

Probabilidade, Estatística e Processos Estocásticos

APOSTILA INE5118

Prof. André W. Zibetti

2026-03-23

Índice

Bem-vindo	3
Sobre este Material	3
Como usar	3
I Módulo 1 - Probabilidade e variáveis aleatórias discretas	4
1 O Modelo	5
1.1 O Problema	5
1.2 Modelo Matemático	5
1.2.1 Experimento Aleatório (E)	6
1.2.2 Espaço Amostral (Ω)	6
1.2.3 Eventos	7
1.3 Algumas relações sobre a teoria dos conjuntos	9
1.4 Visualizando o Espaço Amostral: Diagramas de Venn e Árvores	10
1.4.1 O Diagrama de Venn	11
1.4.2 O Diagrama de Árvore	12
1.4.3 Verificação computacional das regiões do espaço amostral	14
2 Probabilidade	16
2.1 Introdução	16
2.2 Definições de Probabilidade	17
3 Regras para o cálculo de probabilidades	28
3.1 Regra da adição	28
3.2 Probabilidade Condicional	33
3.3 Regra da Multiplicação	34
3.3.1 Eventos Dependentes e Independentes	35
3.4 Regra da Probabilidade Total	36
3.5 Teorema de Bayes	37

Bem-vindo

Bem-vindo ao curso de **Probabilidade, Estatística e Processos Estocásticos**, esta apostila é um material complementar da disciplina e está em constante evolução.

Esta apostila foi desenvolvida para servir como material de apoio para estudantes de Engenharia, combinando teoria com aplicações computacionais em **Python** e **R**.

Sobre este Material

O objetivo é apresentar os conceitos de forma intuitiva, mas sem perder o rigor matemático necessário. A estrutura do curso é dividida em módulos:

1. **Módulo 1:** Fundamentos de Probabilidade e Variáveis Aleatórias Discretas
2. **Módulo 2:** Variáveis Aleatórias Contínuas e Distribuições Amostrais.
3. **Módulo 3:** Inferência Estatística (estimação e testes de hipóteses)
4. **Módulo 4:** Introdução aos Processos Estocásticos.

Como usar

Utilize a barra lateral para navegar entre os capítulos. Os blocos de código são executáveis e podem ser copiados para seus próprios scripts.

Nota: Este material está em constante atualização.

Parte I

Módulo 1 - Probabilidade e variáveis aleatórias discretas

1 O Modelo

Neste capítulo, abordaremos o modelo matemático fundamental para fenômenos estocásticos, essenciais para a Engenharia. O conteúdo cobre desde experimentos aleatórios até a teoria condicional.

1.1 O Problema

Em sistemas críticos de engenharia (como a rede de distribuição de energia elétrica ou o sistema de frenagem ABS de um veículo), a falha de um componente não é **determinística**, mas **probabilística**. Se soubermos a probabilidade de um sensor falhar, como calculamos a probabilidade de falha do sistema inteiro?

Precisamos de uma linguagem matemática baseada na **Teoria de Conjuntos** para modelar esse problema.

1.2 Modelo Matemático

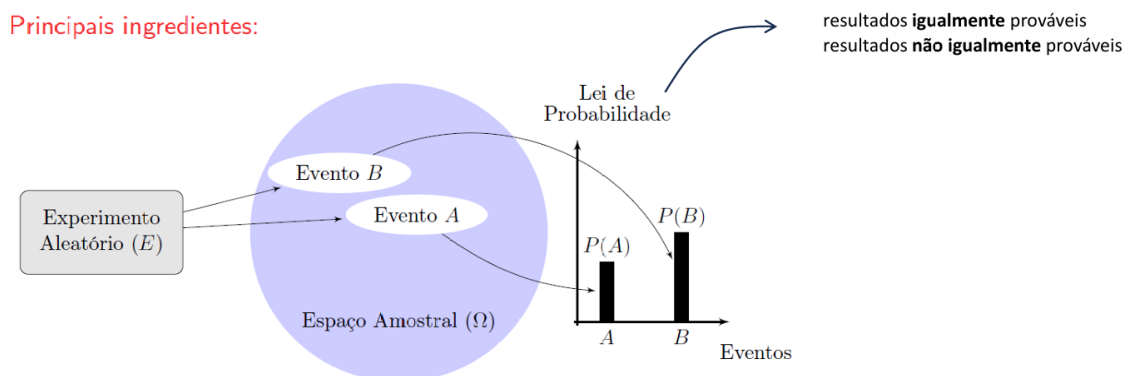


Figura 1.1: Modelo Matemático

O modelo matemático da probabilidade é baseado em três conceitos fundamentais: **experimento aleatório**, **espaço amostral** e **eventos**, esses são os ingredientes necessários para a construção do nosso modelo teórico de probabilidade, afim de obtermos as probabilidades de ocorrência de eventos. A figura Figura 1.1 ilustra esses conceitos, que serão detalhados a

seguir. Para obtermos a noção de probabilidade, precisamos primeiro definir o que é um experimento aleatório, seus resultados e os eventos de interesse.

1.2.1 Experimento Aleatório (E)

Um **experimento** é qualquer ação ou processo cujo resultado está sujeito à incerteza.

Dessa forma, os experimentos aleatórios incluem jogar uma moeda uma ou diversas vezes, selecionar uma ou várias cartas de um baralho, mensurar a massa de um objeto, determinar o tempo de deslocamento da sala de aula ao RU, mensurar a altura de um grupo de indivíduos, medir as resistências de compressão de um lote de vigas de concreto, testar a voltagem de baterias de lanterna, medir a potencia de recarregamento de baterias, avaliar o consumo de carro em litros por cem quilômetros, avaliar o consumo de celular em kWh por mês, entre outros que você pode pensar.

Definição de Experimento Aleatório

Um **experimento aleatório** (E) é qualquer processo de observação cujos resultados não podem ser previstos com certeza absoluta, mesmo quando repetido sob as mesmas condições.

1.2.1.0.1 Exemplo

- E : Medir a vida útil (em horas) de uma lâmpada LED recém-fabricada.
- E : Examinarmos três fusíveis em seqüência e anotarmos o resultado de cada exame (D^c - Não Defeituoso ou D - Defeituoso).
- E : Dois postos de gasolina estão localizados em uma determinada interseção. Cada um possui seis bombas. Considere o experimento em que o número de bombas em uso em determinada hora do dia é determinado para cada posto.

1.2.2 Espaço Amostral (Ω)

O conjunto de todos os resultados possíveis de um experimento aleatório.

Definição de Espaço Amostral

O **espaço amostral** (Ω) é o conjunto de todos os resultados possíveis de um experimento aleatório.

1.2.2.0.1 Exemplo

- E : Medir a vida útil (em horas) de uma lâmpada LED recém-fabricada.
- Ω : O espaço amostral é $\Omega = \{t \in \mathbb{R}^+ \mid t \geq 0\}$, infinitos possíveis resultados.

1.2.2.0.2 Exemplo

- E : Examinarmos três fusíveis em seqüência e anotarmos o resultado de cada exame (D^c - Não Defeituoso ou D - Defeituoso).
- Ω : o espaço amostral é $= \{NNN, NND, NDN, NDD, DNN, DND, DDN, DDD\}$, finitos resultados, $n(\Omega) = 8$.

1.2.2.0.3 Exemplo

- E : Se uma bateria de lanterna nova, tipo D, tiver uma voltagem fora de certos limites, será classificada como falha (F); se a voltagem estiver dentro dos limites especificados, será classificada como sucesso (S). Suponha que um experimento consista em testar cada bateria quando sai de uma linha de montagem até que seja observado um sucesso.
- Ω : O espaço amostral é $= \{S, FS, FFS, FFFS, \dots\}$, que contém um número infinito de resultados possíveis. Apesar de não ser muito provável, um resultado possível é que as primeiras 10 (ou 100 ou 1000 ou ...) sejam F e a próxima seja S . Isto é, para qualquer inteiro n possível, teremos de examinar n baterias antes de obter o primeiro S .

1.2.2.0.4 Exemplo

- E : é registrado o sexo de cada criança recém-nascida, em um horário especificado, até que seja observado o nascimento de uma criança do sexo masculino.
- Ω : O espaço amostral é $= \{F, FM, FFM, FFFM, \dots\}$, a mesma forma abreviada de espaço amostral é apropriada para esse experimento.

O espaço amostral é denotado por Ω e seus resultados ω_i . Exemplo: $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$.

O conjunto pode ser:

- **Finito:** $\Omega = \{\omega_1, \omega_2, \omega_3\}$
- **Infinito:** $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$

Os resultados (ω_i) devem ser mutuamente exclusivos, ou seja, não pode haver resultados iguais, e coletivamente exaustivos, ou seja, todos os resultados devem ser considerados, com certa granularidade.

O espaço amostral Ω e seus resultados ω_i , são representados na Figura 1.2 com o auxílio de um Diagrama de Venn, sendo os pontos todos os possíveis resultados.

1.2.3 Eventos

No estudo de probabilidade, estaremos interessados não apenas nos resultados individuais de Ω , como também em qualquer grupo de resultados de Ω .

