Probabilidade

Jair Donadelli Jr

Probabilidade, ingenuamente, é uma forma quantitativa de expressar o conhecimento ou crença de que um evento ocorra e que está, normalmente, vinculada a interpretações como, por exemplo 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 de ocorrência do evento; a **subjetiva** que atribui uma probabilidade a qualquer evento, mesmo quando não há aleatoriedade envolvida, a qual representa um grau de crença, quanto maior a probabilidade mais plausível é sua ocorrência. Felizmente, para quase todas as interpretações importantes as mesmas regras se aplicam e propriedades matemáticas valem, 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 regras que independem de interpretações, de fato, usaremos a abordagem moderna que trata probabilidade axiomaticamente e as propriedades são deduzidas dos axiomas.

Um jogo de azar marca o começo do estudo sistemático de probabilidades pelos famosos matemáticos franceses Blaise Pascal (2623–1662) e Pierre de Fermat (1601-1665). 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. Os problemas colocados por de Méré levou a uma troca de cartas entre Pascal e Fermat em que os princípios fundamentais de Probabilidades foram formuladas pela primeira vez. Vários historiadores definem esse momento como o início da Teoria das Probabilidades apesar de alguns problemas específicos sobre jogos de azar terem sido resolvidos por alguns matemáticos italianos, como Fra Luca Paccioli (1447–1517) e Geronimo Cardano (1501–1576), nos séculos 15 e 16. Cálculos de probabilidade também aparecem nas obras de Kepler (1571–1630) e Galileo (1564–1642).

O holandês Christian Huygens (1629–1695) publicou o primeiro livro sobre probabilidade em 1657, intitulado *De Ratiociniis no Ludo Aleae*, um tratado sobre problemas relacionados com o jogo que estende o trabalho de Pascal e Fermat. Por causa do apelo inerente aos jogos de azar o assunto se desenvolveu rapidamente durante o século 18. Os principais contribuintes durante este período foram Jakob Bernoulli (1654–1705) e Abraham de Moivre (1667–1754) cujo livro *A Doutrina da Chance*, de 1718, era popular e passou por três edições. Bayes (1671–1746) introduziu o conceito de probabilidade condicional. Em 1812 Pierre de Laplace (1749–1827) introduziu uma série de novas ideias e técnicas matemáticas em seu livro *Théorie des Analytique Probabilités* que permite a disciplina ultrapassar sua primeira fase combinatória. Laplace aplicou ideias probabilísticas em muitos problemas científicos e práticos. Após a publicação do livro de Laplace, o desenvolvimento matemático da Teoria das Probabilidades não teve grandes avanços por muitos anos. Em 1850, vários matemáticos consideraram o método clássico não realista para e redefiniram probabilidade em termos de frequências.

O desenvolvimento da teoria utilizando os métodos de análise ocupa o 19 século e início do século 20, notadamente com base no trabalho de Borel (1871–1956) e Lebesgue (1875–1941) sobre a teoria da medida. A Teoria dos Erros e a Mecânica Estatística (Maxwell (1831–1879), Boltzmann (1844–1906)) são exemplos de aplicações importantes desenvolvidas no século 19. No final do século 19 Galton (1822–1911) e Watson (1827–1903) estudam a evolução do número de indivíduos de uma população durante suas gerações, um exemplo de processo aleatório que evolui ao longo do tempo que será introduzido em toda a sua generalidade por Markov (1856–1922).

Como tantos outros ramos da matemática, o desenvolvimento da Teoria da Probabilidade tem sido estimulado pela variedade das suas aplicações. Muitos matemáticos têm contribuído para a teoria desde a época de Laplace; entre os mais importantes são Chebyshev (1821–1894), Markov (1856–1922) e Kolmogorov (1903–1987).

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 da probabilidade em uma base axiomática, apresentada em 1933 numa monografia escrita pelo matemático russo Andrey Kolmogorov, que forma a base para a teoria moderna. Após o trabalho fundamental de Kolmogorov, Paul Levy (1886–1971) deu o tom para o trabalho moderno sobre processos estocásticos.

Em 2006 Wendelin Werner foi o primeiro matemático que trabalha com 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 que um matemático pode receber.

O seguinte exemplo é a história referida acima da época em que começou-se a estudar probabilidade.

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 quatro 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 é a Lei dos Grandes Números.

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.

Espaço amostral: o espaço amostral de um experimento, denotado por Ω, é um conjunto que representa *todos* os resultados possíveis do experimento. Um elemento de Ω é chamado de *ponto amostral*.

Exemplo 2. São experimentos com os respectivos espaços amostrais

1. o lançamento de um dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$;
2. o lançamento de uma moeda, $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}$;
3. lançamento de moeda até sair Coroa
$$\Omega = \{(\mathrm{Co}), (\mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}, \mathrm{Co}), \dots, (\mathrm{Ca}, \mathrm{Ca}, \mathrm{Ca}\dots)\};$$
4. tempo de vida de uma lâmpada, $\Omega = \{t \in \mathbb{R} \colon t \geqslant 0\}$;
5. a altura em metros de um brasileiro escolhido ao acaso, $\Omega = \{h \in \mathbb{R} \colon h > 0\}$;
6. um ponto escolhido no círculo de raio 1 e centro na origem do plano cartesiano
$$\Omega = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \leqslant 1\}.$$

$\Diamond$

O espaço amostral de um experimento aleatório não é único; no item 6, podemos escrever o espaço amostral como $\{(r,\theta) \in \mathbb{R}^2 \colon 0 \leqslant r \leqslant 1 \text{ e } -\pi < \theta \leqslant \pi\}$. Ainda, o espaço amostral reflete o interesse da observação. Por exemplo, o sorteio de uma loteria com 100 números pode ser representado por $\Omega_1 = \{1, 2, \dots, 100\}$ ou por $\Omega_2 = \{\text{ganhei}, \text{não ganhei}\}$.

Em seguida veremos como atribuir uma probabilidade às coleções de pontos amostrais que estamos interessados em estudar. Para isso ser feito precisamos considerar dois casos de acordo com a cardinalidade do espaço amostral.

Espaço amostral discreto: se o espaço amostral é enumerável (finito ou infinito) então ele é chamado de espaço amostral discreto. De fato, na probabilidade clássica só o caso finito se faz interessante; justificaremos isso mais adiante.

Espaço amostral contínuo: os espaços infinitos não enumeráveis que consideraremos são os de cardinalidade do contínuo, por isso são chamados de espaço amostral contínuo.

Por exemplo, são espaços discretos os espaços amostrais dos experimentos 1, 2 e 3 dados no exemplo 2 acima. Os experimentos 4 e 5 têm espaços contínuos. Notemos que os resultados do experimento 4 estão superestimado no sentido de que, por exemplo, não é possível ocorrência do resultado 20,48 m. O que é *importante* no espaço amostral é que cada resultado esteja associado a um elemento de Ω e resultados diferentes estejam associados a elementos diferentes de Ω.

