Revisão de Probabilidade e Estatística

Preparatória para a disciplina de Econometria I do PPGECO/UFSC

Aishameriane Schmidt

Última revisão: 26 de fevereiro de 2019

Por favor, envie observações, comentários e correções para aishameriane@gmail.com

Sumário

Introdução

Ao longo destas notas iremos estudar tópicos de probabilidade e de estatística que são pré-requisitos para a disciplina de Econometria I do mestrado/doutorado no PPGEco/UFSC. As aulas são mais focadas em aspectos da teoria, porém na medida do possível exemplos computacionais em R serão apresentados, em conjunto com as aulas de R que serão ministradas no contraturno. **Essas notas não são totalmente autorais, muitas das definições foram apenas copiadas de livros, então por favor não distribua elas.**

Alguns exercícios tem solução em formato digitalizado e as soluções serão disponibilizadas conforme demanda. Caso você tenha alguma dúvida ou exercício que queira discutir, me procure. É possível que exercícios estejam repetidos. Nem todos conteúdos e nem todos exercícios serão vistos em sala de aula. *Estas aulas não são oficiais e nem contam nota, mas servem para moldar seu caráter. =P*

Qual a diferença entre probabilidade e estatística?
A probabilidade é, para a estatística, aquilo que o cálculo representa para as engenharias: uma ferramenta. A probabilidade é considerada por muitas pessoas como uma sub-área da matemática. Apesar de estar relacionada com a estatística através de uma certa *noção de incerteza*, a probabilidade não se ocupa de dados, isto é, não há uma amostra, nem uma população e muito menos um problema de inferência. Neste sentido, a estatística é formada de métodos para descrição de amostras, populações e de inferência do primeiro grupo para o segundo. Já a probabilidade acaba lidando com noções mais abstratas e, na sua forma axiomática, não irá tratar de dados coletados.

Iremos começar as aulas revisando as noções matemáticas de conjuntos para então fazer a revisão de probabilidade, que será a maior parte das aulas. O programa das aulas pode ser acessado aqui: `https://github.com/aishameriane/msc-economics/tree/master/revisao-prob`.

1 Noções de teoria dos conjuntos

Esta seção foi retirada de Lima (1982).

Um **conjunto** (ou coleção) é formado por objetos (que são os seus **elementos**). Eles têm como relação básica a relação de **pertinência**, que possibilita relacionar elementos com conjuntos. Quando x (objeto) é um elemento do conjunto A, dizemos que "*x pertence a A*" e denotamos por $x \in A$. Caso contrário, diremos que "*x não pertence a A*" e denotaremos por $x \notin A$. Um conjunto fica bem especificado (definido) quando há uma regra clara que permita avaliar se um elemento arbitrário pertence ou não ao conjunto.

Uma vez que existem conjuntos de tamanho muito grande, se torna difícil sempre que nos referirmos a um conjunto listarmos todos seus elementos. Sendo assim, podemos utilizar uma notação mais compacta que, através de uma regra de pertinência (ou propriedade), deixa o conjunto totalmente especificado. Por exemplo, podemos ter uma situação onde a propriedade P define totalmente o conjunto A: se um objeto x atende P, então $x \in A$, caso contrário, $x \notin A$. Então, podemos escrever:

$$A = \{x : x \text{ tem a propriedade } P\}$$

Os dois pontos na expressão acima fazem o papel da expressão "*tal que*". Outra notação usual, ao invés dos dois pontos, é a barra vertical "|" ou ainda o ponto e vírgula - notação utilizada por (Lima, 1982). Quando estamos falando de um subconjunto $B \subset A$ (B é subconjunto de A ou, equivalentemente, B está contido em A), podemos escrever:

$$B = \{x \in A : x \text{ tem a propriedade } P\}.$$

A expressão acima significa que o conjunto B são os elementos x do conjunto A que satisfazem a propriedade P. Por exemplo, se queremos nos referir aos números reais maiores que 10, podemos definir:

$$A = \{x \in \mathbb{R} : x > 10\},$$

e neste caso, A é um subconjunto do conjunto dos números reais, $\mathbb{R}$, isto é, $A \subset \mathbb{R}$.

Se A e B forem conjuntos, podemos compará-los através da relação de "*inclusão*" ($\subset$). Dizemos que B é subconjunto de A se todo elemento de B também[1] é elemento de A e denotamos por $B \subset A$. Neste caso, também são utilizados os termos: B é *parte* de A, B está *incluído* em A ou ainda B está *contido* em A. Em alguns livros encontramos as notações $B \subset A$ para indicar que B está contido em A (mas não é igual), $B \subseteq A$ para indicar que B está contido e pode ser igual a A ou ainda $B \subsetneq A$ para indicar B está contido mas não é igual a A. Nestas notas, iremos utilizar $B \subset A$ como "B está contido em A e eles podem ser iguais".

Quando não há elemento de A que satisfaça P, o conjunto B não tem nenhum elemento e é denominado *conjunto vazio* ($\emptyset$). Definimos $\emptyset$ da seguinte forma:

$$\forall x, x \notin \emptyset,$$

[1] Observe que não necessariamente a recíproca é verdadeira, não é necessário que todo elemento de A seja elemento de B. Quando a recíproca é de fato verdadeira, os dois conjuntos são iguais.

em que lê-se: qualquer que seja x, x não percence ao vazio. Por exemplo, $\{x \in \mathbb{N} | 1 < x < 2\} = \emptyset$. Perceba que uma implicação importante da relação de pertinência é que qualquer que seja o conjunto A, temos $\emptyset \subset A$. Isso é verdade pois o contrário implicaria que existiria um elemento do conjunto vazio que não pertence ao conjunto A. No entanto, o conjunto vazio não contém elementos, o que implica que ele deve ser subconjunto de todos demais conjuntos.

Uma coisa importante a se notar é que existe diferença entre $\emptyset$ e $\{\emptyset\}$. O primeiro é o conjunto vazio, o segundo é um conjunto cujo único elemento é o conjunto vazio. Para entender melhor a diferença, considere o seguinte exercício:

Exercício 1. Analise se são verdadeiras ou falsas as seguintes sentenças:

1. $\emptyset \in \emptyset$.
2. $\emptyset \in \{\emptyset\}$.
3. $\{\emptyset\} \in \emptyset$.
4. $\emptyset \subset \emptyset$.
5. $\emptyset \subset \{\emptyset\}$.
6. $\{\emptyset\} \subset \emptyset$.

Dado um conjunto A qulquer, podemos definir o conjunto formado pelas partes de A, que é denotado por $\mathscr{P}(A)$. Em outras palavras, se $B \subset A$, então $B \in \mathscr{P}(A)$. Note que este conjunto nunca é vazio pois teremos ao menos $\emptyset \in \mathscr{P}(A)$ e $A \in \mathscr{P}(A)$. Por exemplo:

$$A = \{a, b, c\} \Rightarrow \mathscr{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

1.1 Eventos

Em probabilidade, utilizamos muito a teoria de conjuntos principalmente ao falar de probabilidade de eventos. Vamos introduzir algumas definições básicas e ver alguns exercícios (ainda sem probabilidade). Algumas definições desta seção foram retiradas de Stern and Izbicki (2016).

O modelo probabilístico tem dois "ingredientes" básicos: o *espaço amostral* (S) e a *lei de probabilidade*, para falar deles, precisamos entender a noção de *experimento*.

Por experimento podemos entender qualquer atividade cujo resultado final é desconhecido (porém sabemos quais são as possibilidades resultantes dele, apenas não sabemos a priori o que irá sair). Por exemplo, para o lançamento de uma moeda, podemos ter o resultado cara ou coroa e portanto $S = \{(cara), (coroa)\}$. Não sabemos, antes de lançar a moeda, mas sabemos que será[2] $cara$ ou $coroa$.

