Introdução à Probabilidade

Prof. Leandro Chaves Rêgo

Departamento de Estatística e Matemática Aplicada - UFC

Fortaleza, 11 de outubro de 2021

Experimento Aleatório

- Um *experimento* é qualquer processo de observação. *Experimentos aleatórios* são aqueles em que não somos capazes de predizer seu comportamento em futuras realizações.
- As razões para nossa falta de habilidade para predizer são varias:
  - nós podemos não saber de todas as causas envolvidas;
  - nós podemos não ter dados suficientes sobre as condições iniciais do experimento;
  - as causas podem ser tão complexas que o cálculo do seu efeito combinado não é possível;
  - ou na verdade existe alguma aleatoriedade fundamental no experimento.
- Salvo mencionado em contrário, este curso restringe-se à classe de experimentos aleatórios cujo conjuntos de possíveis resultados seja conhecido.

Observação

É importante ressaltar que frequentemente são encontradas situações práticas onde não se consegue descrever todos os possíveis resultados de um experimento. Uma maneira de contornar este problema é assumir que um resultado possível do experimento é a não ocorrência de qualquer dos resultados descritos, contudo, em problemas práticos, tal suposição pode acarretar em dificuldades quando se tenta elicitar ou deduzir probabilidades.

Caracterização Experimento Aleatório

Os resultados de um experimento aleatório são caracterizados pelos seguintes componentes:

1. o conjunto de resultados possíveis Ω;
2. a coleção de conjuntos de resultados de interesse $\mathcal{A}$;
3. um valor numérico P da verossimilhança ou probabilidade de ocorrência de cada um dos conjuntos de resultados de interesse.

Espaço Amostral

- O conjunto de possíveis resultados de um experimento aleatório é chamado de *espaço amostral*.
- Em um dado experimento aleatório a especificação do espaço amostral deve ser tal que este:
  1. liste todos os possíveis resultados do experimento sem duplicação e o
  2. faça em um nível de detalhamento suficiente para os interesses desejados, omitindo resultados que, embora logicamente ou fisicamente possíveis, não tenham qualquer implicação prática na sua análise.
- Por exemplo, uma única jogada de uma moeda pode ter o espaço amostral tradicional $\Omega = \{cara, coroa\}$, ou podemos considerar que a moeda pode fisicamente ficar equilibrada na borda $\Omega = \{cara, coroa, borda\}$ (SS1). Uma outra possibilidade seria levar em consideração as coordenadas (x, y) do centro da moeda quando ela para após ser jogada no ar.

Espaço Amostral

- Podemos classificar espaços amostrais em dois tipos de acordo com o número de elementos que eles contem. Espaços amostrais podem ser enumeráveis ou não enumeráveis; dependendo se Ω for um conjunto enumerável ou não, respectivamente.
- Em um nível filosófico, pode-se argumentar que só existem espaços amostrais enumeráveis, visto que medidas não podem ser feitas com infinita precisão.
- Enquanto na prática isto é verdadeiro, métodos estatísticos e probabilísticos associados com espaços amostrais não enumeráveis são, em geral, menos complicados que aqueles para espaços amostrais enumeráveis, e proporcionam uma boa aproximação para a situação (enumerável) real.

Evento

Definição

Um *evento* é um subconjunto do espaço amostral, ou seja, é um conjunto de resultados possíveis do experimento aleatório.

- Se ao realizarmos um experimento aleatório, o resultado pertence a um dado evento A, dizemos que A *ocorreu*. Estaremos interessados no estudo da ocorrência de combinações de eventos. Para tanto, utilizaremos as operações Booleanas de conjuntos (complementar, união, intersecção, diferença) para expressar eventos combinados de interesse.

Eventos Disjuntos

Definição

Os eventos A e B são *disjuntos* ou *mutuamente excludentes* ou *mutuamente exclusivos* se não puderem ocorrer juntos, ou, em linguagem de conjuntos, $A \cap B = \emptyset$.

Exemplos

Exemplo

Sejam A, B, e C eventos em um mesmo espaço amostral Ω. Expresse os seguintes eventos em função de A, B, e C e operações Booleanas de conjuntos.

