Estatística Aplicada a Ciências Ambientais

Fundamentos da Teoria de Probabilidades

Daniel Detzel
detzel@ufpr.br

Agenda

Fundamentos de probabilidade – parte I

- conceitos fundamentais
- medidas de probabilidade
- definições: eventos, espaço amostral e variáveis aleatórias

Fundamentos de probabilidade – parte II

- conjuntos
- probabilidade condicional
- independência
- medidas descritivas de variáveis aleatórias

FUNDAMENTOS DE PROBABILIDADE I

conceitos, medidas de probabilidade e definições

Fundamentos de probabilidades | conceitos

Pascal

Fermat

Imagens: https://www.amtonline.com.br/2019/06/qual-lado-cara-ou-coroa-moeda.html
https://www.cheenta.com/an-unexpected-correspondence-cheenta-probability-series-2/

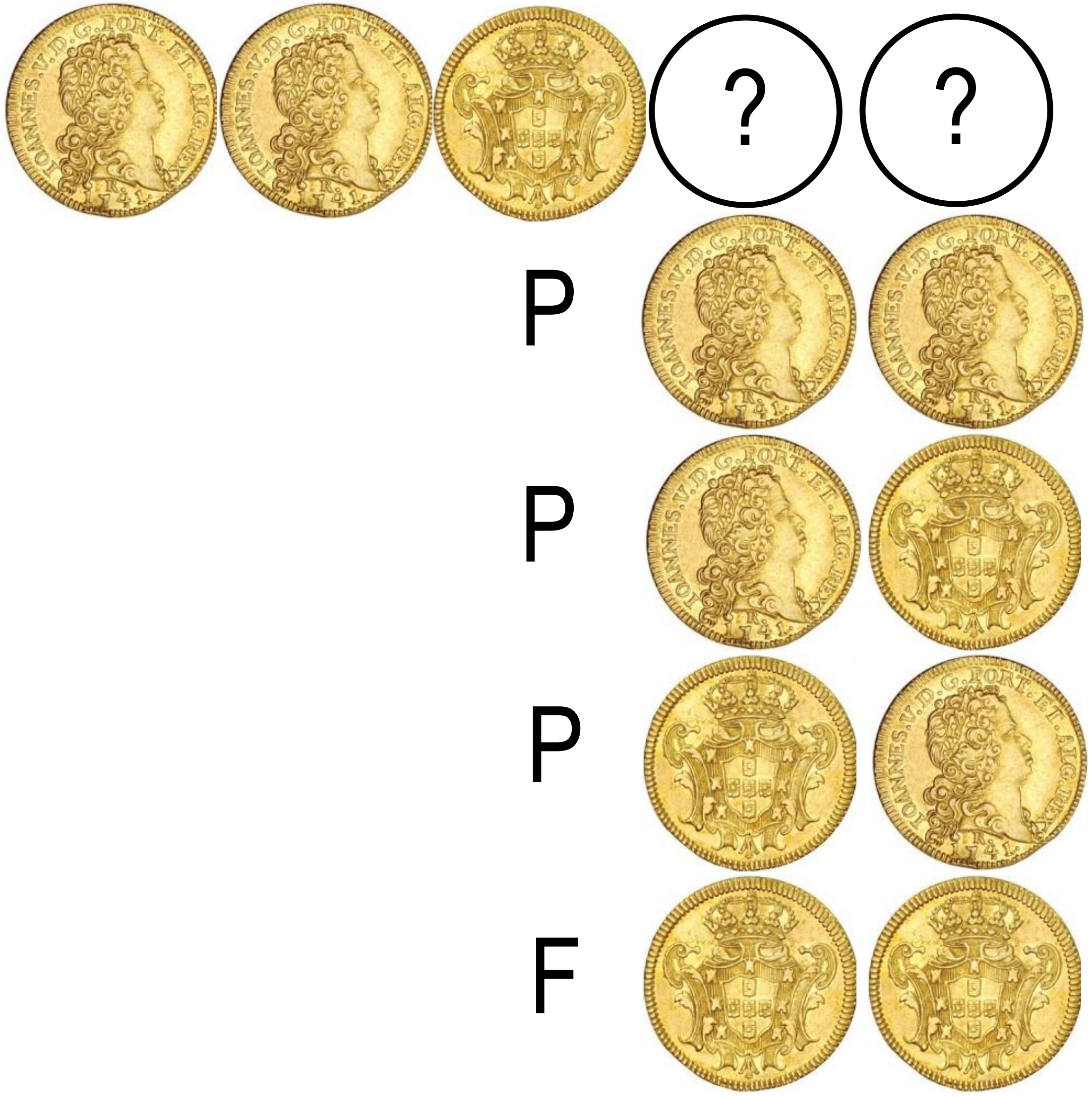
Fundamentos de probabilidades | conceitos

Pascal

Fermat

Imagens: https://www.amtonline.com.br/2019/06/qual-lado-cara-ou-coroa-moeda.html
https://www.cheenta.com/an-unexpected-correspondence-cheenta-probability-series-2/