Definição 1.1.1. Espaço Amostral

O conjunto de todos os possíveis resultados de um experimento particular é chamado de *espaço amostral* e é denotado por Ω (lê-se *ômega*). Este conjunto pode ser: enumerável, finito ou infinito, se houver uma bijeção $f : \Omega \to \mathbb{N}$ ou ainda, pode ser não enumerável (por exemplo, no caso de $\Omega = \mathbb{R}$).

[2]Estamos abstraindo aqui a possibilidade de outros eventos, como por exemplo, um meteoro cair na terra ou um albatroz pegar a moeda antes de observarmos seu resultado.

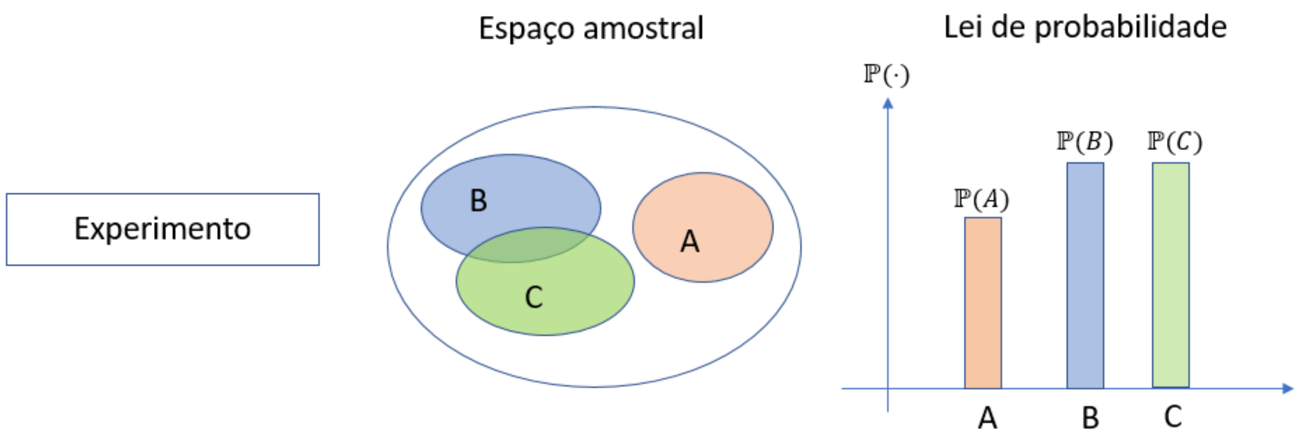

Figura 1: Representação de experimento, espaço amostral e lei de probabilidade.

Exemplo 1.1.2. Espaço amostral do lançamento de uma moeda Imagine que você faz o experimento de lançar uma moeda. Considere que K significa que o resultado foi cara e C significa que o resultado foi coroa. Então, $\Omega = \{K, C\}$ é o espaço amostral do experimento.

Exemplo 1.1.3. Espaço amostral do tempo até uma lâmpada queimar

Considere agora o seguinte experimento: você observa uma lâmpada e está interessado no tempo, em minutos, até a lâmpada queimar[3]. Então, $\Omega = [0, +\infty)$.

Definição 1.1.4. Evento Um *evento* é qualquer coleção de possíveis resultados de um experimento, isto é, qualquer subconjunto de Ω (incluindo o próprio Ω).

Definição 1.1.5. Relação entre eventos

$$A \subset B \iff x \in A \Rightarrow x \in B \tag{1}$$

$$A = B \iff A \subset B \quad e \quad B \subset A \tag{2}$$

A equação 1 significa, em palavras, que *o conjunto A está contido no conjunto B se e somente se qualquer elemento de A está pertence a B*. Já a equação 2 fala da igualdade de dois conjuntos e estabelece que *dois conjuntos A e B serão iguais se e somente se A está contido B e B está contido em A*. Se usarmos 1 em 2, chegamos à conclusão que A e B serão iguais se e somente se qualquer elemento de A está em B e qualquer elemento de B está em A.

Podemos definir ainda três operações de conjuntos: união, interseção e complementar.

Definição 1.1.6. União de dois eventos A união de dois eventos A e B, representada por $A \cup B$, é o conjunto de elementos que percentem a A, B ou ambos:

$$A \cup B = \{x : x \in A \quad ou \quad x \in B\} \tag{3}$$

De maneira mais geral, seja $(A_n)_{n \in \mathbb{N}}$ uma sequência de conjuntos. $x \in \Omega$ é um elemento da união de $(A_n)_{n \in \mathbb{N}}$, denotada por $\cup_{n \in \mathbb{N}} A_n$, se e somente se existe um $n \in \mathbb{N}$ tal que $x \in A_n$. Isto é, $\cup_{n \in \mathbb{N}} A_n = \{x \in \Omega : \textbf{existe } n \in \mathbb{N} \text{ tal que } x \in A_n\}$

Definição 1.1.7. Interseção de dois eventos A interseção de dois eventos A e B, representada por $A \cap B$, é o conjunto de elementos que percentem a A e, ao mesmo tempo, B:

[3] O intervalo será fechado em zero se considerarmos que é possível que a lâmpada já esteja queimada quando iniciamos o teste.

$$A \cap B = \{x : x \in A \quad e \quad x \in B\} \tag{4}$$

Dizemos ainda que A e B são eventos (ou conjuntos) disjuntos se, e somente se, $A \cap B = \emptyset$.

De maneira mais geral, seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de conjuntos. $x \in \Omega$ é um elemento da interseção de $(A_n)_{n\in\mathbb{N}}$, denotada por $\cap_{n\in\mathbb{N}} A_n$, se e somente se para todo $n \in \mathbb{N}$, $x \in A_n$. Isto é, $\cap_{n\in\mathbb{N}} A_n = \{x \in \Omega : \textbf{para todo } n \in \mathbb{N}, x \in A_n\}$.

Definição 1.1.8. Complementar de um evento Seja A um conjunto. x é um elemento de A^c se e somente se $x \notin A$. Isto é, o complemento de A é definido formalmente como $A^c = \{x \in \Omega : x \notin A\}$.

Exemplo 1.1.9. Complementar do espaço amostral
O complementar do espaço amostral (Ω) é o conjunto vazio, $\emptyset$, pois:

- Como o conjunto $\emptyset$ não possui elementos, $\forall \omega \in \Omega$ temos que $\omega \notin \emptyset$;
- Uma vez que $\emptyset$ não possui elementos, não há elemento de $\emptyset$ que pertença a Ω (dizemos que isso ocorre por *vacuidade*).

O exemplo pode parecer confuso, pois também por vacuidade temos que $\emptyset \subset \Omega$. Para mais detalhes, aqui tem uma explicação boa: http://mathcentral.uregina.ca/qq/database/qq.09.06/narayana1.html. Em geral, o complementar não será subconjunto do conjunto original.

Teorema 1.1.10 (Leis de De Morgan)**.** *Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de subconjuntos de Ω. Então, para todo $n \in \mathbb{N}$,*

- $(\cup_{i=1}^n A_i)^c = \cap_{i=1}^n A_i^c$
- $(\cap_{i=1}^n A_i)^c = \cup_{i=1}^n A_i^c$

Além disso,

- $(\cup_{i\in\mathbb{N}} A_i)^c = \cap_{i\in\mathbb{N}} A_i^c$
- $(\cap_{i\in\mathbb{N}} A_i)^c = \cup_{i\in\mathbb{N}} A_i^c$

Definição 1.1.11 (Partição)**.** Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de conjuntos. Dizemos que $(A_n)_{n\in\mathbb{N}}$ particiona Ω se:

- para todo $i, j \in \mathbb{N}$ tal que $i \neq j$, A_i e A_j são disjuntos.
- $\cup_{n\in\mathbb{N}} A_n = \Omega$.

O último conceito que iremos ver antes de fazer alguns exercícios é o de álgebra e de σ-álgebra (lê-se *"sigma-algebra"*).

