Capítulo 2

Probabilidade

2.1 Probabilidades Físicas

A idéia de probabilidades está associada tanto com raciocínio indutivo e julgamentos tais como "Provavelmente João é feliz"ou "Você provavelmente será aprovado em Estatística", quanto a experimentos físicos repetitivos tais como o arremesso de uma moeda ou de um dado. Nesta seção apresentamos a segunda alternativa: as *probabilidades físicas*.

2.1.1 Espaço amostral

Denominamos *fenômeno aleatório* à situação ou acontecimento que não pode ser previsto com certeza. Por exemplo, quando arremessamos uma moeda - honesta ela pode cair com a face Cara (H) ou Coroa (T) voltada para cima (é claro que simplificamos, ignorando o caso raro em que ela fica de pé !). Assim o arremesso de uma moeda é um evento aleatório. Da mesma maneira nosso tempo de vida, o decaimento radioativo ou o resultado da loteria também são fenômenos aleatórios. Os resultados de um experimento envolvendo um fenômeno aleatório são chamados *eventos*. Fazemos ainda a distinção entre *eventos compostos* e *eventos simples*. Por exemplo, se dissermos que o arremesso de dois dados resultou em "soma=6"isso será equivalente a dizermos que o resultado do experimento foi "$(1, 5)$ ou $(2, 4)$ ou $(3, 3)$ ou $(4, 2)$ ou $(5, 1)$". Ou seja podemos decompor o evento composto $\{\text{soma} = 6\}$ em um conjunto com cinco eventos simples $\{(1, 5)ou(2, 4)ou(3, 3)ou(4, 2)ou(5, 1)\}$. Os eventos simples são os resultados possíveis de um experimento idealizado. Cada resultado destes eventos simples é representado por um ponto em um espaço denominado *espaço amostral* Ω (ômega).

2.1.2 Exemplos

1. Distribuição de três bolinhas distinguíveis (a,b e c) em três caixas distinguíveis. Todos os arranjos possíveis são representados na tabela abaixo:

1. $\{abc\|-\|-\}$	10. $\{a\|bc\|-\}$	19. $\{-\|a\|bc\}$
2. $\{-\|abc\|-\}$	11. $\{b\|ac\|-\}$	20. $\{-\|b\|ac\}$
3. $\{-\|-\|abc\}$	12. $\{c\|ab\|-\}$	21. $\{-\|c\|ab\}$
4. $\{ab\|c\|-\}$	13. $\{a\|-\|bc\}$	22. $\{a\|b\|c\}$
5. $\{ac\|b\|-\}$	14. $\{b\|-\|ac\}$	23. $\{a\|c\|b\}$
6. $\{bc\|a\|-\}$	15. $\{c\|-\|ab\}$	24. $\{b\|a\|c\}$
7. $\{ab\|-\|c\}$	16. $\{-\|ab\|c\}$	25. $\{b\|c\|a\}$
8. $\{ac\|-\|b\}$	17. $\{-\|ac\|b\}$	26. $\{c\|a\|b\}$
9. $\{bc\|-\|a\}$	18. $\{-\|bc\|a\}$	27. $\{c\|b\|a\}$

Cada um dos 27 arranjos na tabela corresponde a um evento simples, ou seja, a um ponto no espaço amostral. O evento$A =$“uma caixa é ocupada por mais de uma bolinha”é realizado pela união dos pontos de 1 a 21, ou seja, o evento A é o agregado dos pontos amostrais de 1 a 21. De forma semelhante o evento composto $B =$“a primeira caixa não está vazia”é formado pelos pontos amostrais 1, 4 a 15 e 22 a 27. Podemos ainda combinar eventos definindo um evento $C =$“tanto A quanto B ocorrem”. Pense um pouco no que seria este evento [1]. E o evento $D =$“ou A ou B ocorrem”? [2]. E o evento $E =$“A não ocorre”? [3]. O que dizer então do evento $F =$“primeira caixa vazia e nenhuma caixa com mais de uma bolinha”? [4] Apesar deste exemplo referir-se especificamente a 3 bolas em 3 caixas, poderíamos igualmente falar de r bolsa em n caixas. O mesmo modelo básico pode ser empregado em outras stuações. Por exemplo:

Aniversários. Os possíveis aniversários de r pessoas são exatamente análogos a r bolas em $n = 365$ caixas.

Acidentes. A classificação de r acidentes de acordo com o dia da semana em que eles ocorrerão é equivalente à distribuição de r bolas em $n = 7$ caixas.

Amostragem. A classificação de r pessoas de acordo com sua idade ou profissão é equivalente a alocação de r bolas em um número n de caixas (uma para cada classe).

