O que é a probabilidade? Interpretações da probabilidade.

Azevedo, Cecília
Departamento de Matemática
Universidade do Minho
cecilia@math.uminho.pt

Resumo

Esta comunicação tem como público alvo professores do ensino básico e secundário e pretende resumir conceitos básicos da teoria de probabilidades, apresentando um pouco da história da evolução do conceito de probabilidade e algumas interpretações mais utilizadas naqueles níveis de ensino.

1 Introdução

Podemos afirmar que a probabilidade é tão antiga quanto o Homem. Faz parte da natureza humana indagar acerca da possibilidade das mais diversas ocorrências do dia a dia. De facto, textos muito antigos referem a avaliação da probabilidade na condução da vida, tendo sido estudada pelos Matemáticos chineses do século I. Além disso podemos pensar que o Homem sempre se interessou por jogos de azar, tais como aqueles que envolvem dados e cartas. Os jogos de dados, na forma como hoje a conhecemos, realizam-se desde o Império Romano tanto quanto é conhecido. Os jogos de cartas foram introduzidos na Europa pelos Chineses, Indianos, Egípcios e tornaram-se bastante populares nos finais do sec.XIII.

Problemas envolvendo finanças, seguros e mortalidade surgiram também bem cedo na civilização ocidental.

Gerolamo Cardano (1501-1576) foi um dos primeiros investigadores a estudar problemas envolvendo probabilidade. Cardano era Físico, Astrólogo e Matemático, tendo sido reitor da Universidade de Padova. Era também um exímio jogador de jogos de azar e escreveu uma obra tratando essencialmente desde tipo de jogos, denominada *Liber de Ludo*, onde introduziu, entre outras, a ideia de caracterizar a probabilidade de um acontecimento como um número p entre 0 e 1, mostrando compreender, no contexto de jogos de dados, o conceito de equiprobabilidade de acontecimentos tendo ainda antecipando resultados importantes sobre limites de probabilidades e sobre a distribuição binomial. A sua obra foi publicada a título póstumo em 1633. Foi nesta altura que se considera que o cálculo de probabilidades, como ramo da Matemática, foi iniciado. Blaise Pascal (1623-1662) e Pierre DeFermat (1601-1665) iniciaram, em 1654, troca de correspondência sobre probabilidades e frequências de resultados em jogos de azar. Nesse tempo, em França, era costume os notáveis discutirem assuntos académicos como parte dos seus afazeres sociais e de prazer. Antoine Gombaud, o Chavalier de Méré, propôs a Pascal o problema de dividir justamente o dinheiro apostado num jogo que termina prematuramente. Por exemplo, um jogador aposta a saída de um 6 até que o dado seja lançado 8 vezes. Se o jogo terminar após 3 lançamentos nos quais ocorreu sempre insucesso, quanto deve caber a cada um dos jogadores? Pascal reportou este problema a Fermat, e a sua troca de ideias mostra claramente o entendimento de conceitos tais como repetição de experiências independentes e a diferença entre probabilidade condicional e não condicional. De facto, Pascal propôs que a probabilidade do acontecimento ocorrer na $i-$ésima tentativa é $\frac{1}{6}\left(\frac{5}{6}\right)^{i-1}$ pelo que se se adicionar os termos correspondentes a i desde 4 até 8, obtém-se a probabilidade p de ganhar em 4 ou 5 ou 6 ou 7 ou 8 tentativas, pelo que o dinheiro deverá ser distribuído atribuindo a proporção p do dinheiro ao apostador e $1-p$ aos restantes jogadores. Fermat responde que esta análise só estará correcta se for conhecido o resultado do primeiro dos três lançamentos. Neste caso, a probabilidade do apostador ganhar no próximo lançamento é $\frac{1}{6}$ - noção de probabilidade condicional - pelo que o apostador deverá receber $\frac{1}{6}$ do quinhão independentemente de quando o jogo terminar.

Mais tarde, independentemente um do outro, ambos notaram a importância do coeficiente binomial $\left(\begin{array}{c} n \\ k \end{array}\right)$ no cálculo de probabilidades envolvendo repetições independentes de experiências. Importa referir que o cálculo combinatório tinha já sido usado pelo menos desde 300 a.C. pelos Hindus.

Ainda neste século, mas já na segunda metade, surge o primeiro livro em probabilidade, *De ratiociniis in ludo alea*, de Cristaan Huygens (1629-1695), matemático alemão, que compila e estende os resultados de Pascal e Fermat.