Definição 1.1.12. Álgebra[4]
Seja $\mathscr{F}$ uma família de conjuntos (eventos) de Ω. Diremos que $\mathscr{F}$ é uma *álgebra sobre* Ω se satisfaz as seguintes propriedades:

1. $\Omega \in \mathscr{F}$;
2. $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$ - fechada por complementar
3. $A, B \in \mathscr{F} \Rightarrow A \cup B \in \mathscr{F}$ - fechada por uniões finitas de seus elementos.

[4] Adaptado de https://www.ime.usp.br/~tassio/1-2012/probabilidade-1/notas-de-aula-probabilidade.pdf

Em geral, nos cursos de probabilidade regulares, não se vê a definição de σ-álgebra. Nós não iremos nos aprofundar muito nessas álgebras também, porém essa definição mais tarde vai nos ajudar a deixar algumas definições de probabilidade um pouco mais precisas. A σ-álgebra é um pouco mais geral que a álgebra, pois ela generaliza a propriedade 3 para o caso de uniões infinitas (arbitrárias).

Definição 1.1.13. σ-algebra[5]
Seja $\mathscr{F}$ uma família de conjuntos (eventos) de Ω. Diremos que $\mathscr{F}$ é uma *álgebra sobre* Ω se satisfaz as seguintes propriedades:

1. $\Omega \in \mathscr{F}$;
2. $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$;
3. $A_1, A_2, \ldots \in \mathscr{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathscr{F}$

1.2 Exercícios

Exercício 2. Prove as seguintes propriedades:

a.	Comutatividade	$A \cup B = B \cup A$
		$A \cap B = B \cap A$
b.	Associatividade	$A \cup (B \cup C) = (A \cup B) \cup C$
		$A \cap (B \cap C) = (A \cap B) \cap C$
c.	Leis distributivas	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
		$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
d.	Leis de De Morgan	$(A \cup B)^c = A^c \cap B^c$
		$(A \cap B)^c = A^c \cup B^c$

Dica: Para provar que dois conjuntos são iguais, é necessário utilizar a definição 2. É possível que até aqui você tenha trabalhado nessas provas utilizando os diagramas de Venn, porém eles não são considerados como uma prova formal. Por exemplo, vamos fazer a prova da lei distributiva da interseção $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Queremos mostrar duas coisas:
(1) $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$
(2) $A \cap (B \cup C) \supset (A \cap B) \cup (A \cap C)$

Observe que, utilizando notação de conjuntos,

$$A \cap (B \cup C) = \{x \in \Omega : x \in A \text{ e } x \in (B \cup C)\}$$
$$(A \cap B) \cup (A \cap C) = \{x \in \Omega : x \in (A \cap B) \text{ ou } x \in (A \cap C)\}$$

Vamos mostrar (1). Tome $x \in A \cap (B \cup C)$. Pela definição de interseção, temos que $x \in A$ (*) e, ao mesmo tempo, $x \in (B \cup C)$. Mas isso implica, pela propriedade de união, que $x \in B$ ou $x \in C$. Juntando com (*), temos que $x \in (A \cap B)$ ou $x \in (A \cap C)$. Logo, $x \in (A \cap B) \cup (A \cap C)$.

Agora, vamos mostrar (2).

Seja $x \in (A \cap B) \cup (A \cap C)$. Então, $x \in (A \cap B)$ ou $x \in (A \cap C)$. Abrimos então em dois casos:
Caso a: $x \in (A \cap B)$. Isso significa que $x \in A$ e, ao mesmo tempo, $x \in B$. Logo, x pertence à qualquer união de B, em

[5] Adaptado de `https://www.ime.usp.br/~tassio/1-2012/probabilidade-1/notas-de-aula-probabilidade.pdf`

particular, $x \in (B \cup C)$. Logo, $x \in A \cap (B \cup C)$.
Caso b: $x \in (A \cap C)$. Pelo mesmo argumento anterior, isso significa que $x \in A$ e, ao mesmo tempo, $x \in C$. Logo, x pertence à qualquer união de C, em particular, $x \in (C \cup B)$. Se você provou a propriedade de comutatividade da letra a, pode concluir que $x \in A \cap (B \cup C)$.

Vamos resolver o segundo item da parte (d) $(A \cap B)^c = A^c \cup B^c$.

- **Parte 1:** Queremos mostrar que $(A \cap B)^c \subset A^c \cup B^c$, isto é, mostrar que todo elemento de $(A \cap B)^c$ pertence a $A^c \cup B^c$.
  Tome $a \in (A \cap B)^c$. Então, $a \notin (A \cap B)^c$, pela definição de complementar, de forma que temos três casos possíveis.

Caso 1: $a \in A$ e $a \notin B$. Neste caso, $a \in B^c$ e portanto está em qualquer união de B^c, em particular, está em $A^c \cup B^c$.

Caso 2: $a \notin A$ e $a \in B$. Neste caso, $a \in A^c$ e portanto está em qualquer união de A^c, em particular, está em $A^c \cup B^c$.

Caso 3: $a \notin A$ e $a \notin B$. Neste caso, $a \in A^c$ e $a \in B^c$ e portanto está em $A^c \cup B^c$.

- **Parte 2:** Queremos mostrar que $(A \cap B)^c \supset A^c \cup B^c$, isto é, que todo elemento que está em $A^c \cup B^c$ pertence a $(A \cap B)^c$.
  Tome $a \in A^c \cup B^c$ (apesar de estar usando a mesma letra a, ele não é o mesmo elemento tomado na parte 1!). Então, por se tratar de uma união, novamente temos 3 casos possíveis.

Caso 1: $a \in A^c$. Então, $a \notin A \Rightarrow a \notin A \cap B$, isto é, o fato de a não pertencer ao conjunto A implica que ele não irá pertencer a qualquer interseção de A, pela própria definição de interseção. Podemos concluir, como a não está na interseção, que ele está no complementar, isto é, $a \in (A \cap B)^c$.

Caso 2: O caso 2 é análogo ao caso 1, supondo que $a \in B^c$.

Caso 3: O caso 3 decorre dos casos 1 e 2, isto é, suponha que $a \in A^c$ e $a \in B^c$.

Os outros exercícios são resolvidos de forma similar. Qualquer problema procure a Aishameriane mais próxima.

Exercício 3.

$$\bigcap_{n=1}^{\infty} \left(0, 1 + \frac{1}{n}\right) \triangle \left[\frac{1}{n}, 2 + \frac{1}{n}\right)$$

onde $A \triangle B = (A \cup B) \cap (A^c \cup B^c)$.

Exercício 4. Seja $\Omega = \{a, b, c, d\}$. Será que $\mathscr{F} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ é uma σ-álgebra de Ω?

Exercício 5. Prove as leis de De Morgan do caso geral (teorema 1.1.10).

Exercício 6. Mostre que $A = (A \cap B) \cup (A \cap B^c)$. Este exercício é útil para conseguirmos escrever um conjunto como a união de conjuntos disjuntos.

Exercício 7. Considere A e B dois subconjuntos de Ω. A *diferença simétrica* entre A e B é o conjunto de todos elementos que estão em A ou em B mas que não estão em ambos. Escreva a diferença simétrica formalmente utilizando as operações de união, intersecção e/ou complementar. Mostre que a diferença simétrica entre A e B é igual à diferença simétrica entre A^c e B^c.

