Probabilidade (PPGECD000000001)

Programa de Pós-Graduação em Estatística e Ciência de Dados (PGECD)

Sessão 3

Raydonal Ospina

Departamento de Estatística
Universidade Federal da Bahia
Salvador/BA

σ-álgebras

Em um experimento aleatório nem todos os subconjuntos do espaço amostral são eventos. Para tanto, exigimos que a classe de subconjuntos para os quais estará definida a “chance” de ocorrência seja uma σ-álgebra. Desta forma, exigimos que a álgebra de eventos também seja fechada com relação a um número enumerável de uniões.

Definição 1

Uma σ-álgebra $\mathcal{A}$ é uma álgebra de eventos que também é fechada com relação a uma união enumerável de eventos,

$$(\forall i \in \mathbb{Z}) A_i \in \mathcal{A} \Rightarrow \cup_{i \in \mathbb{Z}} A_i \in \mathcal{A}.$$

Mais formalmente

Definição 2 (σ-álgebra)

Seja $\Omega \neq \emptyset$. Uma classe (família) de eventos $\mathcal{F}$ é uma σ-álgebra sobre Ω, se e somente se, são satisfeitas as as seguintes propriedades:

1. *Se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$,*
2. $\Omega \in \mathcal{F}$,
3. *Se $A_1, A_2, \ldots, \in \mathcal{F}$, então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

Da definição de σ-álgebra é claro que Ω e $\emptyset$ pertencem a qualquer σ-álgebra definida sobre Ω. No caso dos eventos, Ω é chamado de *evento certo* e $\emptyset$ é chamado de *evento impossível*.

Exemplo 1

1. Se $\Omega \neq \emptyset$, então $\{\Omega, \emptyset\}$ é a menor σ-álgebra que pode ser definida sobre Ω. (σ-álgebra trivial).
2. Se $\Omega = \{1, 2, 3\}$, então $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ é uma σ-álgebra sobre Ω.
3. Se $\Omega \neq \emptyset$, então $\mathcal{P}(\Omega)$ a coleção de todos os subconjuntos de Ω (partes de Ω) é uma σ-álgebra (σ-álgebra total). Em particular, seja $\Omega \neq \emptyset$, finito e enumerável (caso discreto), e seja $\mathcal{F}$ uma σ-álgebra que contém todos todos os conjuntos da forma $\{\omega\}$ com $\omega \in \Omega$. Então, $\mathcal{F} = \mathcal{P}(\Omega)$ é uma σ-álgebra.
4. (σ-álgebra gerada). Seja $\Omega \neq \emptyset$ e $\mathcal{L}$ uma coleção de subconjuntos de Ω. Seja

$$\mathcal{M} = \{\mathcal{F} : \mathcal{F} \text{ é uma } \sigma\text{-álgebra que contem a } \quad \mathcal{L}\},$$

então,

$$\sigma(\mathcal{L}) = \bigcap_{\mathcal{F} \in \mathcal{M}} \mathcal{F}$$

é a menor σ-álgebra sobre Ω que contem a $\mathcal{L}$. Esta σ-álgebra é a σ-álgebra gerada por $\mathcal{L}$.

Exercício 1

Verifique os exemplos anteriores.

Exemplo 2

1. Se $\Omega = \mathbb{R}$ é o conjunto dos números reais. Seja $\mathcal{U}$ a coleção formada por uniões disjuntas de intervalos da forma $(a, b]$, então $\mathcal{U}$ é uma σ-álgebra. Em particular, a menor σ-álgebra sobre $\mathbb{R}$ que contém todos os intervalos da forma $(-\infty, a]$ com $a \in \mathbb{R}$ se chama σ-álgebra de Borel e a denotamos por $\mathcal{B}$. Sendo $\mathbb{Q}$ o conjunto dos números racionais, temos que, os conjuntos $(a, \infty) = \mathbb{R} - (-\infty, a]$, $(a, b] = (-\infty, b] \cap (a, \infty)$, $(-\infty, a) = \bigcup_{n=1}^{\infty}(-\infty, a - 1/n]$, $[a, \infty) = \mathbb{R} - (-\infty, a)$, $(a, b) = (-\infty, b) \cup (a, \infty)$, $[a, b] = \mathbb{R} - ((-\infty, a) \cup (b, \infty))$, $\{a\} = [a, a]$, $\mathbb{N} = \bigcup_{n=0}^{\infty}\{n\}$, $\mathbb{Q} = \bigcup_{r\in\mathbb{Q}}\{r\}$ e $\mathbb{Q}^c = \mathbb{R} - \mathbb{Q}$ são conjuntos da σ-álgebra de Borel em $\mathbb{R}$.

Nota 1

Nem todos os subconjuntos de $\mathbb{R}$ pertencem à σ-álgebra de Borel. 0 conjunto de *Cantor* é um exemplo desta classe "exotica"de conjuntos.

Maiores detalhes ver [Gary L. Wise and Eric B. Hall, Counterexamples in Probability and Real Analysis. Oxford University Press, New York 1993]

Definição 3 (Espaço mensurável)

Se $\mathcal{F}$ é uma σ-álgebra não vazia de subconjuntos de Ω, a dupla $(\Omega, \mathcal{F})$ é chamada de espaço mensurável e os subconjuntos de Ω são chamados de conjuntos mensuráveis.

Definição 4 (Medida)

Seja $(\Omega, \mathcal{F})$ um espaço mensurável. Uma **medida** *sobre $(\Omega, \mathcal{F})$ é uma função μ definida sobre $\mathcal{F}$ que assume valores em $\mathbb{R} \cup \{\infty\}$ tais que:*

1. $\mu(A) \geq 0$ *para todo* $A \in \mathcal{F}$,
2. $\mu(\emptyset) = 0$,
3. *(σ-aditividade) Para toda sequência $E_1, E_2, \ldots$ de conjuntos disjuntos dois a dois de $\mathcal{F}$*

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu\left(E_i\right).$$

Intuitivamente, a medida de um conjunto (evento) pode ser interpretada como o seu “tamanho". Neste sentido, uma medida é uma generalização dos conceitos de comprimento, área e volume.

Nota 2

Para cada $A \in \mathfrak{F}$, o número $\mu(A)$ se chama a *medida de A* e a tripla $(\Omega, \mathfrak{F}, \mu)$ recebe o nome de *espaço de medida*. Além disso, dizemos que a medida μ é *finita* se $\mu(\Omega) < \infty$.

Exemplo 3

1. Seja $\Omega \neq \emptyset$ e seja $\mathcal{F}$ uma σ-álgebra em Ω, se fixamos um ponto $\omega \in \Omega$, definimos para cada $E \in \mathcal{F}$:
$$\mu(E) = \delta_\omega(E) = \begin{cases} 1, & \text{se } \omega \in E, \\ 0, & \text{se } \omega \notin E \end{cases}$$
então μ é uma medida finita chamada de *medida de Dirac*.
2. Seja $\Omega = \mathbb{N}$ e $\mathcal{F} = \mathcal{P}(\mathbb{N})$, para cada $E \subseteq \mathbb{N}$ definimos uma medida $\mu(E)$ como o número de elementos de E, se E é finito e como $+\infty$ se E é infinito. Como podemos ver esta medida não é finita. Esta medida é chamada de *medida de contagem* em $\mathbb{N}$.
3. Seja $\Omega = \mathbb{R}$ e $\mathcal{F} = \mathcal{B}$ a σ-álgebra de Borel. Definimos $\mu((a, b]) = b - a$, (comprimento do intervalo $(a, b]$) em que $a, b \in \mathbb{R}$. Esta medida é chamada de *medida de Lebesgue* na reta real.