O primeiro grande período de desenvolvimento da teoria foi o século XVIII onde, em 1713, foi publicada, postumamente, a obra *Ars Conjectandi* de James Bernoulli (1654-1705). Nessa obra Bernoulli estabelece e justifica as bases de análise combinatória subjacentes à teoria das probabilidades. No entanto, Bernoulli não é a fonte original de muitos desses resultados, sendo esse mérito de Gottfried Wilhelm Leibniz (1646-1716).

Em 1718 foi publicada, ainda, *La Doctrine des Chances* de DeMoivre.

Thomas Bayes (1702-1761) é também um dos mais famosos probabilistas anteriores a Laplace, pelo seu trabalho, publicado postumamente em 1763, *An Essay Towards Solving A Problem in the Doutrine of Chances*.

A transição para o sec. XIX foi marcada pelo livro *Théorie Analytique des Probabilités*, publicado em 1812, de Laplace (1749-1827).

No século XX as probabilidades encontravam-se num estado de grande desenvolvimento em teoria e aplicações. De facto, conceitos essenciais à formulação do cálculo de probabilidades só muito recentemente foram enunciados como, por exemplo, o conceito de espaço amostra, de Von-Mises (1883-1953), no primeiro quartel do século XX, e o conceito de $\sigma-$algebra, introduzido por Emile Borel (1871-1956) que foi tão fundamental para a criação do integral de Lebesgue quanto para o desenvolvimento axiomático da probabilidade de Kolmogorov.

O problema da axiomatização da probabilidade foi um dos enumerados por Hilbert, em 1900, no Congresso Internacional de Matemática, e muitos tais como Diogo Pacheco de Amorim, Bernstein e Von-Mises, tentaram resolvê-lo mas sem o mesmo êxito que Kolmogorov. Esta axiomática, perspectivada por Kolmogorov e publicada em 1933 com título original, *Grundegebrieffe des Wahrscheinlichkeistheorie*, é a base da teoria moderna das probabilidades.

Na formalização matemática actual, a probabilidade é uma noção primitiva. Não se define.

As várias tentativas de definição que foram surgindo ao longo dos tempos, deram origem a diversas **interpretações** das quais destacamos:

1. probabilidade clássica;
2. probabilidade frequencista;
3. probabilidade bayesiana.

Quando falamos de probabilidade falamos da possibilidade de um **acontecimento** se realizar.

Os acontecimentos são os objectos aos quais se atribui a probabilidade. Estão associados à realização de **experiências** cujo resultado é incerto.

Definição 1.1 *Uma experiência é um procedimento que permite a obtenção de observações.*

Definição 1.2 *Uma* **experiência aleatória** *é uma experiência em que o resultado observado* ω *é incerto. No entanto, o conjunto dos resultados possíveis - designado por espaço amostra* Ω *- é conhecido antes da realização da experiência.*

Formalmente, um **acontecimento** é uma colecção de resultados possíveis sendo, assim, um **subconjunto de** Ω.

2 Probabilidade Laplaciana (clássica)

Em 1812 Laplace publicou um tratado, *Théorie Analytique des Probabilités*, onde usa métodos analíticos para a teorização da Probabilidade. Na segunda edição, de 1814, é, pela primeira vez, avançada uma definição de probabilidade - o conceito laplaciano de probabilidade - onde são enunciados os princípios que devem guiar a atribuição e cálculo das probabilidades.

O primeiro princípio define a probabilidade de um acontecimento A como o quociente entre o número de casos que lhe são favoráveis e o número de casos possíveis:

$$P(A) = \frac{\text{número de casos favoráveis à ocorrência de } A}{\text{número de casos possíveis}}$$

se os casos possíveis forem equiprováveis, ou seja todos os acontecimentos elementares têm a mesma probabilidade.

A noção de equiprobabilidade ou de acontecimentos "igualmente possíveis"torna a definição de Laplace circular. O *princípio da razão insuficiente* ou *o princípio da indiferença* estabelece que sempre que não há evidência do favorecimento de um acontecimento elementar sobre qualquer outro, todos têm a mesma probabilidade. Ou ainda, se o espaço amostra é constituído por resultados para os quais há uma evidência simétrica e balanceada, aquela definição continua válida.