Fundamentos de probabilidades | conceitos



Pascal

$^{3}/_{4}$

Fermat

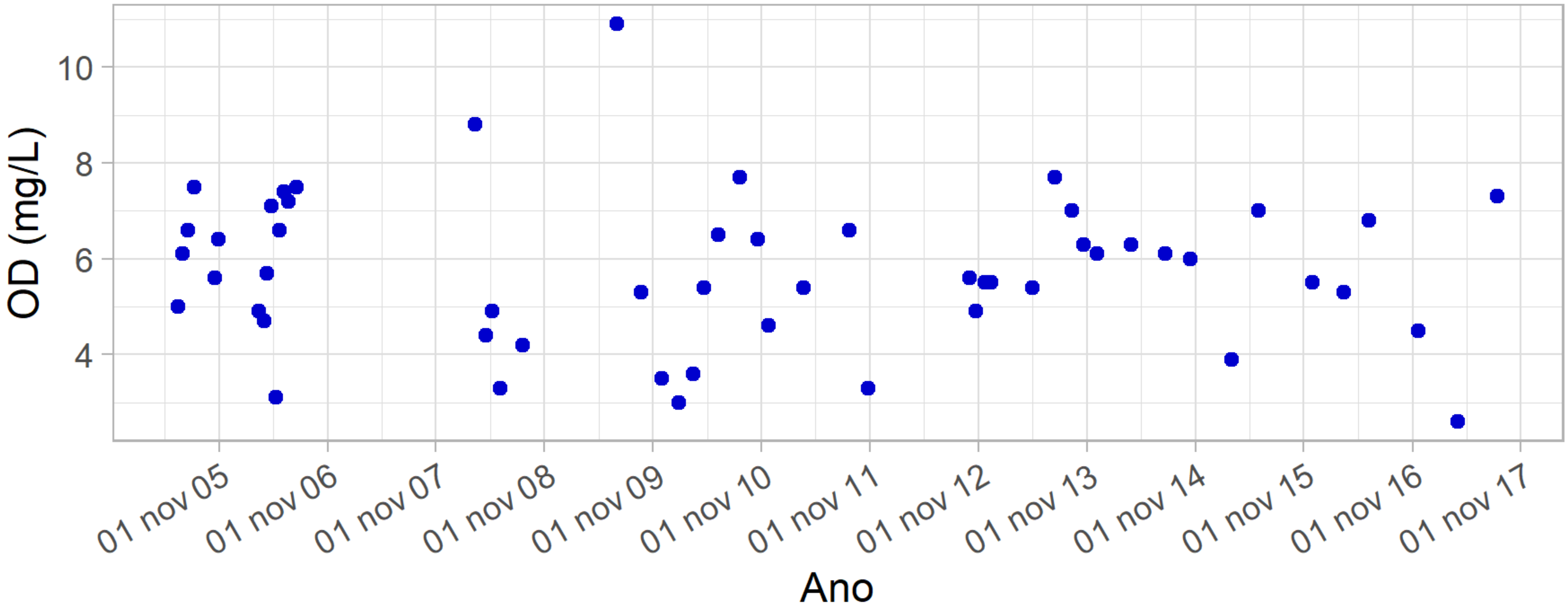
$^{1}/_{4}$

Imagens: https://www.amtonline.com.br/2019/06/qual-lado-cara-ou-coroa-moeda.html
https://www.cheenta.com/an-unexpected-correspondence-cheenta-probability-series-2/

Fundamentos de probabilidades | conceitos

Fermat sugeriu olhar o passado para “prever” o futuro

equivale a olhar o que aconteceu para estimar o que não aconteceu



Qual foi o valor de OD em 01 jul 07?

Fundamentos de probabilidades | medidas de probabilidade

Se um evento aleatório pode ocorrer de n maneiras igualmente possíveis e mutuamente excludentes, e se um evento A é uma dessas maneiras, então a probabilidade de ocorrência do evento A é:

$$\mathrm{prob}(A) = \frac{n_A}{n}$$

onde

n_A número de vezes que A ocorreu

Igualmente possível	$\equiv$	Igualmente provável

É uma definição a priori, por assumir conhecer n e n_A antes de terem sido realizados

Nas ciências naturais, a definição é mais útil quando feita em termos de frequências relativas e limites

afinal, não se conhece a população...

Fundamentos de probabilidades | medidas de probabilidade

Se um evento aleatório ocorre um grande número de vezes n e um evento tem um atributo A em n_A dessas ocorrências, então a probabilidade de ocorrência do evento A é:

$$\mathrm{prob}(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

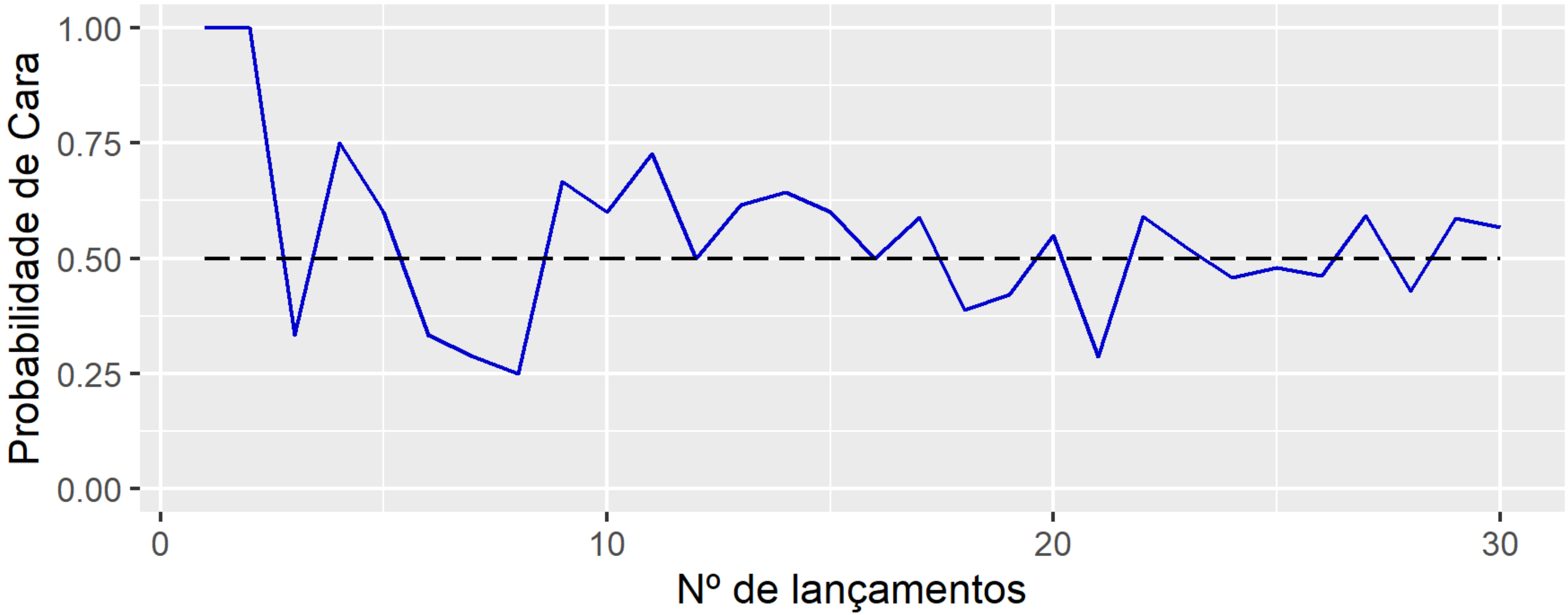
Estimativas de probabilidade baseadas em observações são empíricas