Eventos: um evento aleatório, associado a um experimento aleatório, é um subconjunto do espaço amostral Ω sobre o qual podemos dizer, quando da realização do experimento, se *ocorre* ou *não ocorre*. Na realização de um experimento o evento A *ocorre* se o resultado da execução é um elemento de A, por conseguinte, o evento complementar de A é o evento em que A *não* ocorre. Em especial

- $\emptyset$ é o evento *impossível* e Ω é o evento *certo*;
- o *complemento* do evento A é o evento *não-*A dado por $\overline{A} = \Omega \setminus A = \{\omega \in \Omega \colon \omega \notin A\}$.

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

- $A = \{2, 4, 6\}$, i.e., A é o evento “ocorre número par”;
- $\overline{A} = \{1, 3, 5\}$, i.e., $\overline{A}$ é o evento “não ocorre número par”;
- $B = \{4, 5, 6\}$, i.e., B é o evento “ocorre um resultado > 3”;
- $C = \{4\}$, i.e., C é o evento “ocorre o resultado 4”;

- $B \cap C = \{4\}$, i.e., $B \cap C$ é o evento "ocorre um número > 3 **e** ocorre 4";
- $B \cap A = \{4, 6\}$,i.e., $B \cap A$ é o evento "ocorre número > 3 **e** ocorre número par".

$\Diamond$

Temos o seguinte dicionário de linguagem de teoria de conjuntos para linguagem de probabilidade.

Notação	*Conjunto*	*Evento*
Ω	universo	espaço amostral, evento certo
$\emptyset$	vazio	evento impossível
ω	elemento	ponto amostral
$\{\omega\}$	conjunto unitário	evento elementar
A	subconjunto	ocorre A
$\overline{A}$	complemento	não ocorre A
$A \cap B$	interseção	ocorre A e B
$A \cup B$	união	ocorre A ou B
$A \Delta B$	diferença simétrica	ocorre A ou B, não ambos
$A \subset B$	inclusão	se ocorre A, então ocorre B

Exclusão mútua: A e B são eventos *mutuamente exclusivos* quando não têm elementos em comum, isto é, $A \cap B = \emptyset$. Uma sequência de eventos é dita de eventos mutuamente exclusivos se os eventos são *mutuamente exclusivos quando tomados dois-a-dois*.

Probabilidade em espaços equiprováveis finitos: no modelo probabilístico clássico para espaços amostrais finitos (não vazios) todos os pontos de um espaço amostral são igualmente prováveis de ocorrer como resultado da realização do experimento. Se um experimento tem n desfechos dos quais m correspondem a um determinado evento aleatório, então a probabilidade de ocorrer o evento é m/n, isto é, o subconjunto $A \subset \Omega$ ocorre com probabilidade

$$\mathbb{P}(A) \stackrel{\text{def}}{=} \frac{|A|}{|\Omega|} = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) \tag{1}$$

Para todo $\omega \in \Omega$, a probabilidade de ocorrer $\{\omega\}$ é $1/|\Omega|$. Usualmente, escrevemos $\mathbb{P}(\omega)$ para denotar $\mathbb{P}(\{\omega\})$.

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 $\mathbb{P}\colon 2^{\Omega} \to \mathbb{R}$ como definida em (1).

Decorre das considerações acima que valem as seguintes propriedades para $\mathbb{P}$

Segue da definição em (1) que

$$\mathbb{P}(A) \geqslant 0, \quad \text{para todo evento } A \tag{2}$$

$$\mathbb{P}(\Omega) = 1 \tag{3}$$

ademais, para eventos mutuamente exclusivos A e B,A ou B ocorre com probabilidade

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \tag{4}$$

Dessas propriedades podemos deduzir várias outras, por exemplo, de (4) e (3)

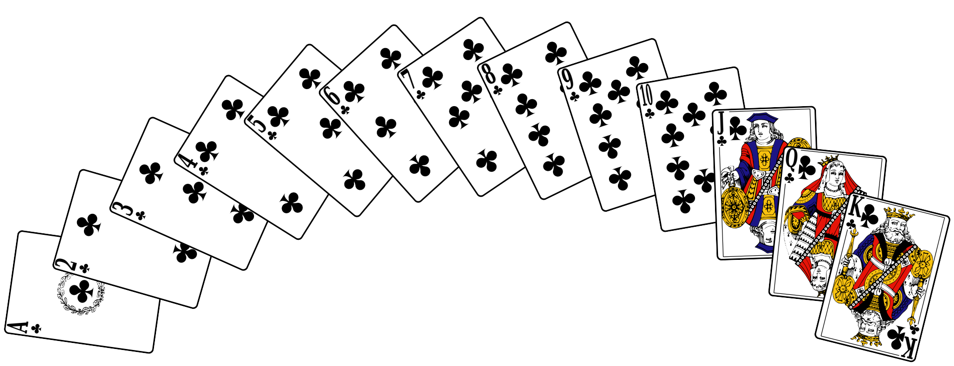
$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A). \tag{5}$$

Notação $\uplus$: em vista de (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 (4) como $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Exemplo 4 (*lançamento de uma moeda equilibrada*). Uma moeda equilibrada é lançada e observamos a face para cima. Um modelo probabilístico clássico para o experimento é $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}$ com $\mathbb{P}(\mathrm{Ca}) = \mathbb{P}(\mathrm{Co}) = 1/2$. $\Diamond$

Exemplo 5 (*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 "o lançamento resulta num número par" é a probabilidade do subconjunto $\{2, 4, 6\}$ que é $1/2$. $\Diamond$

Exemplo 6 (*escolha de uma carta de baralho tradicional*). Um baralho tem 52 cartas divididas em quatro famílias (ou naipes) representadas pelas figuras: $\diamondsuit$ (ouros), $\clubsuit$ (paus), $\heartsuit$ (copas) e $\spadesuit$ (espadas). Cada família traz os números de 2 a 10 e as letras A (ás), J (valete), Q (dama), K (rei).



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. ◊

Observação 7 (*escolha aleatória/sorteio aleatório*). Num espaço amostral finito há uma única maneira de definir probabilidade com todos os elementos do espaço amostral igualmente prováveis. Esta probabilidade define matematicamente a expressão intuitiva "aleatório"(como em *escolha aleatória* de uma carta de baralho, *sorteio aleatório* de uma bola numa urna, *lançamento aleatório* de dado, *escolha aleatória* de um indivíduo de uma população). Uma *escolha aleatória* é um resultado de um experimento idealizado com respostas equiprováveis. ◊

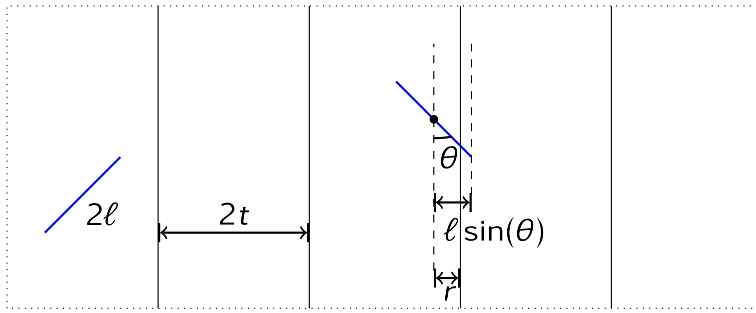
Notemos que a probabilidade de $A \subset \Omega$ não depende da sua natureza, depende apenas da sua cardinalidade $|A|$. O problema de determinar a probabilidade resume-se num problema de contagem, que é estudado no ramo da matemática conhecido como Combinatória.