(a) Pelo menos um deles ocorre:

(b) Exatamente um deles ocorre:

(c) Apenas A ocorre:

(d) Pelo menos dois ocorrem:

(e) No máximo dois deles ocorrem:

(f) Nenhum deles ocorre:

(g) Ambos A e B ocorrem, mas C não ocorre:

Coleção de Eventos

Embora possa-se pensar que, dado um espaço amostral, necessariamente é de interesse analisar todos os seus subconjuntos (e isto eventualmente é verdadeiro), temos três razões para esperar que estejamos apenas interessados em alguns subconjuntos do espaço amostral:

1. O espaço amostral pode conter um grau de detalhamento superior ao que estamos interessados no momento. Por exemplo, ele pode representar uma única jogada de um dado com 6 elementos, mas nós apenas estamos interessados em saber se o resultado é par ou ímpar.
2. Nós vamos querer associar cada evento A com uma probabilidade numérica $P(A)$. Como essas probabilidades estão baseadas em algum conhecimento sobre a tendência de ocorrer do evento ou no grau de nossa crença que determinado evento ocorrerá, nosso conhecimento sobre P pode não estender para todos os subconjuntos de Ω.
3. Condições impostas em P pelos axiomas de Kolmogorov, que estudaremos adiante, podem não permitir que P seja definida em todos os subconjuntos de Ω, em particular isto pode ocorrer quando Ω for não enumerável, mas não iremos demonstrar este fato que está fora do escopo deste curso.

Álgebra de Eventos

Estaremos interessados em uma coleção especial $\mathcal{A}$ de subconjuntos do espaço amostral Ω (note que $\mathcal{A}$ é um conjunto cujos elementos também são conjuntos!) que são eventos de interesse no que se refere ao experimento aleatório $\mathcal{E}$ e os quais temos conhecimento sobre a sua verossimilhança de ocorrência. $\mathcal{A}$ é chamado de uma *σ-álgebra de eventos*. Como veremos, o domínio de uma medida de probabilidade é uma σ-álgebra.

Definição

Uma álgebra de eventos $\mathcal{F}$ é uma coleção de subconjuntos do espaço amostral Ω que satisfaz:

1. não é vazia;
2. fechada com respeito a complementos (se $A \in \mathcal{F}$, então $A^c \in \mathcal{A}$);
3. fechada com respeito a uniões finitas (se $A, B \in \mathcal{F}$, então $A \cup B \in \mathcal{F}$).

σ-Álgebra de Eventos

Definição

Uma σ-álgebra de eventos $\mathcal{A}$ é uma álgebra de eventos de um espaço amostral Ω que também é fechada com relação a uma união enumerável de eventos, ou seja, se $\mathcal{I}$ for um conjunto enumerável:

$$A_i \in \mathcal{A}, \forall i \in \mathcal{I} \Rightarrow \cup_{i \in \mathcal{I}} A_i \in \mathcal{A}.$$

Pelas Leis de De Morgan, vemos que $\mathcal{A}$ é fechada com respeito a intersecções enumeráveis também.

Exemplos

Exemplo

1. A menor σ-álgebra de eventos é $\mathcal{A} = \{\emptyset, \Omega\}$;
2. A maior σ-álgebra de eventos é o conjunto das partes de Ω;
3. Um exemplo intermediário, temos:

$$\Omega = \{1, 2, 3\},\ \mathcal{A} = \{\Omega, \emptyset, \{1\}, \{2, 3\}\}.$$

4. A álgebra de eventos finitos e co-finitos. Seja $\Omega = \boldsymbol{R}$ e

$$\mathcal{A} = \{A \subseteq \boldsymbol{R} : A \text{ é finito}\} \cup \{A \subseteq \boldsymbol{R} : A^c \text{ é finito}\},$$

ou seja, $\mathcal{A}$ consiste dos subconjuntos de $\boldsymbol{R}$ que ou são finitos ou têm complementos finitos. $\mathcal{A}$ é uma álgebra de eventos, mas não é uma σ-álgebra.

Propriedades de álgebra

Lema

Se $\mathcal{A}$ é uma álgebra, então $\Omega \in \mathcal{A}$

Prova: Como $\mathcal{A}$ é não vazio, seja A um elemento qualquer seu. Pela segunda propriedade de álgebras, temos que $A^c \in \mathcal{A}$, e pela terceira propriedade temos que $\Omega = A \cup A^c \in \mathcal{A}$. ∎

Teorema

Sejam $\mathcal{A}_1$ e $\mathcal{A}_2$ álgebras de subconjuntos de Ω e seja $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ a coleção de subconjuntos comuns as duas álgebras. Então, $\mathcal{A}$ é uma álgebra.