convergem para a probabilidade verdadeira (populacional) com o aumento de n

se 2 amostras estão disponíveis, as $\mathrm{prob}(A)$ não são necessariamente iguais entre elas (e nenhuma é igual ao valor populacional)

Fundamentos de probabilidades | medidas de probabilidade

Retornando às moedas... (aplicação em R)



Fundamentos de probabilidades | medidas de probabilidade

A escala de probabilidades varia de zero a um:

$p = 0 \rightarrow$ praticamente impossível

$p = 1 \rightarrow$ quase certo

A probabilidade de ocorrência de um valor exato de uma variável aleatória em uma distribuição de probabilidades contínua é zero

propriedade das distribuições contínuas de probabilidade

Porém, ao selecionar elementos da amostra, o valor pode, ou não, ser observado

por isso o uso dos termos “praticamente” e “quase”

Fundamentos de probabilidades | definições

Em estatística, elementos e conjuntos são usados na definição e manipulação de probabilidades

Um elemento (ou medição) é um ponto específico de um espaço amostral

Um espaço amostral é definido por todas as possíveis realizações de um experimento

Um experimento é qualquer processo que gera valores para uma variável aleatória

Uma variável aleatória é uma função que atribui valores numéricos às realizações de um experimento

Os conjuntos são eventos definidos por coleções de elementos

Fundamentos de probabilidades | definições

Exemplo:

A vazão do rio Iguaçu em Santa Clara é uma variável aleatória

As vazões do rio Iguaçu em Santa Clara entre 01 jan 1981 e 01 jan 1986 constituem um experimento

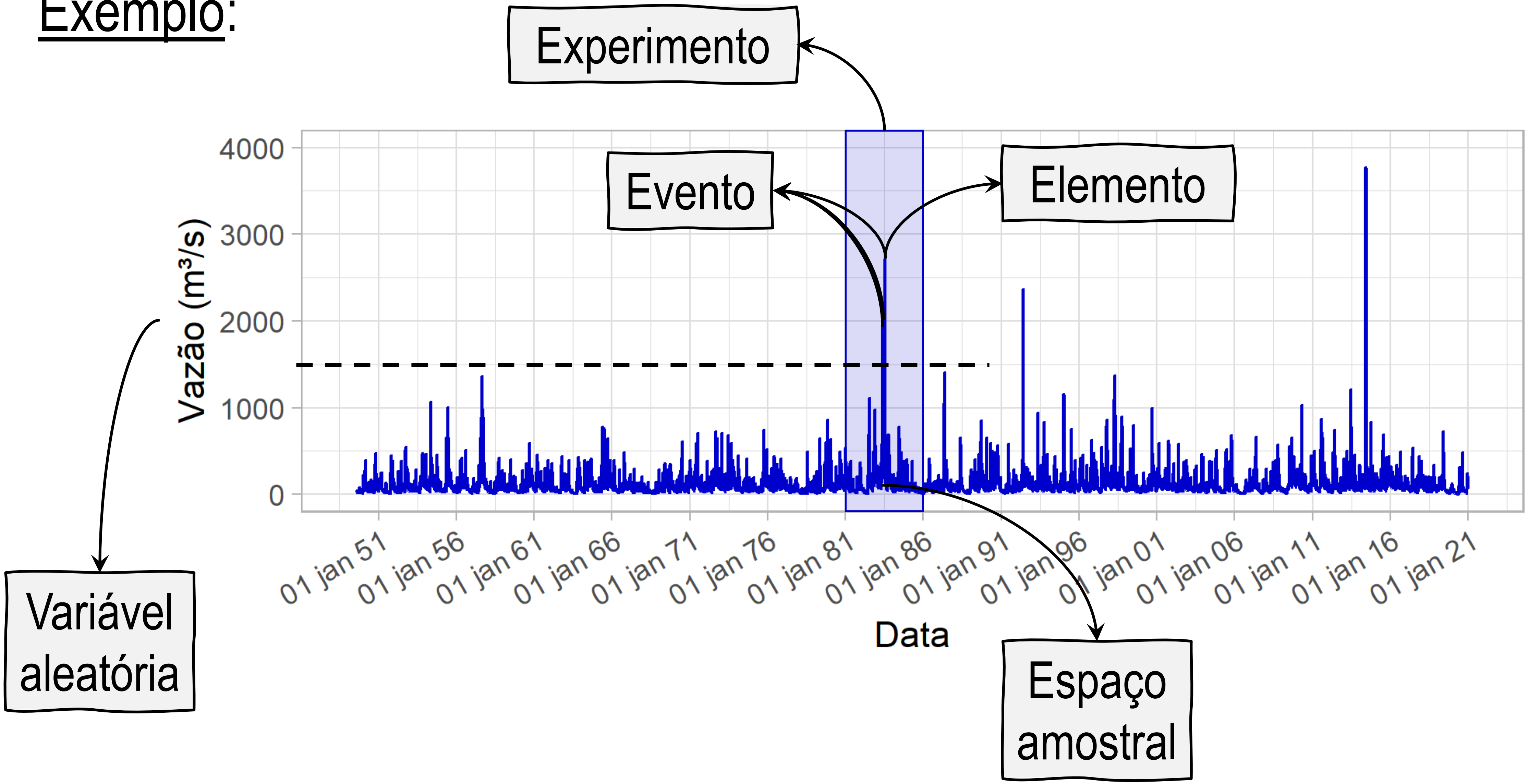
Os números positivos que representam as vazões do rio Iguaçu em Santa Clara entre 01 jan 1981 e 01 jan 1986 formam o espaço amostral