Exercício 8. Seja $A = \{2, 4, 6\}$, $B = \{3, 4\}$ com $\Omega = \{1, 2, 3, 4, 5, 6\}$. Encontre:

a. $A \cup B$

b. $A \cap B$

c. $A \cup B^c$

d. $A \cup A^c$ e $A \cap A^c$

e. $B \cup B^c$ e $B \cap B^c$

f. $(A \cup B)^c$

g. $(A \cap B)^c$

h. $A^c \cap B^c$

i. $A \cup \Omega$

j. $A^c \cup \Omega$

k. $B \cap \Omega$

l. $B^c \cap \Omega$

Exercício 9. Considere que você está analisando para um dia o evento *"A variação do dólar foi negativa em relação ao início do dia"* (N) e *"A variação do dólar foi positiva em relação ao início do dia"* (P). Estamos interessados na variação do dólar em dois dias consecutivos, isto é, se houveram duas variações negativas, duas positivas, etc.

a. Como você escreveria formalmente (notação de conjuntos) o espaço amostral Ω?

b. Como você escreveria formalmente *"Houveram dois dias consecutivos de variação negativa no dólar"*? Chame este evento de A.

c. Como você escreveria formalmente o evento *"Houve pelo menos um dia com variação negativa no dólar"*? Chame este evento de B.

d. Avalie se $A \subset B$.

e. Descreva com suas palavras quem é o conjunto B^c. Escreva-o formalmente.

f. A e B^c são disjuntos? Justifique.

Exercício 10. De um grupo de 25 alunos:

- 14 irão comprar itens para fazer churrasco no final de semana;
- 12 irão comprar um chocolate para o professor de Estatística;
- 5 irão fazer o churrasco e comprar o chocolate para o professor.

Quantos alunos não irão nem comprar os itens de churrasco nem o chocolate?

Exercício 11. Escreva o espaço amostral dos seguintes experimentos:

a. Duas moedas são lançadas simultâneamente e observa-se a sequência de caras e coroas obtida;

b. Um dado é lançado e observa-se a face virada para cima;

c. Duas cartas de um baralho são retiradas e observa-se a sequência dos naipes;

d. Dois dados são lançados e observa-se as faces viradas para cima;

e. Dois dados são lançados e observa-se a soma das duas faces;

f. Uma moeda e um dado são lançados e observa-se a face da moeda e do dado viradas para cima;

g. Três times A,B e C participam de um campeonato de *curling*. Inicialmente o time A joga contra o time B e o vencedor joga com o time C e assim por diante. O campeonato finaliza quando um time ganha duas vezes em seguida ou quando são disputadas, ao todo, quatro partidas. Você está interessado no espaço amostral: resultados possíveis das partidas do torneio.

Exercício 12. Seja o experimento da letra d. do exercício anterior. Defina os seguintes eventos:

- $A =$ a soma das faces dos dois dados é igual a 5;
- $B =$ a face do primeiro dado é menor ou igual a 2.

Determine os eventos A e B em termos de Ω encontrado anteriormente (isto é, explicite quais resultados de Ω pertencem a A e B) e determine os seguintes eventos:

- $A \cup B$;
- $A \cap B$;
- A^c.

2 Probabilidade

Nesta seção iremos introduzir os principais conceitos de probabilidade. Apesar do uso de algumas definições e teoremas, ainda não estamos fazendo uso das definições mais formais, como por exemplo, σ-álgebra. É comum encontrar nos livros específicos de probabilidade esse conteúdo de forma mais aprofundada, porém não é comum que se encontre dessa forma nos textos de estatística econômica. Para quem desejar se aprofundar no assunto de maneira mais rigorosa, recomenda-se a leitura do capítulo 1 de James (2010).

2.1 Definições básicas

Existem três definições de probabilidade que são comumente apresentadas nos livros texto: a probabilidade por frequência relativa, a definição subjetiva e a definição axiomática. As seguintes definições foram retiradas de Mittelhammer (2013):

Definição 2.1.1. Probabilidade por frequência relativa
Seja n o número de vezes que um experimento é repetido sob condições idênticas. Seja A o evento no espaço amostral Ω e defina n_A o número de vezes que o evento A ocorreu. Então, a probabilidade do evento A é igual a:

$$\mathbb{P}(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

Apesar de ser uma definição que tem bases em experimentos *reais* e por isso ser mais familiar, ela não permite desenvolver formalmente muitos resultados, uma vez que não há garantias de que o limite irá convergir em todos os casos. Além disso, não é nem possível observar infinitas repetições para termos certeza deste limite.

Definição 2.1.2. Probabilidade subjetivista
Um número real, $\mathbb{P}(A)$, contido no intervalo $[0, 1]$ e escolhido para expressar o grau de crença pessoal na possibilidade de ocorrência de um evento A, sendo que 1 representa absoluta certeza.

A definição subjetiva está relacionada com a estatística Bayesiana, que incorpora essa subjetividade de maneira a não deixar toda a informação sobre os parâmetros apenas para a amostra. Voltaremos nessa definição na última aula.

A definição axiomática de probabilidade foi desenvolvida por Kolmogorov em 1956 e se baseia em *axiomas*. Axiomas são afirmações que consideramos como verdades (sem a necessidade de prová-los) e a partir disso os teoremas são provados. Por exemplo, temos na geometria Euclidiana, o axioma que diz que *dois pontos podem ser ligados por uma única reta*.

Definição 2.1.3. Probabilidade axiomática
Seja $\mathbb{P}$ uma função que tem como domínio o espaço amostral Ω e contradomínio em $\mathbb{R}$, isto é, temos $\mathbb{P} : \gamma \to \mathbb{R}$, onde γ é o conjunto de todos eventos em Ω, também denominado espaço de eventos. Diremos que $\mathbb{P} : \gamma \to \mathbb{R}$ é uma função de probabilidade (ou medida de probabilidade) se atender aos seguintes axiomas:

1. $\mathbb{P}(A) \geq 0, \ \forall\ A \subset \Omega$ (*Axioma da não-negatividade*);
2. $\mathbb{P}(\Omega) = 1$ (*Axioma da normalização*);
3. Se $(A_n)_{n\in\mathbb{N}}$ é uma sequência de conjuntos disjuntos em Ω, $\mathbb{P}(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_i)$. (*Axioma da aditividade enumerável*).

Iremos diferenciar a *função de probabilidade* da *probabilidade*, sendo que esta última é a imagem de um evento através da função $\mathbb{P}$.

Da definição axiomática decorrem imediatamente cinco lemas, que são comumente utilizados (e quase nunca demonstrados).

Lema 2.1.4. ***A probabilidade de nada ocorrer é zero***
$\mathbb{P}(\emptyset) = 0$

Demonstração. Tome $(\{A_n\})_{n\in\mathbb{N}}$ tal que $A_1 = \Omega$ e $A_n = \emptyset\ \forall\ n > 1$. Note que $\cap_{n\in\mathbb{N}} A_n = \emptyset$, portanto:

$$\mathbb{P}(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

Podemos então escrever:

$$\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \mathbb{P}(\Omega) + \sum_{n>1} \mathbb{P}(\emptyset) \tag{5}$$

Por outro lado, sabemos que $\cup_{n\in\mathbb{N}} A_n = \Omega \cup \emptyset \cup \emptyset \cup \ldots$, de forma que $\mathbb{P}(\cup_{n\in\mathbb{N}} A_n) = \mathbb{P}(\Omega)$. Assim, juntando com 5,

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \sum_{n>1} \mathbb{P}(\emptyset)$$
$$\mathbb{P}(\Omega) - \mathbb{P}(\Omega) = \sum_{n>1} \mathbb{P}(\emptyset)$$
$$0 = \sum_{n>1} \mathbb{P}(\emptyset) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(\emptyset)$$

Note que $\sum_{k=1}^{n} P(\emptyset)$ é não decrescente pois a probabilidade é sempre não negativa. Como o limite é igual a zero e as somas parciais formam uma sequência não decrescente, as somas finitas são zero. Logo, $\mathbb{P}(\emptyset) = 0$. □

Lema 2.1.5. ***A probabilidade da união é a soma das probabilidades se os eventos forem disjuntos***
Se $A_1, A_2, \ldots, A_n$ são disjuntos, então $\mathbb{P}(A_1 \cup A_2 \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.