Probabilidade geométrica: pode ser vista como o modelo clássico em espaços contínuos. No caso de espaço amostral contínuo, temos um pouco mais de trabalho pra definir um modelo precisamente e o principal objetivo nesse momento é chamarmos a atenção para alguns problemas de modelagem probabilística quando tratamos o caso contínuo.

No caso finito a probabilidade de um subconjunto do espaço amostral é proporcional a cardinalidade do subconjunto, num intervalo da reta ou numa região do plano ou do espaço a probabilidade é proporcional ao volume (comprimento ou área) do subconjunto. Podemos definir 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.

Atribui-se a Georges-Louis Leclerc, conhecido como Conde de Buffon (1707–1788), a proposição e resolução do problema que deu 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?

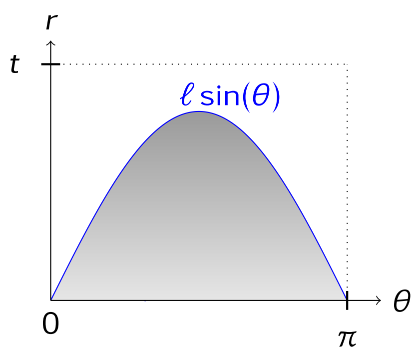
Sejam $r \in [0, t)$ a distância do centro da agulha até a divisória entre tábuas mais próxima, $\theta \in [0, \pi)$ o ângulo com que a agulha cai em relação às linhas paralelas.



O espaço amostral é dado pelos pares $(r,\theta) \in [0,t) \times [0,\pi)$ e os eventos de interesse incluem o produto $D \times A$ de intervalos $D = (d_1, d_2) \subset [0,t)$ e $A = (a_1, a_2) \subset [0,\pi)$; 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 $1/t\pi$ da área da região demarcada no gráfico abaixo, isto é,



$$\mathbb{P}(E) = \frac{1}{t\pi} \int_0^\pi \ell \sin\theta \, d\theta = \frac{2\ell}{t\pi}$$

Notemos que como no caso finito valem as propriedades listadas na página 6, equações (2), (3) e (4), também valem a equação (5).

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[1].

O fato importante para o qual chamamos a atenção é de que

pode ocorrer que nem todo subconjunto admita uma medida de probabilidade

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 (veja *A First Look at Rigorous Probability Theory*, de Jeffrey S Rosenthal, proposição 1.2.6)

[1]B. R. Gelbaum, J. M. H. Olmsted, *Counterexamples in Analysis*, capítulo 11

Paradoxo de Bertrand: diferente do caso finito, no caso contínuo uma *escolha aleatória* não define unicamente o modelo probabilístico, como exemplifica o fato conhecido como *Paradoxo de Bertrand*, no qual três interpretações diferentes para *escolha aleatória* leva a três resultados distintos.

Probabilidade frequentista: nesse modelo, a frequência relativa de ocorrência de um evento observada numa série de repetições de um experimento aleatório é uma media de probabilidade do evento. Isto é, se em n realizações do experimento o evento A ocorreu n_A vezes então $\mathbb{P}(A) \approx n_A/n$. Ademais, 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 $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Axiomas de probabilidade

Observamos anteriormente 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 e na prática podemos nos preocupar somente com os eventos admitem uma medida de probabilidade. A solução que adotamos 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* de subconjuntos de Ω, e que assuma valores reais e que satisfaça

positividade – $\mathbb{P}(A) \geqslant 0$ 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 $A \cup B \in \mathcal{A}$.

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 pra outro tanto de problemas.

Continuidade: vejamos um exemplo em que precisamos de um pouco mais do que descrevemos. Consideremos o seguinte experimento conceitual. Uma moeda é lançada indefinidamente. O espaço amostral do experimento é $\Omega = \{\mathrm{Ca}, \mathrm{Co}\}^{\mathbb{N}}$ e o evento de interesse é A dado por "jamais ocorre o resultado cara".

Seja A_n é o evento "não ocorre cara nos n primeiros lançamentos" cuja probabilidade é 2^{-n}. Notemos que a sequência de eventos $(A_n\colon n \geqslant 1)$ é decrescente, isto é, $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots$ e $A = \bigcap_{n\geqslant 1} A_n$; é natural supormos que

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) = 0$$

e para expressar esse fato precisamos de mais um axioma:

continuidade – para qualquer sequência decrescente de eventos $(A_n\colon n \geqslant 1)$

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

ou, escrevendo $\lim_{n\to\infty} A_n \stackrel{\text{def}}{=} \bigcap_{n=1}^{\infty} A_n$ para uma sequência decrescente de eventos, reescrevemos a condição acima como

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Ao invés de tomarmos positividade, normalização, aditividade finita e continuidade como axiomas para uma medida de probabilidade, usaremos a equivalência: aditividade finita juntamente com a continuidade equivalem a aditividade enumerável (enunciada abaixo). Nesse caso, o espaço de eventos deve incluir, além de Ω e do complemento dos eventos, as interseções enumeráveis de eventos aleatórios.

Espaço de eventos: consideramos como espaço de eventos de um espaço amostral Ω as famílias $\mathcal{A} \subset 2^{\Omega}$ tais que: (i) $\Omega \in \mathcal{A}$, (ii) se $A \in \mathcal{A}$ então $\overline{A} \in \mathcal{A}$ e (iii) se $(A_n\colon n \geqslant 1) \in \mathcal{A}$ então $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$. Nesse caso, $\mathcal{A}$ é uma *σ-álgebra* de subconjuntos de Ω.

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 (positividade) $\mathbb{P}(A) \geqslant 0$;

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 $\{A_i\colon i \geqslant 1\}$ de eventos mutuamente exclusivos.

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 $\mathbb{P}\colon \mathcal{A} \to [0,1]$.

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

$$\mathbb{P}(A) \stackrel{\text{def}}{=} \sum_{\omega_i \in A} \mathbb{P}(\omega_i)$$

Exemplo 8. 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 diferente de ∞ 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 9. 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

$$\sum_{\substack{a \in A \\ a \text{ par}}} \mathbb{P}(\{a\}) = \sum_{k \geqslant 1} \left(\frac{1}{2}\right)^{2k} = \sum_{k \geqslant 1} \left(\frac{1}{4}\right)^k = \frac{1}{3}.$$

◊

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. Um modelo probabilístico de um experimento aleatório também é chamado de *espaço de probabilidade*, isto é, é uma tripla $(\Omega, \mathcal{E}, \mathbb{P})$ composta por um espaço amostral Ω, uma σ-álgebra $\mathcal{E}$ de subconjuntos de Ω e uma medida de probabilidade $\mathbb{P} : \mathcal{E} \to \mathbb{R}$.