A vazão máxima do rio Iguaçu em Santa Clara entre 01 jan 1981 e 01 jan 1986 é um elemento

Todas as vazões que superam 1500 m^3/s representam um evento

Fundamentos de probabilidades | definições

Exemplo:



Fundamentos de Probabilidade II

conjuntos, probabilidade condicional e independência

Fundamentos de probabilidades | conjuntos

Considere as seguintes definições:

S	espaço amostral
E_i	elementos de S, com $i = 1,2, \ldots$
$\mathrm{prob}(E_i)$	probabilidade do elemento E_i
A e B	eventos que ocorrem em S
$\cup$	união ("E/OU")
$\cap$	intersecção ("E")

Assim:

$$0 \leq \mathrm{prob}(E_i) \leq 1$$

Fundamentos de probabilidades | conjuntos

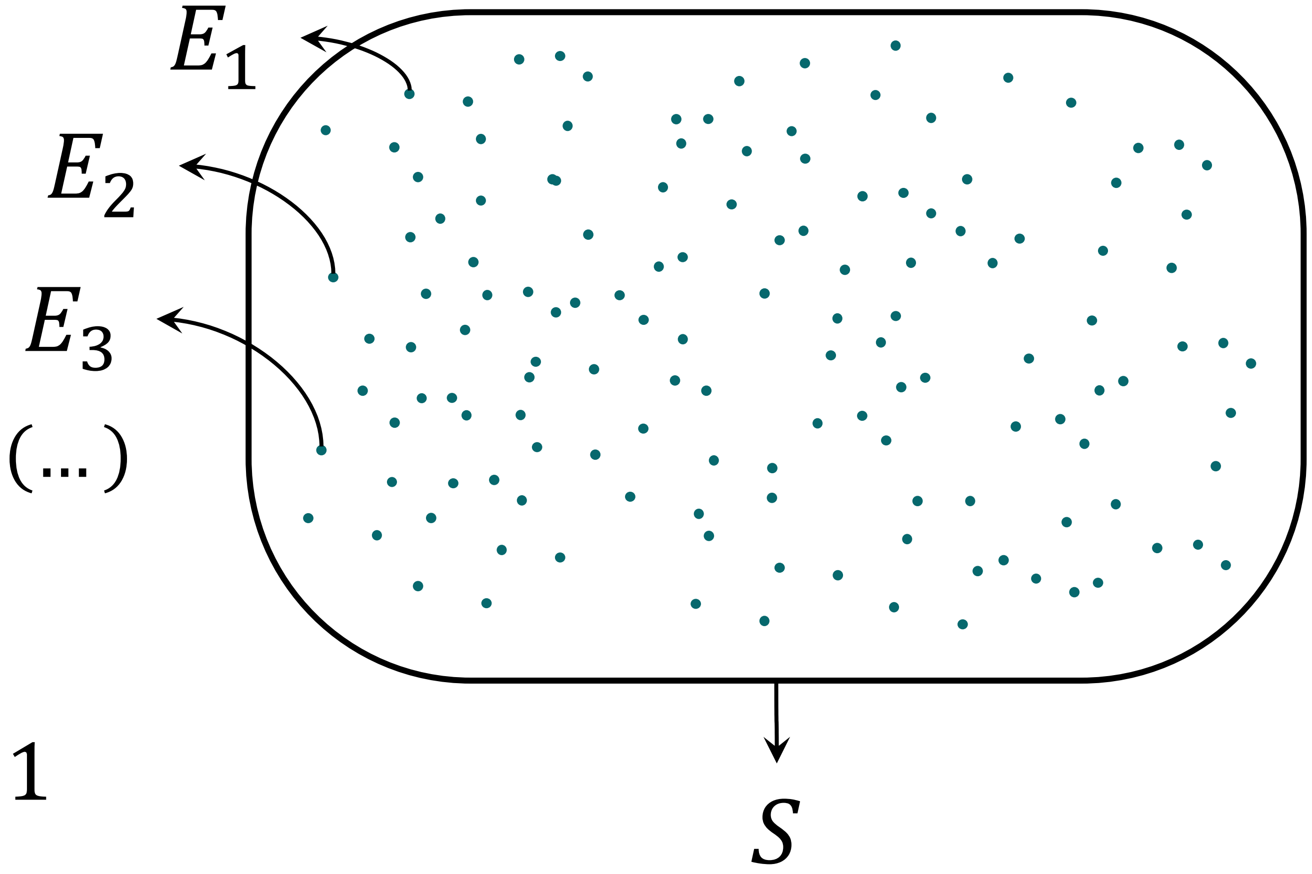
Como o espaço amostral é formado por todos os elementos nele contidos, tem-se:

$$S = \bigcup_{i}^{\infty} E_i$$

Bem como:

$$\mathrm{prob}(S) = \sum_{i}^{\infty} \mathrm{prob}(E_i) = 1$$

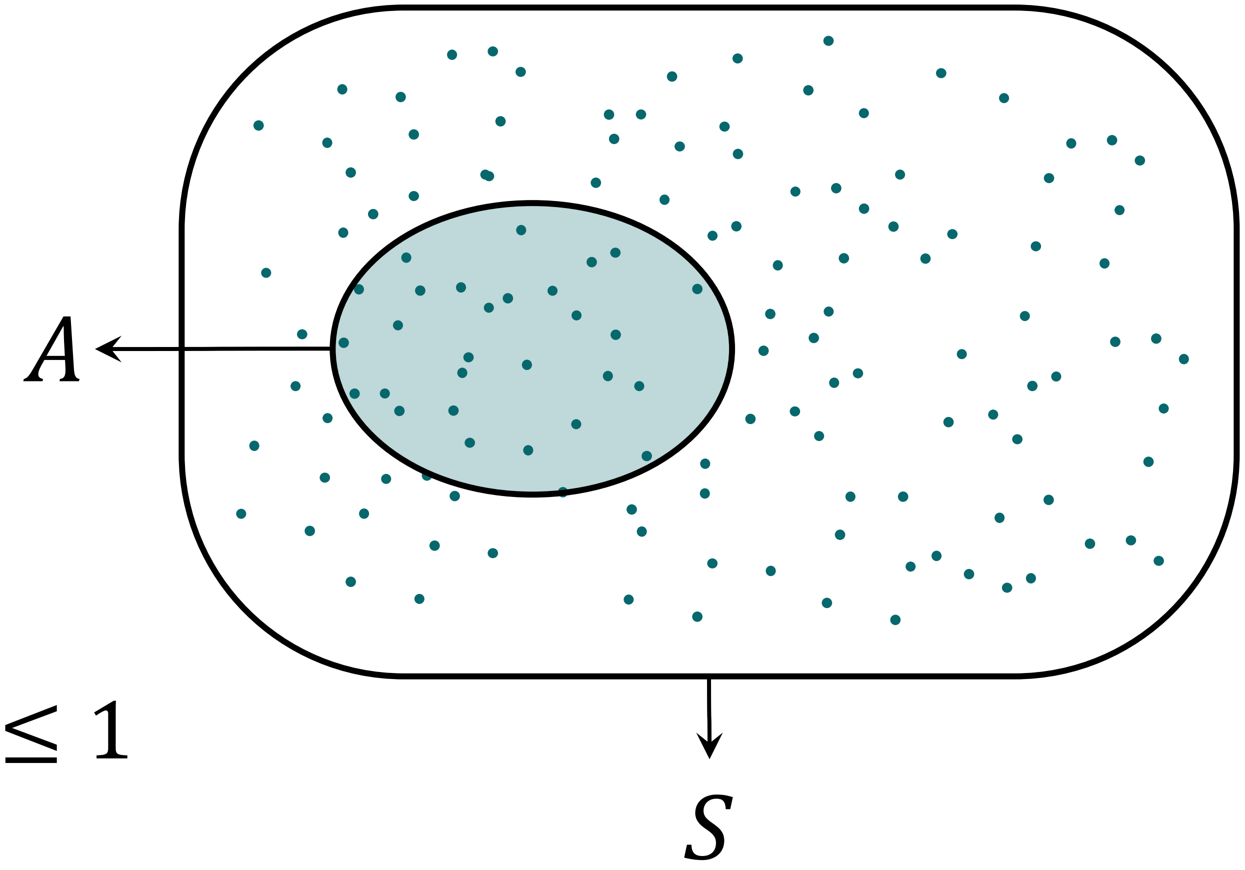
Diagrama de Venn



Fundamentos de probabilidades | conjuntos

Um evento A é formado por um subconjunto dos elementos de S:

$$A = \bigcup_{i=m}^{n} E_i$$



Assim:

$$0 \leq \mathrm{prob}(A) = \sum_{i=m}^{n} \mathrm{prob}(E_i) \leq 1$$

Fundamentos de probabilidades | conjuntos

Os elementos de S que não estão no evento A formam o subconjunto complementar A^c, tal que:

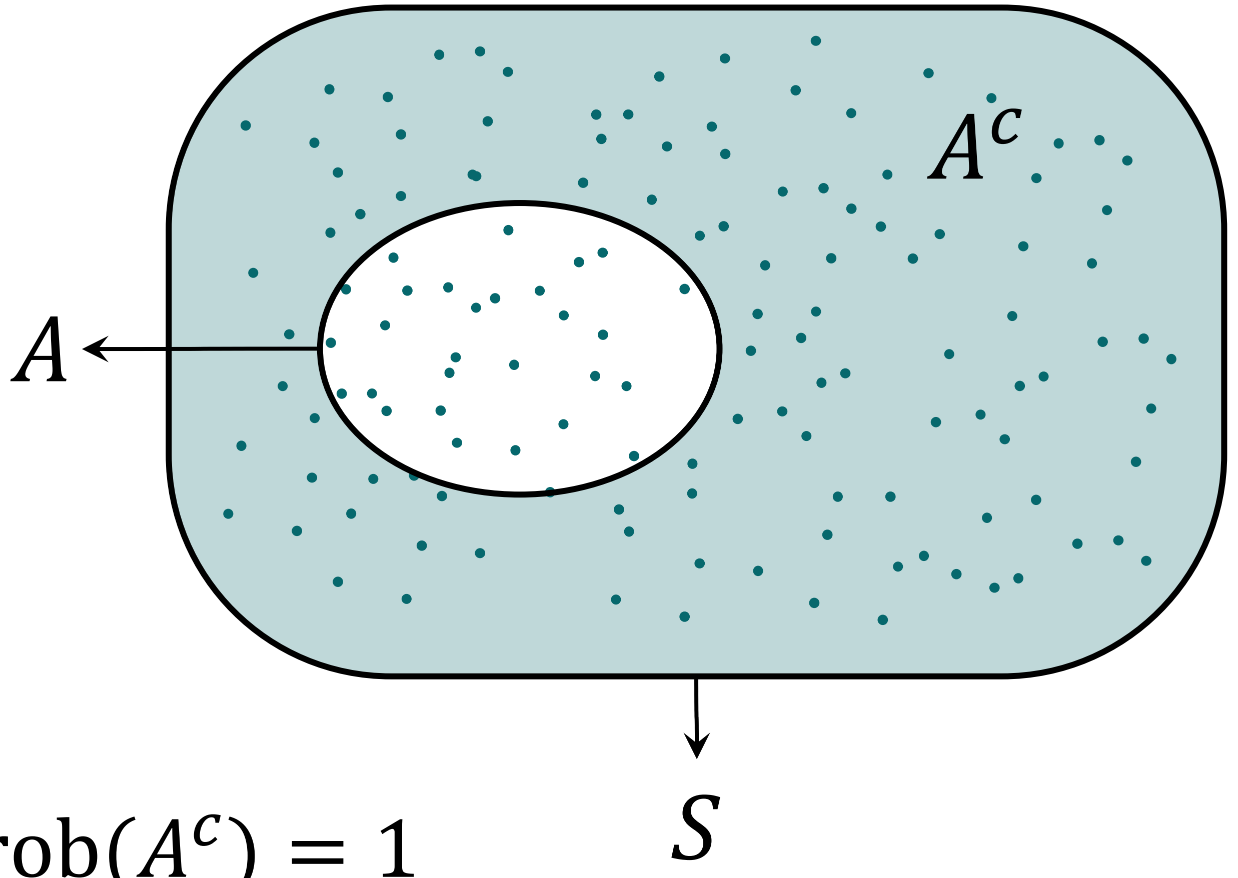
$$A \cup A^c = S$$

Portanto:

$$\text{prob}(A) = 1 - \text{prob}(A^c)$$

Ou:

$$\text{prob}(A \cup A^c) = \text{prob}(A) + \text{prob}(A^c) = 1$$



Fundamentos de probabilidades | conjuntos

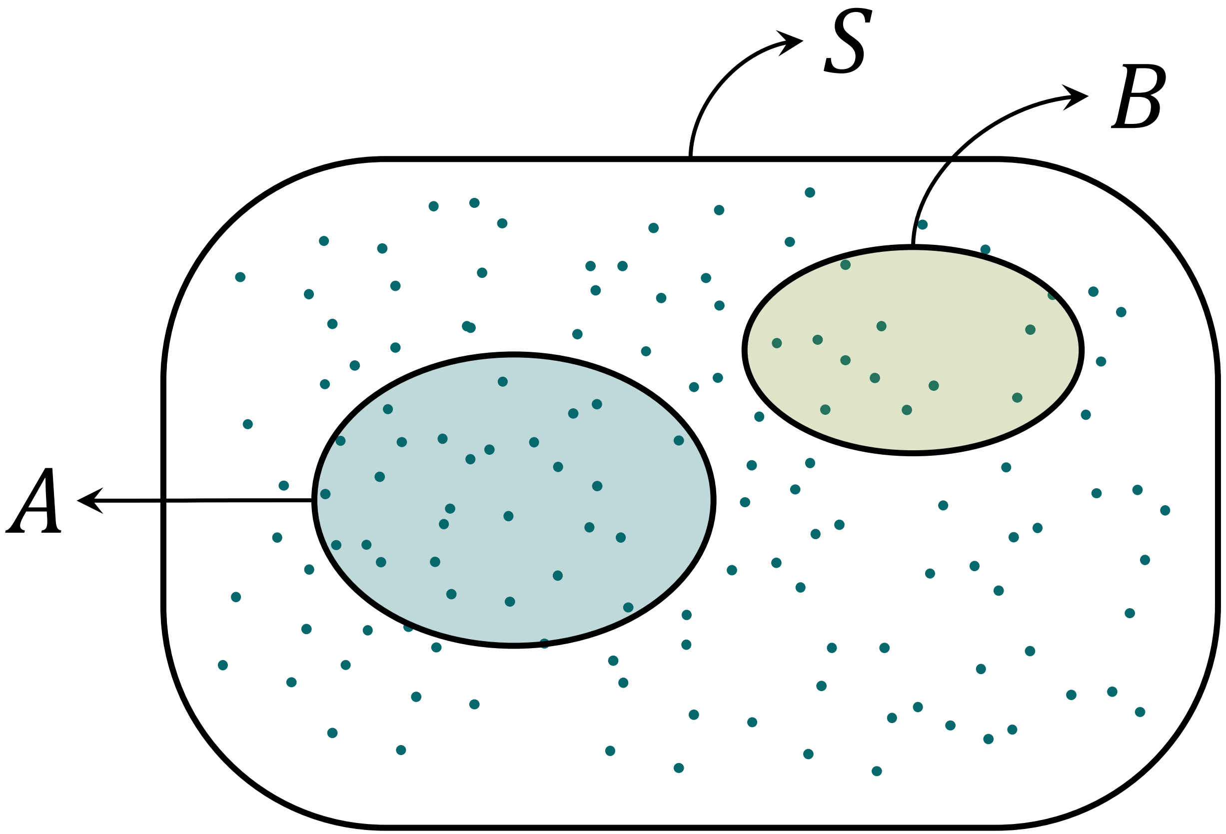
Agora, considerando dos eventos A e B mutuamente excludentes, tem-se:

$$\text{prob}(A \cup B) = \text{prob}(A) + \text{prob}(B)$$

Bem como:

$$A \cap B = 0$$

Ou seja, os eventos não podem ocorrer simultaneamente



Fundamentos de probabilidades | conjuntos

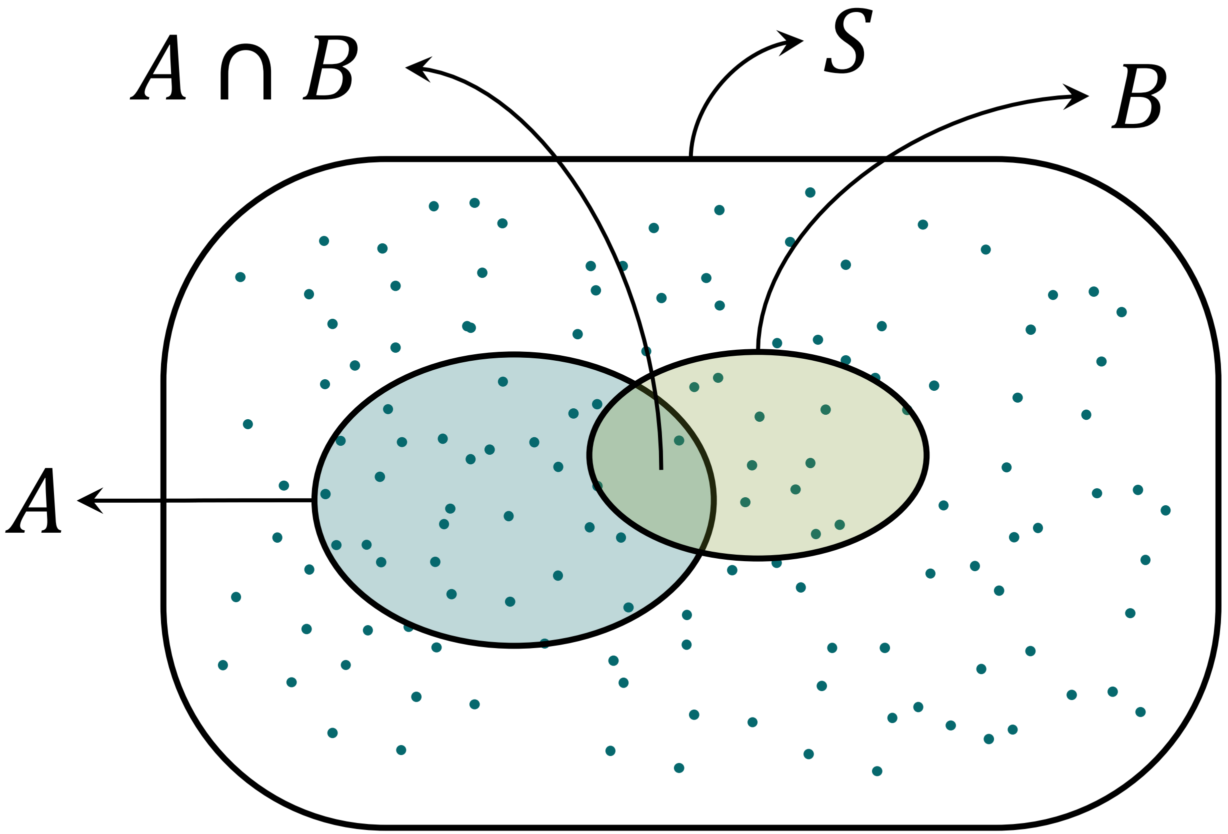
No caso de os eventos A e B não serem mutuamente excludentes, tem-se:

$$\text{prob}(A \cup B) = \text{prob}(A) + \text{prob}(B) - \text{prob}(A \cap B)$$

Pois:

$A \cap B \neq 0$

Assim, os eventos podem (ou não) ocorrer simultaneamente



Fundamentos de probabilidades | conjuntos

Exemplo: Em uma área sujeita a terremotos, dois eventos naturais podem produzir a ruptura de uma barragem: uma cheia catastrófica ou um terremoto destrutivo. Com base em dados anuais observados, estimou-se que a probabilidade de ocorrência desses eventos é 2% e 1%, respectivamente. Estime a probabilidade de a barragem se romper em um ano qualquer.

Solução:

Evento A: cheia catastrófica

Evento B: terremoto destrutivo

Fundamentos de probabilidades | conjuntos

A barragem irá colapsar pela ação do evento A OU pelo evento B. Portanto, o rompimento é causado pela união dos eventos. Assim:

$$\mathrm{prob}(A \cup B) = \mathrm{prob}(A) + \mathrm{prob}(B) - \mathrm{prob}(A \cap B)$$

Não há informações sobre o valor de $\mathrm{prob}(A \cap B)$, portanto assume-se que seja uma probabilidade muito baixa (~ 0). Assim:

$$\mathrm{prob}(A \cup B) = 0{,}02 + 0{,}01 - 0 = 0{,}03$$

A probabilidade de a barragem colapsar em um ano qualquer é de 3%.