Demonstração. Tome $\{B_n\}_{n\in\mathbb{N}}$ tal que $B_i = A_i \; \forall \; i \in \{1, \ldots, n\}$ e $B_i = \emptyset$ para $i > n$. Por construção, os B_i's são disjuntos e portanto:

$$\begin{aligned}
\mathbb{P}(\cup_{i=1}^{n} A_i) &= \mathbb{P}(\cup_{i\in\mathbb{N}} B_i) \\
&= \sum_{i\in\mathbb{N}} \mathbb{P}(B_i) \\
&= \sum_{i=1}^{n} \mathbb{P}(B_i) + \sum_{k=n+1}^{\infty} \mathbb{P}(B_k) \\
&= \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{k=n+1}^{\infty} \mathbb{P}(\emptyset) \\
&= \sum_{i=1}^{n} \mathbb{P}(A_i) + 0 \\
&= \sum_{i=1}^{n} \mathbb{P}(A_i)
\end{aligned}$$

Juntando as duas extremidades, segue o resultado desejado. □

Lema 2.1.6. ***Probabilidade do complementar***
Para todo evento A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Demonstração. Sabemos que A e A^c são disjuntos pois $A \cap A^c = \emptyset$. Por outro lado, temos que $A \cup A^c = \Omega$. Então:

$$\mathbb{P}(\Omega) = 1 = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$$

Segue que $1 = \mathbb{P}(A) + \mathbb{P}(A^c) \Rightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. □

Lema 2.1.7. ***Probabilidade da união***
Para todos os eventos A e B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Demonstração. Note que podemos escrever o conjunto A como a união de dois conjuntos disjuntos:

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

Então:

$$\mathbb{P}(A) = \mathbb{P}[(B^c \cap A) \cup (B \cap A)] = \mathbb{P}(B^c \cap A) + \mathbb{P}(B \cap A)$$

De forma que:

$$\mathbb{P}(B^c \cap A) = \mathbb{P}(A) - \mathbb{P}(B \cap A) \tag{6}$$

Note que B é disjunto de $B^c \cap A$ e que $B \cup (B^c \cap A) = (B \cup B^c) \cap (B \cup A) = \Omega \cap (B \cup A) = (B \cup A)$, portanto,

$$\mathbb{P}(B \cup A) = \mathbb{P}(B) + \mathbb{P}(B^c \cap A) \tag{7}$$

Juntando 6 e 7, temos:

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

□

Lema 2.1.8. ***Probabilidade de subconjuntos***
Se $A \subset B$, então $\mathbb{P}(B) \geq \mathbb{P}(A)$.

Demonstração. Já vimos que $B = (A^c \cap B) \cup (A \cap B)$, que são disjuntos. Então:

$$\mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) \tag{8}$$

Mas $A = (A \cap B^c) \cup (A \cap B) \Rightarrow \mathbb{P}(A) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c)$.

Juntando com 8, temos que $\mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A) - \mathbb{P}(A \cap B^c)$. Mas $A \subset B$, logo, $(A \cap B^c) = \emptyset$, de forma que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$.

Como $\mathbb{P}(A^c \cap B) \geq 0$, temos $\mathbb{P}(B) \geq \mathbb{P}(A)$. □

2.1.1 Exercícios

Exercício 13. Mostre que $\mathbb{P}((A \cap B^c) \cup (A^c \cap B)) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$.

Exercício 14. Prove que

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

Exercício 15. Seja Ω o espaço amostral e A, B e C eventos. Prove que:

a. $\mathbb{P}(A \cup B) = 1 - \mathbb{P}(A^c \cap B^c)$.

b. $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.

c. $\max(\mathbb{P}(A), \mathbb{P}(B)) \leq \mathbb{P}(A \cup B) \leq \min(1, \mathbb{P}(A) + \mathbb{P}(B))$.

d. $\max(0, \mathbb{P}(A) + \mathbb{P}(B) - 1) \leq \mathbb{P}(A \cap B) \leq \min(\mathbb{P}(A), \mathbb{P}(B))$.

Exercício 16. Sejam $A_1, A_2, \ldots \in \gamma$. Mostre que:

$$\mathbb{P}(\cup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n)$$

2.2 Independência de eventos e probabilidade condicional

Quando trabalhamos com probabilidades, é muito comum nos depararmos com situações onde desejamos saber a probabilidade de algo ocorrer *dado* que sabemos algo sobre outro evento passado. Em outras palavras, queremos usar nosso conhecimento prévio para fazer alguma previsão sobre eventos futuros.

O reverendo Thomas Bayes (1701-1761) foi uma das primeiras pessoas a pensar neste tipo de situação e desenvolver uma solução matemática para o problema, que ficou conhecido como o problema das *probabilidades inversas*. Sua contribuição inicial deu origem a uma área inteira da estatística que atualmente é chamada, em sua homenagem, de *estatística bayesiana*[6].

Nesta seção iremos abordar somente o Teorema de Bayes e seu uso em probabilidade, sem explorar mais profundamente outros aspectos da área. Porém é importante destacar que atualmente muitos modelos macroeconométricos e microeconométricos estão baseados em metodologia bayesiana para estimação dos seus parâmetros e cada vez mais a metodologia bayesiana tem ganho espaço nas aplicações em economia. Um exemplo é o modelo atualmente utilizado pelo Banco Central do Brasil, o SAMBA (*Stochastic Analytical Model with a Bayesian Approach*). Para entender um pouco mais dessa relação entre estatística bayesiana e econometria, recomendo este texto do professor Christopher Sims (Nobel de Economia em 2011): `http://sims.princeton.edu/yftp/UndrstndgNnBsns/GewekeBookChpter.pdf`.

Considere para todos as definições e teoremas que existe um espaço de probabilidade $(\Omega, \gamma, \mathbb{P}[\cdot])$; isto é, existe um experimento aleatório para o qual o espaço amostral é Ω, uma coleção de eventos γ e uma função de probabilidade[7] $\mathbb{P}[\cdot]$ que estão bem definidos. As definições a seguir foram retiradas e/ou adaptadas das obras listadas nas referências.

Dados dois eventos, A e B, queremos definir a probabilidade condicional de A ocorrer dados que B ocorreu.

Definição 2.2.1. Probabilidade Condicional
Sejam A e B eventos em um espaço de probabilidade $(\Omega, \gamma, \mathbb{P}[\cdot])$. A *probabilidade condicional* do evento A dado B, denotada por $\mathbb{P}[A|B]$, é definida por:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{, se} \quad \mathbb{P}(B) \neq 0 \tag{9}$$

e não está definida para $\mathbb{P}(B) = 0$.

Observação: 1 Um resultado direto da definição 9 é que $\mathbb{P}(A \cap B) = \mathbb{P}[A|B]\mathbb{P}(B) = \mathbb{P}[B|A]\mathbb{P}(A)$ se ambas probabilidades $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são não nulas (faça a conta em um papel para conferir), de forma a relacionar as condicionais $\mathbb{P}[A|B]$ e $\mathbb{P}[B|A]$ com as probabilidades não-condicionadas de A e B.

Observação: 2 A definição 9 vai ao encontro da noção de probabilidade sob o enfoque frequentista. Suponha que foram observadas N ocorrências de um experimento aleatório (sendo N um número grande) para o qual os eventos A e B estão definidos. Então, $\mathbb{P}[A|B]$ representa a proporção de ocorrências do evento B onde o evento A também ocorreu, isto é:

$$\mathbb{P}[A|B] = \frac{N_{AB}}{N_B}$$

---

[6] Para saber mais sobre a origem da estatística bayesiana, veja McGrayne (2011)

[7] Embora toda a teoria aqui explicitada refira-se a eventos, pode ser extendida naturalmente para variáveis aleatórias tanto discretas como contínuas. Quando necessário, será especificado se a função é de densidade ou distribuição de probabilidade, mas considere como sendo possível adaptar para todos os casos.

Onde N_{AB} representa o número de ocorrências do evento $A \cap B$ e N_B representa o número de ocorrências do evento B nas N observações do experimento. Portanto, $\mathbb{P}[A \cap B] = \frac{N_{AB}}{N}$ e $\mathbb{P}(B) = \frac{N_B}{N}$, de forma que:

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{N_{AB}/N}{N_B/N} = \frac{N_{AB}}{N_B} = \mathbb{P}[A|B]$$

O que é consistente com a definição dada.