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.

Consequências dos axiomas: algumas consequências simples desses axiomas são dadas a seguir. A maioria delas são válidas se considerarmos $\mathbb{P}$ finitamente (ao invés de σ) aditiva.

Proposição 10 (**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 $\mathbb{P}(\emptyset) = 0$. □

Proposição 11 (**probabilidade de uma união finita de eventos disjuntos**). *Para quaisquer* $A_1, A_2, \dots, A_n$ *eventos mutuamente exclusivos*

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

Demonstração. Dados $A_1, A_2, \dots, A_n$ 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 $\mathbb{P}(\emptyset) = 0$. □

Proposição 12 (**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 11 que $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \uplus \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A})$. □

Proposição 13 (**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 11 $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(\overline{A} \cap B)$ e como $\mathbb{P}(\overline{A} \cap B) \geqslant 0$ temos $\mathbb{P}(B) \geqslant \mathbb{P}(A)$. □

Corolário 14. *Para todo evento* A, $\mathbb{P}(A) \leqslant 1$.

Demonstração. Basta notar que $A \subset \Omega$. $\square$

Proposição 15 (teorema da adição). *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 11. A união $A \cup B$ pode ser escrita como a $(A \setminus B) \uplus (B \setminus A) \uplus (A \cap B)$ donde concluímos que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(\overline{A} \cap B). \tag{6}$$

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 em (6) prova a proposição. $\square$

No caso da união de três eventos a probabilidade pode ser facilmente deduzida da proposição 15

$$\begin{aligned} \mathbb{P}((A \cup B) \cup C) &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) \\ &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cap C) \cup (B \cap C)) \end{aligned}$$

agora usamos a proposição 15 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

$$\begin{aligned} \mathbb{P}(A \cup B \cup C \cup D) = {}& \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(A \cap D) - \mathbb{P}(B \cap C) - \mathbb{P}(B \cap D) \\ & - \mathbb{P}(C \cap D) + \mathbb{P}(A \cap B \cap C) + \mathbb{P}(A \cap B \cap D) + \mathbb{P}(B \cap C \cap D) - \mathbb{P}(A \cap B \cap C \cap D). \end{aligned}$$

Proposição 16 (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*

$$\mathbb{P}\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})$$

Proposição 17 (**probabilidade é subaditiva – desigualdade de Boole**)**.** *Sejam* $A_1,A_2,\dots,A_n$ *eventos. Então*

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

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

$$\mathbb{P}\left(\bigcup_{i=1}^{n}A_i\right)=\mathbb{P}\left(A_n\cup\bigcup_{i=1}^{n-1}A_i\right)\leqslant\mathbb{P}(A_n)+\mathbb{P}\left(\bigcup_{i=1}^{n-1}A_i\right)\leqslant\mathbb{P}(A_n)+\sum_{i=1}^{n-1}\mathbb{P}\left(A_i\right)=\sum_{i=1}^{n}\mathbb{P}\left(A_i\right)$$

portanto, a desigualdade de Boole segue do Princípio da Indução Finita. □

Proposição 18 (**probabilidade é subaditiva – desigualdade de Boole**)**.** *Sejam* $A_1,A_2,\dots$ *eventos. Então*

$$\mathbb{P}\left(\bigcup_{i\geqslant 1}A_i\right)\leqslant\sum_{i\geqslant 1}\mathbb{P}(A_i).$$

Demonstração. Definimos a seguinte sequência de eventos mutuamente disjuntos: A_1, $A_2\setminus A_1$, $A_3\setminus(A_1\cup A_2)$, …, de modo geral

$$B_n\stackrel{\text{def}}{=}A_n\setminus\bigcup_{i=1}^{n-1}A_i$$

e temos

$$\mathbb{P}\left(\bigcup_{i\geqslant 1}A_i\right)=\mathbb{P}\left(\biguplus_{i\geqslant 1}B_i\right)=\sum_{i\geqslant 1}\mathbb{P}\left(B_i\right)\leqslant\sum_{i\geqslant 1}\mathbb{P}\left(A_i\right)$$

pois $B_n\subset A_n$ para todo inteiro positivo n. □

Sobre a continuidade $\mathbb{P}$: 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 dos eventos das sequências.

Lema 19. *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

$$\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty} A_n\right) &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \cdots \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_{i+1} \setminus A_i) \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n} (\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)) \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}$$

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

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{A_n}\right) = \lim_{n\to\infty} \mathbb{P}\left(\overline{A_n}\right) = \lim_{n\to\infty} \left(1 - \mathbb{P}\left(\overline{A_n}\right)\right) = 1 - \lim_{n\to\infty} \mathbb{P}\left(A_n\right)$$

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 $\lim_{n\to\infty} \mathbb{P}\left(A_n\right) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right)$ também vale no caso decrescente. □

De fato, a aditividade enumerável segue da aditividade finita mais a aditividade, ou seja, em termos de axiomas, **positividade, normalização e aditividade enumerável** equivale a **positividade, 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 $(A_i \colon i \geqslant 1)$ de eventos disjuntos.

Tomemos $B_n \stackrel{\text{def}}{=} \uplus_{i=1}^{n} A_i$ para cada $n \geqslant 1$ e temos uma sequência monótona de eventos. Por continuidade

$$\mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \mathbb{P}\left(\bigcup_{n\geqslant 1} B_n\right) = \lim_{n\to\infty} \mathbb{P}\left(B_n\right) =$$

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 $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Probabilidade condicional: consideremos uma escolha aleatória de uma família com dois filhos na qual observamos o sexo do filho mais novo e do filho mais velho. Qual é a probabilidade de ambos serem meninos sabendo-se que pelo menos um deles é menino? O espaço amostral é $\Omega = \{(♂,♂),(♂,♀),(♀,♂),(♀,♀)\}$ e sabendo que um dos filhos é menino, a resposta é um elemento do conjunto $\{(♂,♂),(♂,♀),(♀,♂)\}$, portanto a probabilidade de ambos serem meninos é $1/3$.

Agora, suponhamos que alguém bate à porta de uma família que tem dois filhos, um dos filhos 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 $1/2$. Analogamente, se um dos filhos abre a porta e diz "eu sou o filho mais novo" a probabilidade de ambos serem meninos é $1/2$. 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?

Retomemos o modelo clássico para motivar a definição de probabilidade condicional, sejam Ω o espaço amostral e $\mathbb{P}$ a medida de probabilidade de um modelo probabilístico clássico. Se A e B são eventos de Ω, com $B \neq \emptyset$, e se ocorreu um dos $|B|$ pontos amostrais de B, então A ocorre se um dos $|A\cap B|$ pontos amostrais de $A\cap B$ ocorre, de modo que *a probabilidade com que A ocorre se é sabido que B ocorre* é

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

Num modelo probabilístico $(\Omega, \mathcal{A}, \mathbb{P})$ a probabilidade condicional do evento A com respeito ao evento E, em que $\mathbb{P}(E) > 0$, é definida por

$$\mathbb{P}(A \mid E) \stackrel{\text{def}}{=} \frac{\mathbb{P}(A\cap E)}{\mathbb{P}(E)}$$

e $\mathbb{P}(A \mid E)$ é lido como a probabilidade de A dado E.

