gamma_j(k)>0$.

Agora, para termos uma distribuição invariante podemos normalizar o vetor, para todo estado j

$$\pi_j := \frac{\gamma_j}{\sum_{i\in S}\gamma_i}$$

define uma distribuição invariante para a cadeia de Markov. Recordemos que

$$\mu_k = \mathbb{E}\left[T_k \mid X_0=k\right])$$

que é o número esperados de visitas a outros estados, que não o estado k, quando o estado inicial é k até a próxima visita ao estado k, logo $\mu_k=\sum_i \gamma_i$ e

$$\pi_k = \frac{\gamma_k(k)}{\mu_k} = \frac{1}{\mu_k}$$

para qualquer estado $k\in S$.

Para provar a unicidade de $\pi = (\pi_k)_{k\in S}$, suponhamos ν um vetor *qualquer* com $\nu = \nu P$ e todas as entradas positivas. Vamos mostrar que $\nu_j\mu_j = 1$. Primeiro, notemos que

$$\begin{aligned}
\nu_j\mu_j &= \nu_j \sum_{n\geqslant 1} \mathbb{P}(T_j \geqslant n \mid X_0 = j) \\
&= \sum_{n\geqslant 1} \mathbb{P}(T_j \geqslant n \mid X_0 = j)\mathbb{P}(X_0 = j) \\
&= \sum_{n\geqslant 1} \mathbb{P}(T_j \geqslant n, X_0 = j) \\
&= \sum_{n\geqslant 1} \mathbb{P}(X_{n-1} \neq j, \cdots, X_1 \neq j) - \mathbb{P}(X_{n-1} \neq j, \cdots, X_1 \neq j, X_0 \neq j) \\
&= \sum_{n\geqslant 1} \mathbb{P}(X_{n-2} \neq j, \cdots, X_0 \neq j) - \mathbb{P}(X_{n-1} \neq j, \cdots, X_1 \neq j, X_0 \neq j) \\
&= \mathbb{P}(X_0 = j) + \mathbb{P}(X_0 \neq j) - \lim_{n\to\infty} \mathbb{P}(X_n \neq j, \cdots, X_0 \neq j)
\end{aligned}$$

como o limite é 0, por causa da recorrência de j, o resultado é que $\nu_j\mu_j = 1$, ou seja, $\nu_j = 1/\mu_j = \pi_j$, ou seja, .

Até aqui estabelecemos que *uma cadeia de Markov recorrente positiva e irredutível tem uma distribuição estacionária única que satisfaz $\pi_k = 1/\mu_k$ para todo $k \in S$.*

Agora, se uma cadeia irredutível admite uma distribuição estacionária π, então $\pi = \pi P^n$ e podemos escolher n tal que

$$\pi_k = \sum_j \pi_j p_{j,k}^{(n)} > 0$$

pois a cadeia é irredutível. Assim $\lambda_i = \pi_i/\pi_k$, para todo $i \in S$, define um vetor invariantepara P e não-negativo tal que $\lambda_k = 1$. Vamos assumir, por ora que

$$\lambda_j \geqslant \gamma_j(k)$$

para todo j, em que γ_j é como definido no começo da demonstração. Então

$$\mu_k = \sum_i \gamma_i(k) \leqslant \sum_i \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty$$

o que mostra que k é recorrente positivo. □

Exercício 123. Verifique que se P é estocástica então P^n também é estocástica.

Convergência ao equilíbrio. Recordemos que se $\pi^{(0)}$ é a distribuição inicial, então $\pi^{(1)} = \pi^{(0)}P$ e, em geral, para $n \in \mathbb{N}$, a distribuição de X_{n+1} é $\pi^{(n+1)} = \pi^{(n)}P$. Ademais $\pi^{(n+1)} = \pi^{(0)}P^n$. No caso da matriz de transição e transição em dois passos do exemplo 151

$$P=\begin{pmatrix}0{,}7 & 0 & 0{,}3 & 0\\ 0{,}5 & 0 & 0{,}5 & 0\\ 0 & 0{,}4 & 0 & 0{,}6\\ 0 & 0{,}2 & 0 & 0{,}8\end{pmatrix} \text{ e } P^2=\begin{pmatrix}0{,}49 & 0{,}21 & 0{,}21 & 0{,}18\\ 0{,}35 & 0{,}20 & 0{,}15 & 0{,}30\\ 0{,}20 & 0{,}12 & 0{,}20 & 0{,}48\\ 0{,}10 & 0{,}16 & 0{,}10 & 0{,}64\end{pmatrix}$$

e se formos um pouco mais adiante obtemos

$$P^6=\begin{pmatrix}0.285709 & 0.144552 & 0.162561 & 0.407178\\ 0.270935 & 0.147320 & 0.156915 & 0.424830\\ 0.240920 & 0.150912 & 0.147320 & 0.460848\\ 0.226210 & 0.153616 & 0.141610 & 0.478564\end{pmatrix}$$

$$P^{10}=\begin{pmatrix}0.2553440 & 0.1491648 & 0.1519040 & 0.4435872\\ 0.2531733 & 0.1495093 & 0.1511255 & 0.4461919\\ 0.2486079 & 0.1502124 & 0.1495093 & 0.4516704\\ 0.2464373 & 0.1505568 & 0.1487306 & 0.4542752\end{pmatrix}$$

$$P^{20}=\begin{pmatrix}0.2500461 & 0.1499928 & 0.1500164 & 0.4499447\\ 0.2500274 & 0.1499957 & 0.1500098 & 0.4499671\\ 0.2499880 & 0.1500019 & 0.1499957 & 0.4500144\\ 0.2499693 & 0.1500048 & 0.1499890 & 0.4500369\end{pmatrix}$$

$$P^{30}=\begin{pmatrix}0.2500004 & 0.1499999 & 0.1500001 & 0.4499995\\ 0.2500002 & 0.1500000 & 0.1500001 & 0.4499997\\ 0.2499999 & 0.1500000 & 0.1500000 & 0.4500001\\ 0.2499997 & 0.1500000 & 0.1499999 & 0.4500003\end{pmatrix}$$

$$P^{34}=\begin{pmatrix}0.2500001 & 0.15 & 0.15 & 0.4499999\\ 0.2500000 & 0.15 & 0.15 & 0.4500000\\ 0.2500000 & 0.15 & 0.15 & 0.4500000\\ 0.2500000 & 0.15 & 0.15 & 0.4500000\end{pmatrix}$$

$$P^{35}=P^{36}=\cdots P^{40}=\cdots=\begin{pmatrix}0.25 & 0.15 & 0.15 & 0.45\\ 0.25 & 0.15 & 0.15 & 0.45\\ 0.25 & 0.15 & 0.15 & 0.45\\ 0.25 & 0.15 & 0.15 & 0.45\end{pmatrix}$$

Cada coluna dessa matriz é constante e, além disso, na coluna j vale

$$\begin{aligned}(\pi^{(0)}P^n)_j &= \pi_1^{(0)}p_{1,j}^{(n)}+\pi_2^{(0)}p_{2,j}^{(n)}+\pi_3^{(0)}p_{3,j}^{(n)}+\pi_4^{(0)}p_{4,j}^{(n)}\\ &= (\pi_1^{(0)}+\pi_2^{(0)}+\pi_3^{(0)}+\pi_4^{(0)})\pi_j\\ &= \pi_j\end{aligned}$$

para $n \geqslant 35$ pois $p_{i,j}^{(n)} = \pi_j$ para todo i, ou seja, fixado j, para todo i temos $p_{i,j}^{(n)} \to \pi_j$, quando $n \to \infty$ e o vetor $\pi = (\pi_j)_j$ *não depende da distribuição inicial* $\pi^{(0)}$, isto é, independentemente do estado inicial, se n for suficientemente grande então a probabilidade da cadeia de Markov estar em qualquer um dos estados é dada pelo vetor

$$\pi = \begin{pmatrix} 0.25 & 0.15 & 0.15 & 0.45 \end{pmatrix}.$$

Para π e P acima valem

$$\pi P = \pi \quad \text{e} \quad \pi_j = \lim_{n\to\infty} p_{i,j}^{(n)}$$

a distribuição de probabilidade $\pi^{(n)}$ converge, quando $n \to \infty$, para a distribuição invariante.

Obervemos que nem sempre há convergência, como no caso do modelo de Ehrenfest e nos seguintes exemplos.

Exemplo 152. Uma cadeia redutível com matriz de transição

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

tem distribuição estacionária $\pi = \begin{pmatrix} p & 1-p \end{pmatrix}$ para qualquer $p \in (0,1)$ e, por exemplo, para $p = 1/2$ temos que $p_{1,1}^{(n)} \not\to 1/2$. O mesmo acontece com a matriz de transição

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

e $\pi = (\,1/2 \quad 1/2\,)$ é invariante mas não há convergência. A cadeia é periódica. Considere uma cadeia de Markov com quatro estado e matriz de transições

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

nesse caso, $(1,0,1,0)$ e $(0,1,0,1)$ são vetores invariantes.

3.0.36 Teorema. *Uma cadeia de Markov irredutível tem uma distribuição estacionária $\pi = (\pi_j)_j$ se, e somente se, todos os estados são recorrentes positivos. O vetor invariante é único e satisfaz*

$$\pi_j = \frac{1}{\mu_j}.$$

Ainda, se a cadeia for aperiódica então

$$\pi_j = \lim_{n\to\infty} p_{i,j}^{(n)}. \tag{3.0.30}$$

Adiamos a demonstração desse teorema, por enquanto. No caso da cadeia ser periódica de período d então

$$\lim_{n\to\infty} p_{i,i}^{(nd)} = \frac{d}{\mu_i}$$

e ser for recorrente

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} p_{i,j}^{(m)} = \frac{1}{\mu_j} = \pi_j$$

com π (único) vetor de invariante.

3.0.37 Observação. *Se uma sequência converge* $\lim_{n\to\infty} a_n = a$*, então a sequência das médias parciais também converge*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} a_i = a$$

portanto de equação (3.0.30) *temos*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} p_{i,j}^{(m)} = \pi_j$$

e

$$\sum_{m=0}^{n-1} p_{i,j}^{(m)} = \mathbb{E}\left[V_i[0..n-1] \mid X_0 = i\right]$$

em que $V_i[0..n-1]$ *é o número de visistas a* i *no intervalo de tempo* $0,\dots,n-1$ *(veja o lema* 3.0.32*), ou seja,*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[V_i[0..n-1] \mid X_0 = i\right] = \pi_j.$$

Em palavras, π_j *é a fração média de visitas ao estado* i *por unidade de tempo.*

3.0.38 Teorema (**Teorema ergódico**)**.** *Para qualquer cadeia de Markov irredutível*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{V_i[0..n-1]}{n} = \frac{1}{\mu_i}\right) = 1.$$

Exercício 124*.* Uma matriz quadrada com entradas não-negativas é duplamente estocástica se é estocástica e a soma das entradas de cada coluna é 1. Mostre que se M é quadrada de ordem n e duplamente estocástica então o vetor uniforme $(1/n, 1/n, \dots, 1/n)$ é invariante para M.

Cadeias redutíveis: Os resultados da seção anterior aplicam-se a qualquer classe de comunicação recorrente. Se C_ℓ é classe de comunicação recorrente e aperiódica de um processo, então a submatriz formada pelos $i, j \in C_\ell$ é estocástica logo $p_{i,j}^{(n)} \to 1/\mu_j$ quando $n \to \infty$.

Exemplo 153. Por exemplo,

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 2/3 & 1/3 \end{pmatrix}$$

tem as classes recorrentes $\{0, 1\}$ com a submatriz estocástica irredutível

$$P_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

e $\{2, 3\}$ com a submatriz estocástica irredutível

$$P_2 = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}.$$

A matriz P_1 admite o vetor invariante $\pi^{(1)} = \begin{pmatrix} 1/3 & 2/3 \end{pmatrix}$ e a matriz P_2, o vetor invariante $\pi^{(2)} = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix}$ e, além disso

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_0^{(1)} & \pi_1^{(1)} & 0 & 0 \\ \pi_0^{(1)} & \pi_1^{(1)} & 0 & 0 \\ 0 & 0 & \pi_0^{(2)} & \pi_1^{(2)} \\ 0 & 0 & \pi_0^{(2)} & \pi_1^{(2)} \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Ainda, se $i, j \in C_\ell$, com C_ℓ classe de comunicação recorrente e periódica, então

$$\frac{1}{n}\sum_{m=0}^{n-1} p_{i,j}^{(m)} \to \frac{1}{\mu_j}$$

quando $n \to \infty$.

Por outro lado, se j é um estado transiente então $p_{i,j}^{(n)} \to 0$ quando $n \to \infty$ para qualquer estado inicial i.

Exemplo 154. De volta ao exemplo 150, a classe recorrente $\{0, 1\}$ com respeito a matriz P tem distribuição estacionária $\pi = \begin{pmatrix} 1/3 & 2/3 \end{pmatrix}$. Ademais essa classe é atingida a partir do estado 2 com probabilidade $p = 2/3$, portanto $p_{2,0}^{(n)} \to \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$ e $p_{2,1}^{(n)} \to \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$ logo

$$\lim_{n\to\infty} P^n = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 2/9 & 4/9 & 0 & 1/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

No que diz respeito a matriz Q, temos $\{0,1\}$ recorrente aperiódico com $\pi = \begin{pmatrix} 2/4 & 3/5 \end{pmatrix}$, $\{2,3\}$ transiente com probabilidade de sair de 2 (respec., 3) e ir ser absorvido por $\{0,1\}$ sendo $p_2 = 8/17$ (respec., $7/17$), e $\{4,5\}$ recorrente periódico

$$\lim_{n\to\infty} Q^n = \begin{pmatrix} 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ \frac{8}{17}\cdot\frac{2}{5} & \frac{8}{17}\cdot\frac{3}{5} & 0 & 0 & & \\ \frac{7}{17}\cdot\frac{2}{5} & \frac{7}{17}\cdot\frac{3}{5} & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \end{pmatrix}$$

em que os espaços vazios indicam que o limite não existe. Entretanto,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} Q^m = \begin{pmatrix} 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ \frac{8}{17}\cdot\frac{2}{5} & \frac{8}{17}\cdot\frac{3}{5} & 0 & 0 & \frac{9}{17}\cdot\frac{1}{2} & \frac{9}{17}\cdot\frac{1}{2} \\ \frac{7}{17}\cdot\frac{2}{5} & \frac{7}{17}\cdot\frac{3}{5} & 0 & 0 & \frac{10}{17}\cdot\frac{1}{2} & \frac{10}{17}\cdot\frac{1}{2} \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Exemplo: 2-SAT: O 2-SAT (*2-Satisfiability Problem*) é o seguinte problema; dado uma fórmula booleana 2-CNF (cada cláusula é uma conjunção de 2 literais $C_i = \ell_{i_1} \vee \ell_{i_2}$), digamos

$$\Phi = C_1 \wedge C_2 \wedge \cdots \wedge C_k$$

sobre um conjunto de variáveis $V = \{x_1, x_2, \dots, x_n\}$, determinar se existe uma valoração $\nu : V \to \{0,1\}$ que satisfaça Φ, isto é, faça $\Phi = 1$.