Por razões técnicas, fora do escopo deste curso focaremos nossa atenção unicamente numa classe especial de medidas que permitem a formalização axiomática de probabilidade dada por Kolmogorov.

Axiomas de Kolmogorov (Kolmogorov - 1956)

Os axiomas de Kolmogorov não descrevem um único modelo probabilístico, eles apenas determinam uma família de modelos probabilísticos, a escolha de um modelo específico satisfazendo os axiomas é feito pelo analista/estatístico familiar com o fenômeno aleatório sendo modelado.
Motivados pelas propriedades de frequência relativa, impõe-se os primeiros quatro axiomas de Kolmogorov:

- K0. **Inicial.** O experimento aleatório é descrito pelo espaço de probabilidade $(\Omega, \mathcal{A}, P)$ que consiste do espaço amostral Ω, de uma σ-álgebra $\mathcal{A}$, e de uma função de valores reais $P : \mathcal{A} \to \mathbb{R}$.
- K1. **Não-negatividade.** $\forall A \in \mathcal{A}, P(A) \geq 0$.
- K2. **Normalização Unitária.** $P(\Omega) = 1$.
- K3. **Aditividade Finita.** Se A, B são disjuntos, então $P(A \cup B) = P(A) + P(B)$.

É fácil provar (tente!) utilizando indução matemática que K3 é válida para qualquer coleção finita de eventos disjuntos par a par, ou seja, se $A_i, i = 1, 2, \ldots, n$, são eventos disjuntos par a par, então

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

.

Sequências monótonas

- Sejam $A_1, A_2, \ldots$ uma sequência de eventos de $\mathcal{F}$ e suponhamos que $\bigcup_{i=1}^{\infty} A_i = A$, dizemos que os conjuntos A_n formam uma sequência crescente de conjuntos com limite A, se $A_1 \subseteq A_2 \subseteq \ldots$, e o denotaremos por $A_i \uparrow A$.
- Se $A_1 \supseteq A_2 \supseteq \ldots$ e $\bigcap_{i=1}^{\infty} A_i = A$, dizemos que os A_i formam una sequência decrescente de conjuntos com limite A e o denotamos por $A_i \downarrow A$.
- Aplicando as lei de De Morgan a estas sequências temos que se $A_i \uparrow A$, então $A_i^{\complement} \downarrow A^{\complement}$ e se $A_i \downarrow A$, então $A_i^{\complement} \uparrow A^{\complement}$.
- As sequências decrescentes (*crescentes*) de conjuntos as denotamos por *sequências monótonas.*

Um quinto axioma foi proposto por Kolmogorov para garantir um certo grau de continuidade da medida de probabilidade.

K4. **Continuidade Monotônica.** Se para todo $i > 0$, $A_{i+1} \subseteq A_i$ e $\cap_i A_i = \emptyset$, então

$$\lim_{i \to \infty} P(A_i) = 0.$$

Um forma equivalente de K4 é a seguinte:

K4′. σ**-aditividade.** Se $\{A_i\}$ é uma coleção enumerável de eventos disjuntos dois a dois, então

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Teorema 1

Se P satisfaz K0—K3, então P satisfaz K4′ se, e somente se, ela satisfaz K4.

Prova da Equivalência.

Primeiro, vamos provar que K0 – K4 implicam o axioma da σ-aditividade K4$'$. Seja $\{A_i\}$ qualquer sequência enumerável de eventos disjuntos par a par, e defina para todo n

$$B_n = \cup_{i>n} A_i,$$

$$\cup_{i=1}^{\infty} A_i = B_n \cup (\cup_{i=1}^{n} A_i).$$

Claramente, para todo $i \leq n$, temos que A_i e B_n são disjuntos. Por K3, temos

$$P(\cup_{i=1}^{\infty} A_i) = P(B_n) + \sum_{i=1}^{n} P(A_i).$$

Por definição de série numérica,

$$\lim_n \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

K4$'$ segue se conseguirmos mostrar que $\lim_n P(B_n) = 0$. Note que $B_{n+1} \subseteq B_n$, e que $\cap_{n=1}^{\infty} B_n = \emptyset$. Então por K4, temos que o limite acima é zero e K4$'$ é verdadeiro.

$\square$

...continuação

Agora, vamos provar que K0 – K3, K4′ implicam o axioma da continuidade monotônica K4. Seja $\{B_n\}$ qualquer coleção enumerável de eventos satisfazendo as hipóteses do axioma K4: $B_{n+1} \subseteq B_n$ e $\cap_{n=1}^{\infty} B_n = \emptyset$. Defina, $A_n = B_n - B_{n+1}$ e observe que $\{A_n\}$ é uma coleção enumerável de eventos disjuntos par a par. Note que

$$B_n = \cup_{j \geq n} A_j.$$

Então, por K4′ temos que

$$P(B_n) = P(\cup_{j \geq n} A_j) = \sum_{j \geq n} P(A_j).$$

Como por K4′, $\sum_{j=1}^{\infty} P(A_j) = P(\cup_{j=1}^{\infty} A_j) \leq 1$, temos que

$$\lim_n P(B_n) = \lim_n \sum_{j \geq n} P(A_j) = 0,$$

logo K4 é verdadeiro.

Medida de Probabilidade

Definição 5

Uma função que satisfaz as condições K0 – K4 é chamada de uma medida de probabilidade definida sobre $(\Omega, \mathcal{A})$.

1. A terna $(\Omega, \mathcal{A}, P)$ é chamada de **espaço de probabilidade**.
2. Intuitivamente quando se modela uma problema através de probabilidade, basicamente, o que se faz é especificar cada uma das componentes da terna acima.
3. Eventos são os elementos de $\mathcal{A}$, aos quais se pode atribuir probabilidade.
4. Probabilidade é uma função cujo argumento é um conjunto.
5. Portanto, não somente conjuntos, como também as operações sobre eles, têm uma importância fundamental em teoria da probabilidade.

Note que uma vez que $P(\Omega) = 1$, o conjunto Ω é chamado de *evento certo* e $\emptyset$ é chamado de *evento impossível*.

Definição 6 (Espaço de probabilidade completo)

Seja $(\Omega, \mathcal{F}, P)$ *um espaço de probabilidade. Qualquer evento A tal que* $P(A) = 0$ *se chama evento nulo e o espaço* $(\Omega, \mathcal{F}, P)$ *se diz que é completo se todos os subconjuntos de eventos nulos são eventos nulos.*

Exemplos de medidas de probabilidade

Exemplo 4

Seja $\Omega = \{\clubsuit, \heartsuit, \spadesuit\}$, $\mathcal{F} = \{\emptyset, \Omega, \{\clubsuit\}, \{\heartsuit, \spadesuit\}\}$ e P a aplicação definida sobre $\mathcal{F}$ da seguinte forma:

$$P(A) = \begin{cases} 1, & \text{se} \quad \heartsuit \in A, \\ 0, & \text{se} \quad \heartsuit \notin A \end{cases} \tag{1}$$

então, P é uma medida de probabilidade sobre $(\Omega, \mathcal{F})$.