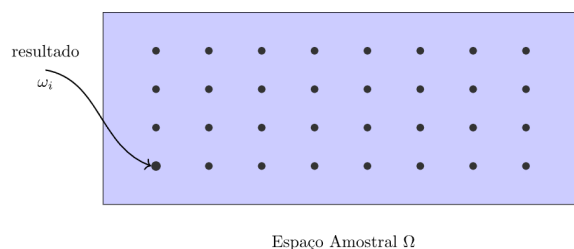


Figura 1.2: Espaço Amostral e seus resultados, representados em um Diagrama de Venn.

Definição de Evento

Um **evento** é qualquer grupo (subconjunto) de resultados contidos no espaço amostral.

O evento é denominado **simples** se consistir um único resultado e **composto** se consistir em mais de um resultado.

Os eventos são denotados por letras latinas maiúsculas (A, B, C, \dots).

Quando um experimento é realizado, determinado evento A ocorre se o resultado experimental estiver contido em A . Em geral, ocorrerá exatamente um evento simples, mas diversos eventos compostos também podem ocorrer simultaneamente. Na Figura 1.3, podemos observar um exemplo de espaço amostral e eventos dois eventos compostos A e B , representados em um Diagrama de Venn.

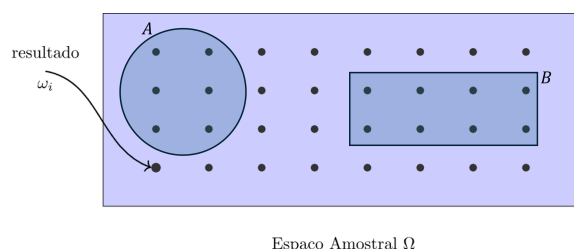


Figura 1.3: Espaço Amostral e Eventos, representados em um Diagrama de Venn.

1.2.3.0.1 Exemplos

Dos exemplos anteriores (três fusíveis e lâmpada de LED), podemos enumerar os seguintes eventos:

Eventos simples

- H : O evento $H = \{DND\}$
- F : O evento $F = \{DDD\}$

Eventos compostos

- A : O evento “A lâmpada dura mais de 10.000 horas” é denotado por $A = \{t > 10000\}$.

- B : O evento “A lâmpada dura menos de 10.000 horas” é denotado por $B = \{t < 10000\}$.
- C : O evento “Exatamente dois fusíveis estão defeituosos” é denotado por $C = \{DND, DDN, NDD\}$.
- E : O evento “Pelo menos um fusível está defeituoso” é denotado por $E = \{DND, DDN, NDD, DDD\}$.

1.3 Algumas relações sobre a teoria dos conjuntos

Um evento é essencialmente um conjunto (subconjunto de Ω), de forma que as relações e resultados da teoria elementar dos conjuntos podem ser usados para o estudo dos eventos. As operações a seguir serão usadas para construção de novos eventos, a partir de eventos já conhecidos.

União

A **união** de dois eventos A e B , representada por $A \cup B$ e lida “A união B”. É o evento que consiste em todos os resultados que estão no evento A ou no B ou em ambos (de forma que a união inclui resultados em que ocorram A e B , bem como aqueles em que exatamente um ocorre), isto é, todos os resultados em ao menos um dos eventos.

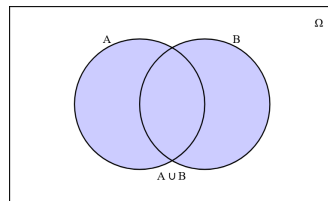


Figura 1.4: União de eventos

Interseção

A **interseção** dos dois eventos A e B , representada por $A \cap B$ e lida “A interseção B”, é o evento que consiste de todos os resultados que estão em ambos A e B .

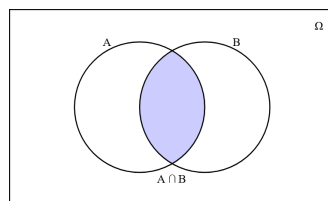


Figura 1.5: Interseção de eventos

Complemento A^c

O **complemento** de um evento A , representado por A^c ou A^c ou A^c , é o conjunto de todos os resultados em que não estão contidos em A .

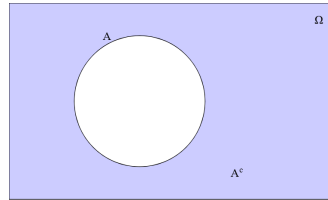


Figura 1.6: Complemento de um evento

1.3.0.0.1 Exemplo

Para o experimento em que é observado o número de bombas em uso em um posto de gasolina de seis bombas, assuma:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$

Os eventos sugeridos são:

- $A = \{1, 2, 3, 4\}$
- $B = \{3, 4, 5, 6\}$
- $C = \{1, 3, 5\}$

Então:

- $A \cup B = \{1, 2, 3, 4, 5, 6\} = \Omega$
- $A \cap B = \{3, 4\}$
- $A \cap C = \{1, 3\}$
- $A^c = \{5, 6\}$
- $(A \cap C)^c = \{2, 4, 5, 6\}$

1.4 Visualizando o Espaço Amostral: Diagramas de Venn e Árvore

Considere um sistema de fornecimento de energia ininterrupta (UPS) em um data center. Este sistema é composto por duas matrizes principais para evitar quedas: painéis solares na rede principal (FS) e geradores a diesel auxiliares (FD). A engenharia de confiabilidade exige calcular a probabilidade de falhas conjuntas. Para isso, o primeiro passo nunca é calcular números, mas organizar as possibilidades de erro de maneira que não deixemos nenhum cenário de fora.

Ao lidar com o espaço amostral Ω , a maneira mais segura de organizar a lógica é tratá-lo visualmente. Pense no espaço amostral como uma caixa contendo absolutamente todos os resultados possíveis do experimento. Desenhando formas dentro dessa caixa, conseguimos mapear onde cada cenário “mora”. Para isso vamos iniciar com o Diagrama de Venn. Começaremos explicitando qual é o espaço amostral e quais são os eventos de interesse.

O espaço amostral desse experimento é $\Omega = \{FSFD, FSFD^c, FS^cFD, FS^cFD^c\}$. Onde FS representa a falha no sistema solar e FD representa a falha no gerador a diesel e o resultado $FSFD$, por exemplo, representa a falha em ambos os sistemas, que também pode ser representado por $FS \cap FD$.

1.4.1 O Diagrama de Venn

Uma representação gráfica de eventos e manipulações de eventos é obtida pelo uso de diagramas de Venn. Para construir um desses diagramas, desenhe um retângulo cujo interior representará o espaço amostral Ω . Então, qualquer evento A é representado como o interior de uma curva fechada (normalmente um círculo, ou retângulo) contido em Ω . A Figura 1.7 mostra um exemplo de diagrama de Venn.

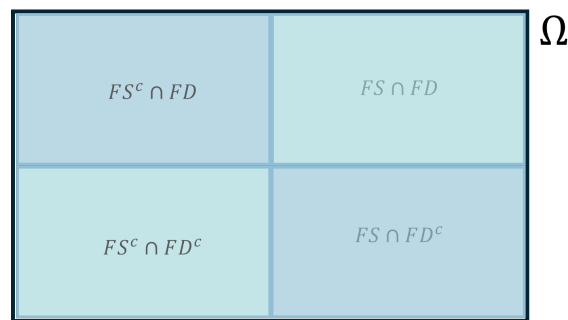


Figura 1.7: Espaço Amostral e Eventos, no contexto de falhas em um sistema UPS

Para analisar as falhas do sistema UPS, temos os 4 possíveis resultados:

- $FSFD$: Falha no sistema solar e falha no gerador a diesel, **ambos falham simultaneamente**.
- $FSFD^c$: Falha no sistema solar e não falha no gerador a diesel.
- FS^cFD : Não falha no sistema solar e falha no gerador a diesel.
- FS^cFD^c : Não falha no sistema solar e não falha no gerador a diesel, **nenhum dos sistemas falha**.

O contorno dessas figuras geométricas tem um papel na teoria dos conjuntos: ele recorta o plano inteiro em subdivisões (frequentemente chamadas de partições ou partições disjuntas). Qualquer dia de operação cairá impreterivelmente em apenas um desses quatro resultados “espaços”.

O espaço amostral desse processo/experimento é $\Omega = \{FSFD, FSFD^c, FS^cFD, FS^cFD^c\}$. Note que $FSFD = FS \cap FD$, e assim por diante.

Podemos nomear eventos de interesse, como por exemplo $A = \text{“Falha no sistema solar”}$ e $B = \text{“Falha no gerador a diesel”}$.

Observe a Figura 1.8 abaixo. O retângulo que delimita a figura é a nossa “caixa” do espaço amostral (Ω) , contabilizando 1, ou 100% das observações. Dentro desse universo, o retângulo laranja, evento intitulado como A agrupa todos os dias em que o sistema solar falhará, e o retângulo azul, evento intitulado como B agrupa os dias em que o sistema solar não falhará.