É sabido[2] que 2-SAT pode ser resolvido em tempo polinomial. A seguir, daremos um algoritmo aleatorizado bem simples, de tempo polinomial e que responde corretamente com alta probabilidade.

Quando existe uma valoração válida o algoritmo não a encontra com probabilidade menor que $(1/2)^t$.

Se não existe valoração que satisfaça todas as cláusulas o algoritmo termina após $2tn^2$ rodadas, caso contrário, seja ν uma valoração que satisfaz Φ e ν_i a valoração construída pelo algoritmo após a i-ésima rodada do laço interno, linha **??**.

Denotemos por X_i o número de valores em comum que tomam as valorações ν e ν_i, isto é, a quantidade de variáveis da fórmula que têm o mesmo valor binário nas duas valorações. O algoritmo termina com $X_i = n$ ou quando encontra alguma valoração diferente que ν que satisfaça Φ.

[2] Aspvall, Plass, Tarjan, *A linear-time algorithm for testing the truth of certain quantified Boolean formulas*, Inform. Process. Lett., 1979, no. 3, pg. 121–123

Notemos que valem $\mathbb{P}(X_{i+1} = 1 \mid X_i = 0) = 1$ e

$$\mathbb{P}(X_{i+1} = j+1 \mid X_i = j) \geqslant 1/2$$
$$\mathbb{P}(X_{i+1} = j-1 \mid X_i = j) \leqslant 1/2$$

($\mathbb{P}(X_{i+1} = j+1 \mid X_i = j)$ vale 1 quando os dois literais envolvem variáveis que discordam de ν). As variáveis $\{X_i\}_i$ não definem uma cadeia de Markov, entretanto, a cadeia do exemplo 139 é um cenário de pior caso, isto é, $\mathbb{E}\left[X_j\right] \leqslant \mathbb{E}\left[Y_j\right]$, onde Y_j é a variável aleatória do exemplo, por isso, usando equação (3.0.27) temos que

$$\mathbb{P}(X_0 \geqslant 2n^2) \leqslant \mathbb{P}(Y_0 \geqslant 2n^2) \leqslant \frac{1}{2}$$

portanto em cada rodada do laço interno a probabilidade de não encontrar uma valoração que satisfaz Φ é no máximo $1/2$. A probabilidade de não ser encontrada uma valoração nas t rodadas do laço externo, linha **??**, é no máximo $(1/2)^t$. Se a fórmula é satisfatível então o laço interno encontrará a valoração certa com alta probabilidade se for repetido $n^2 \log n$ vezes.

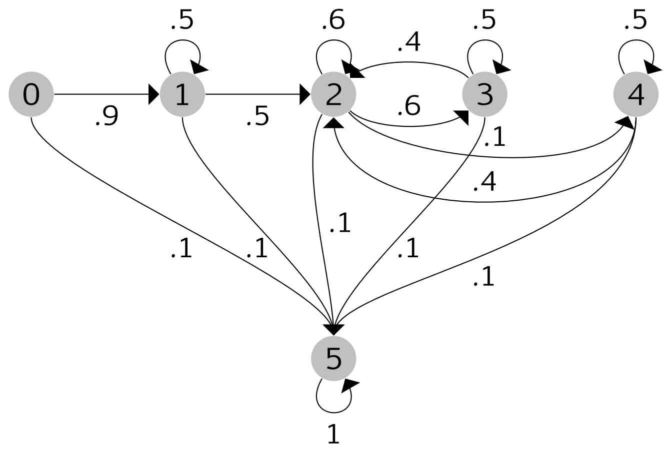
Exercício 125. Por que $\{X_n\}_n$ acima não define uma cadeia de Markov?

Exercício 126. Prove a afirmação $\mathbb{E}\left[X_j\right] \leqslant \mathbb{E}\left[Y_j\right]$ feita no exemplo acima.

Exemplo: Modelo de fecundidade: As mudanças nos padrões sociais afetam a taxa de crescimento populacional. Modelos para analisar os efeitos das mudanças na fecundidade média de uma população levam em conta a idade, a situação conjugal, o número de filhos e várias outras informações a respeito do público feminino da população. Essas características compõem cada estado de um processo cujo interesse é determinar o tempo médio do processo nas categorias de maior fertilidade. As probabilidades de transição são inferidas de dados demográficos.

Num modelo simplificado consideramos os estados $0 \equiv$pré-puberil, $1 \equiv$solteira, $2 \equiv$casada, $3 \equiv$Divorciada $4 \equiv$Viúva, $5 \equiv$morte ou emigração.

$$P = \begin{pmatrix} 0 & 0.9 & 0 & 0 & 0 & 0.1 \\ 0 & 0.5 & 0.4 & 0 & 0 & 0.1 \\ 0 & 0 & 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0 & 0.4 & 0.5 & 0 & 0.1 \\ 0 & 0 & 0.4 & 0 & 0.5 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



Denotemos por e_i o tempo médio da cadeia no estado 2 dado que $X_0 = i$. Então

$$
\begin{cases}
e_0 &= 0.9e_1 + 0.1e_5 \\
e_1 &= 0.5e_1 + 0.4e_2 + 0.1e_5 \\
e_2 &= 1 + 0.6e_2 + 0.2e_3 + 0.1e_4 + 0.1e_5 \\
e_3 &= 0.4e_2 + 0.5e_3 + 0.1e_5 \\
e_4 &= 0.4e_4 + 0.5e_4 + 0.1e_5 \\
e_5 &= 0
\end{cases}
$$

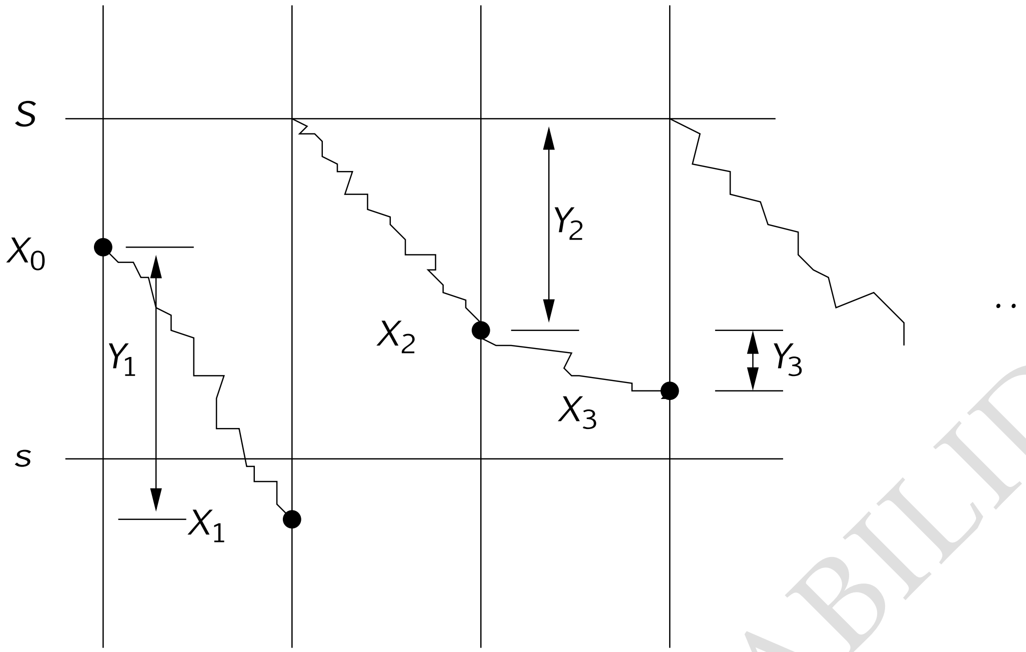
cuja (única) solução é $e_0 = 4.5$, $e_1 = 5$, $e_2 = 6.25$, $e_3 = e_4 = 5$.

Notemos que, como esperado, a distribuição estacionária é $\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix}$.

Exemplo: Controle de estoque: Uma mercadoria é armazenada e ao final dos períodos $n = 0, 1, \dots$ o estoque é reabastecido. A demanda pela mercadoria no período n é uma variável aleatória Y_n com $Y_1, Y_2, \dots$ independentes e identicamentes distribuídas.

A política de reabastecimento envolve escolher dois parâmetros inteiros e positivos s e S de modo que se ao final de um período o estoque é no máximo s então ele é abastecido até atingir S, caso contrário não é abastecido.

A variável aleatória X_n é a quantidade de mercadoria ao fim do período n *antes* do reabastecimento e caso seja negativa a demanda é atendida após o reabastecimento.



Os níveis de estoque em dois períodos consecutivos são relacionados por

$$X_{n+1} = \begin{cases} X_n - Y_{n+1} & s < X_n \leqslant S \\ S - Y_{n+1} & X_n \leqslant s \end{cases}$$

e as probabilidades das transições

$$p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \mathbb{P}(Y_{n+1} = i - j) & s < X_n \leqslant S \\ \mathbb{P}(Y_{n+1} = S - j) & X_n \leqslant s \end{cases}$$

Nesse caso, gostaríamos de saber qual é, a longo prazo, a fração de períodos que a demanda não é atendida

$$\lim_{n\to\infty} \sum_{j<0} \mathbb{P}(X_n = j)$$

e qual é, a longo prazo, o nível médio de estoque

$$\lim_{n\to\infty} \sum_{j>0} j\mathbb{P}(X_n = j)$$

e sob certas condições $\mathbb{P}(X_n = j)$ converge.

Para um exemplo numérico tomemos

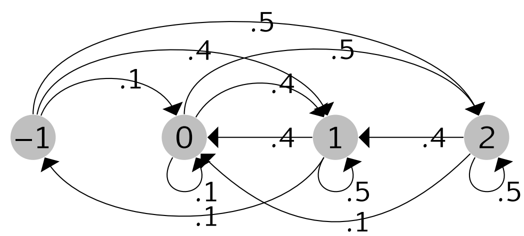
$$\mathbb{P}(Y_n = 0) = 0.5, \quad \mathbb{P}(Y_n = 1) = 0.4, \quad \mathbb{P}(Y_n = 2) = 0.1$$

e $s = 0$ e $S = 2$. Assim, $X_n \in \{-1, 0, 1, 2\}$.

Quando $X_n = 1$ então não é necessário reabastecimento do estoque, logo $X_{n+1} = 0$ caso $Y_{n+1} = 1$, o que ocorre com probabilidade 0.4, logo $p_{1,0} = 0.4$. Agora, se $X_n = 0$ então há reabastecimento para

$S = 2$ e $X_{n+1} = 0$ caso $Y_{n+1} = 2$, logo $p_{0,0} = 0.1$. A matriz fica da seguinte forma, lembrando que os estados são $\{-1, 0, 1, 2\}$

$$P = \begin{pmatrix} 0 & 0.1 & 0.4 & 0.5 \\ 0 & 0.1 & 0.4 & 0.5 \\ 0.1 & 0.4 & 0.5 & 0 \\ 0 & 0.1 & 0.4 & 0.5 \end{pmatrix}$$



que, claramente, é irredutível, recorrente e aperiódica, cujo vetor invariante é

$$\begin{pmatrix} 4/90 & 21/90 & 40/90 & 25/90 \end{pmatrix}.$$

Reversibilidade: Numa cadeia de Markov, dado o estado presente, o futuro e o passado são independentes. Seja $\{X_i\}_{i\geqslant 0}$ uma cadeia de Markov irredutível com respeito a matriz estocástica P e a distribuição inicial π, e π uma distribuição invariante. Tomemos a matriz $Q = q_{i,j}$ dada por

$$q_{i,j} = \frac{\pi_j}{\pi_i} p_{j,i}.$$

Essa matriz é estocástica pois, pela invariância de π

$$\sum_{j\in S} q_{i,j} = \frac{1}{\pi_i} \sum_{j\in S} \pi_j p_{j,i} = 1.$$

Ainda,

$$\sum_{j\in S} \pi_i q_{i,j} = \sum_{j\in S} \pi_j p_{j,i} = \pi_i$$

ou seja, π é invariante com relação a Q.

Se tomarmos $Y_n = X_{N-n}$ para $0 \leqslant n \leqslant N$ então

$$\mathbb{P}(Y_0 = i_0,\ Y_1 = i_1, \dots, Y_N = i_N) = \mathbb{P}(X_0 = i_N,\ X_1 = i_{N-1}, \dots, X_N = i_0) =$$

$$\pi_{i_N} p_{i_N,i_{N-1}} \cdots p_{i_1,i_0} = \pi_{i_0} q_{i_0,i_1} \cdots q_{i_{N-1},i_N}$$

logo $\{Y_n\}_{0\leqslant n\leqslant N}$ é uma cadeia de Markov com relação a Q e π. Ademais,

$$q_{i_N,i_{N-1}} \cdots q_{i_1,i_0} = \frac{1}{\pi_{i_0}} \pi_{i_N} p_{i_0,i_1} \cdots p_{i_{N-1},i_N} > 0$$

portanto, a cadeia é irredutível, chamada *tempo-reverso* da cadeia $\{X_i\}_{0\leqslant i\leqslant N}$.

Exercício 127. Mostre que se P é uma matriz estocástica e λ um vetor não-negativo tal que para todos i, j

$$\lambda_i p_{i,j} = \lambda_j p_{j,i} \tag{3.0.31}$$

então $\lambda P = \lambda$.

Exercício 128. Prove que $\{X_n\}_{n\geqslant 0}$ irredutível, com matriz de transições P e distribuição inicial λ é reversível se e só se vale equação (3.0.31).

Exemplo 155. Consideremos um grafo conexo G com peso positivo $w_{i,j}$ na aresta em cada aresta $\{i, j\} \in E(G)$. Uma partícula move-se sobre o grafo com a seguinte estratégia: se a partícula encontra-se no vértice i no instante t então no instante $t+1$ a partícula estará no estado j com probabilidade

$$p_{i,j} = \frac{w_{i,j}}{\sum_{j\in V(G)} w_{i,j}}$$

de modo que $w_{i,j} = 0$ caso $\{i, j\}$ não seja uma aresta. A equação equação (3.0.31) é equivalente a

$$\pi_i = \frac{\sum_{j\in V(G)} w_{i,j}}{\sum_{i\in V(G)} \sum_{j\in V(G)} w_{i,j}}.$$

No caso mais simples, $w_{i,j} \in \{0, 1\}$ e $w_{i,j} = 1$ se e só se $\{i, j\}$ é aresta,

$$\pi_i = \frac{d(i)}{2|E(G)|}.$$

Veremos que esse sempre é o caso, portanto passeios aleatórios em grafos são reversíveis.

Demonstração do Teorema 3.0.36 no caso finito. Vamos começar considerando o caso em que P é uma matriz estocástica positiva. Digamos que P é $n \times n$ e definimos $\epsilon = \min P > 0$. Vamos provar que $\lim_{k\to\infty} P^k = Q$ com Q uma matriz estocástica com as colunas constantes.