Exemplo 5

Seja $\Omega = \{\clubsuit, \heartsuit\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ e P a aplicação definida sobre $\mathcal{F}$ da seguinte forma:

$$P(A) = \begin{cases} 0, & \text{se} \quad A = \emptyset, \\ 2/5, & \text{se} \quad A = \{\heartsuit\}, \\ 3/5, & \text{se} \quad A = \{\clubsuit\}, \\ 1, & \text{se} \quad A = \{\clubsuit, \heartsuit\}, \end{cases} \tag{2}$$

então, P é uma medida de probabilidade.

Definição 7 (Vetor de probabilidades)

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade com Ω finito ou enumerável e $\mathcal{F} = \mathcal{P}(\Omega)$. Seja $A \in \mathcal{F}$ não vazio. É evidente que $A = \bigcup_{\omega \in A}\{\omega\}$ e portanto, $P(A) = \sum_{\omega \in A} P(\omega)$, em que $P(\omega) = P(\{\omega\})$, i.e., P está totalmente determinada pelas probabilidades $p_j = P(\{\omega_j\})$, em que ω_j, $j = 1, 2, \ldots$ são elementos de Ω. Além disso, o vetor $p = (p_1, p_2, \ldots)$ de dimensão $|\Omega|$, satisfaz:

1. $p_j \geq 0$, *para todo j.*
2. $\sum_{j=1}^{\infty} p_j = 1$,

em que $|\Omega|$ denota o número de elementos de Ω. Um vetor p que satisfaz as condições anteriores se chama vetor de probabilidades.

Definição 8 (Espaço de probabilidade discreto)

Consideramos $\emptyset \neq \Omega = \{\omega_1, \omega_2, \ldots\}$ um conjunto finito ou enumerável, $\mathcal{P}(\Omega)$ a σ-álgebra total sobre Ω e p um vetor de probabilidades de dimensão $|\Omega|$. Facilmente podemos ver que a aplicação P definida sobre $\mathcal{P}(\Omega)$ na forma:

1. $P(\emptyset) = 0$,
2. $P(\omega_j) = p_j$ *com* $j = 1, \ldots,$
3. $P(A) = \sum_{j:\omega_j \in A} p_j$, *para cada* $\emptyset \neq A \subset \Omega$

é uma medida de probabilidade. O espaço de probabilidade $(\Omega, \mathcal{P}(\Omega), P)$ se chama espaço de probabilidade discreto.

Exemplo 6

Escolhemos um inteiro positivo ao acaso, a probabilidade de escolher i é $(\frac{1}{2})^i$. Estendemos a probabilidade a qualquer evento A através de

$$P(A) = \sum_{a \in A} P(\{a\}).$$

A probabilidade de escolher um número par é

$$\sum_{\substack{a \in A \\ a \text{ par}}} P(\{a\}) = \sum_{k \geq 1} \left(\frac{1}{2}\right)^{2k} = \sum_{k \geq 1} \left(\frac{1}{4}\right)^{k} = \frac{1}{3}.$$

Exemplo 7

Um casal é escolhido ao acaso e $(i, j) \in \mathbb{N} \times \mathbb{N}$ representa o número de filhos e o número de filhas do casal. Admitamos que

$$P(\{(i,j)\}) = \frac{1}{2^{i+j+2}}$$

qual é a probabilidade de um casal não ter filho? O evento é dado por $A = \{(0, j) \colon j \in \mathbb{N}\}$e

$$P(A) = \sum_{j \geq 0} \frac{1}{2^{j+2}} = \frac{1}{2}.$$

Espaços amostrais finitos - Espaços de Probabilidade Laplacianos

Definição 9 (Espaço Laplaciano)

Um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ *se chama laplaciano, se* Ω *é finito,* $\mathcal{F} = \mathcal{P}(\Omega)$ *e* $P(\omega) = 1/|\Omega|$ *para todo* $\omega \in \Omega$*. A medida de probabilidade P se chama distribuição Laplaciana (ou uniforme) em* Ω*.*

Nota 3

Se $(\Omega, \mathcal{F}, P)$ é um espaço de probabilidade Laplaciano e $A \subset \Omega$, então

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|},$$

Em outras palavras,

$$P(A) = \frac{\text{“número de casos favoráveis a } A\text{”}}{\text{“número de casos possíveis”}}. \tag{3}$$

A última expressão não é uma definição de probabilidade e sim uma consequência ao supor que todos os resultados são igualmente prováveis. Assim, em um espaço de probabilidade laplaciano o cálculo das probabilidades de um evento se reduz à contagem do número de elementos num conjunto finito, i.e., o cálculo se reduz a um problema de análise combinatória.

Definição 10 (Probabilidade Geométrica)

Seja $(\Omega, \mathcal{F}, P)$ *um espaço de probabilidade e* $A \in \mathcal{F}$ *um evento aleatório. Suponhamos que sobre* $(\Omega, \mathcal{F})$ *seja definida uma medida geométrica* μ *tal como o comprimento, a área, o volume, etc. Definimos a probabilidade do evento A como*

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}. \tag{4}$$

Exemplo 8 (O jogo do ladrilho)

Uma moeda de raio r é lançada ao acaso no chão o qual é coberto por ladrilhos quadrados de lado l $(l > 2r)$ como descrito na Figura 1. As crianças francesas no século XVIII apostavam que a moeda cairia inteiramente dentro de um ladrilho.
Buffon observou, que a probabilidade de a moeda cair inteiramente dentro de um ladrilho era a probabilidade do centro da moeda cair dentro de um quadrado de lado $l - 2r$. Essa probabilidade é a razão entre as áreas do quadrado e do ladrilho, pois a probabilidade do centro da moeda cair em uma região é proporcional à área dessa região.

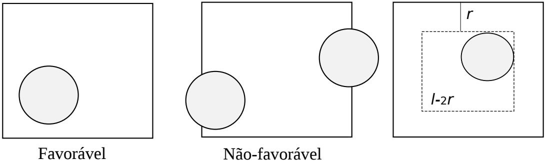

Figura: Representação esquemática do jogo dos ladrilhos.

Portanto, a probabilidade da moeda cair inteiramente dentro de um ladrilho é

$$\frac{(l-2r)^2}{l^2}.$$

Se consideramos um piso formado por quadrados de cerâmica de 30 cm de lado e um disco ("moeda") de raio 5 cm, a probabilidade do disco cair inteiramente dentro de um dos ladrilhos é igual a $(30-10)^2/30^2 = 0,4444$ ou 44,44%.
Nessa situação, o diâmetro d do disco que daria 60% de chances de vitória ao jogador é d = 6,77 cm.

Propriedades da probabilidade

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade então P satisfaz as seguintes propriedades:

P.0 $P(\emptyset) = 0$. De fato, $1 = P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \cup \emptyset \cup \emptyset \cdots)$ $= P(\Omega) + P(\emptyset) + P(\emptyset) + P(\emptyset) + \cdots$, se e somente se, $0 \leq P(\emptyset) \leq 1$ caso contrário contradiz [K.2].