Proposição 20 (**teorema da multiplicação**). *Se A e B são eventos de um modelo probabilístico com $\mathbb{P}(A) > 0$ então* $\mathbb{P}(A\cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$. *Analogamente,* $\mathbb{P}(A\cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$, *se* $\mathbb{P}(B) > 0$. □

No caso geral

Teorema 21 (**teorema 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*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{n-1} \mathbb{P}\left(A_i \,\middle|\, \bigcap_{j=1}^{i-1} A_j\right).$$

Exemplo 22. Numa cômoda há três gavetas 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, sem olhar para o interior da gaveta escolhida, essa gaveta é aberta e um pé de meia é escolhido aleatoriamente dentre os dois pés daquela gaveta; a gaveta escolhida é fechada. O pé de meia retirado é branco. Qual a probabilidade de o outro pé que ficou sozinho na gaveta ser preto?

Se E é o evento "retirou uma meia branca", e A o evento "ficou uma meia preta", então $\mathbb{P}(A \cap E) = 1/6$ e $\mathbb{P}(E) = 1/2$, logo

$$\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{1}{3}.$$

◊

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 $\mathbb{Q} \colon \mathcal{A} \to \mathbb{R}$ por

$$\mathbb{Q}(A) \stackrel{\text{def}}{=} \mathbb{P}(A \mid E).$$

Prova-se que $\mathbb{Q}$ é uma medida de 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). Também, $\mathbb{Q}$ é uma medida de probabilidade sobre $(\Omega, \{A \cap E \colon A \in \mathcal{A}\})$.

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 e $\overline{E}$ ocorre

isto é, $A = (A \cap E) \uplus (A \cap \overline{E})$ donde deduzimos

$$\mathbb{P}(A) = \mathbb{P}\Big((A \cap E) \uplus (A \cap \overline{E})\Big) = \mathbb{P}(A \cap E) + \mathbb{P}(A \cap \overline{E}) = \mathbb{P}(A \mid E)\mathbb{P}(E) + \mathbb{P}(A \mid \overline{E})\mathbb{P}(\overline{E}). \tag{7}$$

Exemplo 23 (*Monty Hall*). o problema de Monty Hall é um problema que 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:

1. Na 1ª etapa o concorrente escolhe uma porta ao acaso (que ainda não é aberta);
2. 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;
3. 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;
4. 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 3) que maximiza a chance de ganhar o carro.

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 $\mathbb{P}(A \mid \overline{E}) = 1$. Usando a lei da probabilidade total, equação (7),

$$\mathbb{P}(A) = \mathbb{P}(A \mid E)\mathbb{P}(E) + \mathbb{P}(A \mid \overline{E})\mathbb{P}(\overline{E}) = \frac{2}{3}$$

portanto, é melhor trocar de porta.

Teorema 24 (**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

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(E_i \cap A) = \sum_{i=1}^{n} \mathbb{P}(A \mid E_i)\mathbb{P}(E_i). \tag{8}$$

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, (8) segue da aditividade e do teorema da multiplicação, proposição 20. □

Exemplo 25 (*urna de Pólya*). Inicialmente, uma urna contém duas bolas, uma branca e uma preta. Em cada instante t, $t = 1, 2, \dots$, 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, para $t \geqslant 1$, 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, por (7)

$$\mathbb{P}(P_2) = \mathbb{P}(P_2 \mid P_1)\mathbb{P}(P_1) + \mathbb{P}(P_2 \mid \overline{P}_1)\mathbb{P}(\overline{P}_1)$$

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, $\mathbb{P}(P_2 \mid \overline{P}_1) = 1/3$, 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 definição de probabilidade condicional

$$\mathbb{P}(P_1 \mid P_2) = \frac{\mathbb{P}(P_2 \mid P_1)\mathbb{P}(P_1)}{\mathbb{P}(P_2)} = \frac{2}{3} = \mathbb{P}(P_2 \mid P_1).$$

Para computar $\mathbb{P}(P_3)$ precisamos de um pouco mais de esforço. Generalizando o teorema da multiplicação

$$\mathbb{P}(P_1 \cap P_2 \cap P_3) = \mathbb{P}(P_1)\mathbb{P}(P_2 \mid P_1)\mathbb{P}(P_3 \mid P_1 \cap P_2) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4}.$$

e por (8)

$$\begin{aligned}
\mathbb{P}(P_3) &= \mathbb{P}(P_1 \cap P_2 \cap P_3) + \mathbb{P}(P_1 \cap \overline{P}_2 \cap P_3) + \mathbb{P}(\overline{P}_1 \cap P_2 \cap P_3) + \mathbb{P}(\overline{P}_1 \cap \overline{P}_2 \cap P_3) \\
&= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} \\
&= \frac{1}{2}
\end{aligned}$$

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.

Para "a n-ésima bola sorteada é preta", consideremos os eventos $X_1, X_2, \dots, X_{n-1}, P_n$ 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

$$\mathbb{P}\left(\left(\bigcap_{t=1}^{n-1} X_t\right) \cap P_n\right) \tag{9}$$

em que a soma é sobre as 2^{n-1} tais sequências de eventos. Usando o caso geral do teorema da multiplicação, teorema 21 na página 17, a probabilidade em (9) 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 (9) 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

$$\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{m}{m+1} \cdot \frac{1}{m+2} \cdot \frac{2}{m+3} \cdots \frac{n-m}{n+1} = \frac{m!\,(n-m)!}{(n+1)!} \tag{10}$$

Agora, usaremos (10) para calcular (9). 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 (10), 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

$$\mathbb{P}(P_j \mid P_k) = \frac{\mathbb{P}(P_j \cap P_k)}{\mathbb{P}(P_k)} = \frac{\mathbb{P}(P_k \mid P_j)\mathbb{P}(P_j)}{\mathbb{P}(P_k)} = \mathbb{P}(P_k \mid P_j).$$

Ainda, usando (10)

$$\mathbb{P}(\text{há } k \text{ bolas pretas após } n\text{-ésimo sorteio}) = \binom{n}{k} \frac{k!\,(n-k)!}{(n+1)!} = \frac{1}{n+1}.$$

Experimentos compostos: um experimento composto é formado por vários experimentos simples realizados de forma consecutiva. Para modelar um experimento composto convém, muitas vezes, fazer um diagrama de árvore que represente todos os resultados. Cada resultado é dado por um caminho no diagrama e a probabilidade desse resultado é o produto das probabilidades dos resultados simples que o formam.