Exemplo 2.2.2. Considere o lançamento de duas moedas. Seja $\Omega = \{(C, C), (C, K), (K, C), (K, K)\}$, onde C significa que a face observada foi cara e K significa que a face observada foi coroa. Assuma que as moedas são honestas. Vamos calcular 1) a probabilidade de saírem duas caras dado que saiu cara na primeira moeda e 2) a probabilidade de duas caras dado que pelo menos uma é cara.

Defina os eventos: $A_1 =$ cara na primeira moeda e $A_2 =$ cara na segunda moeda. Então, a probabilidade de saírem duas caras dado que saiu cara na primeira moeda é:

$$\mathbb{P}[A_1 \cap A_2 | A_1] = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} = \frac{1/4}{1/2} = \frac{1}{2}$$

A probabilidade de saírem duas caras dado que pelo menos uma das moedas é cara será igual a $\frac{1}{3}$ e fica sugerida como exercício.

Quando falamos de probabilidades condicionais do tipo $\mathbb{P}[A|B]$, o que fazemos é definir um novo espaço amostral, Ω_B, onde tomamos apenas as ocorrências do evento B e calculamos a probabilidade de A ocorrer. De certa forma, isso pode ser visto como uma *restrição* do espaço amostral original. No exemplo 2.2.2, temos $\Omega_B = \{(C, K), (C, C)\}$ para o primeiro item e $\Omega_B = \{(C, C), (C, K), (K, C)\}$. Em ambos casos, restringimos o espaço amostral para os eventos que já sabemos que ocorreram e com base nesse novo espaço amostral é que iremos calcular a probabilidade de interesse.

Surge então (ou deveria surgir) a pergunta: para um dado evento B para o qual $\mathbb{P}(B) > 0$, será que $\mathbb{P}[\,\cdot\,|B]$ é uma função de probabilidade que tem γ como seu domínio? Isto é, será que $\mathbb{P}[\,\cdot\,|B]$ satisfaz os três axiomas para ser considerada uma função de probabilidade? Observe que sim, pois:

(i) $\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}(B)} \geq 0 \quad \forall\, A \in \gamma$;

(ii) $\mathbb{P}[\Omega|B] = \frac{\mathbb{P}[\Omega \cap B]}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$;

(iii) Se $A_1, A_2, \cdots$ é uma sequência de eventos mutuamente exclusivos em γ e $\bigcup\limits_{i=1}^{\infty} A_i \in \gamma$, então

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i | B\right] = \frac{\mathbb{P}\left[\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \cap B\right]}{\mathbb{P}(B)} = \frac{\mathbb{P}\left[\bigcup\limits_{i=1}^{\infty} (A_i \cap B)\right]}{\mathbb{P}(B)} = \frac{\sum\limits_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}[A_i|B]$$

Então, $P[\,\cdot\,|B]$ para um dado B que satisfaz $\mathbb{P}(B) > 0$ é uma função de probabilidade, o que justifica chamá-la de probabilidade condicional. $\mathbb{P}[\,\cdot\,|B]$ também apresenta as mesmas propriedades que uma probabilidade não condicionada. Logo, podemos enunciar os seguintes resultados que são similares aos já obtidos para probabilidades não condicionais:

Teorema 2.2.3. *A probabilidade condicional do vazio é zero, isto é,* $\mathbb{P}[\emptyset\,|B] = 0$.

Demonstração.

$$\mathbb{P}(\emptyset|B) = \frac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = \frac{0}{\mathbb{P}(B)} = 0$$

□

Teorema 2.2.4. *Se $A_1, A_2, \cdots, A_n$ são eventos mutuamente exclusivos em γ, então*

$$\mathbb{P}[A_1 \cup \cdots \cup A_n|B] = \sum_{i=1}^{n} \mathbb{P}[A_i|B]$$

Demonstração.

$$\begin{aligned}
\mathbb{P}[A_1 \cup \cdots \cup A_n|B] = & & \text{(definição prop. condicional)} \\
&= \frac{\mathbb{P}[(A_1 \cup \cdots \cup A_n) \cap B]}{\mathbb{P}(B)} & \text{(distributiva)} \\
&= \frac{\mathbb{P}[(A_1 \cap B) \cup \cdots \cup (A_n \cap B)]}{\mathbb{P}(B)} & \text{(independência dos } A_i\text{'s)} \\
&= \frac{\mathbb{P}\left[\sum_{i=1}^{n}(A_i \cap B)\right]}{\mathbb{P}(B)} & \text{(definição prob. condicional)} \\
&= \sum_{i=1}^{n} \mathbb{P}[A_i|B]
\end{aligned}$$

□

Teorema 2.2.5. *Se A é um evento em γ, então*

$$\mathbb{P}[A^c|B] = 1 - \mathbb{P}[A|B]$$

onde A^c é o evento complementar de A.

Demonstração.

$$\begin{aligned}
\mathbb{P}[A^c|B] = & & \text{(def. prob. condicional)} \\
&= \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(B)} & \text{(forma alternativa de } \mathbb{P}(B)\text{)} \\
&= \frac{\mathbb{P}(B) - \mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B)}{\mathbb{P}(B)} - \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= 1 - \mathbb{P}[A|B]
\end{aligned}$$

□

Teorema 2.2.6. *Se A_1 e A_2 pertencem a γ, então*

$$\mathbb{P}[A_1|B] = \mathbb{P}[A_1 \cap A_2|B] + \mathbb{P}[A_1 \cap A_2^c|B]$$

Demonstração.

$$\begin{aligned}
\mathbb{P}[A_1|B] &= && \text{(def. prob. cond.)}\\
&= \frac{\mathbb{P}[A_1 \cap B]}{\mathbb{P}(B)} && \text{(reescrevendo } A_1\text{)}\\
&= \frac{\mathbb{P}[((A_1 \cap A_2) \cup (A_1 \cap A_2^c)) \cap B]}{\mathbb{P}(B)} && \text{(distributiva)}\\
&= \frac{\mathbb{P}[(A_1 \cap A_2) \cap B \cup (A_1 \cap A_2^c) \cap B]}{\mathbb{P}(B)} && \text{(independência)}\\
&= \frac{\mathbb{P}[(A_1 \cap A_2) \cap B] + \mathbb{P}[(A_1 \cap A_2^c) \cap B]}{\mathbb{P}(B)} && \text{(reorganizando)}\\
&= \frac{\mathbb{P}[(A_1 \cap A_2) \cap B]}{\mathbb{P}(B)} + \frac{\mathbb{P}[(A_1 \cap A_2^c) \cap B]}{\mathbb{P}(B)} && \text{(def. prob. cond.)}\\
&= \mathbb{P}[A_1 \cap A_2|B] + \mathbb{P}[A_1 \cap A_2^c|B]
\end{aligned}$$

□

Teorema 2.2.7. *Para quaisquer dois eventos A_1 e $A_2 \in \gamma$,*

$$\mathbb{P}[A_1 \cup A_2|B] = \mathbb{P}[A_1|B] + \mathbb{P}[A_2|B] - \mathbb{P}[A_1 \cap A_2|B]$$

Demonstração.

$$\begin{aligned}
\mathbb{P}[A_1 \cup A_2|B] &= && \text{(def. prob. cond.)}\\
&= \frac{\mathbb{P}[(A_1 \cup A_2) \cap B]}{\mathbb{P}(B)} && \text{(distributiva)}\\
&= \frac{\mathbb{P}[(A_1 \cap B) \cup (A_2 \cap B)]}{\mathbb{P}(B)} && \text{(def. prob. união.)}\\
&= \frac{\mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) - \mathbb{P}(A_1 \cap B \cap A_2 \cap B)}{\mathbb{P}(B)} && \text{(rearranjando)}\\
&= \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(A_2 \cap B)}{\mathbb{P}(B)} - \frac{\mathbb{P}(A_1 \cap A_2 \cap B)}{\mathbb{P}(B)} && \text{(def. prob. cond.)}\\
&= \mathbb{P}[A_1|B] + \mathbb{P}[A_2|B] - \mathbb{P}[A_1 \cap A_2|B]
\end{aligned}$$