P.1 Seja A um evento de $\mathcal{F}$ e $A^{\complement}$ o evento complementar, então

$$P(A) = 1 - P\left(A^{\complement}\right).$$

De fato, sendo Ω o espaço amostral, temos que $\Omega = A \cup A^{\complement}$ onde a união é disjunta, uma vez que $A \cap A^{\complement} = \emptyset$. Utilizando [K.3] segue que

$$P(\Omega) = P(A) + P\left(A^{\complement}\right) \Rightarrow P\left(A^{\complement}\right) = P(\Omega) - P(A) = 1 - P(A).$$

P.2 A probabilidade de ocorrência de eventos associados a um experimento pode ser calculada através da regra da soma da probabilidade para a união de dois eventos. Sejam A e B dois eventos em $\mathcal{F}$ a probabilidade da união destes dois eventos é dada por

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

De fato, se $A \cup B = A \cup (B - A)$ e $A \cap (B - A) = \emptyset$, então $P(A \cup B) = P(A) + P(B - A)$. Agora para $B = (B - A) \cup (A \cap B)$ com $(B - A) \cap (A \cap B) = \emptyset$, então $P(B) = P(B - A) + P(A \cap B)$. Combinando estes dois resultados, temos que

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

P.3 Sejam A, B e C três eventos em $\mathcal{F}$, então

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

De fato, temos que $A \cup B \cup C = (A \cup B) \cup C = (A \cup B) \cup (C - (A \cup B))$ por ser esta união disjunta. Então por [K.3], temos

$$P(A \cup B \cup C) = P(A \cup B) + P(C - (A \cup B)) \tag{5}$$

e utilizando a equação (5) temos
$P(A \cup B \cup C) = P(A) + P(B) - P(A \cap B) + P(C - (A \cup B))$. Mas,

$$C = (C - (A \cup B)) \cup (C \cap (A \cup B)),$$

portanto

$$P(C - (A \cup B)) = P(C) - P(C \cap (A \cup B)). \tag{6}$$

Além disso, temos que $C \cap (A \cup B) = (C \cap A) \cup (C \cap (B - A))$, e esta união é disjunta. Logo,

$$P(C \cap (A \cup B)) = P(A \cup C) + P(C \cap (B - A)). \tag{7}$$

Finalmente, para $C \cap B = (A \cap B \cap C) \cup (C \cap (B - A))$, o que implica que

$$P(C \cap (B - A)) = P(B \cap C) - P(A \cap B \cap C), \tag{8}$$

já que a união é disjunta.

Combinando as equações (5), (6), (7) e (8), concluímos que

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

P.4 Sejam A e B eventos de $\mathcal{F}$ com $A \subset B$, então $P(A) \leq P(B)$. De fato, temos que se $A \subset B$ então $B = A \cup (B - A)$ e $\emptyset = A \cap (B - A)$. Portanto, utilizando [K.3] segue que

$$P(B) = P(A) + P(B - A).$$

Como $P(B - A) \geq 0$, temos então que $P(B) \geq P(A)$.

P.5 Se $A \subset B$ então

$$P(B - A) = P(B) - P(A).$$

De fato, note que $B = A \cup (B - A)$, e ainda que $A \cap (B - A) = \emptyset$. Assim, ao usarmos [K.3] temos $P(B) = P(A \cup (B - A)) = P(A) + P(B - A)$, o que implica

$$P(B - A) = P(B) - P(A).$$

P.6 No caso geral em que $A_n \downarrow A$ temos que $(A_n - A) \downarrow \emptyset$, então, $P(A_n - A) \to 0$. Uma vez que $A \subseteq A_n$, temos $P(A_n - A) = P(A_n) - P(A) \to 0$, ou seja $\lim_{n\to\infty} P(A_n) = P(A)$.

P.7 (Desigualdade de Boole) Sejam $A_1, A_2, \cdots, A_n$ uma sequência de eventos aleatórios, então

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Para mostrar esta propriedade vamos usar indução finita. Inicialmente mostremos que $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$. De fato pela propriedade P.2

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \Rightarrow P(A_1 \cup A_2) \leq P(A_1) + P(A_2),$$

já que $P(A_1 \cap A_2) \geq 0$. Agora vamos supor que esta propriedade seja válida para $n-1$, ou seja, que $P\left(\bigcup_{i=1}^{n-1} A_i\right) \leq \sum_{i=1}^{n-1} P(A_i)$ e mostremos que é válida para n.

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{n-1} A_i \cup A_n\right) = P(C \cup A_n) = P(C) + P(A_n) - P(C \cap A_n) \\ &\leq P(C) + P(A_n), \end{aligned}$$

em que $C = \bigcup_{i=1}^{n-1} A_i$, e ao usarmos a hipótese de indução temos que

$$P(C) + P(A_n) \leq \sum_{i=1}^{n-1} P(A_i) + P(A_n) = \sum_{i=1}^{n} P(A_i).$$

Corolário 1

Para n eventos arbitrários $\{A_1, \ldots, A_n\}$,

$$P(\cap A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

Demonstração.

Utilizando a Lei de De Morgan e a desigualdade de Boole para os eventos $\{A_1^c, \ldots, A_n^c\}$, temos

$$P(\cup_{i=1}^{n} A_i^c) = 1 - P(\cap A_i) \leq \sum_{i=1}^{n} P(A_i^c) = \sum_{i=1}^{n} (1 - P(A_i)).$$

Logo,

$$P(\cap A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

□

P.8 Sejam $A_1, A_2, \cdots$ eventos de $\mathcal{F}$, então $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$.

De fato, se definimos $C_n = \bigcup_{i=1}^{n} A_i$, temos que $C_n \uparrow C$, no qual C é definido como $C = \bigcup_{i=1}^{\infty} A_i$. Por continuidade sabemos que $P(C_n) \uparrow P(C)$. Usando a propriedade P.7 temos que

$$P(C_n) = P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i),$$

por outro lado

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(C) = \lim_{n\to\infty} P(C_n) \leq \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i),$$

Princípio da Inclusão-Exclusão:

O princípio da inclusão-exclusão permite determinar o cardinal da união de vários conjuntos tomando como base os cardinais da cada um deles e todas as suas possíveis interseções. Sejam $A_1, \ldots, A_n$ conjuntos finitos, então:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{i,j\,:\,1\leq i<j\leq n} \left|A_i \cap A_j\right| + \sum_{i,j,k\,:\,1\leq i<j<k\leq n} \left|A_i \cap A_j \cap A_k\right| - \cdots + (-1)^{n-1} \left|A_1 \cap \cdots \cap A_n\right|.$$

Para o caso de dois conjuntos A e B finitos temos
$|(A-B) \cup (B-A) \cup (A \cap B) = |(A-B)| + |(B-A)| + |(A \cap B)$ por serem disjuntos.
Por outro lado $|(A-B)| = |A| + |A \cap B|$ e $|(B-A)| = |B| + |A \cap B|$.
Juntando essa duas expressões temos que
$|A \cup B| = |A| - |A \cap B| + |B| - |A \cap B| + |A \cap B|$, logo

$$|A \cup B| = |A| - |B| - |A \cap B|.$$

Como uma exemplificação verifiquemos a fórmula para o caso de termos três conjuntos A, B e C finitos. Para isto, observemos o diagrama de Venn-Euler (Figura 2) que representa o conjunto $(A \cup B \cup C)$ o qual permitirá construir o argumento para verificar o princípio de inclusão-exclusão para o caso $n = 3$.