Prova: Como $\mathcal{A}_1$ e $\mathcal{A}_2$ são álgebras, ambos contém Ω. Então, $\Omega \in \mathcal{A}$. Se $A \in \mathcal{A}$, então A está em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$. Logo, A^c está em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$, e portanto na sua intersecção $\mathcal{A}$. Se $A, B \in \mathcal{A}$, então eles estão em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$. Consequentemente, $A \cup B$ está em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$ e, portanto, em $\mathcal{A}$. Como $\mathcal{A}$ satisfaz as três condições da definição de álgebra de eventos, $\mathcal{A}$ é uma álgebra de eventos. ∎

Propriedades de álgebra

É fácil ver que a prova do Teorema 3.8 pode ser estendida para o caso de uma intersecção de um número arbitrário de álgebras. O seguinte corolário usa este fato para provar que sempre existe uma menor álgebra contendo uma família qualquer de eventos.

Corolário

Existe uma menor (no sentido de inclusão) álgebra contendo qualquer família dada de subconjuntos de Ω.

Prova: Seja $\mathcal{C}$ uma coleção qualquer de subconjuntos de Ω, defina $\mathcal{A}(\mathcal{C})$ como sendo o conjunto que é igual a intercessão de todas as álgebras de eventos que contém $\mathcal{C}$, isto é:

$$\mathcal{A}(\mathcal{C}) = \bigcap_{\mathcal{A} \supseteq \mathcal{C} : \mathcal{A} \text{ é uma álgebra de eventos}} \mathcal{A}.$$

Pelo Teorema 3.8, $\mathcal{A}(\mathcal{C})$ é uma álgebra de eventos, e consequentemente é a menor álgebra de eventos contendo $\mathcal{C}$. $\mathcal{A}(\mathcal{C})$ é conhecida como a *álgebra de eventos gerada por* $\mathcal{C}$. ∎

Construindo uma álgebra

Observação

Uma maneira de construir uma álgebra de eventos, é primeiro particionar Ω em um número finito subconjuntos e depois considerar álgebra que consiste dos eventos que são uniões finitas dos subconjuntos da partição. ∎

Exemplo

Por exemplo, $\Omega = \{a, b, c, d\}$. Considere a partição, $\{\{a, c\}, \{b, d\}\}$, então considere a coleção de eventos que consiste de uniões finitas dos eventos desta partição: $\mathcal{A} = \{\emptyset, \Omega, \{a, c\}, \{b, d\}\}$. É fácil ver que $\mathcal{A}$ é uma álgebra de eventos.

Álgebra gerada por uma coleção de eventos

Dada uma coleção finita eventos $\mathcal{C} = \{A_1, A_2, \ldots, A_n\}$, define-se um átomo de $\mathcal{C}$ como sendo qualquer evento B da seguinte forma: $B = B_1 \cap B_2 \cap \ldots \cap B_n$, onde $B_i = A_i$ ou $B_i = A_i^c$ para $i = 1, 2, \ldots, n$. Note que existem no máximo $2^{||\mathcal{C}||}$ átomos diferentes e que eles formam uma partição de $\mathcal{C}$ (verifique!). Quando $\mathcal{C}$ for uma coleção finita de eventos, um evento pertencerá a $\mathcal{A}(\mathcal{C})$, se e somente se, for igual a uma união finita de átomos de $\mathcal{C}$. Note que $\mathcal{A}(\mathcal{C})$ terá no máximo $2^{2^{||\mathcal{C}||}}$ elementos (verifique!).

Exemplo

Se $\Omega = \{a, b, c, d, e, f\}$, encontre a álgebra gerada por $\mathcal{C} = \{\{a, b, d\}, \{b, d, f\}\}$. Os átomos de $\mathcal{C}$ são $\{\{a\}, \{f\}, \{c, e\}, \{b, d\}\}$. Logo,

$$\mathcal{A}(\mathcal{C}) = \{\emptyset, \Omega, \{a\}, \{f\}, \{c, e\}, \{b, d\}, \{a, f\}, \{a, c, e\},$$
$$\{a, b, d\}, \{c, e, f\}, \{b, d, f\}, \{b, c, d, e\}, \{a, f, c, e\},$$
$$\{a, f, b, d\}, \{a, b, c, d, e\}, \{b, c, e, d, f\}\}.$$