Exercício 129. Seja P uma matriz estocástica com todas entradas positivas. Tomemos $\epsilon = \min P$ e $\mathbf{c}$ um vetor *coluna* qualquer com $\min \mathbf{c} = a_0$ e $\max \mathbf{c} = b_0$. Sejam $a_1 = \min P\mathbf{c}$ e $b_1 = \max P\mathbf{c}$. Prove que $a_1 \geqslant a_0$, $b_1 \leqslant b_0$ e $b_1 - a_1 \leqslant (1-2\epsilon)(b_0 - a_0)$
(Sugestão: tome $\mathbf{c}'$ o vetor obtido a partir de $\mathbf{c}$ trocando todas as coordenadas por a_0 menos a coordenada b_0; tome $\mathbf{c}''$ o vetor obtido a partir de $\mathbf{c}$ trocando todas as coordenadas por b_0 menos a coordenada a_0. Use o fato de $P\mathbf{c}' \leqslant P\mathbf{c} \leqslant P\mathbf{c}''$.)

Denotemos por $\mathbf{e}_j$ o vetor *coluna* com todas as entradas nulas a menos da posição j que é 1. A seguir, vamos usar o exercício anterior para definir um vetor estocástico π que são as linhas de Q, o qual mostraremos depois ser invariante.

Fixamos $j \in \{1, 2, \dots, n\}$, uma coluna de P, e definimos

$$\begin{aligned} a_k &= \min P^k \mathbf{e}_j \\ b_k &= \max P^k \mathbf{e}_j \end{aligned}$$

o mínimo e o máximo, respectivamente, da j-ésima coluna de P^k, tomamos $a_0 = 0$ e $b_0 = 1$. Para $k \in \mathbb{N}$ temos

$$a_{k+1} = \min P^{k+1}\mathbf{e}_j = \min P(P^k\mathbf{e}_j) \text{ e } a_k = \min(P^k\mathbf{e}_j).$$

Analogamente, temos

$$b_{k+1} = \max P^{k+1}\mathbf{e}_j = \max P(P^k\mathbf{e}_j) \text{ e } b_k = \max(P^k\mathbf{e}_j).$$

Do exercício 129 temos a sequência de máximos e mínimos

$$b_0 \geqslant b_1 \geqslant \cdots \geqslant b_k \geqslant \cdots \geqslant a_k \geqslant a_{k-1} \geqslant \cdots \geqslant a_0$$

de modo que

$$b_k - a_k \leqslant (1 - 2\epsilon)(b_{k-1} - a_{k-1}) \leqslant (1 - 2\epsilon)^k$$

pois $\epsilon > 0$. Da equação anterior $\lim_{k\to\infty} b_k - a_k = 0$ e definimos

$$\pi_j = \lim_{k\to\infty} b_k = \lim_{k\to\infty} a_k.$$

Com isso provamos que todas as entradas da j-ésima coluna de P^k convergem para um valor constante π_j. Feito isso para cada coluna j, o vetor que procuramos é

$$\pi = (\pi_1, \pi_2, \dots, \pi_n)$$

e com isso

$$\lim_{k\to\infty} p_{i,j}^{(k)} = \pi_j \quad (\forall i)$$

e tomamos $Q = \mathbf{1}^\intercal \pi$, a matriz com cada linha igual a π que queríamos provar. Como $P > 0$ temos $a_1 > 0$, portanto $\pi_j > 0$. Também, $\pi_j < 1$.

Exercício 130. Prove que temos $\sum_j \pi_j = 1$ no vetor dado acima.

Estenderemos esse resultado para matrizes não negativas usando o fato de que para a matriz de uma cadeia de Markov irredutível e aperiódica temos P^t positiva para algum $t \in \mathbb{N}$.

3.0.39 Proposição. *Se uma cadeia Markov é irredutível e aperiódica então existe $K_0 \in \mathbb{N}$ tal que para todo $i,\ j \in S$, se $k \geqslant K_0$ então $p_{i,j}^{(k)} > 0$.*

Demonstração. Da cadeia ser irredutível temos que para cada $i, j \in S$ existe $k(i,j)$ tal que $p_{i,j}^{(k(i,j))} > 0$. Da cadeia ser aperiódica temos que para cada $i \in S$ existe $k_0(i)$ tal que $p_{i,i}^{(k)} > 0$ para todo $k \geqslant k_0(i)$. Com essas constantes, se $t \geqslant 0$ então

$$p_{i,i}^{(k_0(i)+t)} > 0 \text{ e } p_{i,j}^{(k(i,j))} > 0$$

portanto, por equação (3.0.26), para todo $t \in \mathbb{N}$

$$p_{i,j}^{(k(i,j)+k_0(i)+t)} = \sum_{\ell \in S} p_{i,\ell}^{((k_0(i)+t))} p_{\ell,j}^{(k(i,j))} \geqslant p_{i,i}^{((k_0(i)+t))} p_{i,j}^{(k(i,j))} > 0.$$

Tomamos

$$K_0 = \max\{k(i,j) + k_0(i) \; i, j \in S\}$$

o que prova a proposição. $\square$

O vetor π é invariante

$$\pi_j = \lim_{n\to\infty} p_{i,j}^{(n)} = \lim_{n\to\infty} \sum_{k \in S} p_{i,k}^{(n)} p_{k,j} = \sum_{k\in S} \lim_{n\to\infty} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in S} \pi_k p_{k,j}.$$

Agora, provaremos que se o estado inicial é algum elemento de $S = \{1, 2, \dots, n\}$ com probabilidade dada por $\rho = \rho^{(0)}$ então $\rho P^t \to \pi$ quando $t \to \infty$, ou seja, qualquer distribuição inicial a longo prazo se aproxima da distribuição estacionária.

Consideremos os vetores de probabilidades

$$\rho^{(0)} = \begin{pmatrix} q_1^{(0)} & \dots & q_n^{(0)} \end{pmatrix}$$

e

$$\rho^{(t)} = \begin{pmatrix} q_1^{(t)} & \dots & q_n^{(t)} \end{pmatrix}$$

e a partir de $\rho^{(t)}$ obtemos a distribuição X_{t+1} por

$$\rho^{(t+1)} = \rho^{(t)} P = \rho P^t.$$

Agora, seja $K_0 = K_0(P)$ a constante dada pela proposição 3.0.39. Assim, P^{K_0} é uma matriz com todas as entradas positivas e para $\epsilon = \min P^{K_0}$ e para $t \in \mathbb{N}$

$$b_{tK_0} - a_{tK_0} \leqslant (1-\epsilon)^t$$

portanto, a subsequência converge, $\lim_{t\to\infty}(b_{tK_0} - a_{tK_0}) = 0$. Ademais, as diferenças $b_t - a_t$ nunca aumentam com t, ou seja, a seqüência $(b_t - a_t)_t$ é não-crescente, portanto converge para 0, portanto, $\lim_{t\to\infty} P^t = Q$

Seja ν um vetor *qualquer* com $\nu = \nu P$. Para mostrar a unicidade de π é suficiente mostrar que ν é um múltiplo escalar de π.

Se $\nu = \nu P$ então $\nu = \nu P^t$, portanto

$$\nu = \lim_{t\to\infty} \nu P^t = \nu \lim_{t\to\infty} P^t = \nu Q$$

logo $v_j = \sum_i v_i \pi_j = \pi_j \sum_i v_i$ e podemos concluir então que $\nu = (\sum_i v_i)\pi$.

Agora, vamos provar que $\lim_{t\to\infty} \rho P^t = \pi$ para qualquer vetor de probabilidades $\rho = (q_1, q_2, \ldots, q_n)$. Como acima,

$$\lim_{t\to\infty} \rho P^t = \rho Q$$

donde $(\rho Q)_j = (\sum_i q_i)\pi_j$ é a j-ésima coordenada de ρQ. Como ρ é um vetor de probabilidades a soma das coordenadas é 1, logo

$$(\rho Q)_j = \pi_j$$

ou seja, $\rho Q = \pi$ como queríamos provar.

Resta provar que $\pi_j = 1/\mu_j$. Como S é finito $\mu_j < \infty$.

Primeiro, definimos $\mu_{i,j} = \mathbb{E}\left[T_j \mid X_0 = i\right]$ para $i \neq j$, e se consideramos o primeiro passo da cadeia temos que vale

$$\mu_{i,j} = p_{i,j} + \sum_{k\neq j} p_{i,k}(\mu_{k,j} + 1) = 1 + \sum_{k\neq j} p_{i,k}\mu_{k,j}.$$

De modo análogo $\mu_{i,i} = \mu_i = 1 + \sum_k p_{i,k}\mu_{k,j}$. Esse conjunto de equações pode ser escrito de forma matricial como

$$M = PM + U - D$$

em que U é a matriz com todas as entradas iguais a 1 e $D = (d_{i,j})$ é a matriz diagonal $d_{i,i} = \mu_i$. Reescrevendo temos

$$D - U = (P - Id)M$$

em que Id é a matriz identidade. Se multiplicarmos por π a direita nos dois lados da igualdade temos, no lado direito, que $\pi(P - Id) = 0$ pois π é invariante, portanto $\pi D = \pi U$, mas $\pi U = \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix}$, logo $(\pi D)_j = 1$ para cada j, isto é,

$$\pi_j \mu_j = 1.$$

Exercícios.

1. Prove que o evento

$$X_{i_1} = s_{i_1},\ X_{i_2} = s_{i_2},\ \ldots, X_{i_r} = s_{i_r}$$

   ocorre com mesma probabilidade em todo espaço das realizações das variáveis com pelo menos r variáveis.

2. Prove que numa cadeia de Markov sobre um conjunto finito de estados, pelo menos um estado é recorrente e todo estado recorrente é positivo

3. Seja s um estado absorvente numa cadeia de Markov tal que para todo estado i da cadeia $p_{i,s}^{(n)} > 0$, para algum $n = n(i)$. Mostre que todos os estados, a não ser s, são transientes.

4. Seja A um subconjunto de estados e defina

$$T_A := \min\{n \geqslant 1 : X_n \in A\}$$

$$\eta_j := \mathbb{P}(T_A < \infty \mid X_0 = j).$$

Mostre que $\eta_j = 1$ se $j \in A$ e

$$\eta_j = \sum_{s \in S} p_{j,s} \eta_s$$

se $j \notin A$. Agora defina

$$\xi_j := \mathbb{E}\left[T_A \mid X_0\right] = j)$$

e mostre que $\xi_j = 0$ se $j \in A$ e que

$$\xi_j = 1 + \sum_{s \in S} p_{j,s} \xi_s$$

se $j \notin A$.

5. Se $\{X_n\}_n$ é uma cadeia de Markov, quais das seguintes sequências é uma cadeia de markov?

   a) $\{X_{m+r}\}_{r \geqslant 0}$, para m fixo;
   b) $\{X_{2m}\}_{m \geqslant 0}$;
   c) $\{(X_n, X_{n+1})\}_{n \geqslant 0}$.

6. Considere a a variável aleatória T_i e a esperança μ_i definidos acima, com a convenção de que se i não é visitado então $T_i = \infty$. Verifique que

$$\mu_i = \mathbb{E}\left[T_i \mid X_0 = i\right].$$

Verifique também que o estado i é transiente se e só se $\mathbb{P}(T_i = \infty \mid X_0 = i) > 0$ e, nesse caso, $\mathbb{E}\left[T_i \mid X_0 = j\right] = \infty$.

7. Considere a variável aleatória V_j definida acima e defina

$$\eta_{i,j} := \mathbb{P}(V_j = \infty \mid X_0 = j).$$

Prove que

$$\eta_{i,i} = \begin{cases} 1, & \text{se } i \text{ recorrente,} \\ 0, & \text{se } i \text{ transiente.} \end{cases}$$

e que

$$\eta_{i,j} = \begin{cases} \mathbb{P}(T_j < \infty \mid X_0 = j), & \text{se } i \text{ recorrente,} \\ 0, & \text{se } i \text{ transiente.} \end{cases}$$

8. Considere a cadeia de Markov com estados $\mathbb{N}$ e transições $p_{0,j} = a_j$ para todo $j \in \mathbb{N}$, e $p_{i,i} = p$, e $p_{i,i-1} = 1 - p$ para todo $i \neq 0$. Classifique os estados e determine os tempos médios de recorrência.

9. Se $\{X_n\}_n$ é uma cadeia de Markov com matriz de transições

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

e f é dada por $f(0) = 0$ e $f(1) = f(2) = 1$, então $Y_n = f(X_n)$ é uma cadeia de Markov?

10. Mostre que a sequência de variável aleatória $\{Y_i\}_{i\geqslant 1}$ do exemplo 136 é de variáveis independentes 2-a-2, conclua que $p_{i,j}^{(n)} = 1/2$ e que vale a identidade de Chapman–Kolmogorov equação (3.0.26) mesmo não sendo uma cadeia de Markov.

§9 Passeio aleatório em grafos. Seja $G = (V, E)$ um grafo finito e conexo. O leitor não familiarizado com as nomenclaturas elementares da Teoria dos Grafos pode consultar estas notas.

Um **passeio aleatório** em G é uma sequência $v_0, v_1, v_2, \ldots$ de vértices de V de modo que v_{i+1} é escolhido uniformemente em $N(v_i) = \{u \in V : \{u, v_i\} \in E\}$, *a vizinhança de* v_i, para todo $i \in \mathbb{N}$. Dizendo de outro modo, é uma cadeia de Markov $\{X_t\}_{t\geqslant 0}$ homogênea com V como conjunto de estados e

$$p_{v,u} = \mathbb{P}(X_{t+1} = u \mid X_t = v) = \frac{1}{d(v)}$$

para todo $u \in N(v)$ e todo $t \in \mathbb{N}$, onde $d(v) = |N(v)|$ é o grau do vértice v.

Se $A = A(G)$ é a matriz de adjacências de G, isto é $A = (a_{v,u})$ com $a_{v,u} \in \{0, 1\}$ e $a_{v,u} = 1$ sse e só se $\{v, u\}$ é aresta, e $D = (d_{u,v})$ é a matriz diagonal $d_{v,v} = 1/d(v)$ então a matriz de transições do passeio aleatório em G é

$$P = AD.$$

Exemplo 156. Seja G o grafo sobre o conjunto de vértice $\{1, 2, \ldots, n\}$ com todas as $\binom{n}{2}$ arestas, chamado de *grafo completo*, e consideremos um passeio aleatório em G, assim

$$p_{v,u} = \frac{1}{n-1}$$

e a matriz de transição é $P = \frac{1}{n-1}A$.

Nesse caso, o número esperado de passos para atingir v a partir de u é $n-1$. De fato, a probabilidade de atingir v em 1 passo é $f_{u,v}^{(1)} = 1/(n-1)$, a probabilidade de atingir v em 2 passos é $f_{u,v}^{(2)} = (n-2)/(n-1)^2$, em 3 passos $f_{u,v}^{(3)} = (n-2)^2/(n-1)^3$, e assim por diante.