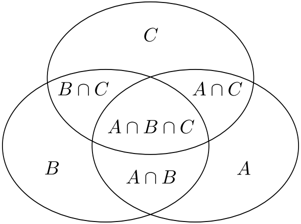

Figura: Representação gráfica do conjunto $(A \cup B \cup C)$ para a verificação do princípio de inclusão-exclusão.

Assim,

$$\begin{aligned}
|A \cup B \cup C| &= |A \cup (B \cup C)| \\
&= |A| + |B \cup C| - |A \cap (B \cup C)| \\
&= |A| + |B| + |C| - |B \cap C| - |(A \cap B) \cup (A \cap C)| \\
&= |A| + |B| + |C| - |B \cap C| - (|A \cap B| + |A \cap C| - |A \cap B \cap C|) \\
&= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.
\end{aligned}$$

O qual verifica a fórmula para $n = 3$.

Exemplo 9

Quantos são os números inteiros positivos menores que 504 e primos com 504? Usando a decomposição em fatores primos temos que $504 = 2^3 \times 3^2 \times 7$. Agora, definimos os conjuntos

$$\begin{aligned} A &= \{1, 2, \ldots, 504\}, \\ A_1 &= \{x \in A : x \text{ é múltiplo de } 2\}, \\ A_2 &= \{x \in A : x \text{ é múltiplo de } 3\}, \\ A_3 &= \{x \in A : x \text{ é múltiplo de } 7\}. \end{aligned}$$

Desejamos calcular a cardinalidade do conjunto $A - (A_1 \cup A_2 \cup A_3)$. Desta forma,
$|A_1| = 504/2 = 252$, $|A_2| = 504/3 = 168$, $|A_3| = 504/7 = 72$,
$|A_1 \cap A_2| = 504/(2 \times 3) = 84$, $|A_1 \cap A_3| = 504/(2 \times 7) = 36$,
$|A_2 \cap A_3| = 504/(3 \times 7) = 24$, $|A_1 \cap A_2 \cap A_3| = 504/(2 \times 3 \times 7) = 12$. Usando o princípio de inclusão-exclusão

$$|A_1 \cup A_2 \cup A_3| = 252 + 168 + 72 - 84 - 36 - 24 + 12 = 360.$$

Assim, existem ao todo 144 números inteiros positivos menores que 504 e primos com 504.

Princípio da Inclusão-Exclusão para probabilidade

O próximo teorema permite que possamos calcular de maneira exata a probabilidade $P(\cup_{i=1}^{n} A_i)$ para n eventos arbitrários.

Teorema 2

Princípio da Inclusão-Exclusão. *Seja I um conjunto genérico de índices que é um subconjunto não-vazio qualquer de $\{1, 2, \ldots, n\}$. Para eventos arbitrários $\{A_1, \ldots, A_n\}$,*

$$P(\cup_{i=1}^{n} A_i) = \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} P(\cap_{i \in I} A_i),$$

onde o somatório é sobre todos os $2^n - 1$ conjuntos de índices excluindo apenas o conjunto vazio.

Demonstração.

A prova é por indução matemática em n.
Para $n = 1$, temos apenas um evento A_1:

$$P(A_1) = \sum_{\emptyset \neq I \subseteq \{1\}} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right) = (-1)^{1+1} P(A_1) = P(A_1),$$

logo vale para $n = 1$.

□

Para $n = 2$, temos eventos A_1 e A_2:

$$
\begin{aligned}
P(A_1 \cup A_2) &= \sum_{\emptyset \neq I \subseteq \{1,2\}} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right) \\
&= (-1)^{1+1} P(A_1) + (-1)^{1+1} P(A_2) + (-1)^{2+1} P(A_1 \cap A_2) \\
&= P(A_1) + P(A_2) - P(A_1 \cap A_2),
\end{aligned}
$$

que é a fórmula para a união de dois eventos que já tínhamos demonstrado, logo para $n = 2$ vale a expressão. Usemos agora a hipótese de Indução, i.e. a fórmula é válida para para $n - 1$ eventos, ou seja:

$$
P\left(\bigcup_{i=1}^{n-1} A_i\right) = \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right).
$$

Note que

$$
P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right).
$$

Logo

$$
P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right).
$$

Usemos agora a hipótese de Indução, i.e. a fórmula é válida para para $n-1$ eventos, ou seja:

$$P\left(\bigcup_{i=1}^{n-1} A_i\right) = \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right).$$

Note que

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right).$$

Logo

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right).$$

Agora, note que

$$\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n = \bigcup_{i=1}^{n-1} (A_i \cap A_n).$$

Então:

$$P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right) = P\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right).$$

□

Note que podemos aplicar a hipótese de indução nos eventos $\{(A_i \cap A_n)\}_{i=1}^{n-1}$, desta forma

$$P\left(\bigcup_{i=1}^{n-1}(A_i \cap A_n)\right) = \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i\in I}(A_i \cap A_n)\right).$$

Note que

$$\bigcap_{i\in I}(A_i \cap A_n) = \left(\bigcap_{i\in I} A_i\right) \cap A_n.$$

Logo, se substituímos estas quantidades nas expressões anteriores, temos:

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\bigcup_{i=1}^{n-1}(A_i \cap A_n)\right) \\
&= \left[\sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i\in I} A_i\right)\right] + P(A_n) \\
&\quad - \left[\sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} P\left(\left(\bigcap_{i\in I} A_i\right) \cap A_n\right)\right].
\end{aligned}$$

$\square$

Reorganizando as expressões e agrupando os termos semelhantes, temos

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_n) + \sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} \left[P\left(\bigcap_{i\in I} A_i\right) - P\left(\left(\bigcap_{i\in I} A_i\right) \cap A_n\right)\right].$$

Agora usando o fato de que para dois conjuntos quaisquer A e B, temos que $P(A) - P(A\cap B) = P(A\setminus B)$, podemos escrever

$$P\left(\bigcap_{i\in I} A_i\right) - P\left(\left(\bigcap_{i\in I} A_i\right) \cap A_n\right) = P\left(\bigcap_{i\in I} A_i \setminus A_n\right).$$

Substituindo está quantidade na expressão anterior, notamos que

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_n) + \sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i\in I} A_i \setminus A_n\right).$$

No entanto, este caminho pode tornar-se complexo (Não vá por este caminho, pode provar mas vá ser uma baita dor de cabeça). Em vez disso, podemos pensar em todas as interseções possíveis que incluem A_n e aquelas que não incluem A_n. Ou seja vamos ver mais de perto esses conjuntos. $\square$