Indução Matemática

Abriremos um parênteses na nossa discussão sobre experimentos aleatórios para discutirmos um método de prova matemática que será bastante utilizado ao longo do curso: a *Indução Matemática*.
Indução Matemática é um método de prova matemática usado para demonstrar a verdade de um número infinito de proposições. A forma mais simples e mais comum de indução matemática prova que um enunciado vale para todos os números naturais n e consiste de dois passos:

- A base: mostrar que o enunciado vale para $n = 1$.
- O passo indutivo: mostrar que, se o enunciado vale para $n = k$, então o mesmo enunciado vale para $n = k + 1$.

Indução Matemática

Esse método funciona provando que o enunciado é verdadeiro para um valor inicial, e então provando que o processo usado para ir de um valor para o próximo é valido. Se ambas as coisas são provadas, então qualquer valor pode ser obtido através da repetição desse processo. Para entender por que os dois passos são suficientes, é útil pensar no efeito dominó: se você tem uma longa fila de dominós em pé e você puder assegurar que:

- O primeiro dominó cairá.
- Sempre que um dominó cair, seu próximo vizinho também cairá. então você pode concluir que todos os dominós cairão.

Indução Matemática (Exemplo)

Exemplo

Suponha que desejemos provar o seguinte enunciado: para todos os números naturais n,

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Esta é uma fórmula simples para a soma dos números naturais de 1 a n. A prova de que o enunciado é verdadeiro para todos os números naturais n é dada a seguir.

Prova: Verificar se o enunciado é verdadeiro para $n = 1$ (base). Claramente, do lado esquerdo da equação fica 1 e do lado direito $1(1+1)/2$, resolvendo dá $1 = 1$. Então o enunciado é verdadeiro para $n = 1$. Podemos definir este enunciado como $P(n)$ e portanto temos que $P(1)$ é verdadeiro.

Indução Matemática (Exemplo)

Agora precisamos mostrar que se o enunciado vale quando $n = k$, então ele também vale quando $n = k + 1$ (passo indutivo). Isto pode ser feito da seguinte maneira:
Assuma que o enunciado é válido para $n = k$, ou seja:

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}.$$

Adicionando k + 1 a ambos os lados:

$$\sum_{i=1}^{k+1} i = \frac{k(k+1)}{2} + k + 1 = \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \frac{(k+2)(k+1)}{2}.$$

Este último é o enunciado para $n = k + 1$. Note que ele não foi provado como verdadeiro: nós assumimos que $P(k)$ é verdadeiro, e desta suposição concluímos que $P(k+1)$ é verdadeiro. Simbolicamente, mostramos que:

$$P(k) \Rightarrow P(k+1)$$

Indução Matemática (Exemplo)

Por indução, no entanto, podemos concluir que o enunciado $P(n)$ vale para todos os números naturais n:

1. $P(1)$ é verdadeiro, logo P(2) é verdadeiro (usando o passo indutivo)
2. Como P(2) é verdadeiro, então P(3) também é
3. Então usando-se o passo indutivo P(N) será verdadeiro e o P(N+1) também

Propriedade de Álgebra

Agora podemos provar que toda álgebra é fechada com respeito a um número finito de uniões.

Teorema

Se $\mathcal{A}$ é uma álgebra de eventos, então

$$A_i \in \mathcal{A}, i = 1, 2, \ldots, n \Rightarrow \cup_{i=1}^{n} A_i \in \mathcal{A}$$

Prova: Para $n = 1$, o resultado é óbvio. Para $n = 2$, o resultado segue diretamente da terceira propriedade na definição de álgebra de eventos. Vamos agora provar o passo indutivo, suponha que

$$A_i \in \mathcal{A}, i = 1, 2, \ldots, k \Rightarrow \cup_{i=1}^{k} A_i \in \mathcal{A}.$$

Vamos agora provar que o caso $n = k + 1$ é verdadeiro. Suponha que $A_i, i = 1, 2, \ldots, k + 1 \in \mathcal{A}$, então como

$$\cup_{i=1}^{k+1} A_i = (\cup_{i=1}^{k} A_i) \cup A_{k+1},$$

temos que utilizando o caso $n = k$, $\cup_{i=1}^{k} A_i \in \mathcal{A}$. Como $\cup_{i=1}^{k} A_i \in \mathcal{A}$ e $A_{k+1} \in \mathcal{A}$, temos que utilizando o caso $n = 2$, $(\cup_{i=1}^{k} A_i) \cup A_{k+1} \in \mathcal{A}$. ∎