O número esperado de passos é

$$\sum_k k f_{u,v}^{(k)} = \sum_{k\geqslant 1} k \frac{1}{n-1}\left(\frac{n-2}{n-1}\right)^{k-1} = \frac{1}{n-1}\frac{1}{\left(1-\frac{n-2}{n-1}\right)^2} = n-1.$$

O número esperado de passos para visitar todos os vértices do grafo pode ser estimado da seguinte forma. Seja t_i o instante em que pela primeira vez temos exatamente i vértices visitados, portanto, $t_{i+1} - t_i$ é uma variável aleatória geométrica que conta o número de passos enquanto espera-se para conhecer um novo vértice, evento que ocorre com probabilidade $(n-i)/(n-1)$, logo $\mathbb{E}[t_{i+1} - t_i] = (n-1)/(n-i)$ e t_n é o número de passos até visitar todos os vértices

$$\mathbb{E}[t_n] = \sum_{i=1}^{n-1} \mathbb{E}[t_{i+1} - t_i] = \sum_{i=1}^{n-1} \frac{n-1}{n-i} = (n-1)\sum_{i=1}^{n-1}\frac{1}{i} = (n-1)H_{n-1}$$

onde H_n denota o n-ésimo número harmônico, $H_n = \sum_{i=1}^{n} 1/i = \ln n + \gamma + \Theta(n^{-1})$, onde $\gamma \approx 0,577$ é a constante de Euler–Mascheroni.

Um passeio aleatório num grafo conexo é uma cadeia de Markov irredutível e em certos casos aperiódica. O próximo resultado caracteriza tais passeios aleatórios aperiódicos.

3.0.40 Teorema. *Se G é conexo com pelo menos dois vértices então um passeio aleatório em G define uma cadeia de Markov irredutível. Ainda, tal cadeia de Markov é periódica se e só se G é bipartido.*

Demonstração. Se G é conexo então para dois vértices u e v quaisquer que estão a distância k vale $p_{u,v}^{(k)} > 0$. Como G é conexo a distância entre quaisquer dois vértices é algum $k \in \mathbb{N}$, portanto a cadeia de Markov definida pelo passeio aleatório é irredutível.

Seja G um grafo conexo e bipartido com bipartição $\{A, B\}$. Então, todos os passeios de u para v têm a mesma paridade no número de arestas, para o caso $u, v \in A$ ou $u, v \in B$ e ímpar caso $u \in A$ e $v \in B$ ou $v \in A$ e $u \in B$. Logo

- se $u, v \in A$ ou $u, v \in B$ então $p_{u,v}^{(k)} > 0$ exceto quando k é ímpar,
- se $u \in A$ e $v \in B$, ou $v \in A$ e $u \in B$, então $p_{u,v}^{(k)} > 0$ exceto quando k é par.

Portanto a cadeia tem período 2.

Agora, suponhamos G não-bipartido, então contém um circuito de comprimento ímpar C. Tomemos P um passeio de u a v, vértices que distam k em G, com o menor número de arestas. Um passeio de u para v com $k + 2r$ arestas existe para todo $r \in \mathbb{N}$, basta repetir r vezes alguma aresta do passeio P. Como G é conexo existe um passeio de algum vértice de P até algum vértice de C, portanto, podemos usar as arestas de C para obter passeios de u para v que têm a paridade oposta a de k. Logo, $p_{u,v}^{(t)} > 0$ para todo t suficientemente grande, ou seja, a cadeia é aperiódica. □

3.0.41 Corolário. *Se $G = (V, E)$ é conexo, não-bipartido e com pelo menos dois vértices então a cadeia de Markov dada por um passeio aleatório em G admite um vetor invariante. Ademais, a distribuição estacionária é única e dada por*

$$\pi = (\pi_v)_{v \in V}, \text{ com } \pi_v = \frac{d(v)}{2|E|}. \tag{3.0.32}$$

Demonstração. A cadeia é irredutível e aperiódica, portanto admite um único vetor invariante pelo teorema 3.0.36. Agora, basta verificar que equação (3.0.32) é invariante.

A soma dos graus dos vértices de um grafo é $2|E|$, logo $\sum_{v \in V} \pi_v = 1$. Resta verificar que $\pi = \pi P$, onde P é a matriz de transição, mas

$$(\pi P)_v = \sum_{u \in V} \pi_u p_{u,v} = \sum_{u \in N(v)} \frac{d(u)}{2|E|} \frac{1}{d(u)} = \sum_{u \in N(v)} \frac{1}{2|E|} = \frac{d(v)}{2|E|} = \pi_v$$

portanto, o vetor dado na equação (3.0.32) é o vetor invariante. □

3.0.42 Observação. *Se $G = (V, E)$ é bipartido então podemos contornar essa propriedade indesejada (no sentido do corolário acima) acrescentando laços aos vértices do grafo com probabilidade de transição $1/2$ (ou seja, $p_{v,v} = 1/2$) e dividir por 2 a probabilidade das outras arestas, ou seja, se P é a matriz de transição da cadeia no grafo original, a nova matriz de transição é*

$$Q = \frac{P + Id}{2}$$

onde Id é a matriz identidade $|V| \times |V|$. Essa transformação apenas "reduz a velocidade" do passeio.

Exemplo 157. De volta ao grafo completo G do exemplo 156, vamos usar a estratégia de acrescentar laço e o número esperado de passos até um passeio aleatório passar por todos os vértices de G. Seja P a matriz de transição da cadeia de Markov do passeio em G e consideremos a modificação acima, acrescentando um laço em cada vértice ficamos com a matriz de transição do novo passeio

$$Q = \frac{P + Id}{2}.$$

Seja T o número de passos para o passeio definido por Q visitar todos os vértice e $q_i = (n - i)/n$ a probabilidade de visitar um vértice novo se outros i vértices já foram visitados, então o número esperado de passos até o passeio visitar o $i + 1$-ésimo vértice é $1/q_i$. Assim,

$$\mathbb{E}[T] = \sum_{i=1}^{n} \frac{n}{i} = nH_n.$$

$s-t$ conexidade em grafos: O problema da $s-t$ conexidade em grafos é: dado um grafo $G=(V,E)$ e dois vértices s,t em G decidir se há um passeio entre esses dois vértices. Esse problema pode ser resolvido em tempo linear no tamanho do grafo, $|V|+|E|$, e usando espaço $\Omega(|V|)$.

A pergunta que interessa nesse caso é se o problema pode ser resolvido usando espaço logarítmico. Essa pergunta foi respondida afirmativamente em [Reingold, *Undirected connectivity in log-space*, J. ACM 58, 2008], até então o único progresso significativo foi o algoritmo aleatorizado que apresentaremos a seguir devido a [Aleliunas, Karp, Lipton, Lovasz, Rackoff, *Random walks, universal traversal sequences, and the complexity of maze problems*, SFCS '79].

Notemos que o espaço extra gasto pelo algoritmo é para manter o vértice atual e o contador de rodadas para o laço, portanto, o espaço utilizado é $O(\log n)$ para um grafo com n vértices.

Se o algoritmo devolve *sim* então o grafo contém um passeio entre os vértices s e t. Agora, se o algoritmo responde *não*, então ou não há caminho, ou o algoritmo não foi capaz de encontrá-lo em tempo. Portanto, se não existe caminho a resposta está correta, caso contrário a resposta pode estar errada e vamos limitar a probabilidade de erro.

Seja T o tempo para um passeio aleatório no grafo G visitar todos os vértices de G. O **tempo de cobertura** de G (*cover time*) é

$$\max_{v\in V(G)} \mathbb{E}\left[T \mid X_0 = v\right].$$

3.0.43 Lema. *O tempo de cobertura de um grafo $G=(V,E)$ conexo é no máximo $4|V||E|$.*

Demonstração. Tomemos um subgrafo acíclico maximal de $G=(V,E)$ (ou, uma árvore geradora) e consideremos um passeio $W = v_0, v_1, \dots, v_{2|V|-2} = v_0$ nas $|V|-1$ arestas dessa árvore de modo que cada aresta seja percorrida exatamente duas vezes, uma vez em cada direção. O tempo de cobertura de G é limitado superiormente pelo tempo esperado para um passeio percorrer a sequência W que é

$$\sum_{i=0}^{2|V|-3} \mathbb{E}\left[T_{v_{i+1}} \,\middle|\, X_0 = v_i\right] < (2|V|-2)(2|E|) < 4|V||E|$$

pois $\mathbb{E}\left[T_{v_{i+1}} \,\middle|\, X_0 = v_i\right] < 2|E|$ para toda aresta $\{v_i, v_{i+1}\}$ em G. De fato, calculando μ_v de dois modos distintos

$$\frac{2|E|}{d(v)} = \frac{1}{\pi_v} = \mu_v = \sum_{u\in N(v)} \left(1 + \mathbb{E}\left[T_v \mid X_0 = u\right]\right)\frac{1}{d(u)}$$

portanto, $2|E| = \sum_{u\in N(v)} \left(1 + \mathbb{E}\left[T_v \mid X_0 = u\right]\right)$. □

Com isso conseguimos estimar a probabilidade de erro da seguinte maneira. Seja G um grafo e s e t dois vértices de G que são ligados por um passeio. O algoritmo erra se em n^4 rodadas não consegue achar um $s-t$ caminho em G. Denotemos por $C(G)$ o tempo de cobertura de G, usando a desigualdade de Markov

$$\mathbb{P}(\text{erro}) = \sum_{k>n^4} f_{s,t}^{(k)} \leqslant \frac{\mathbb{E}\left[T_t \mid X_0 = s\right]}{n^4}$$

logo

$$\mathbb{P}(\text{erro}) \leqslant \frac{C(G)}{n^4} \leqslant \frac{4n|E|}{n^4} < \frac{4}{n}.$$

Passeio aleatório em grafos regulares: Um grafo é regular se todos os vértices têm o mesmo grau. Nesse caso temos um bom tanto de informação que podemos retirar da matriz de transição que é duplamente estocástica (veja o exercício 124). A começar, que essa matriz é simétrica, portanto, todos os seus autovalores são reais, ademais, o Teorema Espectral garante a existência de uma família ortonormal de autovetores, tantos quanto é a dimensão da matriz. Um passeio aleatório num grafo regular converge para a distribuição uniforme (exercício 124) e nesta seção mostraremos que a velocidade dessa convergência é ditada pelo segundo maior autovalor da matriz de transição.

Um grafo $G = (V, E)$ é d-regular, $d \in \mathbb{N}$, se todos os vértice de V têm grau d. No que segue, vamos assumir que todo grafo é sobre $V = \{1, 2, \dots, n\}$, é conexo e d-regular.

Nesse caso, a matriz de transição de um passeio em G é

$$P = \frac{1}{d}A$$

para $A = A(G)$ a martiz de adjacências. A matriz P é simétrica, portanto, seus autovalores são reais. Sejam

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$$

os autovalores de P. Do Teorema de Perron–Frobenius temos que se G é conexo então

- $\lambda_1 > \lambda_2$;
- $\lambda_1 = -\lambda_n$ se e somente se G é bipartido.

Não é difícil constatar que o vetor π com todas as coordenadas iguais a $1/|V|$ é um autovetor de P associado ao autovalor 1, ou seja, $P\pi = 1\pi$ e como P é simétrica

$$\pi = \pi P$$

que, em outras palavras, quer dizer que a distribuição uniforme sobre V é a distribuição estacionária do passeio aleatório em G, como já haviamos dito.

Ainda, se k é tal que $|\pi_k| = \max_v |\pi_v|$ então

$$|\lambda_1 \pi_k| = \left| \sum_{v=1}^{n} \pi_v p_{v,k} \right| \leqslant \sum_{v=1}^{n} |\pi_v||p_{v,k}| \leqslant |\pi_k| \sum_{v=1}^{n} |p_{v,k}| = |\pi_k|$$

ou seja, $\lambda_1 \leqslant 1$, portanto $\lambda_1 = 1$ já que 1 é autovalor.

Seja G um grafo conexo, não-bipartido, d-regular e com n vértices. Fixemos

$$\lambda = \max\{|\lambda_2|, |\lambda_n|\}$$

e $\rho = \rho^{(0)}$ uma distribuição inicial sobre V, seja $\rho^{(t)}$ a distribuição de probabilidade sobre os vértices de G no instante $t \in \mathbb{N}$, isto é, a distribuição de X_t. Vamos mostrar que $\rho p^{(t)}$ converge para π com velocidade controlada por λ.

Seja ξ_i, $1 \leqslant i \leqslant n$, uma base ortonormal de autovetores de P, de modo que ξ_i é autovetor associado ao autovalor λ_i. Notemos que $\xi_1 = \sqrt{n}\pi$.

Se $\rho = a_1\xi_1 + \cdots + a_n\xi_n$ então $\rho^{(t)} = pP^t = a_1\mathbf{v}_1P^t + \cdots + a_n\mathbf{v}_nP^t = a_1\lambda_1^t\xi_1 + \cdots + a_n\lambda_n^t\xi_n$ e

$$\|\rho^{(t)} - a_1\xi_1\|_2^2 = \left(a_1{\lambda_1}^t - a_1\right)^2 + \sum_{i>1}(a_i{\lambda_i}^t)^2 \leqslant \lambda^{2t}\sum_{i>1} a_i^2 \leqslant \lambda^{2t}\|\rho^{(0)}\|_2^2$$

donde segue que

$$\|\rho^{(t)} - a_1\xi_1\|_2 \leqslant \lambda^t\|\rho^{(0)}\|_2 \leqslant \lambda^t\|\rho^{(0)}\|_1 = \lambda^t$$

e como $\lambda < 1$ temos $\rho^{(t)}$ converge para $a_1\xi_1$ quando $t \to \infty$, logo $a_1\xi_1$ é igual a π, pelo teorema 3.0.36.

Com isso, provamos

3.0.44 Lema. *Sejam G um grafo com n vértices, conexo, d-regular e não-bipartido, $P = A(G)/d$ a matriz de transição de um passeio aleatório em G e*

$$\lambda = \max\{|\alpha|\colon \alpha \text{ é autovalor de } P \text{ e } \alpha < 1\}.$$

Então para todo vetor de probabilidades ρ e todo $t \in \mathbb{N}$

$$\|\rho^{(t)} - \pi\|_2 \leqslant \lambda^t$$

onde $\pi = (1/n, 1/n, \dots, 1/n)$ é o vetor invariante do passeio aleatório.

Exercício 131. Seja $\rho = (p_1, \dots, p_n) \in \mathbb{R}^n$ um vetor, então $\|\rho\|_\infty$ é uma norma dada por

$$\|\rho\|_\infty = \max\{|p_i|\colon 1 \leqslant i \leqslant n\}.$$

Prove que $\|\rho\|_\infty \leqslant \|\rho\|_2$ para todo $\rho \in \mathbb{R}^n$.