Irradiação em biologia. Para estudarmos o efeito da radiação sobre os cromossomos imaginamos que cada cromossomo corresponda a uma caixa e cada partícula α (alfa) irradiada corresponda a uma bola.

Dados. As possibilidades de resultado no arremesso de r dados corresponde a r bolas em $n = 6$ caixas.

Erros de impressão. As possíveis distribuições de r erros de impressão em n páginas podem ser interpretadas como r bolas em n caixas.

2. Distribuição de três bolinhas indistinguíveis em três caixas distinguíveis. Neste caso temos os seguintes arranjos possíveis:

[1] $C = \{1, 4, 5, ..., 15\}$

[2] D ocorre com certeza já que contém todos os pontos do espaço amostral *Omega*

[3] $E = \{22, 23, 24, 25, 26, 27\}$

[4] F é impossível, assim, $F = \emptyset$.

1. $\{***\|-\|-\}$	6. $\{*\|**\|-\}$
2. $\{-\|***\|-\}$	7. $\{*\|**\|-\}$
3. $\{-\|-\|***\}$	8. $\{-\|**\|*\}$
4. $\{**\|*\|-\}$	9. $\{-\|*\|**\}$
5. $\{**\|-\|*\}$	10. $\{*\|*\|*\}$

O espaço amostral Ω tem estrutura diferente, se utilizamos o modelo com bolinhas distinguíveis ou indistinguíveis irá depender do particular problema que estivermos analisando.

3. Amostragem. Suponha que quiramos estimar quantas pessoas fumam utilizando uma amostra com 100 pessoas. A única propriedade da amostra que é do nosso interesse é o número x de fumantes. O espaço amostral Ω é constituido por 101 pontos $x = 0, 1, ..., 100$. Cada observação é descrita completamente especificando x. Um evento composto seria, neste caso, algo como "a maioria das pessoas fuma"consistindo do conjunto $x = 51, 52, ...100$.

4. Arremessos de moedas. Para um experimento de arremesso de moedas três vezes o espaço amostral Ω consiste de oito pontos

$$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

, com H representando Cara e T representando Coroa.

2.1.3 Eventos e a teoria de conjuntos

A partir dos exemplos da seção anterior é possível noter que existe uma relação íntima entre eventos compostos e a teoria de conjuntos. Assim, a *união* entre dois eventos, $A \cup B$, representa a ocorrência de, pelo menos, um dos eventos A ou B. A *intersecção* entre dois eventos $A \cap B$, representa a ocorrência simultânea de A e B. Dois eventos A e B são *mutuamente exclusivos* (ou *disjuntos*) quando não têm elementos em comum,isto é , quando $A \cap B = \emptyset$. A e B são complementares se a união é o espaço amostral e sua intersecção é vazia. O complementar de A é representado por A^c e temos $A \cup A^c = \Omega$ e $A \cap A^c = \emptyset$.

Podemos construir uma classe de conjuntos que contenha todos os eventos compostos, essa estrutura recebe o nome de *σ-álgebra de eventos* $\mathcal{F}$ e é definida pelas seguintes propriedades:

1. $\Omega \in \mathcal{F}$. O espaço amostral inteiro é um evento composto.
2. Se $A \in \mathcal{F}$ então $A^c \in \mathcal{F}$. Se A é um evento, seu complementar A, ou seja, não ocorrer A, também é um evento.
3. Se $A_i \in \mathcal{F}$ então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. Se A_i são eventos então A_1 ou A_2 ou $...A_n$ também é um evento para qualquer n.

2.1.4 Axiomas de Kolmogorov

Conforme desenvolvida por Komogorov (1903-1987), a teoria de probabilidades clássica não se preocupa em determinar como seu valor numérico deve ser determinado mas sim com suas propriedades gerais. De forma semelhante à geometria, estabelecem-se propriedades básicas que a medida de probabilidade $P(.)$ deve obedecer. Estas (três) propriedades são conhecidas como os *Axiomas de Kolmogorov*:

1. $P(A) \geq 0$. A probabilidade é um número não-negativo.

2. $P(\Omega) = 1$. O espaço amostral contém todas os possíveis resultados do experimento, assim é um evento certo.

3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ se $A_i, A_j \in \mathcal{F}$ e $A_i \cap A_j = \emptyset$, $i \neq j$. Se dois eventos A_1 e A_2 são mutuamente exclusivos então a probabilidade de A_1 ou A_2 é igual a probabilidade de A_1 somada à probabilidade de A_2. O mesmo vale para qualquer número de eventos mutuamente exclusivos.