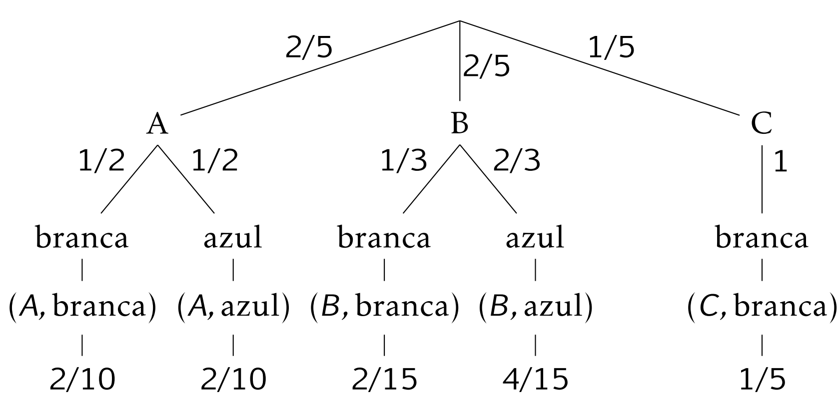
Exemplo 26. Em cada uma de cinco urnas de três tipos diferentes há seis bolas. Duas urnas são do tipo A, nelas temos três bolas brancas e três bolas azuis; duas urnas são do tipo B e nelas temos duas bolas brancas e quatro azuis; uma única urna é do tipo C e todas as bolas nela são brancas. Uma urna é escolhida aleatoriamente e dessa urna uma bola é escolhida aleatoriamente.

Agora, temos um experimento aleatório composto por dois experimentos. Em cada etapa usamos o modelo clássico, de modo que as urnas do tipo A são escolhidas com probabilidade 2/5, assim como as do tipo B; as urnas do tipo C são escolhidas com probabilidade 1/5. Em cada urna, cada bola tem a mesma probabilidade de ser escolhida. Entretanto, o resultado do primeiro experimento determina o modelo do segundo, por exemplo, sorteada a urna A a probabilidade de escolher uma bola branca é 1/2 e sorteada a urna C a probabilidade de escolher uma bola branca é 1.

Cada uma das cinco urnas tem a mesma probabilidade de ser escolhida, portanto, as urnas do tipo A e do tipo B são escolhidas com probabilidade 2/5, que denotamos por $\mathbb{P}_U(A) = \mathbb{P}_U(B) = 2/5$, as urnas do tipo C são escolhidas com probabilidade 1/5, que denotamos por $\mathbb{P}_U(C) = 1/5$. Em cada urna, cada bola tem a mesma probabilidade de ser escolhida mas como a proporção de cada cor é diferente de acordo com o tipo de urna termos as probabilidades das cores condicionada ao tipo de urna, na urna A temos $\mathbb{P}_A(\text{branca}) = \mathbb{P}_A(\text{azul}) = 1/2$; na urna B temos $\mathbb{P}_B(\text{branca}) = 1/3$ e $\mathbb{P}_B(\text{azul}) = 2/3$; na urna C temos $\mathbb{P}_C(\text{branca}) = 1$ e $\mathbb{P}_C(\text{azul}) = 0$. Assim, temos os modelos probabilísticos (discretos) para os sorteios: $(\Omega_U, \mathcal{A}_U, \mathbb{P}_u)$ das urnas; $(\Omega_A, \mathcal{A}_A, \mathbb{P}_A)$ das bolas da urna A; $(\Omega_B, \mathcal{A}_B, \mathbb{P}_B)$ das bolas da urna B; $(\Omega_C, \mathcal{A}_C, \mathbb{P}_C)$ das bolas da urna C.

Para definirmos um modelo probabilístico discreto para o experimento composto podemos tomar o espaço amostral $\Omega = \{A, B, C\} \times \{\text{branca}, \text{azul}\}$ e agora precisamos definir as probabilidades dos pontos amostrais. Se $E = \{A\} \times \{\text{branca}, \text{azul}\}$ é o evento "urna do tipo A" e $F = \{A, B, C\} \times \{\text{branca}\}$ é "bola da cor branca", então $E \cap F = \{(A, \text{branca})\}$ cuja probabilidade é $\mathbb{P}(E \cap F) = \mathbb{P}(F \mid E)\mathbb{P}(E)$, pelo teorema da multiplicação. Ademais $\mathbb{P}(F \mid E)$ é $\mathbb{P}_A(\text{branca}) = 1/2$ e $\mathbb{P}(E)$ é $\mathbb{P}_U(A) = 2/5$. Cada ponto amostral pode ser escrito como uma interseção, como acima, e a probabilidade é calculada com o teorema da multiplicação.

O diagrama de árvore abaixo representa cada etapa do experimento em cada nível da árvore, com respectivas probabilidades nas ramificações correspondentes aos resultados de cada etapa. Uma maneira de atribuir probabilidade a um ponto amostral é tomar o produto das probabilidades no caminho até ele nessa árvore, por exemplo, $\mathbb{P}((A, \text{branca})) = 2/5 \cdot 1/2$,



e estendemos a probabilidade a qualquer evento somando a probabilidade de seus elementos. O evento B definido por "a bola sorteada é branca", ou seja, $B = \{(A, \text{branca}), (B, \text{branca}), (C, \text{branca})\}$ ocorre com probabilidade é $2/10 + 2/15 + 1/5 = 5/15 = 8/15$.

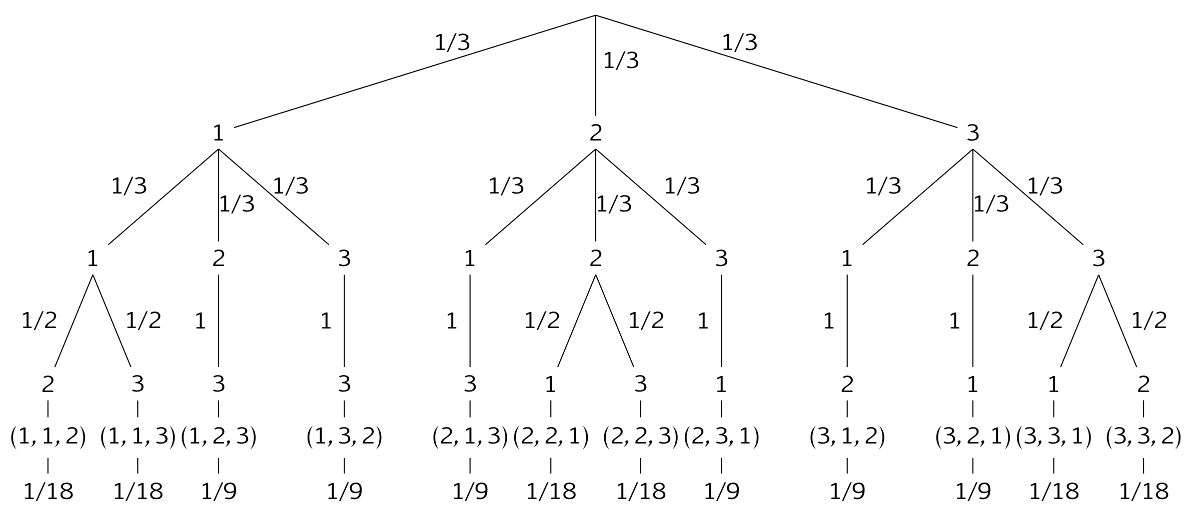
Observemos que essa probabilidade não depende do número de bolas brancas na urna C, portanto, *a probabilidade de B não é a quantidade de bolas brancas dividido pelo número total de bolas*, que é um erro comum nesse caso. $\Diamond$

Considere 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.