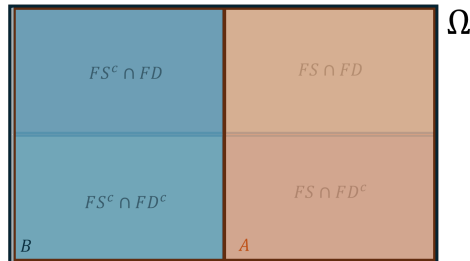


Figura 1.8: Espaço Amostral e Eventos, no contexto de falhas em um sistema UPS

Uma outra forma de representar o mesmo espaço amostral e os seus resultados e eventos é apresentado na Figura 1.9. Note que o resultado é o mesmo, apenas a forma de representar é diferente. Em alguns casos ela pode ser mais intuitiva de entender. Para uma representação mais fidedigna ao espaço amostral, o diagrama de Venn deve ser desenhado de forma que os eventos sejam representados por áreas proporcionais à sua probabilidade.

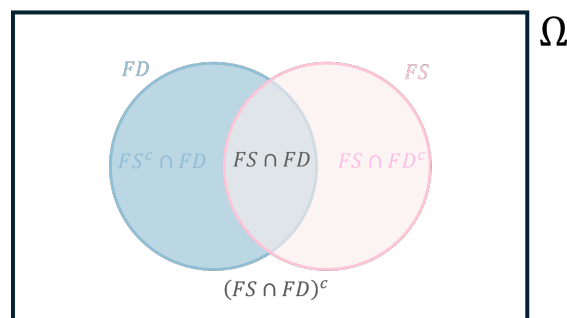


Figura 1.9: O mesmo espaço Amostral e Eventos, no contexto de falhas em um sistema UPS, representado de forma diferente.

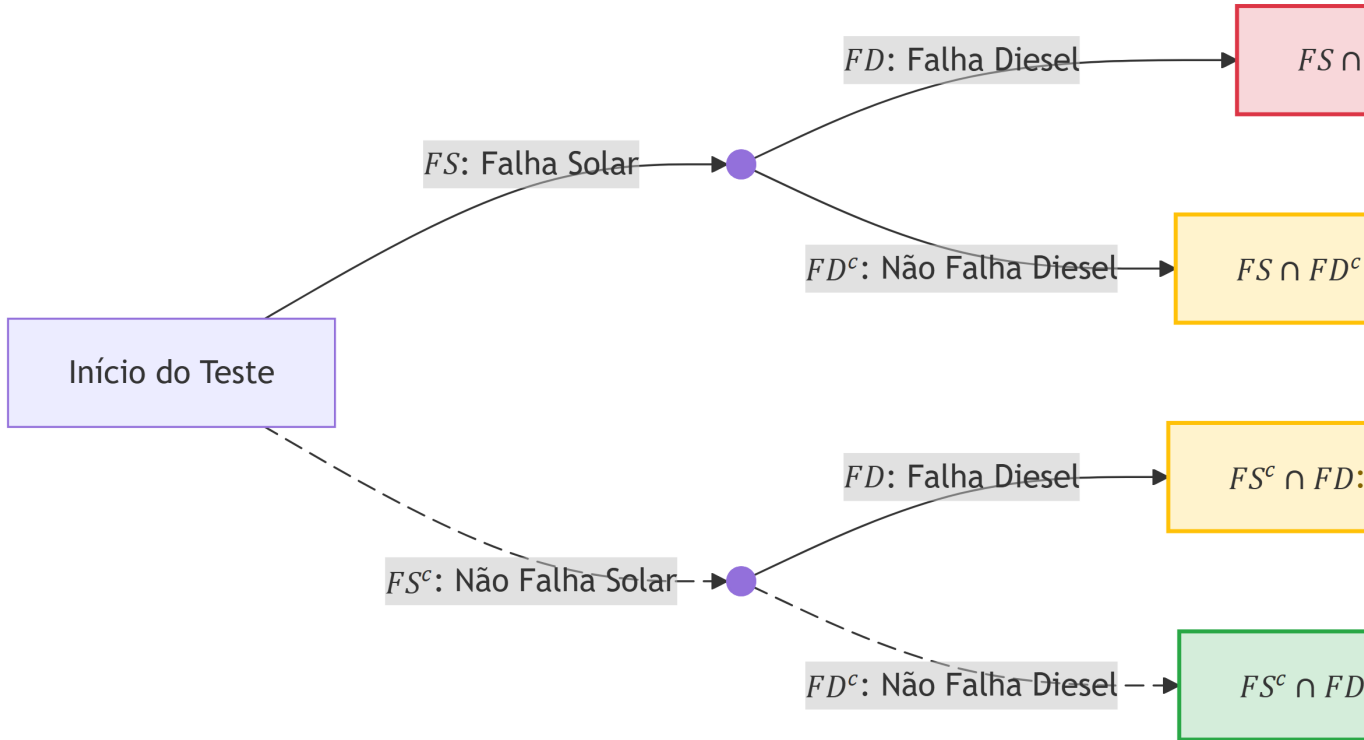
Outra forma de representar o mesmo espaço amostral e os seus eventos é utilizando um diagrama de árvore, como apresentado a seguir.

1.4.2 O Diagrama de Árvore

Em aplicações industriais práticas, o diagnóstico de equipamentos é usualmente serial (passo-a-passo). Primeiro inspecionamos a parte solar, e em seguida validamos o acionamento da retaguarda a diesel. Quando a análise pressupõe uma ordem lógica seqüencial, um diagrama

de árvore funciona muito bem para quebrar o espaço amostral sem perder ou sobrepor vertentes.

Ele simplesmente ramifica o possível resultado de cada teste, percorrido de forma sequencial:



No diagrama de árvore, temos os nós iniciais, intermediários e finais. Os nós iniciais são aqueles que não possuem nenhuma aresta apontando para eles. Os nós intermediários são aqueles que possuem arestas apontando para eles e também arestas apontando para fora deles. Os nós finais são aqueles que possuem arestas apontando para eles, mas nenhuma aresta apontando para fora deles.

Os nós iniciais irão representar o evento FS e FS^c . Os nós intermediários irão representar o evento FD e FD^c . Os nós finais irão representar o conjunto de todos os possíveis resultados do experimento, ou seja, o espaço amostral Ω .

A mesma árvore poderia ser desenhada, iniciando-se com os eventos FD e FD^c , se fizesse sentido. Nesse caso, os nós iniciais iriam representar o evento FD e FD^c . Os nós intermediários iriam representar o evento FS e FS^c . Os nós finais irão representar o mesmo espaço amostral Ω .

Como se pode perceber, os nós finais das árvores capturam as mesmas quatro partições desenhadas pelo diagrama de Venn, que são todos os resultados possíveis do experimento, o espaço amostral Ω . Se percorrermos o fluxo até a marcação superior em verde no final, teremos passado pelo evento complementar da falha no sistema solar FS^c , e depois do evento complementar de falhas do gerador diesel FD^c .

1.4.3 Verificação computacional das regiões do espaço amostral

A matemática garante que as quatro partições geradas limitam de modo que as frequências relativas sempre somem exatamente iguais a 1.0.

Suponhamos um diagnóstico onde falhas nos painéis solares se registrem na faixa dos 15%, enquanto anomalias no combustível afetem a partida em 5%. Em um simples modelo de falhas isoladas onde FS e FD operem independentemente, simular as falhas ao longo de 100 mil dias mapeia empiricamente a partição espacial.

i Verificação em R

```
# Simulação Monte Carlo garantindo a formação do Espaço Amostral
set.seed(42)
N <- 10^5 # Simulação empírica ao longo de 100.000 dias de registros funcionais

# Amostras isoladas geradas analiticamente de FS e FD
falha_S <- rbinom(N, size = 1, prob = 0.15) # gera valores 0 e 1 com prob de 0.15
falha_D <- rbinom(N, size = 1, prob = 0.05) # gera valores 0 e 1 com prob de 0.05

# Obtenção da frequência relativa das 4 partições conjuntas
taxa_apenas_S <- sum(falha_S == 1 & falha_D == 0) / N
taxa_apenas_D <- sum(falha_S == 0 & falha_D == 1) / N
taxa_ambos <- sum(falha_S == 1 & falha_D == 1) / N
taxa_nenhum <- sum(falha_S == 0 & falha_D == 0) / N

cat(sprintf("1. Apenas Solar falhou (FS \U2229 FD^c): %.4f\n", taxa_apenas_S))
cat(sprintf("2. Apenas Diesel falhou (FS^c \U2229 FD): %.4f\n", taxa_apenas_D))
cat(sprintf("3. Falha total: Ambos falharam simultaneamente (FS \U2229 FD): %.4f\n", taxa_ambos))
cat(sprintf("4. Operação Normal: Ambos funcionaram (FS^c \U2229 FD^c): %.4f\n\n", taxa_nenhum))

taxa_total <- taxa_apenas_S + taxa_apenas_D + taxa_ambos + taxa_nenhum
cat(sprintf("Somatório total da cobertura do cenário (\U03A9): %.2f\n", taxa_total))
```



```
1. Apenas Solar falhou (FS \U2229 FD^c): 0.1440
2. Apenas Diesel falhou (FS^c \U2229 FD): 0.0430
3. Falha total: Ambos falharam simultaneamente (FS \U2229 FD): 0.0074
4. Operação Normal: Ambos funcionaram (FS^c \U2229 FD^c): 0.8057

Somatório total da cobertura do cenário (\U03A9): 1.00
```

A execução demonstra por amostragem o princípio regente dos diagramas: A união de resultados mutuamente exclusivos subdivididos compõe o espaço amostral e as ocorrências totais devem agregar até 1.