□

Teorema 2.2.8. *Se A_1 e $A_2 \in \gamma$ com $A_1 \subset A_2$, então*

$$\mathbb{P}[A_1|B] \leq \mathbb{P}[A_2|B]$$

Demonstração. Esse resultado decorre do lema 2.1.8. □

Enquanto que os teoremas de 11 a 16 têm uma correspondência direta com as probabilidades não-condicionais, o seguinte teorema tem aplicação apenas no segundo caso:

Teorema 2.2.9. ***Teorema da probabilidade total***

Para um dado espaço de probabilidade $(\Omega, \gamma, \mathbb{P}[\cdot])$, se $B_1, B_2, \cdots, B_n$ são uma coleção de eventos mutuamente disjuntos (ou exclusivos) em γ satisfazendo $\Omega = \bigcup_{j=1}^{n} B_j$ e $\mathbb{P}(B_j) > 0$ para $j = 1, \cdots, n$[8], então, para todo $A \in \gamma$, vale que

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}[A|B_j]\mathbb{P}(B_j)$$

Demonstração. Observe que $A = \bigcup_{j=1}^{n} (A \cap B_j)$ e que os termos $(A \cap B_j)$ e $(A \cap B_i)$ são mutuamente exclusivos para $i \neq j$. Logo,

$$\mathbb{P}(A) = \mathbb{P}\left[\bigcup_{j=1}^{n}(A \cap B_j)\right] = \sum_{j=1}^{n} \mathbb{P}(A \cap B_j) = \sum_{j=1}^{n} \mathbb{P}[A|B_j]\mathbb{P}(B_j)$$

Na aula, eu fiz a versão acima, que é bem direta. Abaixo tem uma forma mais detalhada que pode ajudar a entender melhor o que está acontecendo.

Observe que

$$A = A \cap \Omega \tag{10}$$

Além disso, como $(B_j)_{j=1}^{n}$ é uma partição de Ω, temos

$$\Omega = \cup_{j=1}^{n}(B_j) \tag{11}$$

Usando a equação 11 na equação 10, obtemos:

$$A = A \cap (\cup_{j=1}^{n} B_j) \tag{12}$$
$$= \cup_{j=1}^{n}(A \cap B_j) \tag{13}$$

e portanto,

$$\mathbb{P}(A) = \mathbb{P}(\cup_{j=1}^{n}(A \cap B_j)) \tag{14}$$

Como $(B_j)_{j=1}^{n}$ é uma partição, é também uma sequência disjunta. Portanto, para todo $i \neq j$, $B_i \cap B_j = \emptyset$ e

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) \tag{15}$$
$$= A \cap \emptyset \tag{16}$$
$$= \emptyset \tag{17}$$

Isto é, $(A \cap B_j)_{j=1}^{n}$ também é uma sequência disjunta. Assim, dos axiomas da probabilidade,

[8] Dizemos, neste caso, que os B_J formam uma *partição* de Ω.

$$\mathbb{P}(\cup_{j=1}^{n}(A \cap B_j)) = \sum_{j=1}^{n} \mathbb{P}(A \cap B_j) \tag{18}$$

Finalmente, do axioma da probabilidade condicional, temos para todo $j \in \{1, \ldots, n\}$, $\mathbb{P}(A \cap B_j) = \mathbb{P}(B_j)\mathbb{P}(A|B_j)$. Usando as equações 14 e 18, concluímos que:

$$\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(B_j)\mathbb{P}(A|B_j) \tag{19}$$

□

O próximo resultado é uma implicação direta do teorema da probabilidade total:

Corolário 2.2.10. *Para um dado espaço de probabilidade $(\Omega, \gamma, \mathbb{P}[\cdot])$, seja $B \in \gamma$ satisfazendo $0 < \mathbb{P}(B) < 1$; então, para todo $A \in \gamma$:*

$$\mathbb{P}(A) = \mathbb{P}[A|B]\mathbb{P}(B) + \mathbb{P}[A|B^c]\mathbb{P}(B^c)$$

Demonstração. O resultado segue diretamente do teorema da probabilidade total. □

Finalmente podemos enunciar a fórmula de Bayes:

Teorema 2.2.11. ***Fórmula de Bayes***
Para um dado espaço de probabilidade $(\Omega, \gamma, \mathbb{P}[\cdot])$, se $B_1, B_2, \cdots, B_n$ são uma coleção de eventos mutuamente disjuntos (ou exclusivos) em γ satisfazendo $\Omega = \bigcup_{j=1}^{n} B_j$ e $\mathbb{P}(B_j) > 0$ para $j = 1, \cdots, n$, então, para todo $A \in \gamma$ para o qual $\mathbb{P}(A) > 0$ vale que:

$$\mathbb{P}[B_k|A] = \frac{\mathbb{P}[A|B_k]\mathbb{P}(B_k)}{\sum\limits_{j=1}^{n} \mathbb{P}[A|B_j]\mathbb{P}(B_j)} \tag{20}$$

Demonstração.

$$\mathbb{P}[B_k|A] = \frac{\mathbb{P}[B_k \cap A]}{\mathbb{P}(A)} = \frac{\mathbb{P}[A|B_k]\mathbb{P}(B_k)}{\sum\limits_{j=1}^{n} \mathbb{P}[A|B_j]\mathbb{P}(B_j)}$$

□

Corolário 2.2.12. *Para um dado espaço de probabilidade $(\Omega, \gamma, \mathbb{P}[\cdot])$, sejam A e $B \in \gamma$ satisfazendo $\mathbb{P}(A) > 0$ e $0 < \mathbb{P}(B) < 1$; então:*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}[A|B]\mathbb{P}(B) + \mathbb{P}[A|B^c]\mathbb{P}(B^c)}$$

Definição 2.2.13. Dois eventos A e B são independentes se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Lema 2.2.14. *A e B são independentes se e somente se $\mathbb{P}(A|B) = \mathbb{P}(A)$. Em outras palavras, A e B são independentes se e somente se a incerteza sobre A não muda se assumirmos que B é verdadeiro.*

Demonstração. Assuma $\mathbb{P}(A|B) = \mathbb{P}(A)$. Da definição de probabilidade condicional, lembramos que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Usando a suposição inicial, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ e A e B são independentes.

Assuma A e B independentes. Portanto, $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Juntando os dois extremos da desigualdade e dividindo por $\mathbb{P}(B)$, $\mathbb{P}(A|B) = \mathbb{P}(A)$. □

Teorema 2.2.15. *Se A e B são eventos independentes, então os eventos A e B^c, A^c e B e A^c e B^c também são independentes.*

Demonstração.

$$\begin{aligned}
\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A)[1 - \mathbb{P}(B)] \\
&= \mathbb{P}(A)\mathbb{P}(B^c)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(B)[1 - \mathbb{P}(A)] \\
&= \mathbb{P}(A^c)\mathbb{P}(B)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A^c \cap B^c) &= \mathbb{P}(A \cup B)^c \\
&= 1 - \mathbb{P}(A \cup B) \\
&= 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)] \\
&= 1 - P(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) \\
&= P(A^c) - \mathbb{P}(B)[1 - \mathbb{P}(A)] \\
&= P(A^c)[1 - \mathbb{P}(B)] \\
&= \mathbb{P}(A^c)\mathbb{P}(B^c)
\end{aligned}$$

□

Definição 2.2.16. Independência Condicional

(surrupiado das notas de aula de Econometria Bayesiana) Seja C um evento qualquer, então dizemos que A e B são condicionalmente independentes se:

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B \cap C) \tag{21}$$

Podemos utilizar a definição de independência não condicional com a regra da multiplicação para obter:

$$
\begin{aligned}
\mathbb{P}(A \cap B|C) &= \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} \quad \text{(prob. cond.)} \\
&= \frac{\mathbb{P}(C)\mathbb{P}(B \cap C)\mathbb{P}(A|B \cap C)}{\mathbb{P}(C)} \quad \text{(regra mult.)} \\
&= \mathbb{P}(B|C)\mathbb{P}(A|B \cap C)
\end{aligned}
$$

Igualando as duas pontas da equação acima, temos:

$$
\begin{aligned}
\mathbb{P}(A|C)\mathbb{P}(B|C) &= \mathbb{P}(B|C)\mathbb{P}(A|B \cap C) \\
\mathbb{P}(A|C) = \frac{P(B|C)\mathbb{P}(A|B \cap B)}{\mathbb{P}(B|C)} &= \mathbb{P}(A|B \cap C)
\end{aligned}
$$

De maneira que $\mathbb{P}(A|C) = \mathbb{P}(A|B \cap C)$.