Um par de propriedades das medidas de probabilidade são corolários imediatos dos Axiomas de Kolmogorov:

Propriedade 1. Como $A \cup A^c = \Omega$, o axioma 2, implica em $P(A \cup A^c) = 1$. Já o axioma 3 implica em $P(A) + P(A^c) = 1$, ou seja,

$$P(A^c) = 1 - P(A). \tag{2.1}$$

Propriedade 2. Da teoria de conjuntos temos que $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$. Onde $A \cap B^c$, $A \cap B$ e $A^c \cap B$ são mutuamente exclusivos (por que ?) [5], pelo axioma 3 temos que

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B).$$

Mas $A = (A \cap B^c) \cup (A \cap B)$ e $B = (A^c \cap B) \cup (A \cap B)$ (por que?). [6] Assim

$$P(A) = P(A \cap B^c) + P(A \cap B)$$

e

$$P(B) = P(A^c \cap B) + P(A \cap B)$$

, onde utilizamos novamente o axioma 3. Substituindo estas expressões na equação acima:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{2.2}$$

Exemplo. Uma moeda é arremessada duas vezes. O espaço amostral é, portanto, $\Omega = \{HH, HT, TH, TT\}$. Se assumirmos que a moeda é honesta

[5] O primeiro conjunto contem pontos do espaço amostral que pertencem a A e não pertencem a B, o segundo conjunto contem pontos que pertencem a A e a B e o terceiro conjunto contem pontos que não pertencem a A e pertencem a B. Um ponto não pode pertencer e não pertencer a B simultaneamente, assim os dois primeiros são disjuntos. Um ponto não pode perntencer e não pertencer a A simultanamente, assim o terceiro conjunto é disjunto aos dois primeiros. Assim, os três conjuntos são disjuntos entre si.

[6] $(A \cap B^c) \cup (A \cap B)$ é o conjunto dos pontos que pertencem a A e não pertencem a B ou os pontos que pertencem a A e pertencem a B. Ou seja, tanto faz se pertencem ou não a B, desde que pertençam a A. De forma equivalente no caso de B.

atribuiremos probabilidade $\frac{1}{4}$ para cada ponto do espaço amostral. Suponhamos agora que A_1 e A_2 sejam, respectivamente, os eventos "Cara no primeiro arremesso" e "Cara no segundo arremesso". Qual seria a probabilidade do evento "Cara no primeiro ou no segundo arremessos"? Este evento equivale a $A_1 \cup A_2$. Pela propriedade 2 teremos que $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. Os eventos são $A_1 = \{HH, HT\}$ e $A_2 = \{HH, TH\}$, a intersecção é $A_1 \cap A_2 = \{HH\}$. Temos então que $P(A_1 \cup A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$. [7]

Dada a medida de probabilidade $P(.)$ para cada ponto do espaço amostral Ω e considerando que os eventos compostos formam uma σ-álgebra é possível calcular a probabilidade de qualquer evento composto utilizando apenas a teoria de conjuntos. Formalmente, um modelo probabilístico para um experimento é composto da terna $\langle \Omega, \mathcal{F}, P \rangle$.

2.1.5 Atribuindo probabilidades

A teoria clássica de probabilidades não se preocupa em como valores numéricos devem ser atribuídos a pontos do espaço amostral. A determinação da medida de probabilidades depende de considerações físicas sobre o fenômeno aleatório ou da suposição de que é possível repetir o experimento infinitas vezes nas mesmas condições. Na prática, mesmo se fosse possível repetir o experimento infinitas vezes isso isso iria requerer quantidade infinita de tempo, assim sendo, a determinação de probabilidades via infinitas repetiçõe deve ser entendida como uma idealização. Formalmente mede-se a probabilidade física do evento A como um limite da razão do número de ocorrências de A (n_A) em um número tendendo a infinito de repetições (n), em símbolos:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}. \tag{2.3}$$

Exemplos. *Bolas distinguíveis.* No exemplo 1 da seção anterior poderíamos, por exemplo, assumir que todos os arranjos são igualmente prováveis, ou seja, a probabilidade de qualquer evento simples é de $\frac{1}{27}$ [8]. Assumindo que $P(E_i) = \frac{1}{27}$ para todo ponto E_i do espaço amostral, podemos calcular a probabilidade de qualquer outro evento composto. Por exemplo, o evento $A =$"uma caixa é ocupada por mais de uma bolinha" é composto $A = \bigcup_{i=1}^{21} E_i$. Como os pontos do espaço amostral são, por definição, mutuamente exclusivos, temos que,

[7] Note que neste caso simples poderíamos simplesmente enumerar diretamente $A_1 \cup A_2 = \{HH, HT, TH\}$. A idéia por trás da dedução de propriedades é justamente sabermos fazer o mesmo cálculo de várias maneiras possíveis, nunca se sabe que tipo de manipulação temos que fazer para resolver cada novo problema. Neste sentido, teoremas funcionam como uma caixa de ferramentas para provarmos novos teoremas cada vez mais complexos.