Fundamentos de Probabilidade

- Raciocínio probabilístico aparece em uma ampla variedade de fenômenos de chance e incerteza, ele é lugar comum em nosso dia-a-dia. Nós expressamos julgamentos probabilísticos tanto através da linguagem como através de nossas ações.
- Ultrapassar um carro em uma estrada com outro carro vindo em direção oposta implica que calculamos as distâncias e velocidades, e calculamos os riscos de uma batida ocorrer e estamos conscientes das graves consequências de erros nos nossos julgamentos, mas os consideramos pequenos o suficiente.
- Raciocínio probabilístico no dia-a-dia enquanto não desenvolvido matematicamente precisa ser levado seriamente em conta se desejamos tomar decisões racionais.
- Nota-se que, em geral, precisamos incorporar conhecimento probabilístico que seja tanto qualitativo e expresso linguisticamente como também o conhecimento quantitativo que pode ser expresso numericamente.

Fundamentos de Probabilidade

Antes de focarmos em uma teoria probabilística, vamos explorar o espaço de alternativas. Nós podemos classificar as formas de raciocínio probabilístico nas seguintes dimensões:

- grau de precisão – o conceito estrutural
- o significado, ou interpretação a ser dada a probabilidade
- estrutura matemática formal de probabilidade dada por um conjunto de axiomas

Classificação do Raciocínio Probabilístico

- O conceito estrutural determina a precisão com que podemos esperar que probabilidade represente fenômenos aleatórios.
- A interpretação proporciona a base com a qual probabilidade deve ser determinada e indica o que podemos esperar aprender com ela, ou seja, o que uma afirmação probabilística significa.
- O conceito estrutural e a interpretação guiam a escolha dos axiomas. O conjunto de axiomas, contudo, pode somente capturar uma parte do que entendemos da interpretação.

Compreensão de fundamentos de probabilidade é importante, pois aplicações de teoria da probabilidade dependem fortemente de seus fundamentos. Por exemplo, os fundamentos influem na escolha dos métodos estatísticos a serem utilizados (Frequentistas, Bayesianos, . . .) e na interpretação dos resultados obtidos.

Exemplos de Motivação

Exemplo

Suponha que Alice tenha uma moeda honesta e que ela e Bob saibam que a moeda é honesta. Alice joga a moeda e olha o resultado. Após a moeda ser jogada, qual a probabilidade de cara segundo Bob? Um argumento diria que a probabilidade ainda é $1/2$, pois Bob não aprendeu nada sobre o resultado da jogada, então ele não deve alterar o valor de sua probabilidade. Um outro argumento, questiona se realmente faz sentido falar sobre probabilidade de cara depois que a moeda foi jogada. Segundo este argumento, a moeda ou caiu cara ou coroa, então o melhor que Bob pode afirmar é que a probabilidade de cara ou é 0 ou é 1, mas ele não sabe discernir entre esses valores.

Exemplos de Motivação

Exemplo

Suponha agora que Alice tenha duas moedas, uma honesta e outra tendenciosa e é duas vezes mais provável dar cara que coroa com esta moeda. Alice escolhe uma das moedas (suponha que ela sabe distinguir as moedas) e está prestes a jogá-la. Bob sabe que uma moeda é honesta e que a outra é tendenciosa e que é duas vezes mais provável cair cara que coroa com a moeda tendenciosa, mas ele não sabe que moeda Alice escolheu nem lhe foi dada a probabilidade com que Alice escolhe a moeda honesta. Qual a probabilidade de cara segundo Bob?

Exemplos de Motivação

Exemplo

Paradoxo de Ellsbergue. Suponha que existam duas urnas cada uma com 60 bolas. A urna 1 contém 30 bolas azuis e 30 bolas verdes. Tudo que se sabe sobre a urna 2 é que ela contém bolas azuis e verdes, mas não sabe-se a distribuição das bolas. Considere que existem duas loteria com prêmios baseados no sorteio de bolas dessas urnas. Loteria L_1 paga R$1.000,00 se uma bola azul for sorteada na urna 1, e R$0,00 caso contrário. Loteria L_2 paga R$1.000,00 se uma bola azul for sorteada na urna 2, e R$0,00 caso contrário. A maioria das pessoas quando questionada se prefere um bilhete da Loteria L_1 ou L_2 prefere um bilhete da loteria L_1.