Note que isso não significa que A e B são independentes. Caso o C seja desconhecido, a nossa probabilidade é alterada. É possível que a informação que C tenha para A seja a mesma que B agregaria e por isso que, condicionado a C, a informação de B para A não seja relevante. Assim, *independência condicional não implica independência incondicional.*

2.3 Exercícios

Exercício 17. Independência do vazio
Na aula, conversamos sobre a independência e o que acontece quando um dos eventos é o evento impossível (ou tem probabilidade zero). O exercício então é mostrar que $\forall B \in \Omega$, B e $\emptyset$ são independentes.

Exercício 18. (Retirado de Viali (2004c)) Calcular a probabilidade de no lançamento de um dado equilibrado[9] obter-se os seguintes eventos:

a. A face observada (o resultado do dado, o número virado para cima) é igual a 5;

b. A face observada é um número ímpar;

c. A face observada é maior que 2;

d. A face observada é um número primo;

e. A face observada é menor que π;

f. A face observada é diferente de 3.

Exercício 19. (Adaptado de Stern and Izbicki (2016)) Um total de 4 bolas numeradas de 1 a 4 é distribuído aleatoriamente em n urnas, também numeradas de 1 a 4, uma de cada vez, de tal modo que cada bola tem exatamente a mesma probabilidade de cair em cada uma das urnas. Assim, cada urna pode receber entre 0 e 4 bolas. Qual é a probabilidade da primeira urna ficar vazia?

Qual é essa probabilidade se você fizer o mesmo experimento, porém com *n* bolas e *n* urnas?

[9] Dizemos que um dado é *equilibrado* quando a probabilidade das faces é sempre a mesma. Para um dado de 6 faces, teremos que a probabilidade de cada uma será $\frac{1}{6}$. Outras formas encontradas na literatura são: dado não viciado; dado não viesado; dado honesto. A mesma nomenclatura se aplica a situações envolvendo moedas.

Exercício 20. Na mochila de sua amiga Misty estão 12 *pokébolas*, sendo que 7 são de *pokémons* do tipo água e as outras 5 são de *pokémons* do tipo fada. Destes 12, um deles é o Togepi[10] e o outro é o Psyduck[11].

Misty irá lutar uma batalha de dois rounds, onde cada treinador irá utilizar 2 *pokémons*, sendo um para cada *round* (isto é, mesmo que vença a batalha, o treinador deve trocar de *pokémon*), selecionados aleatoriamente da mochila. Calcule as seguintes probabilidades:

a. Dos 2 *pokémons* que Misty seleciona serem do tipo água e do tipo fada, nessa ordem;

b. De Misty utilizar seu *pokémon* preferido, Togepi[12], na batalha;

c. De Misty não utilizar seu Psyduck[13] na batalha;

d. De Misty utilizar somente dois *pokémons* do tipo água;

e. De Misty utilizar Togepi e o Psyduck na batalha;

f. De Misty utilizar pelo menos um *pokémon* do tipo fada na batalha;

g. Do segundo *pokémon* escolhido ser de água dado que o primeiro foi o Togepi.

h. Resolva os itens a, b e g considerando agora que após a primeira batalha Misty guarda o *pokémon* utilizado de volta na *pokébola*, vai para o *pokémon* center para curá-lo e então retorna para o segundo round, selecionando o segundo pokémon dentre todas suas 12 *pokébolas*.

Exercício 21. (Retirado de Viali (2004c)) Suponha que A e B sejam eventos tais que $P(A) = a$, $P(B) = b$ e $P(A \cap B) = c$. Escreva as seguintes probabilidades em termos de a, b e c:

a. $P(A \cup B)$;

b. $P(A^c)$;

c. $P(B^c)$;

d. $P(A^c \cup B^c)$;

e. $P(A^c \cup B)$;

f. $P(A^c \cap B^c)$;

g. $P(A \cap B^c)$;

Exercício 22. (Retirado de Viali (2004c)) Uma amostra de 140 clientes de um banco revelou que 80 guardam seu dinheiro na poupança, 30 investem no tesouro direto (TD) e 10 tem aplicações tanto na poupança como no TD. Qual a probabilidade de que um cliente, que tenha sido escolhido ao acaso dos 140 entrevistados, tenha dinheiro na poupança ou no TD?

Exercício 23. (Retirado de Meyer (1973)) Sejam Ω um espaço amostral e A, B e C eventos em Ω. Demonstre as seguintes propriedades:

[10]Togepi é um pokémon do tipo fada a partir da geração VI.
[11]Psyduck é um pokémon do tipo água.
[12]Considere que Togepi está em uma das 12 *pokébolas* e não no colo da Misty.
[13]Desconsidere o fato do Psyduck fugir da *pokébola* em momentos inapropriados.

a. Se $\emptyset$ for o espaço vazio, então $P(\emptyset) = 0$;

b. Se A^c for o evento complementar de A, então $P(A) = 1 - P(A^c)$;

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

d. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$;

e. Se $A \subset B$, então $P(A) \leq P(B)$;

f. $P(A \cup B) = 1 - P(A^c \cap B^c)$;

g. $P(A) = P(A \cap B) + P(A \cap B^c)$. **Obs.:** Este resultado mostra como podemos escrever a probabilidade do evento A de uma forma diferente. Ele é muito utilizado em várias demonstrações e a sua ideia básica é utilizada no teorema de Bayes.

Exercício 24. (Retirado de Meyer (1973)) Mostre que

$$P[(A \cap B^c) \cup (A^c \cap B)] = P(A) + P(B) - 2P(A \cap B)$$

Obs.: *Este resultado trata da probabilidade que exatamente um dos eventos A ou B ocorra, enquanto que a letra c. do exercício anterior pode ser vista como a probabilidade de que pelo menos um deles ocorra. Para fazer a demonstração, você tem a opção de desenvolver um dos lados da igualdade e chegar no outro ou provar que os dois conjuntos são iguais.*

Exercício 25. (Retirado de Magalhães (2011)) Os eventos A e B são independentes. Sendo $P(A \cap B^c) = 0,3$ e $P(A^c \cap B) = 0,2$; calcule $P(A \cap B)$.

Exercício 26. (Retirado de Magalhães (2011)) Sejam os resultados de 3 lançamentos de uma moeda honesta e considere os seguintes eventos: $\alpha = \{$ocorrem pelo menos duas coroas$\}$ e $\beta = \{$ocorre coroa no 1^olançamento$\}$. Determine se α e β são eventos independentes.

Exercício 27. (Adaptado de Schmidt (2011)) Considere Ω, espaço amostral, e A, B eventos em Ω.

a. Calcule $P(A \cap B)$ considerando $P(A) = 0.5$, $P(B) = 0.25$ e A e B mutuamente exclusivos;

b. Caso você saiba que $A \subset B$, é verdade que $P(A|B) \leq P(A)?Justifique.$

c. Considerando $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{4}$, calcule $P(A^c \cap B^c)$.

d. Se $P(A|B) = 0$, então A e B são independentes? Justifique.