[8] Gottfried Leibniz (1646-1716) introduziu o seguinte Princípio da Razão Suficiente: Nada é sem razão suficente para ser como é. Para tudo há uma razão. Laplace adaptou o o princípio para afirmando que devemos atribuir probabilidades idênticas para todos as possibilidades de um fenômento aleatório se não tivermos razão alguma para que elas sejam de outra forma. No entanto, tal princípio não integra a teoria clássica de probabilidades.

utilizando o axioma 3:

$$P\left(\bigcup_{i=1}^{21} E_i\right) = \sum_{i=1}^{21} P(E_i) = \frac{7}{9}.$$

Bolas indistiguíveis: Estatística de Bose-Einstein. Retomemos agora o exemplo 2. Podemos argumentar que o experimento propriamente dito não é afetado por nossa impossibilidade de distinguir as bolas. Assim, nós confundiríamos entre os pontos 7, 8 e 9 do espaço amostral de bolas indistinguíveis, atribuindo a todos três o mesmo ponto 5 do espaço amostral de bolas indistinguíveis. Seguindo este raciocínio, somos levados a atribuir os seguintes probabilidades para cada um dos 10 pontos do espaço amostral (confira !):

$P(1) = \frac{1}{27}$	$P(6) = \frac{1}{9}$
$P(2) = \frac{1}{27}$	$P(7) = \frac{1}{9}$
$P(3) = \frac{1}{27}$	$P(8) = \frac{1}{9}$
$P(4) = \frac{1}{9}$	$P(9) = \frac{1}{9}$
$P(5) = \frac{1}{9}$	$P(10) = \frac{2}{9}$

Antes de deixarmos esta secção é interessante enfatizar que a relação entre teoria e experimento em se tratando de probabilidades é intrincada. Por exemplo, na prática nenhuma moeda é totalmente honesta, ou seja, a atribuição de probabilidades idênticas para Cara e Coroa, $P(H) = P(T) = \frac{1}{2}$ é apenas uma aproximação da realidade. Para determinarmos a "verdadeira" distribuição de probabilidade de uma moeda precisaríamos repetir seu arremesso infinitas vezes. A tabela a seguir mostra o número de Caras obtidas em seqüências de 1000 arremessos.

$0 - 1000$	501
-2000	485
-3000	509
-4000	536
-5000	485
-6000	488
-7000	500
-8000	497
-9000	494
-10000	484

As flutuações observadas na tabela acima dão uma idéia do tipo de dificuldades que se apresentam ao tentarmos medir probabilidades através da repetição de um experimento muitas vezes. Fica claro que técnicas mais avançadas são necessárias, a área da Estatística encarregada da obtenção de probabilidades a partir de dados é a Inferência Estatística que estudaremos com detalhes mais a frente.

2.2 Probabilidade Condicional

Suponha uma população de N pessoas onde N_A são universitários e N_F são do sexo feminino. O evento de que uma pessoa escolhida ao acaso seja universitária é A, o evento de que uma pessoal escolhida ao acaso seja do sexo feminino é F. A probabilidade de A é $P(A) = \frac{N_A}{N}$. A probabilidade de F é $P(F) = \frac{N_F}{N}$. Suponha agora que só olhemos para a população feminina. O número de mulheres universitárias é N_{AF}. A probabilidade de uma mulher ser universitária será $P(A|F) = \frac{N_{AF}}{N_F}$, onde o símbolo $P(A|F)$ é lido como a "probabilidade de A dado F". De forma equivalente, escrevemos:

$$P(A|F) = \frac{N_{AF}}{N_F} = \frac{N_{AF}/N}{N_F/N} = \frac{P(A \cap F)}{P(F)}.$$

Assim definimos a *probabilidade condicional* como:

$$P(A|F) = \frac{P(A \cap F)}{P(F)}. \tag{2.4}$$

Por economia de símbolos trocamos, de agora em diante, $P(A \cap F)$ por $P(AF)$.