3.0.45 Teorema. *Seja $\{X_t\}_{t\in\mathbb{N}}$ um passeio aleatório num grafo G com n vértices, conexo, d-regular e não-bipartido e com segundo maior autovalor λ. Então, para todo $\delta > 0$ existe $k = k(\lambda, \delta)$ tal que*

$$\left|\mathbb{P}(X_k = v) - \frac{1}{n}\right| < \delta$$

para todo vértice v.

Demonstração. Seja G como no enunciado. Vamos mostrar que

$$\|\rho A^k - \pi\|_\infty = \left|\mathbb{P}(X_k = v) - \frac{1}{n}\right| < \delta.$$

Seja $\rho = \rho^{(0)}$ uma distribuição inicial qualquer, então a distribuição de X_k é dada pelo vetor ρP^k,

$$\mathbb{P}(X_k = v) = (\rho P^k)_v$$

em que P é a matriz de transição do passeio aleatório.

Pelo lema 3.0.44 e exercício acima

$$\|\rho A^k - \pi\|_\infty \leqslant \|\rho A^k - \pi\|_2 \leqslant \lambda^{k+1}$$

portanto, se $k > \log_\lambda(\delta/\lambda)$ então $\|\rho A^k - \pi\|_\infty < \delta$. □

Exercício 132. Verifique se vale o resultado anterior para G bipartido, conexo e d-regular com o passeio aleatório dado por $Q = (P + \mathrm{Id})/2$, onde $P = A(G)/d$.

Exercício 133. Sejam G um grafo com n vértices, conexo, d-regular e $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n$ os autovalores da matriz de adjacências de G. Consideremos as matrizes estocásticas $P = A(G)/d$ e $Q = (P + \mathrm{Id})/2$ cujos autovalores são, respectivamente, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots, \lambda_n$ e $\nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_n$. Prove que

$$\lambda_i = \frac{\mu_i}{d} = 2\nu_i - 1$$

para todo inteiro $1 \leqslant i \leqslant n$.

Exercício 134. Sejam G um grafo com n vértices, m arestas e $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n$ os autovalores da matriz de adjacências de G. Prove que

$$\sum_{i=1}^{n} \mu_i^2 = 2m. \tag{3.0.33}$$

(Segestão: $\sum_i \mu_i^2$ e a soma da diagonal princial da matriz A^2 e A é diagonalizável.)

Passeios aleatórios em grafos expansores: Da discussão anterior podemos concluir que quanto menor for o segundo autovalor de um grafo, que denotamos por λ, mais rápido um passeio aleatório converge para a distribuição uniforme, entretanto λ não pode ser arbitrariamente pequeno. da equação (3.0.33) temos

$$(n-1)\lambda^2 + \lambda_1^2 \geqslant 2m$$

portanto, no caso d-regular $\lambda = \Omega(\sqrt{d})$.

No que segue, G^n é um grafo com n vértices, conexo e d-regular para d fixo. Dizemos que G^n é um grafo ϵ-**expansor** se $\lambda \leqslant d - \epsilon$.

Seja W um subconjunto de vértices de G^n. Denotamos por $N(W)$ o subconjunto dos vértices de $V \setminus W$ que são adjacente a algum vértice de W. Dizemos que G é c-**vértice expansor** se para todo W com $|W| \leqslant n/2$ vale que $|N(W)| \geqslant c|W|$, daí vem o adjativo *expansor*. Essas definições são equivalentes no seguinte sentido: *se G^n é ϵ-expansor então é $\epsilon/(2d)$-vértice-expansor. Por outro lado, se G^n é c-vértice-expansor então é $c^2/(4+2c^2)$-expansor*. Logo, grafos expansores são caracterizados combinatorialmente por possuir alta conexidade, que é equivalente a dizer que a *distância espectral* (diferença $d - \lambda$ entre os dois maiores autovalores da matriz de adjacências do grafo) é grande.

Grafos expansores *esparsos* são objetos aparentemente contraditórios, mas a existência desses grafos segue de métodos probabilísticos usuais, como mostrou Pinsker em 1973 que quase todo grafo d-regular, $d \geqslant 3$, é expansor. Embora abundantes, a construção explícita desses grafos não é simples; de fato, em geral a construção e a prova da expansão usam ferramentas profundas e complexas da matemática. Sortear um grafo e testar se é expansor é inviável, o sorteio usa muitos bits aleatórios (uma das principais aplicações desses grafos que veremos aqui é o uso para economizar bits aleatórios em algoritmos aleatorizados) e o problema de decisão está em coNP.

No que segue vamos assumir a possibilidade de construir explicitamente grafos expansores tais que dados um vértice x e $i \in \{1, 2, \dots, d\}$ o vértice $v_i(x)$ pode ser determinado em espaço logarítmico. Para as construções explícitas o leitor pode consultar este documento Uma ótima referência sobre grafos expansores e aplicações é

O nosso interesse agora é no uso desses grafos para desaleatorização em BPP que segue a seguinte estratégia. Suponha que temos um algoritmo que decide $L \in \mathsf{BPP}$ com probabilidade de erro menor que ϵ usando n bits aleatórios para entradas de tamanho fixo. Então, podemos construir um algoritmo que decide a mesma linguagem com probabilidade de erro menor que ϵ^k usando kn bits tomando k rodadas independentes do algoritmo. Agora, suponha que temos um grafo expansor 3-regular com 2^n vértices. Um gerador pseudoaleatório consiste em determinar um vértice desse grafo uniformemente, de modo que temos os n bits aleatórios necessários para uma rodada do algoritmo. Para determinar um desses vértices usamos o lema 3.0.44, que diz que é suficiente começamos um passeio aleatório de comprimento $\approx n$, cada passo do passeio precisa de 2 bits genuinamente aleatórios. Como o algoritmo é executado k vezes, o número de bits aleatórios usados é $O(n + k)$.

Antes de expormos formalmente a idéia acima chamamos a atenção para um resultado conhecido estreitamente relacionado com os resultados discutidos nessa seção.

Expander mixing lemma

É sabido que se $X,\ Y \subseteq V(G^n)$ então a distribuição de arestas entre X e Y em G^n satisfaz

$$||E(X,Y)| - \frac{d}{n}|X||Y|| \leqslant \lambda\sqrt{|X||Y|} \tag{3.0.34}$$

Esboço de prova da equação (3.0.34). Seja ξ_i, $1 \leqslant i \leqslant n$, uma base ortonormal de autovetores de A. Sejam $\mathbf{x}$ e $\mathbf{y}$ os vetores (linha) característicos de X e Y, respectivamente. Então $|E(X,Y)| = \mathbf{x}A\mathbf{y}^{\mathsf{T}}$, onde T denota o transposto. Escrevendo na base de autovetores $|E(X,Y)| = \sum_i \lambda_i\alpha_i\beta_i$, onde α_i e β_i são os produtos escalares $\langle \mathbf{x}, \xi_i\rangle$ e $\langle \mathbf{y}, \xi_i\rangle$, respectivamente. O resultado segue de $\sum_i \lambda_i\alpha_i\beta_i = d|X||Y|/n + \sum_{i\neq 1}\lambda_i\alpha_i\beta_i$. □

Portanto, λ pequeno garante um distribuição de arestas como num grafo aleatório com densidade de arestas d/n. Grafos com λ pequeno são ditos pseudoaleatórios. Linial e Bilu provaram uma recíproca desse resultado, se 3.0.34 vale para algum $\lambda > 0$ então segundo maior autovalor de $A(G)$ é $O(\lambda\log(d/\lambda))$.

da equação (3.0.34) temos

$$\left|\frac{|E(X,Y)|}{nd} - \frac{|X||Y|}{n^2}\right| \leqslant \frac{\lambda}{d}$$

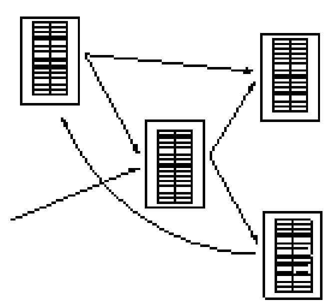
ou seja, a probabilidade de sortear um par de vértices e cair em $X \times Y$ é próxima a probabilidade de sortear uma aresta e cair em $E(X,Y)$. Escrevendo de ouro modo, consideremos a vizinhança de cada vértice de G indexada por $\{1,2,\dots,d\}$ e denotemos por $v_i(x)$ o i-ésimo vizinho do vértice x de acordo com essa indexação. Assim, a equação equação (3.0.34) pode ser interpretada como

$$\left|\mathbb{P}_{(x,i)\in_R V\times[d]}([x\in X]\cap[v_i(x)\in Y]) - \mathbb{P}_{(x,y)\in_R X\times Y}([x\in X]\cap[y\in Y])\right| \leqslant \frac{\lambda}{d}.$$

Passeio aleatório na *WEB*: o *Google PageRank*:

"To test the utility of PageRank for search, we built a web search engine called Google"

O grafo *web* é um grafo dirigido definido pelas páginas *web* e as ligações (*links* ou *hyperlinks*) entre as páginas. Conhecer a estrutura desse grafo é importante para o desenvolvimento de algoritmos eficientes para a *web*, por exemplo.



Um marco no projeto e desenvolvimento desses algoritmos é o algoritmo *PageRank* [Page, Brin, Motwani, Winograd, The PageRank Citation Ranking: Bringing Order to the Web] desenvolvido como parte da ferramenta de busca na *web* batizada *Google* pelos fundadores da empresa com mesmo nome.

O *PageRank* é um algoritmo para classificação (*ranking*) de páginas na *web*. A idéia que o motivou é modelar a importância relativa de uma página de modo que uma busca resulte em resultados com relevância. A própria empresa explica a idéia da seguinte maneira:

> "O coração do nosso software é o PageRank(TM), um sistema para dar notas para páginas na web, desenvolvido pelos nossos fundadores Larry Page e Sergey Brin na Universidade de Stanford. E enquanto nós temos dúzias de engenheiros trabalhando para melhorar todos os aspectos do Google no dia a dia, PageRank continua a ser a base para todas nossas ferramentas de busca na web.
>
> Explicações sobre o PageRank
>
> A classificação das páginas (PageRank) confia na natureza excepcionalmente democrática da Web, usando sua vasta estrutura de links como um indicador do valor de uma página individual. Essencialmente, o Google interpreta um link da página A para a página B como um voto da página A para a página B. Mas o Google olha além do volume de votos, ou links, que uma página recebe; analisa também a página que dá o voto. Os votos dados por páginas "importantes"pesam mais e ajudam a tornar outras páginas "importantes.""
> (http://www.google.com.br/why_use.html)

Assim, uma página tem uma classificação alta se é referenciada por páginas com classificação alta.

O que nos interessa no momento é que o modelo adotado no *PageRank* pode ser interpretado como um passeio aleatório no grafo *web*: um internauta absorto, começa a navegar na *web* a partir de uma página qualquer, e segue a navegação por um dos *links* da página atual escolhido uniformemente; depois de muito tempo nessa tarefa as páginas começam a repetir e o internauta entediado pára o processo e recomeça-o a partir de alguma outra página.

O modelo simplificado do *PageRank* é descrito da seguinte maneira. Seja $G = (V, E)$ o grafo da *web*, ou seja, V é o conjuntos formado pelas páginas *web*, as quais serão consideradas sem perda de generalidade $\{1, 2, \dots, n\} = V$, e $(a, b) \in E$ se na página a há um *link* para a página b. Denotamos por $N^+(a)$ o conjunto dos vértices b tais que (a, b) é uma aresta de G, e por $N^-(a)$ denotamos o conjunto dos vértices b tais que (b, a) é uma aresta de G. No grafo *web* da figura 3.1 $|N^+(3)| = 2$ e $|N^-(3)| = 3$.

A classificação é dada por um vetor $\mathbf{r}$ onde r_a é a classificação da página a e satisfaz

$$r_a = \sum_{b \in N^-(a)} \frac{r_b}{|N^+(b)|} \tag{3.0.35}$$

ou seja, se b aponta para a então b contribui com $1/|N^+(b)|$ de sua relevância para a relevância de a. Seja P a matriz

$$p_{a,b} = \begin{cases} \frac{1}{|N^+(a)|} & \text{se } (a,b) \text{ é aresta} \\ 0 & \text{caso contrário,} \end{cases}$$

então da equação (3.0.35)

$$\mathbf{r} = \mathbf{r}P$$

ou seja, $\mathbf{r}$ é um autovetor à esquerda de P. No grafo *web* da figura 3.1 está descrita a matriz de transição P para aquele grafo.

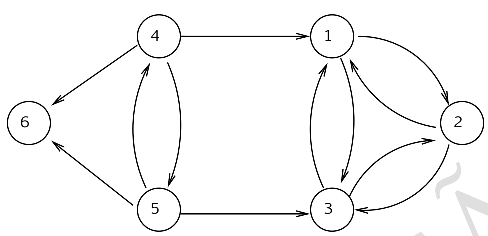
Se P for uma matriz estocástica então o vetor $\mathbf{r}$ é um vetor invariante e dessa forma, $\mathbf{r}$ pode ser calculado escolhendo uma distribuição inicial $\mathbf{r}^{(0)}$ e fazendo $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)}P$. Como vimos, sob certas hipóteses $\mathbf{r}^{(k)}$ converge para $\mathbf{r}$. Esse método, conhecido como método das potências, é usado há muito tempo para calcular autovetores associado ao maior autovalor. Entretanto, na atual situação não sabemos se o vetor converge pois não temos garantia que a matriz P seja irredutível (garante $\lambda_1 > \lambda_2$, portanto convergência) ou estocástica (garante o autovalor $\lambda_1 = 1$, portanto o método converge para o vetor invariante).

Exemplo 158. Por exemplo, para

$$P = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

e $\mathbf{r}^{(0)} = (0,1)$ temos $\mathbf{r}^{(2)} = \mathbf{r} = (0,0)$, No caso de um circuito dirigido, um vetor inicial pode não convergir.

No exemplo da figura 3.1 abaixo a matriz P não é estocástica. Os vértices sem arestas que saem,



$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figura 3.1: Exemplo de grafo *web*.

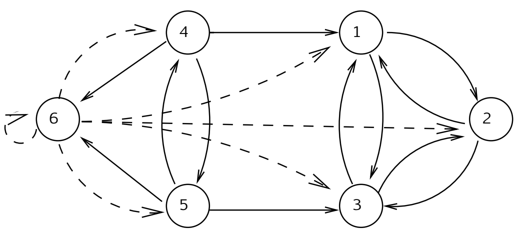
ou seja os vértices v tais que $|N^+(v)| = 0$ são chamados de pendentes (*dangling*) e por causa deles a matriz não é estocástica. No exemplo da figura 3.1 o vértice 6 é pendente. Na *web* de fato há vários vértices pendentes, por exemplo, os documentos em pdf disponíveis em páginas *web* são vértices pendentes do grafo.