Ainda assim continua a existir circularidade na definição de Laplace pois, no primeiro caso a noção de evidência é, ela própria, probabilística e, no segundo caso, parece ser preciso "pesar"a evidência em cada um dos resultados possíveis ...

De qualquer modo recorremos a esta definição na "modelação de jogos de azar", em particular quando trabalhamos com moedas equilibradas, dados perfeitos, esferas de extracção de lotaria ou totoloto, cartas bem baralhadas ... Esta definição também é apropriada na dedução de distribuições de probabilidade de grande importância teórica.

A teoria clássica de probabilidade é restrita a espaços de resultados de dimensão finita. Quando o espaço amostra é **infinito numerável**, o princípio desta teoria pode ser aplicado aplicando ao *princípio da máxima entropia*, que é uma generalização do princípio da indiferença.

A entropia é uma medida da falta de informação de uma distribuição de probabilidade. Quanto mais concentrada é a distribuição, menor é a entropia (maior é a informação).

O *princípio da máxima entropia* permite seleccionar de uma família de distribuições consistentes a distribuição que maximiza esta quantidade.

Quando o espaço amostra é infinito não numerável, os pontos (resultados elementares) de tal espaço são equiprováveis - têm todos probabilidade nula. Surgem probabilidades não triviais quando associamos um número não numerável de pontos em conjuntos maiores. Se houver um número finito destes conjuntos, a teoria clássica pode ser aplicada desde que não haja evidência de favorecimento de um conjunto sobre qualquer outro.

Examplo 2.1 *Se pensarmos que se divide o intervalo* $[0,1]$ *num ponto escolhido ao acaso, a probabilidade desse ponto estar no intervalo* $[0,0.25]$ *é* 0.25*, uma vez que há* 4 *intervalos com a mesma amplitude:* $[0,0.25]$*,* $[0.25,0.5]$*,* $[0.5,0.75]$ *e* $[0.75,1]$*.*

Aqui entra o **paradoxo de Bertrand**. Vejamos o seguinte exemplo adaptado de van Frassen, 1989:

Examplo 2.2 *Uma fábrica produz cubos com comprimento de lado entre* 0 *e* 1 *metro.*

Qual é a probabilidade que um cubo seleccionado aleatoriamente tenha lado de comprimento entre 0 *e* $0.5m$*?*

A resposta será 0.5 *se imaginarmos que o processo de produção é tal que o comprimento de lado de cada cubo é uniformemente distribuído.*

Mas a questão poderia ser posta de outro modo (equivalente):

Uma fábrica produz cubos com área de face entre 0 *e* 1 m^2. *Qual a probabilidade de um cubo escolhido aleatoriamente ter área de face entre* 0 *e* $0.25m^2$*?*

Agora a resposta é 0.25 *se se admitir que o processo de produção é tal que a área de face de cada cubo é uniformemente distribuída.*

Não podemos ter o mesmo acontecimento com probabilidades diferentes.

Existem infinitas possibilidades de reestruturar a questão de modos equivalentes (em termos de volume e de todas as potências de expoente real não nulo do comprimento), dando todas origem a probabilidades diferentes.

O paradoxo de Bertrand surge do facto da expressão "escolher ao acaso"não ser inequívoca. No primeiro caso escolhe-se um comprimento, no segundo uma área, no terceiro um volume, ...

O princípio da indiferença pode, assim, ser usado de formas diferentes (e incompatíveis). Não há qualquer evidência de favorecimento do acontecimento "o comprimento de lado pertence ao intervalo $[0, 0.5]$"em detrimento de pertencer ao intervalo $[0.5, 1]$ ou vice-versa, pelo que aquele princípio permite atribuir probabilidade 0.5 a cada um dos acontecimentos que constituem o espaço amostra. Mas também não há qualquer evidência da área do lado de cada cubo estar em qualquer um dos intervalos $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$ e $[0.75, 1]$ com favorecimento para qualquer um destes, pelo que se atribui probabilidade 0.25 a cada um.

Há, assim, necessidade de definir claramente o espaço amostra associado à experiência aleatória que se está a realizar.

3 Probabilidade frequencista

Do ponto de vista da teoria clássica, a probabilidade de nascer uma criança do sexo feminino é 0.5. De facto, na formação de células seminais o processo de meiose origina, a partir de cada célula do progenitor com um par de cromossomas XY um par de espermatozóides, um transmissor de X (feminino) e um transmissor de Y (masculino).