Exemplos de Motivação

Exemplo

Suponha agora que temos duas outras loterias L_3 e L_4, onde a primeira paga R\$1.000,00 somente se uma bola verde for sorteada da urna 1, e a segunda para R\$1.000,00 somente se uma bola verde for sorteada da urna 2. Também, é verificado que a maioria das pessoas que preferiram a loteria L_1 a loteria L_2 preferem a loteria L_3 a loteria L_4. Com estas preferências, não é possível que o decisor possua uma única distribuição de probabilidade subjetiva sobre as cores das bolas na urna 2, pois a primeira preferência (L_1 sobre L_2) indica que o decisor considera que existam mais bolas verdes que azuis na urna 2, e a segunda (L_3 sobre L_4) indica que o decisor considera que existam mais bolas azuis que verdes na urna 2. Esse fenômeno é conhecido na literatura como *aversão a ambiguidade*, e pode-se modelar a incerteza do decisor por um conjunto de medidas de probabilidade ao invés de uma única medida de probabilidade.

Nós discutiremos uma variedade de conceitos estruturais e interpretações de probabilidade. Depois nós focaremos na probabilidade numérica tradicional que satisfaz os famosos axiomas de Kolmogorov e em uma interpretação baseada em frequências de ocorrência.

Hierarquia de Conceitos Estruturais de Probabilidade

Os seguintes são exemplos de uma variedade de conceitos estruturais de probabilidade:

Possivelmente. "Possivelmente A" é o conceito mais rudimentar e menos preciso, e o usado pelos antigos Gregos para distinguir entre o que era necessário e o que era contingente. Existe um número de conceitos de possibilidade que incluem os seguintes:

- **possibilidade lógica,** no sentido que não se contradiz logicamente;
- **possibilidade epistêmica,** segundo a qual ocorrência de A não contradiz nosso conhecimento, que inclui, mas estende mais que mera lógica;
- **possibilidade física,** a ocorrência de A é compatível com leis físicas, contudo ela pode ser extremamente improvável — por exemplo, uma moeda parando e ficando equilibrada na borda em uma superfície rígida;
- **possibilidade prática,** a noção do dia-a-dia segundo a qual A é praticamente possível se ele tem pelo menos uma verossimilhança não tão pequena de ocorrer.

Hierarquia de Conceitos Estruturais de Probabilidade

Provavelmente. Provavelmente A é um fortalecimento da noção de possibilidade que significa mais que provável que não. Enquanto ela pode corresponder ao caso que a probabilidade numérica de A seja maior que $1/2$, este conceito não requer nenhum comprometimento com probabilidade numérica nem com o preciso estado de conhecimento que probabilidade numérica requer.

Probabilidade Comparativa. "A é pelo menos tão provável quanto B". A probabilidade comparativa inclui "provavelmente A" através de "A é pelo menos tão provável quanto A^c". Pode ser relacionada com probabilidade numérica através de $P(A) \geq P(B)$; embora como nos dois exemplos anteriores, probabilidade comparativa não requer nenhum comprometimento com probabilidade numérica.

Probabilidade Intervalar. "A tem probabilidade intervalar, ou probabilidade inferior e superior $(\underline{P}(A), \overline{P}(A))$". Isto permite um grau de indeterminação variável sem nenhum comprometimento com que exista um "verdadeiro" valor no intervalo.

Hierarquia de Conceitos Estruturais de Probabilidade

Probabilidade Numérica. "A probabilidade de A é o número real $P(A)$." Este é o conceito usual como qual nos ocuparemos neste curso. Enquanto este conceito absorveu quase toda atenção de pessoas envolvidas com fenômenos de chance e incerteza e provou ser frutífero na prática científica, este não é o único conceito utilizado em linguagem ordinária e no raciocínio probabilístico do dia-a-dia. É duvidoso que probabilidade numérica seja adequada a todas as aplicações que ela é utilizada, e é provável que ela tenha inibido o desenvolvimento de teorias matemáticas apropriadas para outros fenômenos aleatórios.

De agora em diante focaremos no conceito estrutural mais utilizado e preciso que é a probabilidade numérica.