Exemplo. *Amostragem sem reposição.* A partir de uma população com n elementos $1, 2, ..., n$ tomamos uma amostra ordenada. Suponha que i e j são dois elementos diferentes. Assumindo que i é o primeiro elemento amostrado, qual é a probabilidade de que o segundo elemento seja j. Temos que a probabilidade de sortearmos primeiro i e j é $P(ij) = 1/n(n-1)$. A probabilidade de sortearmos i como primeiro elemento é $P(i) = 1/n$. A probabilidade de sortearmos j logo após o sorteio de i é então $P(j|i) = P(ij)/P(i) = 1/(n-1)$. Isso simplesmente expressa o fato de que tinhamos primeiro n possibilidades para o sorteio, após o sorteio de i passamos a ter $n-1$ possibilidades.

Exemplo. *Distribuição de sexos.* Cosidere familias com exatamente dois filhos. Chamemos de b se menino e g se menina. Temos quatro possibilidades $\{bb, bg, gb, gg\}$, onde a primeira letra indica o filho mais velho (desconsideremos os gêmeos para efeito de exemplo). Associemos a probabilidade de $\frac{1}{4}$ a cad ponto do espaço amostral. Dado que a familia já tem um menino, qual é a probabilidade de que as duas crianças sejam meninos ? Temos que $P(b|b) = P(bb)/P(b) = P(bb)/(P(bb) + P(bg) + P(gb)) = 1/3$.

No exemplo acima utilizamos a seguinte propriedade de distribuições de probabilidade.

Propriedade 3. Eventos $C_1, C_2, ..., C_k$ formam uma *partição* do espaço amostral se $C_i \cap C_j = \emptyset$ para $i \neq j$ e $\bigcup_{i=1}^k C_i = \Omega$. Assim $P(A) = \sum_{i=1}^k P(AC_i)$. Tente demosntrar.[9]

Da Propriedade 3 e da definição de probabilidades condicionais segue:

[9]Se $\{C_i\}$ representa uma partição de Ω, então $A \cap C_i$ são eventos disjuntos e $A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^k C_i\right) = \bigcup_{i=1}^k (A \cap C_i)$. Do axioma 3 segue a propriedade que queriamos demonstrar.

Propriedade 4.

$$P(A) = \sum_{i=1}^{k} P(AC_i) = \sum_{i=1}^{k} P(A|C_i)P(C_i).$$

2.3 Independência Estatística

A noção de *independência estatística* tem um significado físico mais ou menos bem definido. Dois fenômenos aleatórios são independentes se o conhecimento do resultado de um deles não diz nada sobre o resultado do outro. Por exemplo, a data de nascimento de João e a temperatura em Plutão. Do ponto de vista da teoria das probabilidades dosi eventos A e B são estatisticamente independentes se $P(A|B) = P(A)$, ou seja, a probabilidade da ocorrência de A não depende de B ter ocorrido. Usando a definição de probabilidade condicional temos que

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

ou seja

$$P(AB) = P(A)P(B). \tag{2.5}$$

. E se tivermos três eventos A, B e C como podemos saber se são ou não independentes? Será que basta termos $P(AB) = P(A)P(B)$, $P(AC) = P(A)P(C)$ e $P(BC) = P(B)P(C)$? Vejamos um exemplo.

Exemplo. Dois dados são arremessados. Definimos três eventos: A="O resultado do primeiro dado é ímpar"; B="O resultado do segundo dado é ímpar"; C="A soma dos dois dados é ímpar". Cada um dos 36 pontos do espaço amostral tem probabilidade $\frac{1}{36}$. Os eventos A e B são independentes, visto que os dados são independnetes, assim vale $P(AB) = P(A)P(B)$. Os eventos A e C ou B e C também são independentes, já que a ocorrência da "soma ímpar"(um dado par e outro ímpar) nada diz sobre ter sido o primeiro (A) ou o segundo (B) dado a dar resultado ímpar. Até aqui tudo bem. No entanto, note que A, B e C **não** podem ocorrer ao mesmo tempo. Assim, sabemos que se A e B ocorreram então C não ocorreu. Da mesma forma se soubermos A e C ou B e C, sabemos que o terceiro evento não pode ocorrer. Em outras palavras, o conhecimento de dois dos eventos **determina** o terceiro e $P(C|AB) \neq P(C)$. Em conclusão, para que três (ou mais) eventos sejam independentes é necessário que seus pares e ternas (ou quaisquer grupos possíveis) sejam independendentes.