Vamos expressar o somatório sobre todos os subconjuntos não vazios de $\{1, 2, \ldots, n\}$ como a soma de dois termos, i.e., *Subconjuntos que não contêm o n-ésimo*: Todos os subconjuntos de $\{1, 2, \ldots, n-1\}$ e os *Subconjuntos que contêm o n-ésimo* ou seja, subconjuntos da forma $I' \cup \{n\}$, em que $I' \subseteq \{1, 2, \ldots, n-1\}$.
Com isto na cabeça, vejam que, se abusamos com a notação temos que o somatório total é:

$$\sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} = \sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} + \sum_{I' \subseteq \{1,\ldots,n-1\}} \left(I = I' \cup \{n\}\right).$$

Desta forma, podemos rescrever:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right) + \sum_{I' \subseteq \{1,\ldots,n-1\}} (-1)^{|I'|+1} P\left(\bigcap_{i \in I'} A_i \cap A_n\right).$$

É fácil ver que , para $I' = \emptyset$, temos $P\left(\bigcap_{i \in \emptyset} A_i \cap A_n\right) = P(A_n)$ e $|I'| = 0$, então o coeficiente é $(-1)^{0+1} = -1$ que tem o correspondente termo $-P(A_n)$. □

Agora, se adicionamos as duas somatórias e ajustamos os sinais, podemos ver que

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \left[\sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i\in I} A_i\right)\right] + \left[-P(A_n) + \sum_{\emptyset \neq I' \subseteq \{1,\ldots,n-1\}} (-1)^{|I'|+1} P\left(\bigcap_{i\in I'} A_i \cap A_n\right)\right].$$

Note que $(-1)^{|I'|+1+1} = (-1)^{|I'|+2} = (-1)^{|I'|}$. Logo, se reorganizarmos os termos temos

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \left[\sum_{\emptyset \neq I \subseteq \{1,\ldots,n-1\}} (-1)^{|I|+1} P\left(\bigcap_{i\in I} A_i\right) + (-1)^{1+1} P(A_n)\right] + \left[\sum_{\emptyset \neq I' \subseteq \{1,\ldots,n-1\}} (-1)^{(|I'|+1)+1} P\left(\bigcap_{i\in I'} A_i \cap A_n\right)\right].$$

Assim,

$$\sum_{\emptyset \neq I' \subseteq \{1,\ldots,n-1\}} (-1)^{|I'|} P\left(\bigcap_{i\in I'} A_i \cap A_n\right).$$

□

Portanto, a expressão completa é:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{I\subseteq\{1,\ldots,n\},\ \emptyset\neq I,\ n\notin I} (-1)^{|I|+1} P\left(\bigcap_{i\in I} A_i\right) + \sum_{I\subseteq\{1,\ldots,n\},\ n\in I} (-1)^{|I|+1} P\left(\bigcap_{i\in I} A_i\right)$$

e as somas cobrem todos os subconjuntos não vazios de $\{1, \ldots, n\}$, correspondendo ao enunciado do teorema. Assim, o teorema é válido para $n-1$, então é válido para n. FIM

(Graças a Deus, Zeus, Ganesh, Thor, Buda, ... e meus alunos do PGECD) $\square$

Extra: Técnicas para contagem

Variações com repetição: Amostras com ordem e com substituição.

Suponhamos que temos uma gaveta com n objetos distintos. Desejamos realizar k extrações ao acaso de um objeto ao mesmo tempo. Ao efetuar uma extração, registramos o objeto escolhido (marcamos o objeto, isto é o distinguimos) e o devolvemos à gaveta, desta forma o objeto pode ser selecionado várias vezes. Em cada extração temos n objetos possíveis para serem escolhidos e efetuamos k extrações. Assim pelo princípio da multiplicação o número total de arranjos que podem ser obtidos desta gaveta ao se fazer k extrações é

$$\underbrace{n \cdot n \cdots n}_{n \text{ vezes}} = n^k.$$

Este número é chamado de *ordenações com repetição*.

Exemplo 10

Suponhamos que temos um conjunto de 60 caracteres distintos. este conjunto contém todas as letras minúsculas e as letras maiúsculas do alfabeto, os dez dígitos e alguns caracteres especias como: %, @, \$, #, etc. Quantas senhas de comprimento 6 podem ser construídas usando estes 60 caracteres?
Como cada caracter dos 60 disponíveis pode ser escolhido para ser colocado em cada uma das seis posições da senha, então podemos construir
$60 \times 60 \times 60 \times 60 \times 60 \times 60 = 60^6 = 4.6656 \times 10^{10}$ distintas senhas.

Variações sem repetição: Amostras com ordem e sem substituição.

Temos uma gaveta com n objetos e dos quais se devem extrair, um a um k objetos. Suponhamos que nesta situação a amostra é *sem substituição*, isto quer dizer, uma vez selecionado um objeto este não é devolvido à gaveta. O total de arranjos distintos que podemos obter é

$$n(n-1)(n-2)\cdots(n-k-1). \tag{9}$$

Pelo princípio da multiplicação notamos que há k fatores na expressão anterior. O primeiro fator é n e isto é devido ao fato que temos qualquer dos n objetos para serem colocados na primeira posição, para a segunda posição temos $(n-1)$ objetos, para a terceira posição $(n-2)$ objetos, e assim por diante. este raciocínio termina ao escolher o k-ésimo objeto para o qual temos unicamente $(n-k+1)$ possibilidades. A expressão anterior pode ser escrita como

$$P(n,k) = \frac{n!}{(n-k)!}, \quad k \leq n$$

e se chama de *permutação de n em k*.

Exemplo 11

De quantas formas podem ser atribuídos ou primeiro, segundo e terceiro prêmio em uma rifa de 10 boletos numerados de 1 até 10? Notemos que de fato este problema é um problema de ordenação sem repetição de 10 objetos em que devem ser extraídos 3 objetos de estes. Daí, existem $10 \times 9 \times 8 = 720$ distintas atribuições para os três primeiros números na rifa.

Permutações: Amostras exaustivas com ordem e sem substituição.

A pergunta básica a respeito do número total de formas em que podemos colocar em um ordem linear (um elemento após o outro e portanto sem repetição) n objetos distintos tem com resposta o chamado *fatorial* de| n, o qual é denotado por

$$n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1.$$

Por definição temos que $0! = 1$.

Exemplo 12

Suponhamos que queremos acomodar algumas crianças em uma fila, quatro meninas e três meninos. Se os meninos e meninas podem ser alocados em qualquer ordem então há $7! = 5040$ formas de acomodar as crianças. Agora se queremos que os meninos e as meninas fiquem alternados na fila então há $(4\times 3)\times(3\times 2)\times(2\times 1)\times 1 = 144$ formas de organizá-los. Se desejamos que os meninos formem um grupo e as meninas formem outro grupo na fila temos $2\times 4!\times 3! = 288$ formas de acomodá-los.