Interpretações de Probabilidade

Parece não ser possível reduzir probabilidade a outros conceitos; ela é uma noção em si mesma. O melhor que podemos fazer é relacionar probabilidade a outros conceitos através de uma interpretação. Os cinco mais comuns grupos de interpretação são os seguintes:

1. **Lógica:** grau de confirmação da hipótese de uma proposição que "A ocorre" dada uma evidência através da proposição que "B ocorreu". Esta interpretação está ligada a um sistema lógico formal e não, digamos, ao mundo físico. Ela é usada para tornar o raciocínio indutivo quantitativo. Quando as evidências ou premissas são insuficientes para deduzir logicamente a hipótese ou conclusão, podemos ainda medir quantitativamente o grau de suporte que uma evidência da a uma hipótese através de probabilidade lógica.
2. **Subjetiva:** se refere ao grau de crença pessoal na ocorrência do evento A e é medida através da interpretação comportamental de disposição a apostar ou agir.

Interpretações de Probabilidade

3. **Frequentista:** se refere ao limite da frequência relativa de ocorrência do evento A em repetidas realizações não relacionadas do experimento aleatório $\mathcal{E}$. Note que limites de frequência relativas são uma idealização, pois não se pode realizar infinitas realizações de um experimento.
4. **Propensidade:** tendência, propensidade, ou disposição para um evento A ocorrer. Por exemplo, considerações de simetria, podem levar a conclusão que um dado tem a mesma propensão ou tendência a cair em qualquer uma de suas faces.
5. **Clássica:** baseada em uma enumeração de *casos igualmente prováveis*.

Frequências Relativas

Resta-nos discutir o terceiro elemento para modelagem do raciocínio probabilístico, a associação de uma medida numérica a eventos que representam a probabilidade com que eles ocorrem. As propriedades desta associação são motivadas em grande parte pelas propriedades de frequência relativas. Considere uma coleção de experimentos aleatórios $\mathcal{E}_i$ que possuem a mesma σ-álgebra de eventos $\mathcal{A}$ e tem resultados individuais não necessariamente numéricos $\{\omega_i\}$. Fixando uma dada sequência de resultados $\{\omega_i\}$, se estamos interessados na ocorrência de um dado evento A, a frequência relativa de A nada mas é que uma média aritmética da função indicadora de A calculada em cada um dos termos da sequência $\{\omega_i\}$, ou seja,

Definição

A frequência relativa de um evento A, determinada pelos resultados $\{\omega_1, \ldots, \omega_n\}$ de n experimentos aleatórios, é

$$r_n(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(\omega_i) = \frac{N_n(A)}{n}.$$

Propriedades da Frequência Relativa

Propriedades chaves da frequência relativa são:

FR0. $r_n : \mathcal{A} \to \boldsymbol{R}$.

FR1. $r_n(A) \geq 0$.

FR2. $r_n(\Omega) = 1$.

FR3. Se A e B são disjuntos, então $r_n(A \cup B) = r_n(A) + r_n(B)$.

FR4. Se $A_1, A_2, \cdots A_n, \cdots$ é uma sequência de eventos disjuntos dois a dois, então $r_n(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} r_n(A_i)$.

Regularidade Estatística

Nós prosseguiremos como se existisse algo que garanta que $r_n(A) \to P(A)$, embora que o sentido de convergência quando n cresce só será explicado pela Lei dos Grandes Números, que não será discutida neste curso. Esta tendência da frequência relativa de estabilizar em um certo valor é conhecida como *regularidade estatística*. Deste modo, P herdará propriedades da frequência relativa r_n.

Axiomas de Kolmogorov

- Questões de probabilidade em situações práticas basicamente constituem-se, como seria o esperado, em como calcular probabilidades. Aí é onde a situação se complica. Portanto a construção axiomática da teoria da probabilidade, abstrai o cálculo de probabilidades de casos particulares e nos provê de um método formal para resolver problemas probabilísticos.
- Os axiomas que descreveremos a seguir não descrevem um único modelo probabilístico, eles apenas determinam uma família de modelos probabilísticos, com os quais poderemos utilizar métodos matemáticos para descobrir propriedades que serão verdadeiras em qualquer modelo probabilístico. A escolha de um modelo específico satisfazendo os axiomas é feito pelo analista/estatístico familiar com o fenômeno aleatório sendo modelado.