Quando a ocorrência de um evento A depende de outro evento B, diz-se que a probabilidade de A está condicionada à probabilidade de B

a probabilidade de A é alterada pela ocorrência do evento B

Usa-se a notação $\mathrm{prob}(A|B)$, na qual o símbolo “|” representa “dado que”

lê-se: a probabilidade de A dado que B aconteceu

Assim:

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

Fundamentos de probabilidades | probabilidade condicional

Alternativamente:

$$\text{prob}(A \cap B) = \text{prob}(B) \cdot \text{prob}(A|B)$$

Agora, se $\text{prob}(A|B) = \text{prob}(A)$, B não influencia A. Nesse caso, são eventos independentes e:

$$\text{prob}(A \cap B) = \text{prob}(B) \cdot \text{prob}(A)$$

Exemplo: Considerando a barragem do exemplo anterior, assuma que as ocorrências de uma cheia catastrófica e de um terremoto destrutivo sejam eventos independentes. Recalcule a probabilidade de ruptura da barragem.

Solução: Em sendo eventos independentes, tem-se:

$$\mathrm{prob}(A \cap B) = \mathrm{prob}(B) \cdot \mathrm{prob}(A) = 0{,}01 \cdot 0{,}02 = 0{,}0002$$

Ou seja, a probabilidade de ocorrer uma cheia catastrófica e um terremoto destrutivo é de 0,02%.

Fundamentos de probabilidades | probabilidade condicional

Portanto:

$$\text{prob}(A \cup B) = 0{,}02 + 0{,}01 - 0{,}0002 = 0{,}00298$$

A probabilidade (atualizada) de a barragem colapsar em um ano qualquer é de 2,98%.

Exemplo: Um estudo sobre as precipitações diárias em Taiamã (MS), mostrou que em novembro a probabilidade de um dia chuvoso ser seguido por outro dia chuvoso é 0,33, um dia seco seguido de outro dia seco é de 0,71, um dia chuvoso ser seguido por um dia seco é de 0,29 e de um dia seco ser seguido por um dia chuvoso é de 0,67. Se um dia qualquer de novembro de 2023 for chuvoso, qual é a probabilidade de que os próximos dois dias sejam também chuvosos?

Solução:

Condição inicial: chuva (hoje)

Evento B: dia 1 chuvoso (amanhã)

Evento A: dia 2 chuvoso (depois de amanhã)

O que se pede é $\mathrm{prob}(A \cap B)$ (os dois próximos dias chuvosos)

$$\mathrm{prob}(A \cap B) = \mathrm{prob}(B) \cdot \mathrm{prob}(A|B)$$

Em sendo a condição inicial chuva, tem-se:

$$\mathrm{prob}(B) = \mathrm{prob}(\mathrm{chuva}|\mathrm{chuva}) = 0{,}33$$

Agora, a probabilidade condicional $\mathrm{prob}(A|B)$ se refere à ocorrência do evento A (chuva depois de amanhã), dado que o evento B (chuva amanhã) ocorreu.

$$\mathrm{prob}(A|B) = \mathrm{prob}(\mathrm{chuva}|\mathrm{chuva}) = 0{,}33$$

Assim:

$$\mathrm{prob}(A \cap B) = 0{,}33 \cdot 0{,}33 \cong 0{,}11$$

Portanto, a probabilidade de um dia chuvoso ser seguido por outros dois dias chuvosos em Taiamã, no mês de novembro de 2023, é de 11%.

Fundamentos de Probabilidade

medidas descritivas de variáveis aleatórias

As definições dadas a seguir se aplicam a medidas descritivas populacionais de variáveis aleatórias

Assim como as estatísticas descritivas resumem as características de uma amostra, as medidas descritivas populacionais resumem as características de uma população

valor esperado
variância populacional
assimetria populacional

Para todas as definições, considera-se a variável aleatória X

Valor esperado (momento central de primeira ordem):

O valor esperado de X equivale à sua média populacional.

Para variáveis aleatórias discretas:

$$E[X] = \mu_X = \sum_{\text{todos } x_i} x_i \cdot p_X(x_i)$$

onde

$p_X(x_i)$ função massa de probabilidades

Valor esperado: (cont.)

Para variáveis aleatórias contínuas:

$$E[X] = \mu_X = \int_{-\infty}^{+\infty} x \cdot f_X(x) \cdot dx$$

onde

$f_X(x)$ função densidade de probabilidades

$p_X(x_i)$ e $f_X(x)$ são assuntos a serem tratados oportunamente

Fundamentos de probabilidades | medidas descritivas de v.a.

<u>Variância populacional</u> (momento central de segunda ordem):

Empregada para avaliar a dispersão das funções $p_X(x_i)$ e $f_X(x)$

$$VAR[X] = \sigma_X^2 = E[(X - \mu_X)^2] = E[(X - E[X])^2]$$
$$VAR[X] = E[X^2] - (E[X])^2$$

onde

$E[X]$ valor esperado de X

Assimetria populacional (momento central de terceira ordem):

Empregada para avaliar a assimetria das funções $p_X(x_i)$ e $f_X(x)$

$$\gamma = \frac{\mu_X^3}{(\sigma_X)^3} = \frac{E[(X-\mu_X)^3]}{(\sigma_X)^3}$$

onde

σ_X desvio padrão populacional de X

Em ciências naturais não se conhece a população, portanto não é possível calcular as medidas descritivas explicitadas

Ainda assim, o conhecimento dessas medidas permite o desenvolvimento de técnicas de estimação de parâmetros das distribuições de probabilidade

Revisão

A probabilidade é uma definição a priori cuja medida cresce em precisão conforme se aumenta o tamanho da amostra (o tamanho ideal da amostra será tema futuro)

a escala de probabilidades varia de 0 a 1

Definições úteis (e necessárias!)

elementos: pontos da amostra
espaço amostral: toda a amostra
experimento: o que nos interessa na amostra
variável aleatória: dá nome ao fenômeno natural de interesse

Probabilidade condicional e independência estatística

Medidas descritivas populacionais

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

PPGERHA

Estatística Aplicada a Ciências Ambientais

Daniel Detzel
detzel@ufpr.br