Combinações: Amostras sem ordem e sem substituição

Suponhamos que temos um conjunto de n objetos distinguíveis e estamos interessados em obter uma amostra (subconjunto) de tamanho k. Suponhamos que as amostras agora devem ser *sem ordem e sem repetição*. Lembremos que quando a ordem importa temos $n!/(n-k)!$ possibilidades. Agora, como não estamos interessados na ordem observamos que um dos arranjos desta fórmula está sendo contado $k!$ vezes. As vezes em que os mesmos k elementos podem ser permutados uns com os outros, uma vez que o conjunto de dados é o mesmo. Assim, para obter arranjos em que a ordem não importa devemos então dividir por $k!$. Está fórmula se chama de *combinações de n em k* a qual é denotada por

$$C_n^k = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}, \tag{10}$$

em que n é o total de elementos e k é o número de elementos escolhidos.
Note que da equação 10 pode ser deduzido que

$$\binom{n}{k} = \binom{n}{n-k}. \tag{11}$$

De fato, se o número $\binom{n}{k}$ representa o número de subconjuntos de k elementos de um conjunto de n elementos, então, se inicialmente escolhemos k objetos estamos deixando de lado $n-k$ objetos, que é equivalente a escolher $n-k$ objetos que logo serão deixados de lado.

Exemplo 13

Megassena Uma importante aplicação de combinação é nas loterias: Megassena, quina entre outras. A megassena consiste em uma cartela de 60 números dentre os quais devemos acertar 6 (prêmio principal). Calcule a quantidade total de resultados possíveis para o prêmio principal. Para marcar um cartão, precisamos escolher 6 entre 60 números, em que a ordem de escolha não interfere no cartão que será marcado. Trata-se portanto, de acordo com a definição, de um problema de combinação (devemos combinar 60 números, em grupos de 6 números, ou seja, queremos subconjuntos de 6 elementos de um conjunto de 10 elementos). O número de cartões é

$$\binom{60}{6} = 50063860$$

Exemplo 14

Seja A um conjunto finito com n elementos, então A possui 2^n subconjuntos. De fato, sabemos que o coeficiente $\binom{n}{k}$ representa o número de subconjuntos de k elementos de um conjunto de n elementos. Então, se somamos em k ($k = 0$ elementos, $k = 1$ elementos, e assim por diante ate $k = n$ elementos) obtemos o número de subconjuntos do conjunto A.

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = (1 + 1)^n = 2^n.$$

Extra: Exercícios - Exemplos

Exemplo 15

Modelos de gavetas Suponhamos que em uma caixa há N bolas do mesmo tipo, mas de cores diferentes, a saber: R bolas vermelhas e $N - R$ brancas. Se extraem ao acaso n bolas. Qual é a probabilidade de extrair exatamente $k \leq n$ bolas vermelhas?

Caso I

Suponhamos que as bolas se encontram enumeradas de 1 até N e que a enumeração das bolas vermelhas vá de 1 até R. Nesta situação devemos lembrar que é necessário distinguir dois casos. A extração é feita com substituição e a extração é feita sem substituição. No primeiro caso podemos considerar duas alternativas:

i) As bolas são retiradas uma após a outra : As n bolas são extraídas uma a uma da caixa e deixadas fora da caixa. Neste caso o espaço amostral está dado por:

$$\Omega = \{(a_1, a_2, \ldots, a_n) : a_j \in \{1, 2, \ldots, N\}, a_i \neq a_j, i \neq j, j = 1, 2, \ldots, n\}$$

Então, definimos os eventos

A_k="exatamente $k \leq n$ bolas vermelhas são extraídas".

Note que A_k é uma n-upla em Ω, que contém exatamente k componentes menores ou iguais a R. Temos desta forma que:

$$
\begin{aligned}
|\Omega| &= N(N-1)...(N-(n-1)) = (N)_n \\
|A_k| &= \binom{n}{k} R(R-1)...(R-k+1)(N-R)...(N-R-(n-k)+1) \\
&= \binom{n}{k} P(R,k)P(N-R,n-k).
\end{aligned}
$$

Agora, ao supormos que o experimento é Laplaciano, tem-se

$$P(A_k) = \frac{|A_k|}{|\Omega|} = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

ii) As bolas são todas retiradas ao mesmo tempo: As n bolas são todas extraídas ao tempo; neste caso temos que $\Omega = \{T : T \subseteq \{1, 2, \ldots, N\}, \text{com } |T| = n\}$ e A_k definido como em i) consta de todos os subconjuntos de $\{1, 2, \ldots, N\}$ que contém exatamente k componentes menores ou iguais a R. Desta forma $|\Omega| = \binom{N}{n}$ e $|A_k| = \binom{R}{k}\binom{N-R}{n-k}$. Ao supor novamente que o experimento é Laplaciano

$$P(A_k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

Aqui, $p_k = P(A_k)$ define uma medida de probabilidade sobre o conjunto $\Omega' = \{0, 1, \ldots, n\}$, chamada de *distribuição hipergeométrica* com parâmetros n, R e N, a qual é denotada por $H_g(n, R, N)$.

Caso II

No segundo caso, a extração é com substituição, cada bola extraída é devolvida imediatamente à caixa; depois de misturar as bolas, extrai-se aleatoriamente a seguinte bola e assim por diante. Neste caso, temos que o espaço amostral é igual a:

$$\begin{aligned}\Omega &= \{(a_1, a_2, \ldots, a_n) : a_j \in \{1, 2, \ldots, N\},\ j = 1, 2, \ldots, n\} \\ &= \{1, \ldots, N\} \times \{1, \ldots, N\} \times \cdots \times \{1, \ldots, N\} = \{1, \ldots, N\}^n,\end{aligned}$$

e o evento A_k consta de todos os elementos $(a_1, a_2, \ldots, a_n) \in \Omega$ com exatamente k componentes menores ou iguais a R. Temos assim, $|\Omega| = N \cdot N \cdots N = N^n$ e

$$|A_k| = \binom{n}{k} R^k (N - R)^{n-k}.$$

Portanto, se supursemos que todas as bolas têm a mesma chance de serem extraídas, então

$$p_k = P(A_k) = \binom{n}{k} \frac{R^k (N-R)^{n-k}}{N^n} = \binom{n}{k} p^k q^{n-k},$$

em que $p = R/N$ e $q = 1 - p$. Se estamos interessados unicamente no número k de bolas vermelhas entre as n bolas extraídas, temos que

$$p_k = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, 2, \ldots, n, \qquad 0 < p < 1, q = 1 - p$$

define una medida de probabilidade sobre o conjunto $\Omega' = \{0, 1, 2, \ldots, n\}$ chamada de *distribuição binomial* com parâmetros n e p a qual é denotada por $\mathcal{B}(n, p)$.

Exercício 2

Se A, B e C forem eventos mutuamente excludentes, com $P(A) = 0{,}2$, $P(B) = 0{,}3$ e $P(C) = 0{,}4$, determine:

(a) $P(A \cap B \cap C)$.

(b) $P(A^c \cup (B \cup C))$.

(c) $P((A \cup B) \cap C)$.

Exercício 3

Se A, B e C forem eventos mutuamente excludentes, será possível obter $P(A) = 0{,}3$, $P(B) = 0{,}4$ e $P(C) = 0{,}5$? Justifique.

Exercício 4

Se $\Omega = \{a, b, c\}$, e a álgebra $\mathcal{A}$ é o conjunto das partes de Ω, e a medida de probabilidade P é parcialmente definida por

$$P(\{a, b\}) = 0.5,\ P(\{b, c\}) = 0.8,\ P(\{a, c\}) = 0.7,$$

então complete a especificação de P para todos os eventos em $\mathcal{A}$.