Para iniciarmos o estudo adicionando probabilidades ao nosso modelo, precisamos entender as definições de probabilidade e suas relações com o espaço amostral e os eventos.

2 Probabilidade

2.1 Introdução

Neste capítulo, abordaremos as definições de probabilidade e suas relações com o espaço amostral e os eventos. Os objetivos deste capítulo são:

1. Conhecer as interpretações de probabilidade.
 2. Conhecer as regras para o cálculo de probabilidade e suas relações
- [Regra da adição](#)
 - [Probabilidade condicional](#)
 - [Regra da multiplicação](#)
 - [Eventos dependentes e independentes](#)
 - [Teorema de Bayes](#)

2.2 Definições de Probabilidade



Figura 2.1: Pierre-Simon Laplace

A **definição clássica** de probabilidade foi proposta por Pierre-Simon Laplace em 1812, no livro “*Théorie analytique des probabilités*”. Laplace definiu a probabilidade de um evento como a razão entre o número de casos favoráveis ao evento e o número total de casos possíveis, desde que todos os casos sejam igualmente prováveis. A definição clássica de probabilidade foi questionada por diversos autores do século XIX e XX, pois ela pressupõe que o espaço amostral é finito e que todos os resultados são equiprováveis, o que nem sempre é verdade. Além disso, a definição clássica de probabilidade não é aplicável a experimentos aleatórios com um número infinito de resultados possíveis.



Figura 2.2: Richard von Mises

A **definição frequentista** de probabilidade foi proposta Richard von Mises em 1928, no livro “*Wahrscheinlichkeit, Statistik und Wahrheit*”. Mises definiu a probabilidade de um evento como a razão entre o número de vezes que o evento ocorre e o número total de vezes que o experimento é repetido, desde que o número de repetições seja grande o suficiente. A definição frequentista de probabilidade foi questionada por diversos autores do século XX, pois ela pressupõe que o experimento pode ser repetido um número infinito de vezes, o que nem é possível na prática. A probabilidade frequentista define a probabilidade de um evento (a probabilidade a longo prazo) como o limite da sua frequência relativa em infinitas tentativas.



Figura 2.3: Andrey Kolmogorov

A **definição axiomática** de probabilidade foi proposta por Andrey Kolmogorov em 1933, no livro “*Grundbegriffe der Wahrscheinlichkeitsrechnung*”. Kolmogorov definiu a probabilidade de um evento como uma medida de probabilidade que satisfaz três axiomas: não-negatividade, normalidade e aditividade. Os axiomas da probabilidade não especificam nem pressupõem qualquer interpretação particular de probabilidade, mas podem ser motivados partindo de uma definição filosófica de probabilidade e argumentando que os axiomas são satisfeitos por essa definição

A probabilidade de uma evento pode ser interpretada e definida de diferentes formas, de acordo com os seus autores:

Definição Clássica (Laplace)

Quando o espaço amostral é **finito** e todos os seus elementos são **estruturalmente equiprováveis**, por simetria ou por razão lógica, *antes de qualquer observação*, a probabilidade de um evento A é:

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

Expressão por extenso:

Probabilidade teórica do evento A = $\frac{\text{número de vezes que o evento A ocorre no espaço amostral}}{\text{número de elementos do espaço amostral}}$

onde $|A|$ é o tamanho ou número de resultados favoráveis ao evento A e $||$ é o tamanho ou número de elementos no espaço amostral.

Esta é chamada de **probabilidade a priori**: é determinada pela estrutura do experimento, não por dados coletados.

Premissas:

- Número finito de possíveis resultados
- Hipótese de equiprobabilidade de resultados - **resultados igualmente prováveis**

Deficiências:

- Não faz sentido para espaço amostral infinito
- Não é capaz de definir a probabilidade de eventos supostamente não equiprováveis - **resultados não igualmente prováveis**

Definição Frequentista (Von Mises)

Baseada na observação empírica. Se um experimento é repetido n vezes e o evento A ocorre n_A vezes, a probabilidade é o limite:

$$P(A) = \lim_n \frac{n_A}{n}$$

Expressão por extenso:

Probabilidade empírica do evento A = $\lim_n \frac{\text{número de vezes que o evento A ocorreu}}{\text{número de repetições do experimento}}$

Premissas:

- Número “suficientemente” grande de repetições do experimento aleatório
- Condições uniformes para realização do experimento
- Ideal para **resultados não igualmente prováveis**

Deficiências:

- Definição de um número “suficientemente” grande
- Não é capaz de definir a probabilidade de eventos que não podem ser repetidos

A outra interpretação de probabilidade é o do conceito subjetivo, porém esse não será tratado neste material.

As interpretações de probabilidades não são suficientes para a formulação rigorosa da probabilidade como é o caso de outras disciplinas da matemática, dessa forma Kolmogorov apresentou um conjunto de **axiomas** para definir probabilidade.

Definição Axiomática (Kolmogorov)

$P(A)$

A função de probabilidade toma um argumento A e retorna um valor numérico.

Uma medida de probabilidade matemática estrita satisfaz:

1. **Axioma 1 - Não-negatividade:** Para qualquer evento A , $P(A) \geq 0$
2. **Axioma 2 - Normalidade:** $P(\Omega) = 1$
3. **Axioma 3 - Aditividade:** Se A_1, A_2, A_3, \dots é uma sequência infinita de eventos disjuntos (mutuamente exclusivos), então: $P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$
 - Ou seja, para dois eventos mutuamente exclusivos, teríamos $P(A \cup B) = P(A) + P(B)$.

A seguir temos as consequências dos axiomas de Kolmogorov.

i Consequências dos Axiomas (Clique para expandir)

- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$, onde A^c é o evento complementar.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Limites da probabilidade: $0 \leq P(A) \leq 1$
- Monotonicidade: Se $A \subseteq B$, então $P(A) \leq P(B)$

Probabilidade do conjunto vazio: $P(\emptyset) = 0$

Demonstração:

Sabemos que

=

Como \emptyset e Ω são disjuntos, pela aditividade:

$$P(\emptyset) + P(\Omega) = P(\Omega)$$

Subtraindo $P(\Omega)$ dos dois lados:

$$P(\emptyset) = 0$$

Regra do Complemento: $P(A^c) = 1 - P(A)$

Demonstração:

Observe que

$$= A \cup A^c$$

e que $A \cap A^c = \emptyset$.

Logo,

$$P(\Omega) = P(A) + P(A^c)$$

Como $P(\Omega) = 1$:

$$P(A^c) = 1 - P(A)$$

Limites da Probabilidade: $0 \leq P(A) \leq 1$

Demonstração:

Pelo axioma da não-negatividade, $P(A) \geq 0$.

Além disso, como

$$= A \cup A^c$$

temos

$$1 = P(A) + P(A^c)$$

Como $P(A^c) \geq 0$, segue que

$$P(A) \leq 1$$

Monotonicidade: Se $A \subset B$, então

$$P(A) \leq P(B)$$

Demonstração:

Se $A \subset B$, podemos escrever

$$B = A \cup (B \setminus A)$$

com união disjunta.

Logo,

$$P(B) = P(A) + P(B \setminus A)$$

Como $P(B \setminus A) \geq 0$, segue que

$$P(B) \geq P(A)$$

Fórmula da União de Dois Eventos: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Demonstração:

Podemos decompor B como

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

com união disjunta. Assim,

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

Agora escrevemos

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

Também é união disjunta, então

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

Substituindo

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

obtemos

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

Subaditividade: $P(A \cup B) \leq P(A) + P(B)$

Justificativa:

Da fórmula anterior,

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

Como $P(A \cap B) \geq 0$, concluímos que

$$P(A \cup B) \leq P(A) + P(B)$$

Comentário Final

Essas propriedades mostram que a probabilidade se comporta como uma medida de tamanho para conjuntos.

A partir apenas desses três axiomas, toda a teoria da probabilidade é construída: probabilidade condicional, independência, Teorema de Bayes, variáveis aleatórias e resultados assintóticos.

2.2.0.0.1 Exemplo

- E : Lançamento de uma moeda, uma única vez (sem especificar se é justa ou não).
- Ω : O espaço amostral é $\Omega = \{H, T\}$. H cara e T coroa.

Os axiomas especificam que $P(\Omega) = 1$, de forma que, para completar a atribuição de probabilidade, falta apenas determinar $P(H)$ e $P(T)$.

Já que H e T são os resultados do experimento, são também eventos simples e naturalmente mutuamente exclusivos, assim, $H \cap T = \emptyset$, o Axioma 3 implica que $1 = P(\Omega) = P(H) + P(T)$.

Essa expressão implica que $P(T) = 1 - P(H)$ ou $P(H) = 1 - P(T)$.

A única liberdade permitida pelos axiomas nesses experimentos é a escolha de um valor para a probabilidade atribuída a um dos eventos simples, por exemplo H .

Uma possível atribuição de probabilidades é $P(H) = 0.5$, $P(T) = 0.5$, enquanto outra atribuição possível é $P(H) = 0.75$, $P(T) = 0.25$, e assim por diante.

De fato, representar p por qualquer número fixo entre 0 e 1, assim:

- $P(H) = p$
- $P(T) = 1 - p$

é uma atribuição consistente com os axiomas.