No entanto sabe-se que há preponderância de nascimentos de rapazes. Na década de 70 em Portugal, a frequência relativa de recém-nascidos rapazes foi 0.517 (ver Pestana, Velosa).

Há, então, quem prefira atribuir a probabilidade de um acontecimento A à frequência relativa da ocorrência de A.

Venn(1876), Reichenbach (1949) e von Mises (1957) entre outros notáveis, identificaram a probabilidade de um acontecimento com o limite da frequência relativa desse acontecimento: $\mathbb{P}(A) =$ limite da frequência relativa com que se observa A

Este limite não é o conceito usual da análise matemática, mas sim o **limite em probabilidade**. Dizer que

$$\mathbb{P}(A) = \lim_{n \to +\infty} f_n(A)$$

equivale a dizer que para todo $\epsilon > 0$,

$$\mathbb{P}\left(\left|\mathbb{P}(A) - f_n(A)\right| > \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 0$$

.

Em análise, dizer que $\lim\limits_{n\to+\infty} f_n(A) = \mathbb{P}(A)$ significa que $\left|f_n(A) - \mathbb{P}(A)\right| < \epsilon$ para todo $n > n_0(\epsilon)$, enquanto que na convergência em probabilidade $\left|f_n(A) - \mathbb{P}(A)\right|$ pode exceder qualquer valor dado, por maior que este seja (< 1) embora com probabilidades cada vez mais próximas de zero.

É usual escrever-se

$$f_n(A) \xrightarrow{P} \mathbb{P}(A).$$

4 Probabilidade bayesina

Em muitas aplicações importantes de probabilidade como, por exemplo, económicas, não faz sentido admitir equiprobabilidade dos acontecimentos elementares e, por outro lado, não se esta perante um esquema de experiências repetidas. Nestes casos a única forma de atribuir probabilidades a acontecimentos é com base em convicções subjectivas e experiência de quem a avalia.

De qualquer modo, a necessidade de reavaliar a probabilidade quando a informação muda é evidente. Várias são as situações em que a equiprobabilidade é apenas um ponto de partida - uma probabilidade "a priori".

Se, por exemplo, em 100 lançamentos de uma moeda se observar 80 vezes "cara"é difícil continuar a considerar que a moeda é equilibrada e que os acontecimentos "cara"e "coroa"são equiprováveis...

Na probabilidade bayesiana estamos a quantificar um grau de credibilidade subjectiva de acontecimentos incertos. Este tipo de fundamento funciona sem contradições e tem tido bastante sucesso em aplicações que envolvem avaliação de riscos. A regra de avaliação da probabilidade de um acontecimento A face a uma nova evidência E, conhecida como condicionamento estabelece que

$$\mathbb{P}(A|E) = \frac{\mathbb{P}(A\cap E)}{\mathbb{P}(E)}, \text{ com } \mathbb{P}(E) > 0$$

O Teorema de Bayes permite reavaliar a probabilidade, obtendo uma probabilidade "a posteriori".

5 Axiomatização da probabilidade

Da análise dos vários conceitos de probabilidade surge a necessidade de se proceder a uma axiomatização da mesma.

Como já foi dito, no final do séc.XIX, Hilbert apresentou uma listagem dos grandes problemas em aberto que incluia o problema da axiomatização da probabilidade.

Em 1933, Kolmogorov em "Foundations of the theory of probability"apresenta uma axiomática que tem como base três conceitos fundamentais:

- experiência aleatória
- σ-álgebra dos acontecimentos
- medida de probabilidade

Vamos recordar a definição de experiência aleatória.

Definição 5.1 *Experiência aleatória é uma experiência que satisfaz às seguintes condições:*

(i) *pode ser repetida em condições análogas*

(ii) *é conhecido o conjunto Ω de resultados possíveis da experiência (espaço de resultados, espaço amostra)*

(iii) *em cada realização da experiência não se sabe qual dos resultados irá ocorrer.*

A descrição do espaço de resultados Ω é feita indicando os **acontecimentos elementares**, cada um dos quais é um resultado possível de cada realização da experiência aleatória. Os **acontecimentos** são subconjuntos de Ω e as operações elementares sobre conjuntos levam a novos acontecimentos.