A definição mais geral de independência é, portanto:

Independência Estatística. Os eventos A_1, A_2,...A_n são mutuamente independentes se para todas as combinações possíveis de índices $1 \leq i < j < k < ... \leq n$ vale

$$P(A_iA_j) = P(A_i)P(A_j)$$
$$P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k)$$

$$...$$
$$P(A_1A_2...A_n) = P(A_1)P(A_2)...P(A_n)$$

2.4 Teorema de Bayes

Teorema de Bayes. Seja $\{C_j\}$ uma partição do espaço amostral. Uma propriedade de grande importância das probabilidades condicionais é:

$$P(C_j|H) = \frac{P(H|C_j)P(C_j)}{\sum_{j=1}^{k} P(H|C_j)P(C_j)}. \tag{2.6}$$

Demonstração: A definicção de probabilidade condicional nos permite escrever

$$P(C_j|A) = \frac{P(C_jA)}{P(A)} = \frac{P(AC_j)}{P(A)} = \frac{P(A|C_j)P(C_j)}{P(A)}.$$

Utilizando a Propriedade 4 completamos a demonstração.

Exemplo. Suponha que um fabricante de sorvete recebe 20% de todo o leite que utiliza de uma fazendo F_1, 30% de uma fazenda F_2 e 50% de F_3. Um órgão de fiscalização inspecionou as fazendas de surpresa e observou que 20% do leite produzido em F_1 estava adulterado por adição de água, enqunato que para F_2 e F_3, essa proporção era de 5% e 2%, respectivamente. Na indústria de sorvetes os galões de leite são armazenados em um refrigerador sem identificação das fazendas. Se analisarmos um galão ao acaso e constatarmos que está adulterado (evento A) qual é a probabilidade de que o leite seja, por exemplo, proveniente da fazenda F_1? A probabilidade do leite ser proveniente da fazenda F_1 dado que está adulterado é, segundo o teorema de Bayes:

$$P(F_1|A) = \frac{P(A|F_1)P(F_1)}{P(A|F_1)P(F_1) + P(A|F_2)P(F_2) + P(A|F_3)P(F_3)}.$$

Temos que $P(A|F_1) = 0,2$, $P(A|F_2) = 0,05$ e $P(A|F_3) = 0,02$. Temos também que $P(F_1) = 0,2$, $P(F_2) = 0,3$ e $P(F_3) = 0,5$. Fazendo as contas chegamos a $P(F_1|A) = 0,615$.

2.5 Probabilidades Subjetivas

2.5.1 Atualizando crenças com a fórmula de Bayes

No começo deste capítulo falamos sobre dois tipos de uso para o conceito de probabilidades: como probabilidades físicas, associadas à freqüência de ocorrência de um certo fenômeno aleatório (por ex.: "Qual é a probabilidade do arremesso de um dado resultar em 6, três vezes seguidas?"); e como crenças associadas ao raciocínio indutivo (por ex.: "Qual é a chance do Brasil ganhar o hexacampeonato mundial de futebol na Alemanha?"). Quando um investigador de policia diz algo como "a chance de um determinado crime ter sido encomendado é de

80%"ele certamente não se refere a um experimento que quando repetido um número muito grande de vezes retorna o crime em 80 a cada 100 vezes. O que ele quer dizer é que a experiência dele (I) e as evidências do caso permitem que ele atribua à hipótese do crime a probabilidade de 80. Dito isso, tentemos agora olhar para o Teorema de Bayes de outra maneira:

$$P(\text{Crime}|\text{Evidencia}, I) \propto P(\text{Evidencia}|\text{Crime}, I)P(\text{Crime}|I), \tag{2.7}$$

aqui omitimos do denominador o termo $P(\text{Evidencia})$ que apareceria na fórmula de Bayes. Mas o que significa esta expressão? Antes de coletar as evidências o investigator possuia uma crença na possibilidade de crime baseada em sua experiência prévia, ou *a priori*, $P(\text{Crime}|I)$. Após a coleta de evidências, esta crença foi atualizada para o *posterior* $P(\text{Crime}|\text{Evidencia}, I)$. A expressão define como a nova informação contida nas evidências deve ser utilizada na atualização das crenças do investigador, ou seja, basta multiplicar a distribuição a priori pela *verossimilhança* $P(\text{Evidencia}|\text{Crime}, I)$, que define a chance do crime produzir as evidências coletadas.

Mas, como seria possível formalizar matematicamente crenças humanas?

2.5.2 Axiomas de Cox: probabilidades como lógica indutiva

Na lógica dedutiva partimos de um conjunto de axiomas A_1, A_2,...A_n e, utilizando algumas poucas regras, chegamos a uma conclusão T. As regras de dedução garantem que se os axiomas são verdadeiros então a conclusão tmbém o será, em qualquer lugar e a qualquer tempo. Na indução temos acesso a evidências E_1, E_2, ...,E_n. A partir destas evidências procuramos construir suas causas H. Do ponto de vista da lógica dedutiva, a indução utiliza um argumento inválido (ou uma falácia) conhecido por afirmar o conseqüente. Algo como: Todos os humanos são mamíferos. Bigode é um mamífero, portanto, Bigode é humano. Bigode poderia set, por exemplo, um gato!