2.2.0.0.2 Exemplo

- E : Considere um sistema de cinco componentes idênticos ligados em série, conforme ilustrado na Figura 2.4. Represente um componente que falha por F e o que não falha por S (de sucesso). Represente por A o evento em que o sistema falha
- Ω : O espaço amostral é $\Omega = \{SSSSS, SSSSF, SSSF, \dots\}$. S sucesso e F falha.

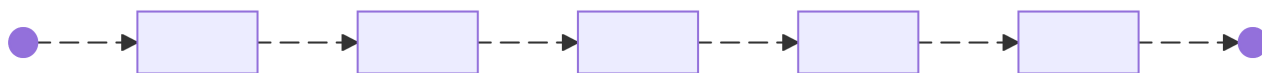


Figura 2.4

Para que A ocorra, ao menos um dos componentes individuais deve falhar. Os resultados de A incluem $SSFSS$ (1, 2, 4 e 5 funcionam, mas 3 não), $FFSSS$ e assim por diante. Na verdade há 31 diferentes resultados em A . Entretanto, A^c , o evento em que o sistema funciona, consiste em um único resultado $SSSSS$.

Na próxima seção veremos que se 90% de todos esses componentes não apresentarem falhas e se componentes diferentes apresentarem falhas independentemente um do outro, $P(A^c) = P(SSSSS) = 0.90^5 = 0.59$. Portanto, $P(A) = 1 - 0.59 = 0.41$. Dessa forma, em um grande número de tais sistemas, cerca de 41% apresentarão falhas.

Para o cálculo de probabilidades, devemos, com base no experimento aleatório, definir o espaço amostral e os eventos de interesse e a **lei de probabilidade**.

- **Lei de equiprobabilidade**: quando todos os resultados (ou eventos simples) tem a mesma probabilidade de ocorrer.
- **Lei de não equiprobabilidade**: quando os resultados (ou eventos simples) não tem a mesma probabilidade de ocorrer.

2.2.0.0.3 Exemplo

- E : Lançamento de uma moeda justa, uma única vez.
- Ω : O espaço amostral é $\Omega = \{H, T\}$. H cara e T coroa.
- Eventos simples: $A = \{H\}$ e $B = \{T\}$.
- Lei de **equiprobabilidade**: resultados igualmente prováveis.
- Definição clássica de probabilidade:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{2}$$

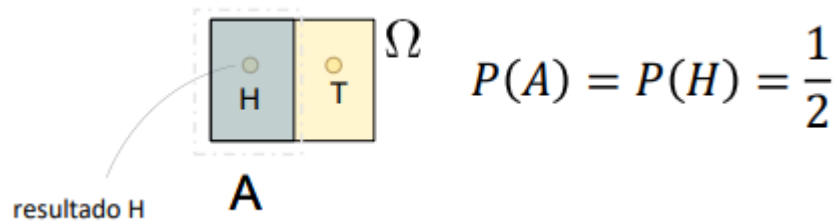
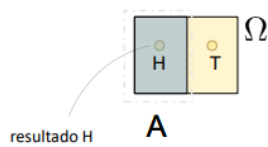


Figura 2.5: Exemplo de lançamento de moeda justa

2.2.0.0.4 Exemplo

- E : Lançamento de uma moeda não justa, uma única vez.
- Ω : O espaço amostral é $\Omega = \{H, T\}$. H cara e T coroa.
- Lei de **não equiprobabilidade**: resultados **não** igualmente prováveis. $P(H) = p, P(T) = 1 - p$
- Definição frequentista de probabilidade:

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$$



Repetições do experimento: (ex: 10 repetições)

H H T T T T T H T H ...

$P(A) = ?$

Figura 2.6: Exemplo de lançamento de moeda viesada

Para obtermos a probabilidade desse resultado devemos executar o experimento um grande número de vezes. Faremos então uma simulação para estimar essa probabilidade.

```

library(ggplot2)

set.seed(42)

# Imagine que a "verdadeira" frequencia de H é 0.60
p_H <- 0.60

# Simular N lançamentos individuais
N_max <- 2000
resultados <- rbinom(N_max, size = 1, prob = p_H)
freq_rel <- cumsum(resultados) / seq_along(resultados)
df_conv <- data.frame(n = 1:N_max, freq = freq_rel)

ggplot(df_conv, aes(x = n, y = freq)) +
  geom_line(color = "#2c7bb6", linewidth = 0.8) +
  geom_hline(yintercept = p_H, linetype = "dashed",
            color = "#d7191c", linewidth = 1) +
  annotate("text", x = 100, y = p_H + 0.05,
         label = sprintf("P(H) %.2f", p_H),
         color = "#d7191c", size = 4) +
  labs(
    x = "Número de lançamentos (N)",
    y = "Frequência relativa de caras",
    title = "Convergência Empírica: Lei dos Grandes Números",
    subtitle = "Lançamento de uma moeda viesada"
  ) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1),
                    limits = c(0, 1)) +
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(face = "bold"),
    panel.grid.minor = element_blank() )

```

Convergência Empírica: Lei dos Grandes Números

Lançamento de uma moeda viesada

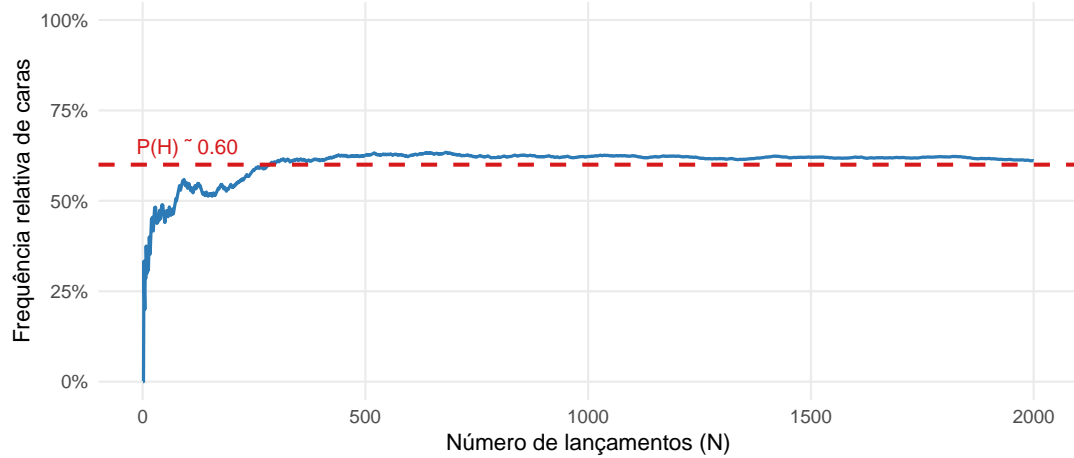


Figura 2.7: Convergência da frequência relativa de caras (linha vermelha pontilhada = estimativa estabilizada de 60%). Com poucas amostras a frequência oscila; com N grande ela converge.

Lei dos Grandes Números: informa que quanto mais observações são coletadas, a proporção de ocorrência de um resultado particular converge na probabilidade daquele resultado.

3 Regras para o cálculo de probabilidades

Agora veremos as regras para o cálculo de probabilidades. Elas auxiliam no cálculo de probabilidades de eventos mais complexos.

3.1 Regra da adição

A regra da adição é utilizada para calcular a probabilidade da união de dois ou mais eventos. No caso de dois eventos ela se refere a probabilidade de *A* **ou** *B* **ou** ambos ocorrerem.

Se os eventos forem mutuamente exclusivos, a regra da adição é simplificada para:

$$P(A \cup B) = P(A) + P(B)$$

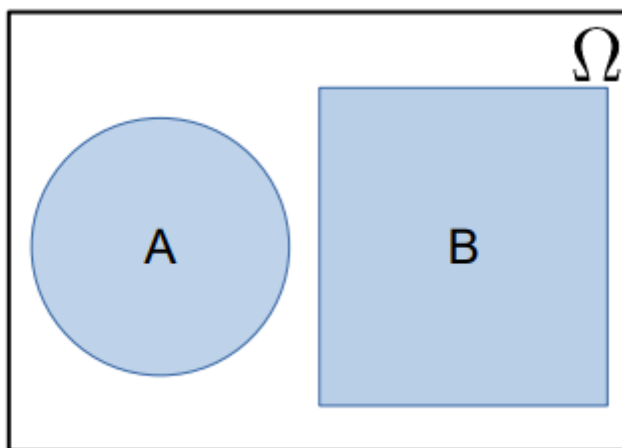


Figura 3.1: União de dois eventos mutuamente exclusivos

Se os eventos não forem mutuamente exclusivos, a regra da adição é:

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

Para três eventos:

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

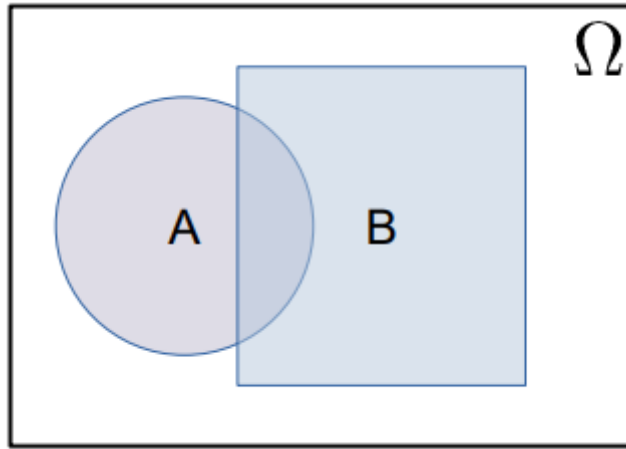


Figura 3.2: União de dois eventos não mutuamente exclusivos

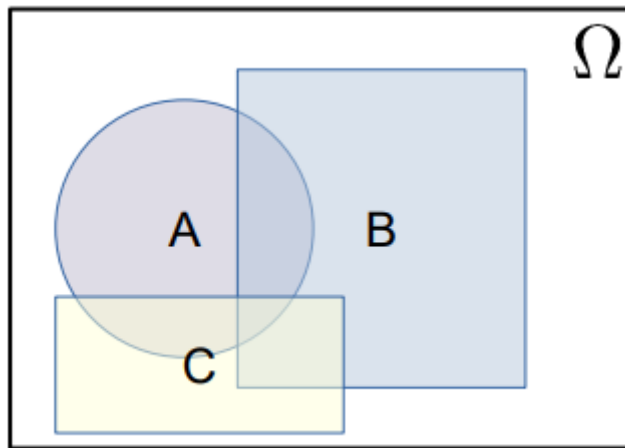


Figura 3.3: União de três eventos não mutuamente exclusivos

A união pode se dar sobre um conjunto finito ou infinito de eventos, segundo o Axioma 3 de Kolmogorov.

Regra da adição

De maneira simplificada, a regra da adição:

Para dois eventos:

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

Para três eventos:

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

Onde, para n eventos, o cálculo da probabilidade da união é bastante demandante.

Uma lei muito útil para o cálculo de probabilidades da união de múltiplos eventos é a lei de DeMorgan's, que diz o seguinte:

Lei de DeMorgan's

A lei explicita que é possível calcular a união de múltiplos eventos calculando a intersecção dos eventos complementares.

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

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

e assim por diante. Aplicada ao cálculo de probabilidade, temos:

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

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

$$P(A \cup B \cup C) = 1 - P(A^c \cap B^c \cap C^c)$$

3.1.0.0.1 Exemplo

- E : Em um determinado bairro residencial, 60% de todos os lares assinam o Netflix (N), 80% assinam o Disney+ (D) e 50% de todos os lares assinam os dois. Se um lar for selecionado aleatoriamente, qual será a probabilidade de ele assinar (1) ao menos um dos serviços de streaming e (2) exatamente um dos dois serviços

de streaming?

- : O espaço amostral é o conjunto de todos os lares do bairro, $\Omega = \{N^c D^c, ND^c, N^c D, ND\}$.

As probabilidades obtidas pelo enunciado são:

- $P(N) = 0.60$
- $P(D) = 0.80$
- $P(N \cap D) = 0.50$

Perguntas

1. Qual será a probabilidade de ele assinar ao menos um dos serviços de streaming, A ?
2. Qual será a probabilidade de ele assinar exatamente um dos dois serviços de streaming, B ?

Respostas

1. $P(A) = P(N \cup D) = P(N) + P(D) - P(N \cap D) = 0.60 + 0.80 - 0.50 = 0.90$
2. $P(B) = P(ND^c \cup N^c D)$, como os eventos ND^c e $N^c D$ são mutuamente exclusivos, temos:
 - $P(B) = P(ND^c) + P(N^c D) = P(N \cap D^c) + P(N^c \cap D)$
 - $P(N \cap D^c) = P(N) - P(N \cap D) = 0.60 - 0.50 = 0.10$
 - $P(N^c \cap D) = P(D) - P(N \cap D) = 0.80 - 0.50 = 0.30$
 - $P(B) = P(N \cap D^c) + P(N^c \cap D) = 0.10 + 0.30 = 0.40$

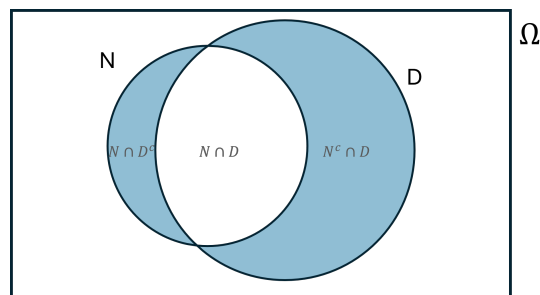


Figura 3.4: Diagrama de Venn para o exemplo de serviços de streaming

3.1.0.0.2 Exemplo

- E : Uma determinada fábrica opera em três turnos diferentes. No ano anterior, ocorreram 200 acidentes na fábrica. Alguns deles podem ser atribuídos em parte

a condições de trabalho inseguras, enquanto os outros não estão relacionados a condições de trabalho. A tabela a seguir fornece as porcentagens de acidentes que se encaixam em cada categoria de turno de trabalho.

- : O espaço amostral é o conjunto de todos os acidentes ocorridos na fábrica.

Tabela 3.1: Tabela de probabilidades por turno e tipo de condição

Turno	Condições Inseguras	Não relacionado às condições
<i>Dia</i>	10%	35%
<i>Alternado</i>	8%	20%
<i>Noite</i>	5%	22%

Suponha que um dos 200 relatórios de acidente seja selecionado aleatoriamente de um arquivo de relatórios e sejam determinados o tipo de acidente e o turno.

Perguntas

- Quais são os eventos simples?
- Qual é a probabilidade de que o acidente selecionado seja atribuído a condições inseguras?
- Qual é a probabilidade de que o acidente selecionado não tenha ocorrido no turno do dia?

Respostas

Sejam S_1, S_2 e S_3 os turnos: *dia*, *alternado* e *noite*, respectivamente.

Sejam C_1 e C_2 as *condições inseguras* e *não relacionadas às condições*, respectivamente.

- Os eventos simples, ou os 6 possíveis resultados, são $\{S_1C_1\}, \{S_1C_2\}, \{S_2C_1\}, \{S_2C_2\}, \{S_3C_1\}, \{S_3C_2\}$, pela notação considere S_iC_j como a intersecção dos eventos, $S_i C_j$, exemplo $S_1C_1 = S_1 \cap C_1$.
- $P(C_1) = P(S_1C_1 \cup S_2C_1 \cup S_3C_1) = 0.10 + 0.08 + 0.05 = 0.23$
- $P(S_1^c) = 1 - P(S_1C_1 \cup S_1C_2) = 1 - (0.10 + 0.35) = 0.55$

3.1.0.0.3 Exemplo

- E : Considere o tipo de secadora de roupas (a gás ou elétricas) comprada por cinco clientes diferentes em uma loja.

Perguntas

- Se a probabilidade de no máximo um desses clientes fazer uma compra de uma secadora elétrica for 0.428, qual será a probabilidade de ao menos dois clientes comprarem uma secadora elétrica?
- Se $P(\text{os cinco comprarem a gás}) = 0.116$ e $P(\text{os cinco comprarem elétricas})$

$= 0.005$, qual será a probabilidade de haver uma compra de ao menos uma de cada tipo?

Respostas

Assumindo que as compras dos clientes são eventos independentes, gás G e elétrica E .

- $= \{GGGGG, GGGGE, GGGE, \dots, EEEEE\}$

a. Seja A o evento de que no máximo um cliente faça uma compra de uma secadora elétrica. Então A^c é o evento de que pelo menos dois clientes comprem uma secadora elétrica.

- $P(A^c) = 1 - P(A) = 1 - 0.428 = 0.572$

b. Seja $B = \{GGGGG\}$ o evento de que os cinco comprem a gás. Seja $C = \{EEEEEE\}$ o evento de que os cinco comprem elétricas. Todos os outros resultados possíveis são aqueles em que pelo menos um de cada tipo é comprado. Assim, a probabilidade desejada é:

- $P(\text{uma compra de ao menos uma de cada tipo}) = 1 - P(B) - P(C) = 1 - 0.116 - 0.005 = 0.879$

3.2 Probabilidade Condicional

Para entender melhor essa regra do cálculo de probabilidade, imagine o cenário em que o evento B já ocorreu, e neste cenário você gostaria de calcular a probabilidade de A ocorrer. Para isso, você precisa “re-escalar” o espaço amostral para o evento B , ou seja, considerar que o evento B é o novo espaço amostral. Assim, a probabilidade de A ocorrer, **dado que** B ocorreu, é a probabilidade de $A \cap B$ ocorrer, re-escalada no universo de B .