Ω é ele próprio um acontecimento: o acontecimento certo. $\varnothing \subseteq \Omega$ é o acontecimento impossível. Qualquer subconjunto de Ω é um acontecimento. O conjunto de todos os acontecimentos é o conjunto das partes de Ω. O crescimento de $\#\mathcal{P}(\Omega)$ com $\#\Omega$ é muito rápido. Em muitas situações interessam-nos padrões por exemplo, verificar se se observa o acontecimento elementar A :"face"no lançamentos de uma moeda equilibrada. Neste caso, o conjunto de todos os acontecimentos possíveis será simplesmente $\{\varnothing, \Omega, A, (\Omega - A)\}$.

O objectivo é descrever um espaço de acontecimentos em que as operações usuais sobre conjuntos levem de acontecimentos a acontecimentos, mas que não seja demasiado sobrecarregado.

6 O espaço de acontecimentos

Definimos espaço de acontecimentos como o par $(\Omega, \mathcal{A})$ onde Ω é o espaço de resultados e os seus subconjuntos estão em $\mathcal{A}$ (álgebra de acontecimentos ou subconjuntos probabilizáveis) verificando os axiomas da **álgebra de conjuntos**:

(E_1) $\Omega \in \mathcal{A}$

(E_2) $A \in \mathcal{A} \Rightarrow \overline{A} = (\Omega - A) \in \mathcal{A}$

(E_3^*) $A \in \mathcal{A}, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

Estas três exigências arrastam imediatamente que $\mathcal{A}$ é fechada para uniões finitas, intersecções finitas, diferenças, diferenças simétricas e, na generalidade, para todas as operações internas usuais sobre conjuntos.

O axioma (E_3^*) não abrange situações em que temos uma infinidade numerável de acontecimentos. Há que considerar um axioma de $\sigma-$aditividade:

$$(E_3)\ A_1, A_2 \cdots \in \mathcal{A} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$$

Nota 6.1 *De* $(E_1), (E_2)$ *e* (E_3) *decorre imediatamente que se*
$A_1, A_2 \cdots \in \mathcal{A} \Rightarrow \bigcap_{k=1}^{\infty} A_k \in \mathcal{A}$

Por outro lado, considerando $A_{n+1} = A_{n+2} = \cdots = \varnothing$*, então* $\bigcup_{k=1}^{n} A_k \in \mathcal{A}$ *e, em particular* $(E_3) \Rightarrow (E_3^*)$.

Definição 6.2 *Chamamos* $\sigma-$*álgebra a uma família* $\mathcal{A}$ *de subconjuntos de* Ω *(acontecimentos) que verifique os axiomas* $(E_1), (E_2), (E_3)$.

Note-se que existe sempre uma $\sigma-$álgebra nestas condições porque $\mathcal{P}(\Omega)$ é uma $\sigma-$álgebra.

No caso de Ω ser finito, pode tomar-se $\mathcal{A} = \mathcal{P}(\Omega)$, embora em muitas circunstâncias baste tomar espaços de acontecimentos muito mais simples tais como vimos anteriormente $\{\varnothing, \Omega, A, \overline{A}\}$.

No caso de Ω ser infinito não numerável, é em geral conveniente - ou mesmo imprescindível - construir a $\sigma-$álgebra dos acontecimentos de forma adequada. Usamos frequentemente a $\sigma-$álgebra dos borelianos, $\mathcal{B}$ gerada pelos subconjuntos abertos de Ω. Mais tarde retomaremos este assunto.

7 O espaço de probabilidade

Define-se o espaço de probabilidade como o terno $(\Omega, \mathcal{A}, \mathbb{P})$, em que a medida de probabilidade $\mathbb{P}$ é uma função de conjunto cujo domínio é $\mathcal{A}$, e que verifica os seguintes axiomas:

(P_1) $\mathbb{P}(A) \geq 0$, qualquer que seja o acontecimento $A \in \mathcal{A}$ (positividade)

(P_2) $\mathbb{P}(\Omega) = 1$ (normalização)

(P_3^*) Se A e B são acontecimentos disjuntos, i. e. $A \cap B = \varnothing$, então
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (aditividade finita)

Também nesta axiomática não são contempladas as situações em que temos de considerar uma infinidade numerável de acontecimentos. Em vez de (P_3^*) temos, nesse caso, de considerar um axioma mais forte:

(P_3) Se $A_1, A_2, \ldots$ são acontecimentos disjuntos dois a dois, então

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k) \ \text{(aditividade numerável)}$$

O axioma (P_3) estabelece uma condição suficiente para ser válida a "lei de adição"do cálculo de probabilidades clássico.

