Dois Exemplos de Espaços Amostrais

Nelson Kuhl - IME/USP

11 de abril de 2022

Em dois exemplos discutidos em aula do livro *The Pleasures of Probability*, de Richard Isaac, foram calculadas algumas probabilidades sem fazer menção ao espaço amostral. É um exercício interessante explicitá-lo, tanto para fixar conceitos como para a percepção de que é possível fazer cáculos sem explicitá-lo.

Exemplo 1

Suponha que testes clínicos para um exame laboratorial de uma doença determinaram que o resultado é positivo em 95% dos casos nos quais a doença está presente. Dizemos então que 0.95 é uma estimativa para a *sensibilidade* do teste definida como

$$\mathbb{P}(\text{resultado é positivo}|\text{doença está presente}),$$

onde $\mathbb{P}$ é a medida de probabilidade. Testes clínicos em uma população saudável determinaram também uma frequência de 5% de falsos positivos. Sabe-se também que a doença está presente em 0.3% da população. Determine

$$\mathbb{P}(\text{doença está presente}|\text{resultado é positivo}).$$

Não iremos refazer as contas, mas simplesmente apresentar um espaço amostral. Foram definidos dois eventos:

$$A = \{\text{doença está presente}\}, \quad B = \{\text{resultado é positivo}\}.$$

Que espaço amostral Ω podemos definir de modo que A e B sejam eventos, ou seja, subconjuntos de Ω? Pensando nas possibilidades de uma pessoa, ela poderia estar doente e testar positivo ($D+$), estar doente e testar negativo ($D-$), estar saudável e testar positivo ($S+$) e estar saudável e testar negativo ($S-$).

Podemos então definir o espaço amostral como

$$\Omega = \{D+, D-, S+, S-\}.$$

Os eventos A e B serão então os subconjuntos de Ω

$$A = \{D+, D-\} \quad \text{e} \quad B = \{D+, S+\}.$$

Os dados do problema são

$$\begin{aligned}
&\mathbb{P}(B|A) = 0.95 \text{ (sensibilidade)} \\
&\mathbb{P}(B|A^c) = 0.05 \text{ (falso positivo)} \\
&\mathbb{P}(A) = 0.003
\end{aligned}$$

Queremos calcular $\mathbb{P}(A|B)$, que pode ser obtida dos dados acima pela fórmula de Bayes.

Exemplo 2

Duas bolas são sorteadas cegamente a partir de grandes quantidades iguais de bolas brancas e pretas e colocadas em uma urna. Iremos selecionar bolas da urna *com reposição* segundo o procedimento: misture bem, selecione uma bola, anote a sua cor e a recoloque na urna, misture bem, selecione uma bola, e assim por diante. Suponha que nas duas primeiras vezes foram selecionadas bolas brancas. Qual é então a probabilidade de se obter uma bola branca na terceira vez?

Foram definidos os seguintes eventos:

$$\begin{aligned}
D &= \text{ as bolas da urna têm cores diferentes} \\
B &= \text{ as bolas da urna são brancas} \\
P &= \text{ as bolas da urna são pretas} \\
B2 &= \text{ as duas primeiras seleções resultaram em bolas brancas} \\
C &= \text{ a bola é branca na terceira seleção}
\end{aligned}$$

Qual espaço amostral usar? Uma descrição completa usaria seleções ilimitadas. Como estamos interessados em ir até a terceira seleção, vamos definir um espaço amostral associado ao experimento de selecionar 3 vezes. O resultado final não muda se considerarmos $N > 3$ seleções, ou mesmo uma infinidade de seleções.

Podemos então considerar

$$\Omega = \Big\{(bb,(b,b,b)),(bp,(b,b,b)),(bp,(b,b,p)),(bp,(b,p,b)),(bp,(b,p,p)),$$
$$(bp,(p,b,b)),(bp,(p,b,p)),(bp,(p,p,b),(bp,(p,p,p)),(pp,(p,p,p))\Big\},$$

onde bb, bp e pp denotam as bolas nas urnas sem ordenação e as triplas ordenadas $(*,*,*)$ denotam as primeira, segunda e terceira seleções. Então,

$$D = \Big\{(bp,(b,b,b)),(bp,(b,b,p)),(bp,(b,p,b)),(bp,(b,p,p)),$$
$$(bp,(p,b,b)),(bp,(p,b,p)),(bp,(p,p,b),(bp,(p,p,p))\Big\}$$
$$B = \Big\{(bb,(b,b,b))\Big\}$$
$$P = \Big\{(pp,(p,p,p))\Big\}$$
$$B2 = \Big\{(bb,(b,b,b)),(bp,(b,b,b)),(bp,(b,b,p))\Big\}$$
$$C = \Big\{(bb,(b,b,b)),(bp,(b,b,b)),(bp,(b,p,b)),(bp,(p,b,b)),(bp,(p,p,b)\Big\}$$

Por hipótese, $\mathbb{P}(B) = \mathbb{P}(P) = 1/4$ e $\mathbb{P}(D) = 1/2$. Queremos calcular $\mathbb{P}(C|B2)$. Como curiosidade, calcule $\mathbb{P}(B|B2)$, $\mathbb{P}(D|B2)$ e $\mathbb{P}(C|B2)$ usando contagem e as probabilidades atribuídas inicialmente. Verifique que os resultados são idênticos ao resultados vistos em aula obtidos usando-se a fórmula de Bayes.

Não é necessário a todo momento explicitar o espaço de probabilidade (o que inclui especificar o espaço amostral), mas deve-se estar atento a não cometer erros especificando-se probabilidades que não satisfazem os axiomas (e que portanto não são probabilidades). Poer exemplo[1], se uma pessoa afirmar que

- a chance de chover hoje é de 30%,
- a chance de chover amanhã é de 40%,
- a chance de que se chova hoje e amanhã é de 20%,
- a de chover hoje ou amanhã é de 60%,

estes valores não são consistentes com os axiomas de uma probabilidade (por que?).

Durante o curso, não iremos mais prestar muita atenção a espaços de probabilidade associados a experimentos. Para variáveis aleatórias, a especificação de suas distribuições de probabilidade é o que basta.

[1]Ver S. Ross, *A First Course in Probability.*