A lógica indutiva, no entanto, não visa o estabelecimento absoluto da verdade, mas sim a atualização de crenças dadas evidências. Por exemplo, saber que Bigode é um mamífero apenas *aumenta as chances* de ele ser um humano. Precisariamos de outras evidências, por exemplo, saber que Bigode tem quatro patas tornaria altamente improvavel que ele fosse um humano. As idéias das probabilidades como crenças começaram com Bayes no século 18 (veja a seção sobre história no capítulo introdutório) se desenvolveram com Laplace no século 19 e foram depois esquecidas com as ondas de desenvolvimento da Estatística capitaneadas por Galton, Pearson e Fisher, na passagem do século 19 ao 20. Estas idéias foram retomadas por Jeffreys (colega de Fisher em Cambridge) e formalizadas na década de 1940 por Richard T. Cox em seu livro *A Álgebra da Inferência Provável.* Mais recentemente tem havido um renascimento destas idéias com aplicações importantes em áreas como Reconhecimento de Padrões, Mineração de Dados e Inteligência Artificial.

Em seu livro Cox demonstra que são necessários dois axiomas sobr e crenças e as regras da álgebra Booleana para reobtermos **todas** as propriedades das probabilidades físicas de Kolmogorov. Os Axiomas de Cox são:

Axioma 1. *A probabilidade de uma inferência dada determinada evidência determina a probabilidade da seu oposto dada a mesma evidência.* Assim, a probabilidade de chover dado que o céu está nublado determina a probabilidade de não chover dado que o céu está nublado.

Axioma 2. *A probabilidade de que duas inferências (I1 e I2) sejam simultaneamente verdadeiras dada certa evidência (E) é determinada pela probabilidade de que I1 seja verdadeira dada E e, separadamente, pela probabilidade de I2 ser verdadeira dada E e I1.* Por exemplo, a probabilidade do Brasil ser hexacampeão na Alemanha e Adriano ser o artilheiro, dada a experiência que temos em assistir jogos da seleção, é determinada pela probabilidade do Brasil ser hexa, dados os jogos que assistimos, e pela probabilidade de Adriano ser o artilheiro, dados os jogos que assistimos e se o Brasil for hexacampeão. É claro que o evento composto só pode ocorrer se, pelo menos, o Brasil for primeiro hexacampeão.

Exemplo. *Paradoxo de Monty Hall* Imagine o seguinte jogo envolvendo você e um apresentador, que chamaremos de Silvio. Um prêmio é colocado atrás de uma de três portas (A, B e C) que permanecem fechadas. Você deve escolher uma das portas (por exemplo, A). Silvio então abrirá uma das portas que não é a sua e também não contem o prêmio (por exemplo, B) e perguntará se voce quer trocar de porta. Pergunta-se, você deve abrir a sua porta (A) ou deve trocar pela porta restante (C)?

Utilizemos um raciocínio bayesiano para resolvermos o problema. Qual é a probabilidade a priori do prêmio estar atrás de qualquer uma das três portas ($X = A, B$, ou C)? $P(X, I) = 1/3$, onde I representa nosso conhecimento das regras do jogo. Qual seria então a probabilidade de Silvio abrir a porta B (evento que chamaremos de SB) dado que o prêmio está justamente atrás da porta que escolhi (A)? $P(SB|A, I) = 1/2$. E qual seria a probabilidade de Silvio abrir a porta B se o prêmio estivesse atrás dela própria? $P(SB|B, I) = 0$ pelas regras do jogo. Finalmente, qual seria a probabilidade de Silvio escolher a porta B dado que o prêmio está em C? Seria $P(SB|C, I) = 1$, novamente pelas regras do jogo. Temos agora que escolher entre nossa primeira escolha A e a outra porta restante C. Para isso temos que calcular $P(A|SB, I)$ e $P(C|SB, I)$, ou seja, a chance do prêmio estar respectivamente em A e em C dado que Silvio apontou a porta B e as regras do jogo. Utilizemos então o teorema de Bayes:

$$\begin{aligned} P(A|SB,I) &= \frac{P(SB|A,I)P(A|I)}{P(SB|A,I)P(A|I)+P(SB|B,I)P(B|I)+P(SB|C,I)P(C|I)} \\ &= \frac{1/2\times 1/3}{1/2\times 1/3+0\times 1/3+1\times 1/3} \\ &= 1/3. \end{aligned} \tag{2.8}$$