Vejamos um exemplo de um espaço de probabilidade:

Examplo 7.1 *Experiência aleatória: "Lançamento de um dado equilibrado"*

$$\Omega = \{1,2,3,4,5,6\}, \quad \mathcal{B} = \{A : A \in \Omega\}, \quad \mathbb{P}(A) = \frac{1}{6}\#A$$

Vejamos, agora, um exemplo de uma outra medida de probabilidade definida em P(Ω):

Examplo 7.2 *Se Ω é um conjunto e $\omega \in \Omega$, então a função*
$\delta_\omega : \mathcal{P}(\Omega) \to [0,1]$ *dada por*

$$\delta_\omega(A) = \begin{cases} 1 & se\ \omega \in A \\ 0 & se\ \omega \notin A \end{cases}$$

é uma probabilidade sobre as partes de Ω, denominada delta de Dirac em ω. Pode imaginar-se medidas como distribuições de massa no espaço Ω. No caso da medida de Dirac toda a massa está concentrada num só ponto $\omega \in \Omega$.

<u>Consequências desta axiomática</u>

(1) $A_1, \dots, A_n$ disjuntos dois a dois,

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

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

(3) $A \subseteq B \Rightarrow \mathbb{P}(A) \le \mathbb{P}(B)$

(4) $\mathbb{P}(A) \in [0,1], \quad \forall A$

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

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

(7) $\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1^n} \mathbb{P}(A_k) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) +$
$\sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1}\mathbb{P}(\bigcap_{k=1}^{n} A_k)$

(8) Desigualdade de Boole:

$$\mathbb{P}(A \cap B) \ge 1 - \mathbb{P}(\overline{A}) - \mathbb{P}(\overline{B})$$

(9) Mais geralmente,

$$\mathbb{P}\left(\bigcap_{k_1}^{n} A_k\right) \ge 1 - \sum_{k=1}^{n} \mathbb{P}(\overline{A_k})$$

(10) Desigualdade de Benferroni:

Dados n acontecimentos $A_1, A_2, \dots, A_n$

$$\sum_{k=1}^{n} \mathbb{P}(A_i) - \sum_{k>j} \mathbb{P}(A_k \cap A_j) \le \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Nota 7.3 *É natural considerar que um acontecimento com probabilidade* 1 *se realiza "quase certamente" e que um acontecimento com probabilidade* 0 *é "quase certamente impossível". Mas é preciso ter em atenção que o impossível, nesta acepção, pode realizar-se. Considere-se a experiência de escolher ao acaso um número entre* 0 *e* 10*. A probabilidade de marcarmos nesse segmento o ponto* π *é zero e, genericamente, a probabilidade de marcarmos um ponto qualquer é zero. No entanto, quando marcamos* $\{x_0\}$ *estamos a observar um acontecimento* $\{x_0\}$ *de probabilidade nula.*

Continuidade da medida de probabilidade

A $\sigma-$aditividade tem como consequência a *continuidade* da medida de probabilidade, isto é, a possibilidade de calcular a probabilidade do *limite* de certas sucessões de acontecimentos, calculando o limite das probabilidades dos elementos da sucessão.

Seja $\{A_n\}_{n\geq 1}$ uma sucessão de subconjuntos (em particular uma sucessão de acontecimentos) de Ω. O **limite superior da sucessão** é o conjunto

$$\lim_{n\to\infty} \sup A_n = \overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

e o **limite inferior da sucessão** é o conjunto

$$\lim_{n\to\infty} \inf A_n = \underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k \right)$$

Assim,

$\underline{\lim} A_n$ é o conjunto dos pontos que pertencem a todos os A_n, à excepção de um número finito de valores de n.

$\overline{\lim} A_n$ é o conjunto dos pontos que pertencem aos A_n para um número infinito de valores de n.

Consequentemente,

$$\underline{\lim} A_n \subseteq \overline{\lim} A_n$$

.

Se $\underline{\lim} A_n = \overline{\lim} A_n = \lim A_n$ a sucessão tem limite.

No caso particular em que $\{A_n\}_{n\geq 1}$ **é uma sucessão monótona de acontecimentos**, temos:

- $\lim\limits_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$ se a sucessão é não decrescente;
- $\lim\limits_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$ se a sucessão é não crescente.