Seja n o número de vértices em G. Definimos uma matriz auxiliar A pondo para cada vértice pendente a, o que corresponde a uma página sem *link*s, a linha a com entradas $1/n$ e tomamos

$$Q = P + A$$

isso significa que um passeio aleatório que chega numa página sem saída continua em qualquer outra página com igual probabilidade.

A matriz Q é estocástica. A matriz Q referente a matriz P da figura 3.1 é dada a seguir na figura 3.2 ao lado do grafo *web* que corresponde à modificação no grafo que reflete a modificação na matriz P. No grafo da figura 3.2 as arestas tracejadas são as arestas incluídas e correspondem à possibilidade de navegação do internauta que chega a uma página sem saída e recomeça a navegação de qualquer lugar uniformemente.



$$Q = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

Figura 3.2: Modelo de grafo *web* modificado, sem vértices pendentes. As arestas tracejadas são as arestas incluídas artificialmente.

No exemplo dado na figura 3.2, o maior autovalor da matriz Q é 1 com multiplicidade 1 e autovetor associado $(1, 1, 1, 0, 0, 0)$. Notemos que os vértices $4,\ 5,\ 6$ tem classificação 0. Isso decorre do fato de não haver aresta que sai de $\{1, 2, 3\}$ e chega em qualquer outro vértice diferente desses. Esse conjunto é chamado de sorvedouro (*rank sink*).

Para lidar com esses sorvedouros basta garantir que a matriz seja irredutível pois se $p_{a,b}^{(k)} > 0$ para algum k, para todo $a,\ b \in V$ então não há sorvedouros no grafo. Para garantir uma matriz irredutível[3] tomamos $p \in (0, 1)$ e consideramos um passeio aleatório que segue as transições de Q com probabilidade p ou que com probabilidade $1 - p$ vai pra qualquer outra página *web* (como o comportamento do internauta absorto que ficou entediado), ou seja,

$$R = pQ + (1 - p)\frac{1}{n}\mathbf{1}$$

onde $\mathbf{1}$ é a matriz com todas as entradas iguais a 1. A matriz obtida de Q do exemplo da figura 3.2

$$R = \begin{pmatrix} \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} \\ \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} \\ \frac{p}{2} + \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} \\ \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} \\ \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} \\ \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} \end{pmatrix}$$

É sabido que o método iterativo para computar o vetor invariante descrito, o *método das potências*, converge com velocidade $|\lambda_2/\lambda_1|$ onde $\lambda_1 > \lambda_2$ são os dois maiores autovalores da matriz R. Ainda, é sabido que $\lambda_1 = 1$ e que $|\lambda_2| \leqslant p$.

[3] Em linguagem de Teoria dos Grafos, o grafo dirigido tem que ser fortemente conexo, caso contrário o vetor invariante pode ter todas as coordenadas nulas fora de uma componente fortemente conexa do grafo.

O parâmetro p é conhecido como fator de amortecimento (*damping factor*). No trabalho que originou o algoritmo os autores do PageRank estabeleceram o valor $p = 0,85$ após testes. Atualmente, a Google não divulga o valor desse fator. No mesmo trabalho, os autores reportam de 50 a 100 iterações do método das potências até a condição de parada do método das potências. 0 critério tradicional de parada é da forma $\|\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}\|_2 < \epsilon$ para uma tolerância $\epsilon > 0$ pequena; notemos que não é necessário conhecer as grandezas do vetor invariante, só é preciso determinar a ordem das coordenadas, o que pode ser usado para diminuir o número de iterações. Para $p = 0,85$ foi reportado que com 29 iterações $\|\pi^{k+1} - \pi^k\|_2 < 10^{-2}$, e no caso de 50 a 100 iterações a tolerância é de 10^{-3} a 10^{-7}.

Para concluirmos esta seção vão mostrar uma alternativa para escrever essa matriz, uma vez que atordoantes 1 trilhão de páginas indexadas foi reportado pela Google em julho de 2008. A matriz P é esparsa pela natureza da *web*: páginas com poucos *link*s, 52 em média, e muitos vértices pendentes, a matriz Q é mais densa e a matriz R é positiva. A matriz A pode ser escrita como $\mathbf{u}^\mathsf{T}\mathbf{a}$ onde $\mathbf{u}$ é o vetor com todas as entradas iguais a $1/n$. De fato, pode ser qualquer vetor de probabilidades, o vetor uniforme foi a escolha original, mas atualmente sabe-se que favorece *link spamming* e esse parâmetro também não é divulgado pela Google. O vetor $\mathbf{a}^\mathsf{T}$ é o transposto do vetor característico dos vértices pendentes. Com essas definições

$$\begin{aligned}\mathbf{p}^{(t)} = \;& \mathbf{p}^{(t-1)}R = p\mathbf{p}^{(t-1)}Q + (1-p)\tfrac{1}{n}\mathbf{p}^{(t-1)}\mathbf{1} = p\mathbf{p}^{(t-1)}Q + (1-p)\tfrac{1}{n}\mathbf{1} = \\ & p\mathbf{p}^{(t-1)}Q + p\mathbf{p}^{(t-1)}\mathbf{u}^\mathsf{T}\mathbf{a} + (1-p)\mathbf{u}^\mathsf{T}\mathbf{e}\end{aligned}$$

onde $\mathbf{e} = (1,1,\dots,1)$, portanto só precisamos armazenar a matriz P e os vetores $\mathbf{a}$, $\mathbf{e}$ e $\mathbf{u}$.

Exercício 135. Mostre que a matriz R é irredutível.

Exercício 136. O grafo dirigido com vértices $\{0,1,\dots,n-1\}$ e arestas $(i, i+1 \mod n)$ é um circuito dirigido com n vértices. Analise o comportamento de $\mathbf{r}^{(k)}$ para $k \in \mathbb{N}$ com $\mathbf{r}^{(0)} = (1,0,\dots,0)$.

Exercícios complementares.

Exercício 137. Seja P uma matriz de transições de uma cadeia de Markov. Defina a matriz $P' = (P + \mathrm{Id})/2$. Prove que a cadeis com matriz de transições P' é aperiódica.

Exercício 138. Considere a generalização natural do algoritmo para o 2-SAT, seção 3.0.27, para 3-SAT. Prove que tal algoritmo é exponencial, $O(2^n)$ em que n é o número de variáveis.

Exercício 139. Por dia, uma dentre n pessoas é requisitada, a pessoa i requisitada com probabilidade p_i. As pessoas são organizadas numa lista ordenada e a cada requisição a pessoa escolhida é colocada no topo da lista, o resto da lista fica inalterada.

1. Numa modelagem com cadeia de Markov, quais são os estados?

2. Qual é a distribuição invariante? (Dica: Não precisa fazer conta.)

Exercício 140. Uma pessoa tem n guarda-chuvas que usa no trajeto de casa ao trabalho e vice-versa. Se ela está em casa no começo de dia e está chovendo então escolhe um guarda-chuva para ir ao escritório, caso haja disponível. Se ela está no escritório no fim do dia e está chovendo então escolhe um guarda-chuva para ir para casa, caso haja disponível. Quando não está chovendo ela não escolhe usa guarda-chuva. Assumindo que no início do dia e no fim do dia chove de modo independente e com probabilidade p, qual é a fração das viagens que ela faz sob chuva e sem guarda-chuva?

Exercício 141. Um grupo de n processadores é arranjado em lista ordenada e quando uma tarefa chega é atribuída ao primeiro da lista; se ele está ocupado tenta-se o segundo da lista, e assim por diante. Uma vez que a tarefa é processada ou que não é encontrado um processador disponível pra ela, ela deixa o sistema. Nesse momento, podemos reordenar os processadores antes que a próxima tarefa chegue e a estratégia é: se um processador atendeu a tarefa ele avança uma posição na lista, trocando-o com o processador imediatamente a frente; se todos falharam ou o primeiro atendeu a ordem permanece a mesma. Assumamos que no processador i atende a tarefa com probabilidade p_i.

1. Defina uma cadeia de Markov para analisar o problema.
2. Mostre que a cadeia é reversível.
3. Descubra as probabilidades limite.

Exercício 142. Num tabuleiro de xadrez o cavalo começa num dos quatro cantos e se move com igual probabilidade para qualquer uma das posições legais, de acordo com as regras do seu movimento. Qual é o número esperado de movimentos até voltar para a posição de saída?

Exercício 143. Como podemos gerar valores para vinte variáveis aleatórias independentes e com distribuição uniforme em $(0, 1)$ condicionadas a soma delas serem menor que dois?

Exercício 144. Uma coloração de um grafo é uma atribuição de cores aos vértices do grafo. Um grafo é k-colorível se há um coloração com k cores sem que vértices adjacentes recebam a mesma cor. Seja G um grafo 3-colorível.

Mostre que G tem uma coloração com 2 cores sem que haja triângulo com os três vértices da mesma cor.

Considere o seguinte algoritmo para achar a coloração afirmada no parágrafo acima: o algoritmo começa com uma coloração de duas cores arbitrária. Enquanto houver triângulo monocromático, escolha um deles e sorteie um de seus vértices, mude a cor do vértice sorteado. Determine um limitante superior para o número esperado de recolorações até oal goritmo terminar.

Exercício 145. Seja X_n a soma dos resultados de n lançamentos independentes de um dado. Mostre que para $k \geqslant 2$

$$\lim_{n\to\infty} \mathbb{P}(k \text{ divide } X_n) = \frac{1}{k}.$$

Exercício 146. Numa cidade com $n+1$ habitantes, uma pessoa conta um boato para uma segunda pessoa, que por sua vez conta o boato para uma terceira pessoa e assim por diante. Em cada passo o ouvinte do boato é escolhido aleatoriamente de maneira uniforme dentre as outras pessoas na cidade. Como esse processo pode ser modelado como uma cadeia de Markov? Qual a probabilidade do boato ser contado r vezes sem repetir nenhuma pessoa? Escreva um código em R para determinar o número de rodadas até que todos tenham ouvido o boato com probabilidade 0.999, para $n = 128$.

Exercício 147. Um gato e um rato passeiam aleatoriamente num grafo conexo e não-bipartido de modo independente. Eles começam ao mesmo tempo em vértices diferentes e ambos dão um passo a cada instante, também de modo independente. O gato come o rato se em algum instante eles estão no mesmo vértice. Mostre que o número esperado de passos até els se encontrarem é $O(m^2 n)$, em que n é o número de vértice e m o número de arestas de G. (Dica: considere a cadeia de Markov com estados (u, v) sendo as posições de cada animal no grafo, num instante.)

Exercício 148. O grafo *lollipop* é um grafo completo com n vértices e um caminho com n vértices e um único vértice u em comum. Chamamos de v o úico vértice de grau um, o vértice final do caminho.

Mostre que se um passeio aleatório começa em v então o número esperado de passos até visitar todos os vértices é $O(n^2)$.

Mostre que se um passeio aleatório começa em u então o número esperado de passos até visitar todos os vértices é $O(n^3)$.

§10 Percolação. O modelo de percolação apareceu em meados da década de 1950, foi introduzido por Broadbent e Hammersley no estudo de fenômenos de transporte de um fluido através de um meio poroso como, por exemplo, o petróleo através de uma rocha, a água através de um filtro de areia, a água quente através do pó de café, a difusão de uma epidemia num povoado. Aqui, ao contrário do passeio aleatório num grafo, não é o movimento que é aleatório, mas o meio. O meio é representado por grafo $G = (V, E)$ que modela uma rede de poros ligados por canais onde escoam um fluido. Na terminologia da área os vértices do grafo são chamados de *sítios* e representam os poros do meio e as arestas do grafo são chamados de *elos* e representam os canais.

O modelo probabilístico para o problema consiste em considerar que cada elo pode, aleatoriamente, estar *aberto*, o que permite a passagem do fluido, ou *fechado* não permitindo a passagem do fluido. Cada elo e tem associado uma probabilidade p_e de estar aberto. Um caminho no grafo é *aberto* se todos os elos desse caminho estão abertos. Dois sítios estão *conectados* se há um caminho aberto que os contém. Um subconjunto maximal de sítios conectados é um *aglomerado* e o *tamanho* de um aglomerados é o número de vértices nele. O problema fundamental é determinar a probabilidade de haver caminhos atravessando o meio, especificamente, queremos saber se ocorre um aglomerado

C de tamanho infinito, isto é, se ocorre *percolação* o que significa dizer que o meio se torna permeável ao fluido.

Nos grafos $G = (V, E)$ que consideramos aqui V têm uma quantidade infinita enumerável de vértices, mas cada vértice tem quantidade finita de vizinhos, isto é, os vértices têm grau finito, portanto E é enumerável. Os elos estão abertos independentemente um dos outros e com a mesma probabilidade p. O *subgrafo aberto* é o subgrafo induzido pelos elos abertos e é denotado G_p. Se C_x é o aglomerado de G_p que contém o vértice x, estamos interessado em

$$\theta_x(p) := \mathbb{P}(|C_x| = \infty)$$

que, claramente, também depende de G e que chamamos de *probabilidade de percolação*.

As seguintes notações são usuais em textos sobre percolação: escrevemos $x \leftrightarrow y$, para quaisquer dois vértices x e y, se eles estão conectados, isto é, ligados por um caminho aberto. Para qualquer $A \subset V$ a fronteira de A, denotada ∂A, é o subconjunto dos vértices de A que têm vizinhos em $\overline{A}$.

O modelo probabilístico:: o espaço amostral é $\{0, 1\}^E$, a σ-álgebra é a gerada pelos cilindros e a medida de probabilidade $\mathbb{P}_p$ é a medida produto (como definidos na página 115).

O aglomerado C_x tem tamanho infinito no subgrafo aberto se, e só se, existe um caminho a partir de x com um número infinito de vértices, o que denotamos por $x \to \infty$. Claramente, se $x \to \infty$ então $|C_x| = \infty$; a recíproca decorre do grau ser finito e de caminho não repetir vértice. Agora, seja A_n o evento "x está conectado a um vértice de G_p a distância n", então

$$[|C_x| = \infty] = [x \to \infty] = \bigcap_{n>0} A_n$$

portanto, $[|C_x| = \infty]$ é de fato um evento aleatório do modelo.

Exemplo 159. Tomemos $Z = (\mathbb{Z}, \{\{i, j\} : |i - j| = 1\})$ com as arestas abertas independentemente uma das outras e com a mesma probabilidade p. Notemos que $|C_x|$ e $|C_y|$ são variáveis aleatórias identicamente distribuídas.

Seja C o aglomerado que contem a origem, $C_+ \subset C$ e $C_- \subset C$ são, respectivamente, dados pelos sítios inteiros positivos e pelos inteiros negativos de C. O tamanho $|C_+|$ é uma variável aleatória e $\mathbb{P}_p(|C_+| \geqslant k) = p^k$, logo $\mathbb{P}_p(|C_+| = \infty) = 0$ se $p < 1$. Como C_- é independente de C_+ e tem a mesma distribuição, vale a mesma conclusão. Como $|C| = |C_+| + |C_-| + 1$ não há aglomerados infinitos quando $p < 1$. Se $p = 1$, então $C = \mathbb{Z}$.