Da mesma forma temos que

$$\begin{aligned} P(C|SB,I) &= \frac{P(SB|C,I)P(C|I)}{P(SB|A,I)P(A|I)+P(SB|B,I)P(B|I)+P(SB|C,I)P(C|I)} \\ &= \frac{1\times 1/3}{1/2\times 1/3+0\times 1/3+1\times 1/3} \\ &= 2/3. \end{aligned} \tag{2.9}$$

Ou seja, a chance de ganhar **trocando** de porta é o duas vezes maior do que permanecendo! Você deveria trocar de porta. A resposta parece contraintuitiva, para vermos que não é basta considerarmos uma versão modificada do jogo. Agora temos, por exemplo, 1000 portas. Você escolhe uma e não abre. Silvio então abre todas as portas restantes menos uma (ou seja, abre 998 portas). Você trocaria agora?

2.6 Exercícios

1. Uma moeda é viciada de modo que a probabilidade de sair cara é 4 vezes maior que a de sair coroa, Para 2 lançamentos independentes dessa moeda, determinar: (a) o espaço amostral; (b) a probabilidade de sair somente uma cara; (c) a probabilidade de sair pelo menos uma coroa; (d) a probabilidade de dois resultados iguais.

2. Considere um conjunto de 4 números dos quais nenhum deles é zero, dois são positivos e dois são negativos. Sorteamos ao acaso, com reposição, 2 números desse conjunto. Determine a probabilidade de: (a) um deles ser negativo; (b) o quociente ser negativo; (c) os dois números terem o mesmo sinal.

3. Uma caixa contem três moedas com Caras nas duas faces, quatro moedas com Coroas nas duas faces e duas moedas honestas. Se uma destas nove moedas for selecionada ao acaso e arremessada uma vez, qual é a probabilidade de que uma Cara seja obtida.

4. O Palmeiras ganha com probabilidade 0,7 se chove e com 0,8 se não chove. Em Setembro a probabilidade de chuva é de 0,3. O Palmeiras ganhou a partida em Setembro, qual a probabilidade de ter chovido nesse dia?

5. Uma escola de ensino médio do interior de São Paulo tem 40% de estudantes do sexo masculino. Entre estes, 20% nunca viram o mar, ao passo que, entre as meninas, essa porcentagem é de 50%. Qual a probabilidade de que um aluno selecionado ao acaso seja: (a) do sexo masculino e nunca tenha visto o mar; (b) do sexo feminino ou nunca tenha visto o mar.

6. Mostre que se A e B são independentes, então A^c e B^c também são independentes.

7. Um médico desconfia que um paciente tem tumor no abdômen, pois isto ocorreu em 70% dos casos similares que tratou. Se o paciente de fato tiver o tumor, o exame ultra-som o detectará com probabilidade 0,9. Entretanto, se ele não tiver o tumor, o exame pode, erroneamente, indicar que tem com probabilidade 0,1. Se o exame detectou um tumor, qual é a probabilidade do paciente tê-lo de fato?

2.7 Referências

Se quiser ler bons livros de divulgação científica relacionados ao problema do acaso:

- Ruelle, D., Acaso e Caos, Editora Unesp, 1993.
- Bennett, D.J., Aleatoriedade, Martins Fontes, 2003.

Se quiser fazer mais exercícios procure por (alguns dos exercícios do texto foram extraídos de lá):

- Magalhães M.N., de Lima A.C.P., Noções de Probabilidade e Estatística, Edusp,2004.

Exemplos sobre espaços amostrais, sobre probabilidades condicionais e independência foram extraídos de:

- Feller, W. An introduction to Probability Theory and Its Applications, Volume I, John Wiley & Sons, 1950.
- DeGroot, M., Probability and Statistics, Addison-Wesley, 1975.

Para se aprofundar na teoria clássica de probabilidades:

- Kolmogorov,A.N., Foundations of the Theory of Probability.

Sobre a noção matemática de independência:

- Kac, M., Statistical Independence in Probability, Analysis and Number Theory.

Sobre probabilidades subjetivas:

- Cox, R.T., The Algebra of Probable Inference, The Johns Hopkins Press, 1961.
- Jeffreys, H., Theory of probability, Oxford Press, 1967.

Para ler sobre aplicações da Inferência Bayesiana:

- Russel, S., Norvig P., Inteligência Artificial, Ed. Campus, 2003.

Sobre o Dilema de Monty Hall há o seguinte artigo na Wikipedia:

- http://en.wikipedia.org/wiki/Monty_Hall_problem