Teorema 7.4 *Seja* $\{A_n\}_{n\ \geq 1}$ *uma sucessão monótona de acontecimentos. Então,*

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

8 Probabilidade Condicional e Independência

Sendo a probabilidade uma forma de quantificar a incerteza, naturalmente que temos de atender que nova informação pode alterar, e por vezes de forma muito significativa, a avaliação da probabilidade. É natural que se estabeleça um processo de reavaliar a probabilidade por forma a incorporar nova informação. É o conceito de probabilidade condicional.

Definição 8.1 *A probabilidade condicional de A dado B é*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \textit{se } P(B) > 0.$$

Invertendo a expressão acima obtém-se a *regra do produto*

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

Também se tem

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

e

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

Nota 8.2 *As probabilidades absolutas e as probabilidades condicionais são equivalentes. De facto,*

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

Invertendo agora esta expressão temos

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

que é o denominado Teorema de Bayes ou Teorema da probabilidade inversa. Retomaremos este teorema mais adiante.

Teorema 8.3 *Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade, B um acontecimento fixo tal que $P(B) > 0$.*

Definindo a medida de probabilidade condicional a B, $P_B(A) = P(A|B)$, para todo $A \in \mathcal{A}$, também $(\Omega, \mathcal{A}, P_B)$ é um espaço de probabilidade.

Há, no entanto, acontecimentos B que não contêm informação relevante para a reavaliação da probabilidade de A.

Definição 8.4 *O acontecimento A é independente de B se $P(A|B) = P(A)$. Neste caso*

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

Como quer a intercessão quer o produto são operações comutativas, entre conjuntos e entre números reais, respectivamente, dizer que A é independente de B é equivalente a afirmar que B é independente de A.

Nota 8.5 *Dois acontecimentos incompatíveis (mutuamente exclusivos) não podem ser independentes, excepto se um deles for o acontecimento impossível.*

Generalização da probabilidade da intercessão de acontecimentos

Vejamos em primeiro lugar

$$P(A_1 \cap A_2 \cap A_3) = P((A_1 \cap A_2) \cap A_3) = P(A_3|(A_1 \cap A_2))P(A_1 \cap A_2),$$

pelo que

$$P(A_1 \cap A_2 \cap A_3) = P(A_3|(A_1 \cap A_2))P(A_2|A_1)P(A_1).$$

Mais geralmente para a intercessão de n acontecimentos, convencionando denotar uma cadeia de acontecimentos

$$A_1 \cap A_2 \cap \cdots \cap A_n = A_1 A_2 \dots A_n,$$

temos a regra da multiplicação ou regra da probabilidade de uma cadeia

$$P(A_1 A_2 \dots A_n) = P(A_n|A_1 \dots A_{n-1})P(A_{n-1}|A_1 \dots A_{n-2})$$
$$\dots P(A_2|A_1)P(A_1).$$

Definição 8.6 *Os acontecimentos $A_1, A_2, \dots, A_n$ são* **mutuamente independentes** *se para todo o natural k e para toda a escolha de $i_1, i_2, \dots, i_k$ distintos*

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(Ai_k).$$

Nota 8.7 *O conceito de probabilidade condicional pode simplificar a atribuição da probabilidade.*

No contexto de experiência sequenciais, a expressão geral da probabilidade de uma cadeia é de utilização muito simples recorrendo a árvores de probabilidade.

Quando consideramos experiências aleatórias sequenciais no tempo, originando espaços amostra sequenciais com o número "razoável"de resultados possíveis, os acontecimentos podem ser indexados pela ordem da experiência em que ocorrem, e dispostos num grafo.

9 Vantagens/Odds

Uma forma alternativa de exprimir probabilidades, como uma medida de incerteza, é através das vantagens.

Definição 9.1 *A vantagem é a razão entre a probabilidade de A e a probabilidade de $\overline{A}$:*

$$r(A) = \frac{\mathbb{P}(A)}{\mathbb{P}(\overline{A})} = \frac{\mathbb{P}(A)}{1 - \mathbb{P}(A)}.$$

Consequentemente,

$$\mathbb{P}(A) = \frac{r(A)}{1 + rA}.$$

Se usarmos a definição clássica de probabilidade

$$\mathbb{P}(A) = \frac{\text{número de casos favoráveis à ocorrência de } A}{\text{número de casos possíveis}}$$

$$r(A) = \frac{\text{número de casos favoráveis à ocorrência de } A}{\text{número de casos desfavoráveis à ocorrência de } A}$$

Examplo 9.2 *Seja A : retirar um ás de um baralho honesto de 52 cartas.*

$$\mathbb{P}(A) = \frac{4}{52}$$

sendo

$$r(A) = 4 : 48 \quad \textit{ou ainda} \quad 1 : 12.$$

- No lançamento de uma moeda honesta, a vantagem da saída de cara é de 1 : 1.