Para G fixo e $0 \leqslant p < p' \leqslant 1$ podemos construir um acomplamento das medidas de probabilidade $\mathbb{P}_p$ e $\mathbb{P}_{p'}$: tomamos variáveis aleatórias independentes $U_e \sim$ Uniforme$([0, 1])$ do espaço de probabilidade $([0, 1], \mathcal{B}([0, 1]), \mathbb{P}_{\mathcal{U}})$ (definido na página 95). Um elo e é p-aberto se, e somente se, $U_e < p$, caso contrário é p-fechado. Assim, para todo $e \in E$

$$\mathbb{P}_{\mathcal{U}}(e \text{ é } p\text{-aberto}) = \mathbb{P}_{\mathcal{U}}(U_e < p) = p = \mathbb{P}_p(e \text{ é aberto}),$$

podemos realizar G_p como o subgrafo de G gerado pelos elos p-abertos. Ademais, se e é p-aberto então é p'-aberto, pois $U_e < p < p'$, de modo que

$$G_p \text{ é subgrafo de } G_{p'}$$

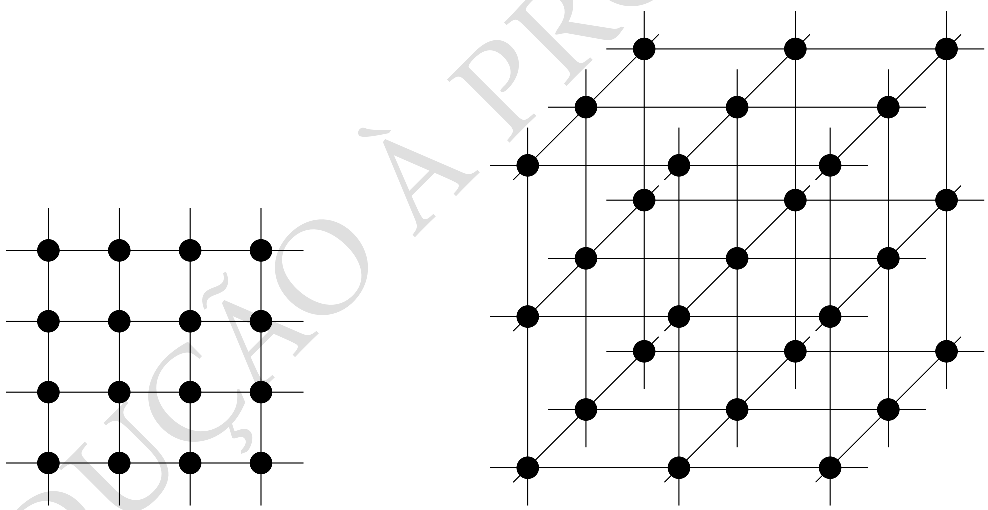
portanto

$$\theta_x(p) = \mathbb{P}_p(|C_x| = \infty) \leqslant \mathbb{P}_{p'}(|C_x| = \infty) = \theta_x(p')$$

ou seja,

3.0.46 Proposição. *$\theta(p)$ é não decrescente como função de p.* □

Modelo hipercúbico: No modelo de Broadbent e Hammersley, $\mathcal{Z}(d) = (\mathbb{Z}^d, E^d)$, os sítios são os pontos do $\mathbb{Z}^d$, com $d = 1, 2, \dots$ e os elos $E^d = \{\{x, y\} : x \in \mathbb{Z}^d,\ y \in \mathbb{Z}^d,\ \|x - y\|_1 = 1\}$. O caso $d = 1$ está no exemplo 159 acima. Abaixo representamos uma região no caso $d = 2$ e $d = 3$. No $\mathbb{Z}^2$, por exemplo, os sítios são os pontos (i, j) com coordenadas inteiras e os seus vizinhos são $(i, j - 1)$, $(i, j + 1)$, $(i - 1, j)$ e $(i + 1, j)$.



Notemos que a medida de probabilidade $\mathbb{P}_p$ é invariante por translação nesse modelo ($\mathcal{Z}(d)$ é isomorfo a $\mathcal{Z}(d) + (a, b)$) de modo que C_x e C_y têm a mesma distribuição o que nos faz referirmos sempre ao aglomerado C que contém a origem. Assim, a probabilidade de percolação θ_x não depende de x, mas depende de d, de modo que $\theta_d(p) = \mathbb{P}_p(|C| = \infty)$.

Transição de fase:: uma característica desse modelo para percolação é a existência de um limiar, ou valor crítico, $p_c \in (0, 1)$ onde ocorre transição de fase, ou seja, $\theta_d(p) = 0$ para $p < p_c$ e $\theta_d(p) > 0$ para $p > p_c$ $(d \geqslant 2)$.

A probabilidade de $\mathcal{Z}(d)_p$ ter um caminho de comprimento n a partir da origem pode ser calculado da seguinte maneira: a partir da origem há $2d$ possíveis primeiro passo, e a partir desses há no máximo $2d-1$ sítios possíveis para o próximo passo, portanto são no máximo $2d(2d-1)^{n-1}$ candidatos a caminho e cada um deles ocorre com probabilidade p^n no subgrafo aberto, de modo que a probabilidade de ocorrer algum deles é no máximo

$$2d(2d-1)^{n-1}p^n = \frac{2d}{2d-1}((2d-1)p)^n$$

e se $p < 1/(2d)$ então

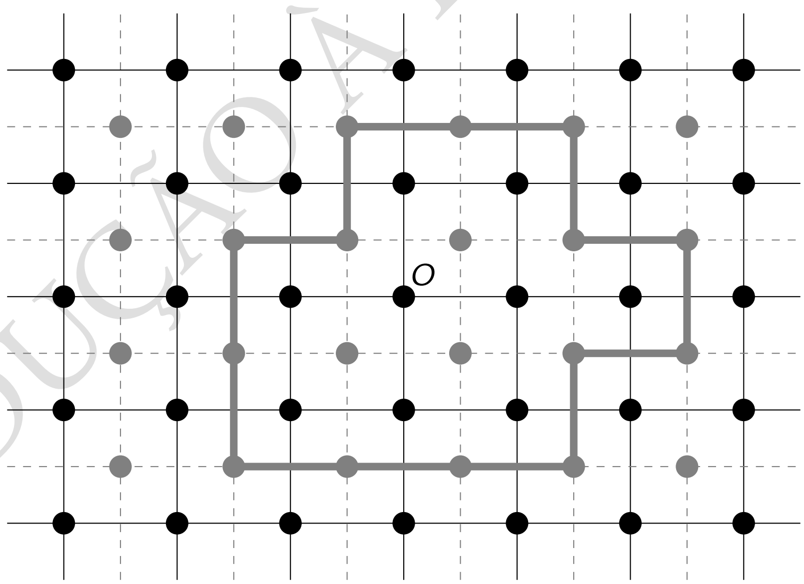
$$\frac{2d}{2d-1}((2d-1)p)^n < \frac{2d}{2d-1}\left((2d-1)\frac{1}{2d}\right)^n = \left(\frac{2d-1}{2d}\right)^{n-1}$$

que, como a $(2d-1)/(2d) < 1$, a probabilidade tende a 0 quando $n \to \infty$. Logo, se p for suficientemente pequeno vale que $\theta_d(p) = 0$, isto é, não há percolação.

3.0.47 Proposição. *Para todo* $d \geqslant 1$, *existe* $p^* \in (0,1]$ *tal que se* $p < p^*$ *então* $\theta_d(p) = 0$. □

O próximo passo é mostrar que $\theta_d(p) > 0$ para p suficientemente grande.

Começaremos com o caso $d = 2$. Aqui será conveniente considerarmos o grafo dual de $\mathcal{Z}(2)$ que denotamos $\mathcal{Z}(2)^*$; a maneira mais prática de definirmos esse grafo é tomando o conjunto de vértices (sítios) $\mathbb{Z}^2 + (1/2, 1/2)$ e as correspondentes arestas transladadas de E^2 por $(1/2, 1/2)$. Uma representação parcial desse grafo está na figura a seguir em cinza com elos tracejados (em preto temos $\mathcal{Z}(2)$).



Não haverá percolação no $\mathcal{Z}(2)_p$ se, e somente se, há algum circuito C em $\mathcal{Z}(2)^*$ que contorna a origem e tal que todas os elos de $\mathcal{Z}(2)$ que encontram C estão fechados.

Notemos que tais circuitos que contornam a origem têm comprimento par. Para especificar cada um desses circuitos com $2n$ elos ($n \geqslant 2$) basta: (i) tomarmos a coordenada $x + 1/2$ mais a direita na abcissa onde o circuito a cruza (na figura acima $x = 2$), verifique que $x \in \{0, 1, \ldots, n-2\}$ — são

$n-1$ possíveis casos — (ii) a partir desse ponto $x+1/2$ especificar as $2n$ direções dos elos no sentido anti-horário, a primeira é *para cima* e as restantes têm três possíveis opções, totalizando 3^{2n-2} casos. Dessa dedução, o número de tais circuitos é no máximo $(n-1)3^{2n-2}$; como cada aresta do circuito encontra exatamente um elo de $\mathcal{Z}(2)$ tais $2n$ elos de $\mathcal{Z}(2)$ devem estar fechados, o que ocorre com probabilidade $(1-p)^{2n}$, logo a probabilidade de não haver percolação é no máximo

$$\sum_{n\geqslant 2}(n-1)3^{2n-2}(1-p)^{2n} = \frac{1}{3^2}\sum_{n\geqslant 2}(n-1)(3(1-p))^{2n}.$$

Para avaliar essa soma usaremos que $\sum_{n\geqslant 0} x^n = 1/(1-x)$ para todo $x \in (0,1)$. Façamos $x = (3(1-p))^2$ e temos para todo $p > 2/3$

$$\frac{1}{3^2}\sum_{n\geqslant 2}(n-1)(3(1-p))^{2n} = \frac{1}{3^2}\sum_{n\geqslant 2}(n-1)x^n = \frac{1}{3^2}\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{n\geqslant 0} x^n\right) = \frac{1}{3^2}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right) = \left(\frac{x}{3(1-x)}\right)^2$$

a função $\left(\frac{x}{3(1-x)}\right)^2$ é crescente em $(0,1)$ e vale 1 para $x = 3/4$, portanto, a probabilidade de não haver percolação é menor que 1 para $(3(1-p))^2 < 3/4$, isto é, $p > 1 - 1/(2\sqrt{3})$ que vale, aproximadamente, $0{,}72$. Assim, a probabilidade de percolação é positiva para todo $p > 0{,}72$.

3.0.48 Proposição. *Existe $p^* \in (0,1)$ tal que se $p > p^*$ então $\theta_2(p) > 0$.* □

O que podemos dizer para $d > 2$? Que o mesmo vale e isso decorre do fato de $\theta_d(p)$ também ser não decrescente no parâmetro d (e p fixo). Basta observarmos que em $\mathcal{Z}_{d+1}$ se considerarmos os subgrafo Z induzido pelos sítios $(z_1,\dots,z_{d+1}) \in \mathbb{Z}^{d+1}$ com$z_{d+1} = 0$ temos que Z é isomorfo a $\mathcal{Z}_d$ e contém a origem, portanto $\theta_{d+1}d(p) \leqslant \theta_d(p)$. Com isso, provamos

3.0.49 Teorema. *Para todo $d \geqslant 2$, existe $p_c \in (0,1)$ tal que*

(3.0.36)
$$\theta_d(p)\begin{cases} = 0, & \text{se } p < p_c \\ > 0, & \text{se } p > p_c. \end{cases}$$

Demonstração. Fixamos $d \geqslant 2$ e fazemos $p_c = \sup\{p : \theta_d(p) = 0\}$. O resultado em equação (3.0.36) segue de $\theta_d(p)$ não decrescente em p. Ademais, as deduções que levaram às proposições acima implicam em $0 < 1/(2d) \leqslant p_c \leqslant 1 - 1/(2\sqrt{3}) < 1$. □

Percolação em árvore: Processo de ramificação $p = 1/2$

Cadeias de Markov na Biologia. Modelos probabilísticos são apropriados para sistemas de alta complexidade, como os frequentemente encontrados na biologia. A seguir, mostraremos alguns exemplos onde cadeias de Markov têm sido utilizadas para modelar processos biológicos, no estudo de crescimento populacional, epidemias e herança genética. Destacamos que alguns destes modelos são simplificados de maneira a deixá-los matematicamente tratáveis. Entretanto, por prover um entendimento quantitativo de vários fenômenos, eles fornecem uma contribuição significativa.

Processos de ramificação: Originalmente, processos de ramificações foram considerados por Galton e Watson nos anos de 1870 quando estes procuravam uma explicação quantitativa para o fenômeno do desaparecimento de sobrenomes, mesmo em uma população crescente. Sob a suposição de que cada homem em uma dada família tem a probabilidade p_k de ter k filhos, então desejamos determinar a probabilidade que após n gerações um indivíduo não tenha descendentes machos.

O modelo básico de processos de ramificação tem muitas aplicações em problemas de crescimento populacional, e também no estudo de reações em cadeia na química e na fissão nuclear. Suponhamos que no tempo $n = 0$ há um indivíduo que morre e é substituído no tempo $n = 1$ por um número aleatório de descendentes N. Suponhamos agora que esses também morrem e são substituídos no tempo $n = 2$, cada um, independentemente, por um número aleatório de novos descendentes que possuem a mesma distribuição de N, e assim por diante. Podemos construir o processo tomando para cada $n \in \mathbb{R}$ uma sequência de variáveis aleatórias

$$\{N_k^n\}_{k\in\mathbb{N}}$$

cada uma com a mesma distribuição de N, definindo $X_0 = 1$ e definindo, indutivamente, para $n \geq 1$

$$X_n = N_1^n + \ldots + N_{X_{n-1}}^n.$$

Então X_n nos dá o tamanho da população na n-ésima geração. O processo X_n, $n \geq 0$ é uma cadeia de Markov em $I = \{0, 1, 2, \ldots\}$ com estado absorvente 0. O caso onde $\mathbb{P}(N = 1) = 1$ é trivial então excluímos ele. Então temos

$$\mathbb{P}(X_n = 0 \mid X_{n-1} = i) = \mathbb{P}(N = 0)^i$$

então se $\mathbb{P}(N = 0) > 0$ temos i tendendo para 0, e cada estado $i \geq 1$ é transiente. Se $\mathbb{P}(N = 0) = 0$ então $\mathbb{P}(N \geq 2) > 0$, então para cada $i \geq 1$, i tende à j para algum $j > i$, e j não tende para i, uma vez que i é transiente em qualquer caso. Deduzimos que com probabilidade 1 que $X_n = 0$ para algum n ou $X_n \to \infty$ quando $n \to \infty$.