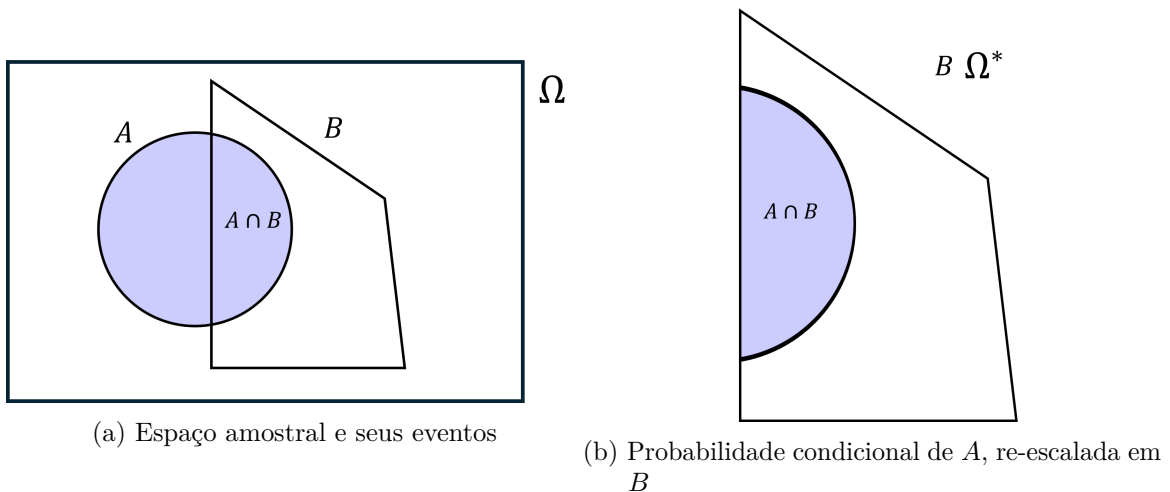


Figura 3.5: Probabilidade Condicional

Probabilidade Condicional

A probabilidade condicional é a probabilidade de um evento ocorrer, **dado que** um outro evento já ocorreu. A probabilidade condicional de um evento A dado um evento B é definida como:

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

Lê-se probabilidade condicional de A dado B .

3.2.0.0.1 Exemplo

- Suponha que, de todos os indivíduos que compram um determinado celular, 60% incluem um cartão de memória opcional na compra, 40% incluem uma capa protetora extra e 30% incluem um cartão e uma capa protetora. Considere a seleção aleatória de um comprador e sejam A {compra de cartão de memória} e B {compra de capa protetora}. Dessa forma, $P(A) = 0.60$, $P(B) = 0.40$ e $P(\text{compra de ambos}) = P(A \cap B) = 0.30$. Dado que o indivíduo selecionado comprou uma capa, a probabilidade de compra de um cartão opcional é

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.30}{0.40} = 0.75$$

Isto é, de todos os que compraram uma capa, 75% compraram um cartão de memória extra. De forma análoga,

$$P(\text{capa}|\text{cartão de memória}) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.30}{0.60} = 0.50$$

Observe que $P(A|B) \neq P(A)$ e $P(B|A) \neq P(B)$

3.3 Regra da Multiplicação

A regra da multiplicação é uma consequência direta da definição de probabilidade condicional. Ela é usada para calcular a probabilidade da intersecção de dois eventos, ou seja, a probabilidade de que ambos os eventos ocorram (simultaneamente). É a probabilidade conjunta de eventos, de um evento e outro evento ocorrer.

Regra da Multiplicação

A probabilidade da intersecção de dois eventos A e B é dada por:

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

Podemos ler como a probabilidade de A ocorrer e B ocorrer, dado que A já ocorreu.

Por simetria, podemos escrever a regra da multiplicação de uma outra forma, como mostrado abaixo.

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

A regra da multiplicação pode ser estendida para mais de dois eventos. Por exemplo, para três eventos A , B e C , a regra da multiplicação é dada por:

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

Por simetria, podemos escrever a regra da multiplicação de outras formas, como mostrado abaixo.

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

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

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

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

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

3.3.1 Eventos Dependentes e Independentes

Existem uma relação entre os eventos, essa relação pode ser classificada como a dependência ou independência entre os eventos.

Se houver uma **dependência entre os eventos**, então a probabilidade de um evento ocorrer é afetada pela probabilidade do outro evento ocorrer, é onde a regra da **probabilidade condicional** é utilizada.

Se houver **independência** entre os eventos, então a probabilidade de um evento ocorrer não é afetada pela probabilidade do outro evento ocorrer, ou seja, dado que um evento ocorreu, a probabilidade do outro evento ocorrer não se altera. Logo a probabilidade condicional é igual a probabilidade simples, ou seja, por exemplo, $P(A|B) = P(A)$ e $P(B|A) = P(B)$.

Eventos Independentes

Se dois eventos são **independentes**, implica que a probabilidade de um não se altera com a ocorrência do outro, logo a probabilidade da intersecção (ou conjunta) de dois

eventos, A e B , é dada por:

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

Se e somente se, A e B forem independentes.

Equivalentemente: $P(A|B) = P(A)$ e $P(B|A) = P(B)$.

Para três eventos independentes A , B e C , a probabilidade da intersecção (ou conjunta) de três eventos é dada por:

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

E assim por diante para n eventos independentes.

3.4 Regra da Probabilidade Total

Imagine que você gerencia a linha de montagem de um componente eletrônico crítico, alimentada por três fornecedores diferentes: A_1 , A_2 e A_3 . Cada fornecedor entrega uma fração diferente do inventário total e possui uma taxa de defeito inerente distinta.

Se um componente é retirado aleatoriamente da esteira final de produção, qual é a probabilidade global de que ele seja defeituoso (B)?

Como a produção mistura peças de várias origens, não podemos simplesmente usar a taxa de um único fornecedor. Precisamos particionar o problema em partes mais gerenciáveis e combiná-las formalmente.

A **Regra da Probabilidade Total** lida com isso particionando o espaço amostral em cenários exaustivos e mutuamente exclusivos. Se os eventos A_1, A_2, \dots, A_n formam uma partição do espaço amostral (isto é, $A_i \cap A_j = \emptyset$ para $i \neq j$), a probabilidade de um evento B pode ser expressa como a soma das probabilidades de B ocorrer em intersecção com cada A_i :

$$P(B) = \sum_{i=1}^n P(B \cap A_i)$$

Utilizando a regra da multiplicação vista anteriormente ($P(B \cap A_i) = P(A_i) P(B | A_i)$), reescrevemos a probabilidade total de B como:

$$P(B) = \sum_{i=1}^n P(A_i) P(B|A_i)$$

De forma prática, a probabilidade global de um efeito $P(B)$ é a média das probabilidades condicionais daquele efeito diante de cada cenário $P(B|A_i)$, ponderadas pela chance de cada cenário acontecer, $P(A_i)$.

Podemos ilustrar a regra particionando o espaço amostral de um diagrama de Venn.

```

\usetikzlibrary{shapes.geometric, arrows, backgrounds}
\begin{tikzpicture}
% Draw the sample space
\draw[thick] (0,0) rectangle (8,5) node[below left] {\Omega};
% Draw partitions
\draw[thick] (2.5, 0) -- (2.5, 5);
\draw[thick] (5.5, 0) -- (5.5, 5);

\node[font=\Large\bfseries] at (1.25, 4.5) {A_1};
\node[font=\Large\bfseries] at (4, 4.5) {A_2};
\node[font=\Large\bfseries] at (6.75, 4.5) {A_3};

% Draw event B
\draw[thick, fill=blue, fill opacity=0.3] (4,2.5) ellipse (3.5cm and 1.5cm);

% Nodes for intersections
\node at (1.5, 2.5) {B \cap A_1};
\node at (4, 2.5) {B \cap A_2};
\node at (6.5, 2.5) {B \cap A_3};
\end{tikzpicture}

```

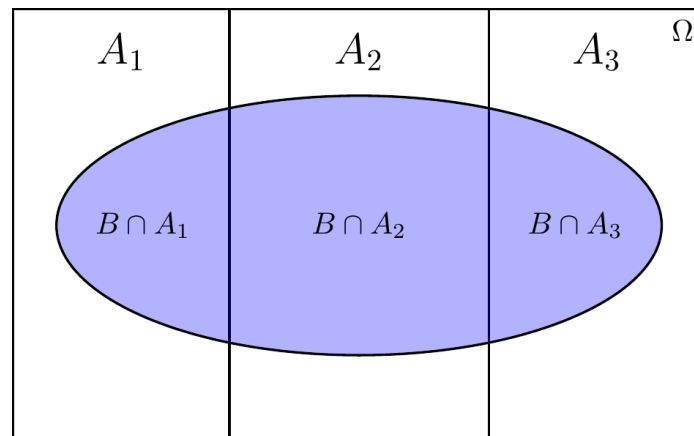


Figura 3.6: Diagrama ilustrando a Regra da Probabilidade Total. O evento B é a união de suas fatias ($B \cap A_i$) sobre o espaço particionado.

3.5 Teorema de Bayes

Enquanto a Probabilidade Total calcula a chance matemática do efeito B ocorrer, o **Teorema de Bayes** é utilizado reversamente como ferramenta de **diagnóstico de causa**.

Retornando ao exemplo logístico: constatou-se que um componente na saída está defeituoso. Qual a confiança de que ele tenha vindo do fornecedor A_1 ?

O que procuramos é $P(A_1|B)$, revertendo a condicional que costumamos ter catalogada como métrica de qualidade ($P(B|A_1)$).

O Teorema de Bayes expressa essa lógica de reversão analiticamente:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

O numerador representa a intersecção do fato com um cenário causal, ao passo que o denominador normaliza todas as vias por intermédio da Propabilidade Total do evento.