Exercício 5

Se $\{A_i\}$ for uma partição enumerável de Ω e $P(A_i) = ab^i$, $i \geq 1$, então quais as condições que a e b devem satisfazer para que P seja uma medida de probabilidade?

Exemplo 16

Em um grupo de r pessoas qual a probabilidade de haver pelo menos duas pessoas que façam aniversário no mesmo dia, assumindo que a distribuição de aniversários é uniforme ao longo do ano e desprezando a existência de anos bissextos?

Solução: O número de resultados possíveis para os aniversários de r pessoas é 365^r. O número de casos possíveis onde todas as pessoas fazem aniversário em dias diferentes é dado por $365 \times 364 \times \cdots \times (365 - (r - 1))$. Portanto, o número de casos possíveis onde pelo menos duas pessoas fazem aniversário no mesmo dia é a diferença entre o número total de aniversários possíveis e o número de casos onde as pessoas têm aniversários em datas diferentes, ou seja, é igual a

$$365^r - 365 \times 364 \times \cdots \times (365 - (r - 1)).$$

Logo, a probabilidade deste evento é:

$$1 - \frac{365 \times 364 \times \cdots \times (365 - (r - 1))}{365^r}.$$

Para $r = 23$, temos que essa probabilidade é aproximadamente igual a $0,51$. E para $r = 50$, essa probabilidade é igual a $0,97$.

Exemplo 17

Em uma loteria de N números há um só prêmio. Salvador compra n $(1 < n < N)$ bilhetes para uma só extração e Sílvio compra n bilhetes, um para cada uma de n extrações. Qual dos dois jogadores têm mais chances de ganhar algum prêmio?

Solução: A probabilidade de Salvador ganhar algum prêmio é $\frac{n}{N}$. O número total de n extrações possíveis é N^n. O número de casos onde Sílvio não ganha nenhum prêmio é $(N-1)^n$, logo o número de casos onde Sílvio ganha algum prêmio é igual a $N^n - (N-1)^n$. Logo, a probabilidade de Sílvio ganhar algum prêmio é $1 - \frac{(N-1)^n}{N^n}$. Vamos provar por indução que Salvador tem mais chance de ganhar, ou seja, $\frac{n}{N} > 1 - \frac{(N-1)^n}{N^n}$, que equivale a

$$\frac{(N-1)^n}{N^n} > 1 - \frac{n}{N}.$$

Para $n = 2$, temos:

$$\frac{(N-1)^2}{N^2} = 1 - \frac{2}{N} + \frac{1}{N^2} > 1 - \frac{2}{N}.$$

Solução (cont.) Suponha que para $n = k$, temos que

$$\frac{(N-1)^k}{N^k} > 1 - \frac{k}{N}.$$

Multiplicando esta expressão por $\frac{N-1}{N}$, obtemos:

$$\frac{(N-1)^{k+1}}{N^{k+1}} > (\frac{N-1}{N})(1 - \frac{k}{N})$$
$$= 1 - \frac{1}{N} - \frac{k}{N} + \frac{k}{N^2} > 1 - \frac{k+1}{N}.$$

Exemplo 18

Doze pessoas são divididas em três grupos de 4. Qual é a probabilidade de duas determinadas dessas pessoas ficarem no mesmo grupo?

Solução: O número total de divisões de doze pessoas em 3 grupos de 4 é igual a $\binom{12}{4}\binom{8}{4}\binom{4}{4}$. Vamos agora contar o número de casos favoráveis ao nosso evento. Existem 3 opções de escolhermos em qual grupo as duas pessoas determinadas podem ficar.

Solução (cont.)

Das 10 pessoas restantes, temos que escolher mais duas para estarem neste grupo, o que podemos fazer de $\binom{10}{2}$ maneiras diferentes. E temos $\binom{8}{4}\binom{4}{4}$ maneiras diferentes de dividir as outras 8 pessoas nos dois grupos restantes. Portanto, a probabilidade de duas determinadas pessoas ficarem no mesmo grupo é:

$$\frac{3\binom{10}{2}\binom{8}{4}\binom{4}{4}}{\binom{12}{4}\binom{8}{4}\binom{4}{4}} = \frac{3}{11}.$$

Exemplo 19

Suponha que temos em uma sala n mães cada uma com um filho. Suponha formemos duplas aleatoriamente, onde cada dupla contém uma mãe e um filho, qual a probabilidade de que pelo menos uma mãe forme uma dupla com seu próprio filho?

Solução: Seja A_i o evento que a i-ésima mãe forma dupla com seu filho. Queremos determinar

$$P(\cup_{i=1}^{n} A_i).$$

Vamos calcular esta probabilidade utilizando a fórmula da inclusão exclusão.

Solução: (cont.) Note que:

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n} \text{ para todo } i \in \{1, 2, \ldots, n\}$$
$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \text{ para } i \neq j$$

e em geral, para um grupo $I \in \{1, 2, \ldots, n\}$ de mães temos que

$$P(\cap_{i \in I} A_i) = \frac{(n-|I|)!}{n!}.$$

Como existem $\binom{n}{|I|}$ grupos de mães com cardinalidade $||I||$, temos que

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} \frac{(n-i)!}{n!}$$
$$= \sum_{i=1}^{n} (-1)^{i+1} \frac{1}{i!}$$

Note que quando $n \to \infty$, temos que esta probabilidade tende a $1 - \frac{1}{e}$.

Exemplo 20

Demonstre que se $P(A_i) = 1$ para $i = 1, 2, \ldots$, então $P(\cap_{i=1}^{\infty} A_i) = 1$.

Solução: Como $P(A_i) = 1$, temos que $P(A_i^c) = 1 - P(A_i) = 0$. Logo pela não-negatividade e pela desigualdade de Boole, temos $0 \leq P(\cup_{i=1}^{\infty} A_i^c) \leq \sum_{i=1}^{\infty} P(A_i^c) = 0$. Portanto, como pela Lei de De'Morgan, $\cap_{i=1}^{\infty} A_i = (\cup_{i=1}^{\infty} A_i^c)^c$, temos que $P(\cap_{i=1}^{\infty} A_i) = 1 - P(\cup_{i=1}^{\infty} A_i^c) = 1$.

Exemplo 21

Demonstre: se $A_1, A_2, \ldots$ e $B_1, B_2, \ldots$ são eventos aleatórios do mesmo espaço de probabilidade tais que $P(A_n) \to 1$ e $P(B_n) \to p$, então $P(A_n \cap B_n) \to p$.

Solução: Note que

$$\begin{aligned} P(A_n \cap B_n) &= 1 - P((A_n \cap B_n)^c) = 1 - P(A_n^c \cup B_n^c) \\ &\geq 1 - P(A_n^c) - P(B_n^c) = P(A_n) + P(B_n) - 1. \end{aligned} \tag{12}$$

Como $P(A_n) + P(B_n) - 1 \to p$, temos que $\liminf P(A_n \cap B_n) \geq p$. Por outro lado, como $P(A_n \cap B_n) \leq P(B_n)$ e $P(B_n) \to p$, temos que $\limsup P(A_n \cap B_n) \leq p$. Portanto, $\lim P(A_n \cap B_n) = p$.