Exemplo 27. *Monty Hall* No caso do problema de Monty Hall 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 $\Omega =$

$$\Big\{(1,1,2),(1,1,3),(1,3,2),(1,2,3),(2,2,1),(2,2,3),(2,3,1),(2,1,3),(3,3,1),(3,3,2),(3,1,2),(3,2,1)\Big\}$$

e definimos uma medida de probabilidade de acordo com o diagrama de árvore a seguir



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 esta na porta escolhida inicialmente"

$$A = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$$
$$B = \{(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\}$$

O jogador ganha sem trocar de porta se ocorre B e $\mathbb{P}(B) = 1/3$; e ganha quando troca de porta se ocorre A e $\mathbb{P}(A) = 2/3$. $\lozenge$

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 $\mathbb{P}(B \mid \overline{A}) = 0{,}03$ e a de um falso negativo $\mathbb{P}(\overline{B} \mid A) = 0{,}02$. 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 temos

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid \overline{A})\mathbb{P}(\overline{A})} = \frac{0{,}98 \cdot 0{,}001}{0{,}98 \cdot 0{,}001 + 0{,}03 \cdot 0{,}999} = 0{,}031664.$$

A chance de ter a doença dado que o teste foi positivo é ligeiramente menor que $3,2\%$.

A dedução feita acima pode ser facilmente generalizada para provar o seguinte.

Teorema 28 (**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 $1 \leqslant j \leqslant m$*

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{m} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

Demonstração. A prova é imediata da definição de probabilidade condicional $\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\mathbb{P}(B)}$ e do teorema da probabilidade total $\mathbb{P}(B) = \sum_{i=1}^{m} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)$. $\square$

Exemplo 29. Temos 3 moedas e apenas uma delas, não sabemos qual, é desbalanceada; qualquer uma é a desbalanceada com mesma probabilidade. Essa moeda desbalanceada resulta cara com probabilidade 2/3. Consideremos os eventos E_i definidos por "a i-ésima moeda é a desbalanceada"; logo $\mathbb{P}(E_i) = 1/3$ para todo i. Realizamos lançamento de cada uma delas resulta em (Ca, Ca, Co).

Consideremos o evento B definido por "o resultado dos lançamentos é Ca, Ca, Co".

$$\mathbb{P}(B \mid E_1) = \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{6}$$
$$\mathbb{P}(B \mid E_2) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{6}$$
$$\mathbb{P}(B \mid E_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{12}$$

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 $\mathbb{P}(E_1) = 1/3$, logo "aprendemos mais" sobre as moedas depois do lançamento. Também, $\mathbb{P}(E_2| B) = 2/5$ mas $\mathbb{P}(E_3 \mid B) = 1/5$.

Independência: o conhecimento da ocorrência de um evento possível B afeta a probabilidade de ocorrência de 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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{11}$$

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

Proposição 30. *Todo evento aleatório A de um modelo probabilístico é independente do evento certo Ω e do evento impossível $\emptyset$.*

Proposição 31. *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. $\square$

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

$$\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid B) = \mathbb{P}(A \mid C) = \mathbb{P}(A). \tag{12}$$

Consideremos a definição acima de que A é independente de B se vale a equação (11). Assumamos, como definição, que A é independente de B e C se vale a equação (12). Nesse caso, 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 parágrafo anterior podemos definir que A é independente de B e C se for independente de todo evento do menor espaço de eventos que contém $\{B, C\}$. Ademais, notemos que essa definição é compatível com a definição de "A independente de B" dada anteriormente pois, pela proposição 30 e proposição 31, 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 $B \cap C$*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B \cap C).$$

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C), \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Exemplo 32. 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, $\mathbb{P}(E_1 \cap E_2 \cap E_3) = 1/9$. $\lozenge$

Exemplo 33. 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 lançamentopor 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.

Independência condicional: dizemos que A_1 e A_2 são *condicionalmente independentes* com respeito a B (ou dado B) se

$$\mathbb{P}(A_2 \mid B \cap A_1) = \mathbb{P}(A_2 \mid B) \tag{13}$$

que equivale a

$$\mathbb{P}(A_2 \cap A_1 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B) \tag{14}$$

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,

$$\mathbb{P}(A_1) = \mathbb{P}(A_1 \mid A)\mathbb{P}(A) + \mathbb{P}(A_1 \mid \overline{A})\mathbb{P}(\overline{A}) = 0{,}4 \cdot 0{,}3 + 0{,}2 \cdot 0{,}7 = 0{,}26$$

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, $\mathbb{Q}(\cdot) = \mathbb{P}(\cdot \mid A_1)$. Dessa forma, pelo teorema da probabilidade total

$$\mathbb{Q}(A_2) = \mathbb{Q}(A_2 \mid A)\mathbb{Q}(A) + \mathbb{Q}(A_2 \mid \overline{A})\mathbb{Q}(\overline{A})$$

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

$$\begin{aligned}
&\mathbb{Q}(A_2) = \mathbb{Q}(A_2 \mid A)\mathbb{Q}(A) + \mathbb{Q}(A_2 \mid \overline{A})\mathbb{Q}(\overline{A}) \Leftrightarrow\\
&\mathbb{P}(A_2 \mid A_1) = \mathbb{P}(A_2 \mid A)\mathbb{P}(A \mid A_1) + \mathbb{P}(A_2 \mid \overline{A})\mathbb{P}(\overline{A} \mid A_1) \Leftrightarrow\\
&\mathbb{P}(A_2 \mid A_1) = 0{,}4 \cdot \frac{6}{13} + 0{,}2 \cdot \frac{1}{13} = 0{,}29.
\end{aligned}$$

Repetições independentes de um experimento: o número de placas de automóveis 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 $\Omega = \bigtimes_{i=1}^{n} \Omega_i$.

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 $p_i \overset{\text{def}}{=} \mathbb{P}(\omega_i)$. 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).

Exemplo 34. 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 35. Consideremos uma sequência infinita de lançamentos de uma moeda, cada lançamento independente dos outros. Em cada realização o evento cara ocorre com probabilidade $p \in (0,1)$. 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

$$\mathbb{P}(B_n) = \mathbb{P}(E_1)\mathbb{P}(E_2)\cdots\mathbb{P}(E_n) = (1-p)^n$$

por causa da independência das realizações. Portanto, pelo menos uma ocorrência de cara nas n primeiras realizações tem probabilidade $1-(1-p)^n$. Notemos que $B_1 \supset B_2 \supset B_3 \supset \cdots$ logo

$$\mathbb{P}\left(\bigcap_{n\geqslant 1} B_n\right) = \mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty}(1-p)^n = 0$$

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? ◊

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 desses experimentos da seguinte forma.