Tudo que envolva processos em estágios, como Causa Efeito, é também melhor organizado usando um diagrama de árvore de probabilidades.

```
\usetikzlibrary{trees,matrix,calc}
\begin{tikzpicture}[
  grow=right,
  level 1/.style={sibling distance=3cm, level distance=3.5cm},
  level 2/.style={sibling distance=1.5cm, level distance=3.5cm},
  every node/.style={fill=white, inner sep=2pt}
]
\node[inner sep=0pt] (root) {}
  child {node (A3) {$A_3$}
    child {node (A3Bc) {$B^c$} edge from parent node[below] {$P(B^c|A_3)$}}
    child {node (A3B) {$\boldsymbol{B}$} edge from parent node[above] {$P(B|A_3)$}}
    edge from parent node[below] {$P(A_3)$}
  }
  child {node (A2) {$A_2$}
    child {node (A2Bc) {$B^c$} edge from parent node[below] {$P(B^c|A_2)$}}
    child {node (A2B) {$\boldsymbol{B}$} edge from parent node[above] {$P(B|A_2)$}}
    edge from parent node[above, near start] {$P(A_2)$}
  }
  child {node (A1) {$A_1$}
    child {node (A1Bc) {$B^c$} edge from parent node[below] {$P(B^c|A_1)$}}
    child {node (A1B) {$\boldsymbol{B}$} edge from parent node[above] {$P(B|A_1)$}}
    edge from parent node[above] {$P(A_1)$}
  };

\matrix [matrix of math nodes,
  nodes in empty cells,
  row sep={1.5cm,between origins},
  column sep=5mm,
  anchor=mat-2-1.west
]
```

```

(mat) at ($(A1B.east)+(1,0)$)
{
  \omega      & P(\omega)      & B      \\
  \{A_1\cap \boldsymbol{B}\} & P(A_1)P(B|A_1) & \bullet \\
  \{A_1\cap B^c\} & P(A_1)P(B^c|A_1) & \\
  \{A_2\cap \boldsymbol{B}\} & P(A_2)P(B|A_2) & \bullet \\
  \{A_2\cap B^c\} & P(A_2)P(B^c|A_2) & \\
  \{A_3\cap \boldsymbol{B}\} & P(A_3)P(B|A_3) & \bullet \\
  \{A_3\cap B^c\} & P(A_3)P(B^c|A_3) & \\
  & & P(B) \\
};

% Horizontal lines for the table
\foreach \x/\y in {1/2, 2/3, 3/4, 4/5, 5/6, 6/7, 7/8}
  {\draw [-] ($(mat-\x-1.west -| mat-2-1.west)!0.5!(mat-\y-1.west)$) --
    ($(mat-\x-3.east -| mat-1-3.east)!0.5!(mat-\y-3.east -| mat-1-3.east)$);}

% Indicate summation by double rule
\draw [double, shorten >=-1mm, shorten <=-1mm]
  ($(mat-7-2.west)!0.5!(mat-8-2.west)$) --
  ($(mat-7-2.east)!0.5!(mat-8-2.east)$);

\end{tikzpicture}

```

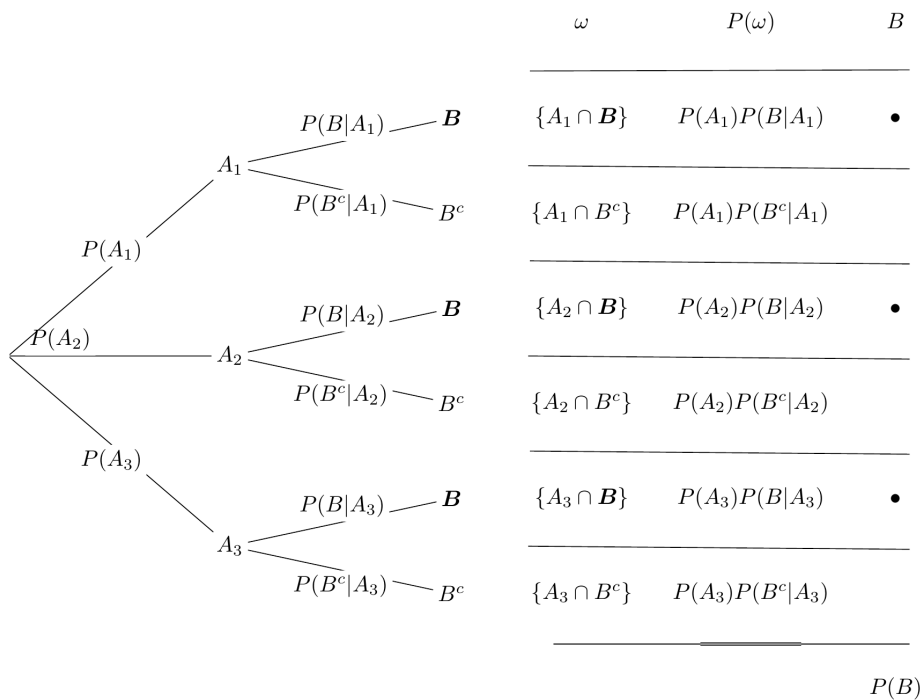


Figura 3.7: Árvore de probabilidades dos caminhos de causa de A em direção aos efeitos B e manifestação global.

Teorema de Bayes

O Teorema de Bayes expressa uma probabilidade condicional, e para ser obtida é necessário conhecer a probabilidade condicional do evento inverso e as probabilidades marginais dos eventos. Para calcular a probabilidade de B faz-se o uso da regra da probabilidade total.

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

Relembrando: Regra da probabilidade total

$$P(B) = \sum_{i=1}^n P(A_i) P(B|A_i)$$

3.5.0.0.1 Exemplo

Vamos estruturar um caso prático na área de qualidade.

Problema: Uma fábrica tem 3 máquinas (M_1, M_2, M_3) produzindo uma mesma peça mecânica nas proporções de carga de 30%, 20%, e 50%. As taxas de defeito históricas de cada máquina são rigidamente calculadas em 2%, 4%, e 1%.

1. Se a expedição escolhe aleatoriamente uma peça recém acabada, qual a sua predição teórica da peça possuir um defeito (D)?
2. Recebido o laudo de **Defeito** daquela peça, calcule o grau de suspeita de ter sido processada primariamente na Máquina 1.

Solução:

O texto define as origens produtivas como partições exaustivas:

- $P(M_1) = 0.30, P(M_2) = 0.20, P(M_3) = 0.50$

E as taxas de falha (condicionais):

- $P(D|M_1) = 0.02, P(D|M_2) = 0.04, P(D|M_3) = 0.01$

Aplicando (1) **Regra da Probabilidade Total**:

$$P(D) = P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)$$

$$P(D) = (0.02 \text{ E } 0.30) + (0.04 \text{ E } 0.20) + (0.01 \text{ E } 0.50) = 0.006 + 0.008 + 0.005 = 0.019$$

Encontramos uma probabilidade teórica de 1.9% da fábrica produzir uma peça com defeito.

Aplicando (2) **Teorema de Bayes** usando o total acima para descobrir $P(M_1|D)$:

$$P(M_1|D) = \frac{P(D|M_1)P(M_1)}{P(D)} = \frac{0.006}{0.019} \approx 0.31578$$

Temos a inferência que existe 31.58% de credibilidade empírica à hipótese que a Máquina 1 seja o agente causador.

Verificação em R - Lei dos Grandes Números Aplicada

```

# Simulação de confiabilidade fabril usando inferência estatística (Monte Carlo)
set.seed(42)
N <- 10^6

# 1. Simular o maquinário escolhido em linha com os pesos de produção fabril
maquinas <- sample(c("M1", "M2", "M3"), size = N, replace = TRUE, prob = c(0.3, 0.2, 0.5))

# 2. Inserir defeitos de acordo com a premissa individual e condicional computacional
defeitos <- numeric(N)
defeitos[maquinas == "M1"] <- rbinom(sum(maquinas == "M1"), 1, 0.02)
defeitos[maquinas == "M2"] <- rbinom(sum(maquinas == "M2"), 1, 0.04)
defeitos[maquinas == "M3"] <- rbinom(sum(maquinas == "M3"), 1, 0.01)

# Validação: Taxa de probabilidade Total - P(D)
prob_simulada_D <- sum(defeitos == 1) / N

# 3. Restringir observações APENAS ao subset classificado no defeito
amostra_defeituosa <- maquinas[defeitos == 1]

# Validação: Proporcionalidade reversa das causas ou Bayes - P(M1|D)
prob_simulada_M1_dado_D <- sum(amostra_defeituosa == "M1") / length(amostra_defeituosa)

cat(sprintf(" - Probabilidade Total Analítica(1.9%): %.4f | Simulada: %.4f\n",
  prob_simulada_D, prob_simulada_D))
cat(sprintf(" - Teorema de Bayes Analítico(31.58%): %.4f | Simulado: %.4f\n",
  prob_simulada_M1_dado_D, prob_simulada_M1_dado_D))

```

```

- Probabilidade Total Analítica(1.9%): 0.0190 | Simulada: 0.0190
- Teorema de Bayes Analítico(31.58%): 0.3158 | Simulado: 0.3184

```

O resultado computacional nos auxilia a entender a derivação analítica. Simular 10^6 rodadas de um equipamento auxilia na compreensão dessas fórmulas para avaliar as reais responsabilidades por quebras em qualquer malha de produção ou processamento contínuo.

Dica: Construa uma árvore de probabilidades para auxiliar na resolução de problemas de probabilidade condicional.