Podemos ainda falar em desvantagem de A : $\widetilde{r}(A)$. Neste caso teremos,

$$\widetilde{r} = \frac{\text{número de casos desfavoráveis à ocorrência de } A}{\text{número de casos favoráveis à ocorrência de } A}$$

agora a relação entre probabilidade e desvantagem é dada por

$$\mathbb{P}(A) = \frac{1}{1+\widetilde{r}}$$

- Lançamento de um dado *honesto*
- $P(\text{sair face3}) = 1/6$, a desvantagem da face 3 é 5 : 1

Vantagens à posteriori Vejamos o seguinte exemplo:

Uma corrida na qual participam 5 cavalos: A, B, C, D e E em que a probabilidade de cada um deles ficar em primeiro lugar é dada por

$P(A_1) = 0.8, P(B_1) = 0.1, P(C_1) = 0.05$, $P(D_1) = 0.03$ e $P(E_1) = 0.02$.

Calculemos as vantagens à priori:

- $r(A_1) = \frac{80}{20} = 4:1$, $r(B_1) = \frac{10}{90} = 1:9$, $r(C_1) = \frac{5}{95} = 1:19$
  $r(D_1) = \frac{3}{97} = 3:97$, $r(E_1) = \frac{2}{98} = 1:49$
- Qual a probabilidade de B ficar em segundo lugar?

$$P(B_2) = ?$$

(1) Se A ganhar, qual a probabilidade de B_2?

- $r(B_2|A_1) = 10 : (5+3+2) = 1:1$
- $P(B_2|A_1) = 0.5$

  Pelo que,

$$P(A_1 \cap B_2) = 0.5 \times 0.8 = 0.4$$

(1) Se C ganhar, qual a probabilidade de B_2?

- $r(B_2|C_1) = 10 : (80+3+2) = 10:85$
- $P(B_2|C_1) = \dfrac{10/85}{1+10/85} = \dfrac{10}{95}$

  Tendo-se,

$$P(C_1 \cap B_2) = 0.05 \times \frac{10}{95} = 0.052632$$

(3) Qual a probabilidade de D ganhar e B ficar em segundo lugar?

- $r(B_2|D_1) = 10 : (80 + 5 + 2) = 10 : 87$
- $P(B_2|D_1) = \dfrac{10/87}{1 + 10/87} = \dfrac{10}{97}$

  Assim,

$$P(D_1 \cap B_2) = 0.03 \times \frac{10}{97} = 0.0030928$$

(4) Qual a probabilidade de E ganhar e B ficar em segundo lugar?

- $r(B_2|E_1) = 10 : (80 + 5 + 3) = 10 : 88$
- $P(B_2|E_1) = \dfrac{10/88}{1 + 10/88} = \dfrac{10}{98}$

  Assim

$$P(E_1 \cap B_2) = 0.02 \times \frac{10}{98} = 0.0020408$$

  Finalmente

- $P(B_2) = 0.4103968$

A título de nota, referimos ainda os chamados **payoffs odds** que é a nomenclatura normalmente utilizada para as possibilidades de lucro em jogos de azar.

Consideremos o popular jogo da roleta de Nevada onde existem 36 entradas numeradas de 1 a 36 (vermelhos e pretos e dois sinais 0 e 00 (a verde). Numa jogada simples as possibilidades de lucros oferecidas são $35 : 1$ sendo a (des)vantagem correcta de $37 : 1$. Assim o ganho fixo da casa por jogada é

$$\frac{2}{30} \times 100\% = 5.26\%$$

Referências

[1] Papoulis, A.; Probability, Random Variables, and Stochastic Processes, *Mc Graw Hill,* 1991.

[2] Pestana, D., Velosa, S.; Introdução à Probabilidade e à Estatística, *Fundação Caloustre Gulbenkian*, 2002.

[3] Pitman, J.; Probability, *Springer*, 1997.