O espaço amostral da realização conjunta desses experimentos é dada 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 $A_1 \times A_2 \in \mathcal{A}_1 \times \mathcal{A}_2$, o qual passamos a denotar por $\mathcal{A}$. No produto definimos a medida de probabilidade

$$\mathbb{P}(A_1 \times A_2) \stackrel{\text{def}}{=} \mathbb{P}_1(A_1)\mathbb{P}_2(A_2).$$

O espaço de probabilidade $(\Omega_1 \times \Omega_2, \mathcal{A}, \mathbb{P})$ é chamado de espaço produto. Isso pode ser estendido para o produto de vários espaços.

Eventos cilíndricos: para definir um modelo probabilístico para o experimento do exemplo 35, 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 Ω é o evento obtido fixando um número finito de coordenadas, para quaisquer $(a_0,\dots,a_{n-1}) \in \{0,1\}^n$ e para qualquer $F \subset \{0,1\}^n$, com $n \in \mathbb{Z}^+$

$$C_{a_0,\dots,a_{n-1}} = \Big\{(\omega_0,\omega_1,\dots) \in \Omega \colon \omega_i = a_i \text{ para todo } i,\ 1 \leqslant i \leqslant n\Big\} \text{ e} \tag{15}$$

$$C_F = \biguplus_{(a_0,\dots,a_{n-1}) \in F} C_{a_1,\dots,a_n} \tag{16}$$

serão chamados de cilindros de ordem n; dizemos simplesmente cilindro quando a ordem não for relevante.

A família formada pelos cilindros C_F com $F \subset \Omega$ finito é uma álgebra de subconjuntos de Ω. De fato, comecemos por notar que $\Omega = C_{\{0,1\}^n}$ e que $\emptyset = C_\emptyset$. Ainda, para os cilindros da forma (15) temos que a interseção de dois deles $C_{a_0,\dots,a_{n-1}} \cap C_{b_0,\dots,b_{m-1}}$, supondo sem perda de generalidade que $n \leqslant m$, é um cilindro da forma (15) se $a_i = b_i$ para todo $0 \leqslant i < n$ ou é vazio, caso contrário, isto é

$$C_{a_0,\dots,a_{n-1}} \cap C_{b_0,\dots,b_{m-1}} = \begin{cases} C_{a_0,\dots,a_{n-1}}, & \text{se } n \leqslant m \text{ e } a_i = b_i \text{ para todo } i < n \\ C_{b_0,\dots,b_{m-1}}, & \text{se } m \leqslant n \text{ e } a_i = b_i \text{ para todo } i < m \\ \emptyset, & \text{caso contrário.} \end{cases} \tag{17}$$

Ademais, se $F = \{0,1\}^n \setminus \{(a_0,\dots,a_{n-1})\}$, então

$$\overline{C_{a_0,\dots,a_{n-1}}} = \biguplus_{(b_0,\dots,b_{n-1}) \in F} C_{b_0,\dots,b_{n-1}} \tag{18}$$

Até o momento, (17) e (18) garantem que a família dos cilindros da forma (15) formam uma semiálgebra de subconjuntos de Ω. É sabido que o conjunto formado por uniões finitas de elementos de uma semiálgebra é uma álgebra, mas vamos verificar esse fato. Só é preciso mostrar que $\mathcal{C}$ é fechada para complemento e união finita.

Para o complemento de elementos de $\mathcal{C}$ notemos que $\overline{C_F} = C_{\overline{F}}$. Para a união de C_F com C_E cilindros de ordem n e m, respectivamente, temos, supondo sem perda de generalidade que $n \leqslant m$ e considerando F' como o conjuntos das sequências $(a_0,\ldots,a_{m-1}) \in \{0,1\}^m$ tais que $(a_0,\ldots,a_{n-1}) \in F$, temos

$$C_F = \biguplus_{(a_0,\ldots,a_{m-1})\in F'} C_{a_0,\ldots,a_{m-1}}, \quad \text{e} \quad C_F \cup C_E = \biguplus_{(a_0,\ldots,a_{m-1})\in F'\cup E} C_{a_0,\ldots,a_{m-1}}. \tag{19}$$

portanto $\mathcal{C}$ é estável sob complementação e união finita. Com isso temos que $\mathcal{C}$ é uma álgebra de subconjuntos de Ω de modo que qualquer medida de probabilidade definida em $\mathcal{C}$ admite uma única extensão para a σ-álgebra gerada por $\mathcal{C}$ (esse é o conhecido teorema de extensão de Caratheodory).

Em $\{0,1\}$ temos $\mathbb{P}_0(1) = p$ e $\mathbb{P}_0(0) = q \stackrel{\text{def}}{=} 1-p$. Vamos definir $\mathbb{P}_1 : \mathcal{C} \to \mathbb{R}$. Nos cilindros tomamos a medida produto, nos da forma (15)

$$\mathbb{P}_1(C_{a_0,\ldots,a_{n-1}}) \stackrel{\text{def}}{=} p^{|\{i:a_i=1\}|}q^{|\{i:a_i=0|\}}$$

e nos cilindros da forma (16)

$$\mathbb{P}_1(C_F) \stackrel{\text{def}}{=} \sum_{(a_0,\ldots,a_{n-1})\in F} p^{|\{i:a_i=1\}|}q^{|\{i:a_i=0|\}}.$$

É possível provar que $\mathbb{P}_1$ é consistente no sentido de que se um conjunto puder ser escrito de mais de uma maneira como um cilindro, em todas elas a medida coincide.

Claramente, $\mathbb{P}_1(C) \geqslant 0$ para todo cilindro $C \in \mathcal{C}$ e

$$\mathbb{P}_1(\Omega) = \sum_{(a_0,\ldots,a_{n-1})\in\{0,1\}^n} p^{|\{i:a_i=1\}|}q^{|\{i:a_i=0|\}} = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k} = (p+q)^n = 1.$$

Ainda, $\mathbb{P}_1$ é aditiva (finita), usando (19)

$$\mathbb{P}_1(C_F \uplus C_E) = \sum_{(a_0,\ldots,a_{m-1})\in F'\cup E} \mathbb{P}_1(C_{a_0,\ldots,a_{n-1}}) = \mathbb{P}_1(C_F) + \mathbb{P}_1(C_E)$$

e, por fim, é possível provar que $\mathbb{P}_1$ é σ-aditiva, o que não faremos aqui, de modo que existe $\mathbb{P}$ definida em $\sigma(\mathcal{C})$ que coincide com $\mathbb{P}_1$ nos cilindros e $(\Omega, \sigma(\mathcal{C}), \mathbb{P})$ é um modelo probabilístico para lançamento de moeda infinitas vezes.