Mais informações sobre $(X_n)_{n\geq 0}$ é obtida explorando a estrutura de ramificações. Considere a função de geração de probabilidade

$$\phi(t) = \mathbb{E}(t^N) = \sum_{k=0}^{\infty} t^k \mathbb{P}(N = k),$$

definida para $0 \leq t \leq 1$. Condicionalmente em $X_{n-1} = k$ temos

$$X_n = X_1^n + \cdots + N_k^n$$

então

$$\mathbb{E}(t^{X_n} | X_{n-1} = k) = \mathbb{E}(t^{N_1^n + \cdots + N_k^n}) = \phi(t)^k$$

e então

$$\mathbb{E}(t^{X_n}) = \sum_{k=0}^{\infty} \mathbb{E}(t^{X_n}|X_{n-1}=k)\mathbb{P}(X_{n-1}=k) = \mathbb{E}(\phi(t)^{X_{n-1}}).$$

Portanto, por indução, encontramos que $\mathbb{E}(t^{X_n}) = \phi^{(n)}(t)$, onde $\phi^{(n)}$ é a composição de n-termos $\phi \circ \cdots \circ \phi$. Em princípio, isto nos dá toda a distribuição de X_n, onde $\phi^{(n)}$ pode ser uma função bastante complicada. Algumas quantidades são facilmente deduzidas: temos

$$\mathbb{E}(X_n) = \lim_{t\uparrow 1} \frac{d}{dt}\mathbb{E}(t^{X_n}) = \lim_{t\uparrow 1} \frac{d}{dt}\phi^{(n)}(t) = \left(\lim_{t\uparrow 1}\phi'(t)\right)^n = \mu^n,$$

onde $\mu = \mathbb{E}(N)$; ainda

$$\mathbb{P}(X_n = 0) = \phi^{(n)}(0)$$

então, uma vez que 0 é absorvente, temos

$$q = \mathbb{P}(X_n = 0 \text{ for some } n) = \lim_{n\to\infty} \phi^{(n)}(0).$$

Agora $\phi(t)$ é uma função convexa com $\phi(1) = 1$. Façamos $r = \inf\{t \in [0,1] : \phi(t) = t\}$, então $\phi(r) = r$ por continuidade. Uma vez que ϕ está crescendo e $0 \le r$, temos $\phi(0) \le r$ e, por indução, $\phi^{(n)}(0) \le r$ para todo n, desde que $q \le r$. Por outro lado,

$$q = \lim_{n\to\infty} \phi^{(n+1)}(0) = \lim_{n\to\infty} \phi(\phi^{(n)}(0)) = \phi(q)$$

desde que $q \ge r$. Portanto $q = r$. Se $\phi'(1) > 1$ então precisamos ter $q < 1$, e se $\phi'(1) \le 1$ uma ve que $\phi'' = 0$ ou $\phi'' > 0$ em todo intervalo $[0,1)$ precisamos ter $q = 1$. Mostramos que a população sobrevive com uma probabilidade positiva se e só se $\mu > 1$, onde μ é a média da distribuição dos descendentes.

Epidemias: Muitas doenças infecciosas persistem com uma baixa intensidade em uma população por longos períodos. Ocasionalmente, um grande número de casos ocorrem simultaneamente na forma de uma epidemia.

Para responder algumas questões quantitativas, que são importantes na previsão do comportamento de epidemias, podemos supor uma população idealizada onde todos os pares de indivíduos fazem contato aleatoriamente e independentemente à uma taxa comum, estejam estes indivíduos infectados ou não. Para uma doença idealizada, podemos supor que um indivíduo se torna infectado ao fazer contato com um indivíduo infectado, e assim se mantém por um tempo aleatório exponencial, após o qual ele irá morrer ou se recuperar. Este modelo idealizado é claramente não-realista, mas é o modelo matemático mais simples para incorporar as características básicas de uma epidemia.

Vamos denotar o número de indivíduos susceptíveis por S_t e o número de infectados por I_t. No modelo idealizado, $X_t = (S_t, I_t)$ compõe uma cadeia de Markov em $(\mathbb{Z}^+)^2$ com taxas de transição

$$q_{(s,i)(s-i,i+1)} = \lambda s i, \quad q_{(s,i)(s,i-1)} = \mu i$$

para algum $\lambda, \mu \in (0, \infty)$. Desde que $S_t + I_t$ não aumente, temos um espaço de estados finitos. Os estados $(s, 0)$ para $s \in \mathbb{Z}^+$ são todos absorventes e todos os outros estados são transientes; de fato, todas as classes de comunicação são únicas. A epidemia precisa eventualmente acabar, e as propabilidades de absorção da distribuição devem fornecer o número de indivíduos susceptíveis que escapam da infecção. Podemos calcular estas probabilidade explicitamente quando $S_0 + I_0$ é pequeno.

O comportamento da epidemia em uma população grande, digamos, de tamanho N, é de grande interesse. Consideremos as proporções $S_t^N = S_t/N$ e $i_t^N = I_t/N$ e suponhamos que $\lambda = \nu/N$, onde ν é independente de N. Consideremos agora uma sequencia de modelos com $N \to \infty$ e escolhemos $s_0^N \to s_0$ e $i_0^N \to i_0$. Pode ser mostrado que à medida que $N \to \infty$ o processo (s_t^N, i_t^N) converge para a solução (s_t, i_t) das equações diferencias

$$(d/dt)s_t = -\nu s_t i_t$$

$$(d/dt)i_t = \nu s_t i_t - \mu i_t$$

iniciando em (s_0, i_0). Aqui, convergência significa que $\mathbb{E}[|(s_t^N, i_t^N) - (s_t, i_t)|] \to 0$ para todo $t \geq 0$.

Considere o caso onde $S_0 = N - 1, I_0 = 1, \lambda = 1/N$ e $\mu = 0$. Podemos ter a seguinte interpretação: um rumor é iniciado por um único indivíduo que diz ele para todos que ele conhece; eles, por sua vez, contam o rumor para todo mundo que eles conhecem. Vamos assumir que cada indivíduo conhece outro aleatoriamente com um tempo que segue um processo de Poisson de taxa 1. Quanto tempo passa até que todo mundo conheça o rumor? Se i indivíduos conhecem o rumor, então $N - i$ não conhecem, e a taxa à qual o rumor é passado adiante é

$$q_i = i(N - i)/N.$$

Então, o tempo esperado até que todos tenham conhecido o rumor é

$$\sum_{i=1}^{N-1} q_i^{-1} = \sum_{i=1}^{N-1} \frac{N}{i(N-i)} = \sum_{i=1}^{N-1} \left(\frac{1}{i} + \frac{1}{N-i} \right) = 2 \sum_{i=1}^{N-1} \frac{1}{i} \sim 2 \log N$$

a medida que $N \to \infty$. Este nãoé um limite propriamente dito, mas, mais do que isto, uma equivalência assintótica. O fato de que o tempo esperado cresce com N esta relacionado ao fato de que não alteremos I_0 conforme N: quando um rumor é conhecido por um pequeno grupo ou por praticamente todos, a proporção de "infectados"se altera muito lentamente.

Modelo de Wright–Fisher: Este modelo é uma cadeia de Markov de tempo discreto com probabilidades

$$p_{ij} = \binom{m}{0ptj} \left(\frac{i}{m} \right)^j \left(\frac{m-i}{m} \right)^{m-j}.$$

Em cada geração há m alelos, alguns do tipo A e outros do tipo a. Os tipos de alelos na geração $n + 1$ são encontrados escolhendo aleatoriamente (com reposição) a partir dos tipos da geração n. Se X_n

denota o número de alelos do tipo A na geração n, então $(X_n)_{n\geq 0}$ é uma cadeia de Markov com as probabilidades de transição acima.

Isto pode ser interpretado como um modelo de herança para um gene particular com dois alelos A e a. Suponhamos que cada indivíduo tennha dois genes, então as possibilidades são AA, Aa e aa. Façamos m ser oar com $m = 2k$. Supomos que indivíduos da próxima geração são obtidos pegando aleatoriamente pares de indivíduos da geração atual e que os descendentes herdam um alelo de cada indivíduo pai. Vamos permitir que ambos os pais talvez sejam o mesmo e não fazer nenhuma exigência de que os pais tenha sexos opostos. Então, se a geração n é, por exemplo

$$AA \quad aA \quad AA \quad AA \quad aa,$$

então cada gene da geração $n+1$ é A com probabilidade 7/10 e a com probabilidade 3/10, todos de maneira independente. Podemos ter, por exemplo,

$$aa \quad aA \quad Aa \quad AA \quad AA.$$

A estrutura de pares de genes é irrelevante para a cadeia de Markov $(X_n)_{n\geq 0}$, que simplesmente conta o número de alelos do tipo A.

As classes de comunicação de $(X_n)_{n\geq 0}$ são $\{0\}, \{1, \dots, m-1\}, \{m\}$. Os estados 0 e m são absorventes e $\{1, \dots, m-1\}$ são transientes. A probabilidade de ocorrência do estado m (AA puro) é dada por

$$h_i = \mathbb{P}_i(X_n = m \text{ para algum } n) = i/m.$$

Isto é trivial quando notamos que $(X_n)_{n\geq 0}$ é um martingal; alternativamente podemos verificar que

$$h_i = \sum_{j=0}^{m} p_{ij} h_j.$$

De acordo com este modelo, a diversidade genética eventualmente desaparece. É conhecido, entretanto, que para $p \in (0,1)$, à medida que $m \to \infty$

$$\mathbb{E}_{pm}(T) \sim -2m\{(1-p)\log(1-p) + p\log p\}$$

onde T é o tempo de ocorrência de $\{0, m\}$, então em uma grande população a diversidade não desaparece rapidamente.

São possíveis ainda modificações para se modelar outros aspectos da teoria genética. Poderíamos modelar a vantagem seletiva de determinados genótipos ou, ainda, permitir que ocorram mutações nos genes.

Modelo de Moran: O modelo de Moran é uma cadeia do tipo nascimento-e-morte em $\{0, 1, \ldots, m\}$ com probabilidades de transições:

$$p_{i,i-1} = i(m-i)/m^2, \quad p_{ii} = (i^2 + (m-i)^2)/m^2, \quad p_{i,i+1} = i(m-i)/m^2.$$

Uma possível interpretação genética: uma população consiste de indivíduos de dois tipos, a e A; escolhemos aleatoriamente um indivíduo da população no tempo n, e adicionamos um novo indivíduo do mesmo tipo; então escolhemos, novamente aleatoriamente, um indivíduo da população no tempo n e removemos ele; então obtemos a população no tempo $n+1$. O mesmo indivíduo pode ser escolhido em cada tempo, para gerar um novo indivíduo e para morrer, de tal maneira que não haja mudanças na composição da população. Agora, se X_n denota o número de indivíduos do tipo A na população no tempo n, então $(X_n)_{n\geq 0}$ é uma cadeia de Markov com a matriz de transição P.

A estrutura básica da cadeia de Markov é a mesma do modelo de Wright-Fisher, com classes de comunicação $\{0\}, \{1, \ldots, m-1\}, \{m\}$, estados absorventes 0 e m e classes transientes $\{1, \ldots, m-1\}$. O modelo de Moran é reversível e é um martingal. As probavilidades de ocorrências são dadas por

$$\mathbb{P}_i(X_n = m \text{ para algum } n) = i/m.$$

Podemos também calcular explicitamente o tempo médio de absorção

$$k_i = \mathbb{E}_i(T)$$

onde T é o tempo de ocorrência de $\{0, m\}$. O método mais fácil é primeiro fixar j e escrever as equações para o tempo médio k_i^j decorrido em j, iniciando de i, antes da absorção:

$$k_i^j = \delta_{ij} + (p_{i,i-1}k_{i-1}^j + p_{ii}k_i^j + p_{i,i+1}k_{i+1}^j) \text{ para } i = 1, \ldots, m-1$$

$$k_0^j = k_m^j = 0.$$

Então, para $i = 1, \ldots, m-1$

$$k_{i+1}^j - 2k_i^j + k_{i-1}^j = -\delta_{ij} m^2/j(m-j)$$

de maneira que

$$k_i^j = \begin{cases} (i/j)k_j^j & \text{para } i \leq j \\ ((m-i)/(m-j))k_j^j & \text{para } i \geq j \end{cases}$$

onde k_j^j é determinado por

$$\left(\frac{m-j-1}{m-j} - 2 + \frac{j-1}{j}\right)k_j^j = -\frac{m^2}{j(m-j)}$$

que nos dá $k_j^j = m$. Pois

$$k_i = \sum_{j=1}^{m-1} k_i^j = m\left\{\sum_{j=1}^{i}\left(\frac{m-i}{m-j}\right) + \sum_{j=i+1}^{m-1}\frac{i}{j}\right\}.$$

Estamos interessados no caso em que m é grandem e $i = pm$ para algum $p \in (0,1)$. Então

$$m^{-2}k_{pm} = (1-p)\sum_{j=1}^{mp}\frac{1}{m-j} + p\sum_{j=mp+1}^{m-1}\frac{1}{j} \to -(1-p)\log(1-p) - p\log p$$

à medida que $m \to \infty$. Então, quando $m \to \infty$

$$\mathbb{E}_{pm}(T) \sim -m^2\{(1-p)\log(1-p) + p\log p\}.$$

Exercício 149. Considere o processo de ramificações com imigração definido por

$$X_n = N_1^n + \cdots + N_{X_{n-1}}^n + I_n$$

onde $(I_n)_{n\geq 0}$ é uma sequencia de variáveis aleatórias independentes em $\mathbb{Z}^+$ com função geradora $\psi(t) = \mathbb{E}(t^{I_n})$. Mostre que, se $X_0 = 1$, então

$$\mathbb{E}(t^{X_n}) = \phi^{(n)}(t)\prod_{k=0}^{n-1}\psi(\phi^{(k)}(t)).$$

No caso em que o número de imigrante em cada geração segue uma distribuição de Poisson de parâmetro λ, e onde $\mathbb{P}(N=0) = 1-p$ e $\mathbb{P}(N=1) = p$, encontre a proporção de tempo do evento durante a qual a população é zero.

Exercício 150. Uma espécie de planta pode ter três genótipos: AA, Aa e aa. Uma única planta de genótico Aa é cruzada com ela mesma, de tal forma que os descendentes têm genótipos AA, Aa ou aa com probabilidades 1/4, 1/2 e 1/4, respectivamente. Quanto tempo em média demora para se obter uma muda pura, isto é, AA ou aa? Suponha que desejamos obter uma planta AA. O que devemos fazer? Quantos cruzamentos são necessários em média?

Exercício 151. No modelo de Moran podemos introduzir um fator de seleção fazendo com que o indivíduo do tipo a tenha o dobro de chance de ser escolhido para morrer, se comparado com o indivíduo do tipo A. Então em uma população de tamanho m contendo i indivíduos do tipo A, a probabilidade de que alguns do tipo A sejam escolhidos para morrer é agora $i/(i+2(m-i))$. Suponhamos que inicialmente temos apenas um indivíduo do tipo A. Qual a probabilidade de que eventualmente a população inteira seja do tipo A?

Índice Remissivo