Axiomas de Kolmogorov

Motivados pelas propriedades de frequência relativa, impõe-se os primeiros quatro axiomas de Kolmogorov:

K1. **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 $P : \mathcal{A} \to \boldsymbol{R}$.

K2. **Não-negatividade.** $\forall A \in \mathcal{A}, P(A) \geq 0$.

K3. **Normalização Unitária.** $P(\Omega) = 1$.

K4. **Aditividade Finita.** Se A, B são disjuntos, então $P(A \cup B) = P(A) + P(B)$.

É fácil provar (tente!) utilizando indução matemática que K4 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 $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Axiomas de Kolmogorov

Um quinto axioma, embora não tenha significado em espaços amostrais finitos, foi proposto por Kolmogorov para garantir um certo grau de continuidade da medida de probabilidade.

K5. **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 K5 é a seguinte:

K5′. σ-**aditividade.** Se $\{A_i\}$ é uma coleção enumerável de eventos disjuntos dois a dois, então

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

Equivalência entre K5 e K5′

Teorema

Se P satisfaz K1—K4, então P satisfaz K5′ se, e somente se, ela satisfaz K5.

Prova: Primeiro, vamos provar que K1—K5 implicam o axioma da σ-aditividade K5′. 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 K4, temos

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

Por definição de série numérica,

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

K5′ 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 K5, temos que o limite acima é zero e K5′ é verdadeiro.

Equivalência entre K5 e K5′

Agora, vamos provar que K1—K4, K5′ implicam o axioma da continuidade monotônica K5. Seja $\{B_n\}$ qualquer coleção enumerável de eventos satisfazendo as hipóteses do axioma K5: $B_{n+1} \subseteq B_n$ 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 K5′ temos que

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

Como por K5′,

$$\sum_{j=1}^{\infty} P(A_j) = P(\cup_{j=1}^{\infty} A_j) \leq 1,$$

temos que

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

logo K5 é verdadeiro. ∎

Definição de Probabilidade

Definição

Uma função que satisfaz K1—K5 é chamada de uma medida de probabilidade.

A terna $(\Omega, \mathcal{A}, P)$ é chamada de **espaço de probabilidade**. Intuitivamente quando se modela um problema através de probabilidade, basicamente, o que se faz é especificar cada uma das componentes da terna acima.
Eventos são os elementos de $\mathcal{A}$, aos quais se pode atribuir probabilidade.
Probabilidade é uma função cujo argumento é um conjunto.

Linguagem Probabilística

Portanto, não somente conjuntos, como também as operações sobre eles, têm uma importância fundamental em teoria da probabilidade. Entretanto, é preciso que a linguagem de conjuntos seja traduzida para a linguagem de probabilidade. A tabela, a seguir, exibe algumas dessas traduções. A idéia subjacente é que um experimento aleatório foi realizado e aconteceu algum evento.

Ω	conjunto universo	espaço amostral, evento certo
ω	elemento	resultado do experimento
A	conjunto A	evento A
$\emptyset$	conjunto vazio	evento impossível
A^c ou $\overline{A}$	complemento de A	não ocorreu o evento A
$A \cap B$	A intersecção B	os eventos A e B ocorreram
$A \cup B$	A união B	os eventos A ou B ocorreram
$\cap_n A_n$	intersecção dos conjuntos A_n	todos os eventos A_n ocorreram
$\cup_n A_n$	união dos conjuntos A_n	ao menos um dos eventos A_n ocorreu

Tabela: Interpretações interessantes

Exemplos de Medidas de Probabilidade

Exemplo

Se Ω for um conjunto finito, então temos que a probabilidade clássica que assume que todos os resultados são igualmente prováveis, é um exemplo de uma medida de probabilidade. Neste caso, temos que

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

definido para qualquer subconjunto A de Ω. O fato que $0 \leq ||A|| \leq ||\Omega||$ e que

$$||A \cup B|| = ||A|| + ||B|| - ||A \cap B||,$$

permitem que verifiquemos que P satisfaz os axiomas de Kolmogorov.

Exemplo

Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ um conjunto finito, e seja $P(\{\omega_i\}) = p_i$, onde $p_i \geq 0, i \geq 1$ e $\sum_{i=1}^{n} p_i = 1$, e $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$. Neste caso, também é fácil verificar que P é uma medida de probabilidade verificando os axiomas.

Propriedades de uma Medida de Probabilidade

Teorema

Se P é uma medida de probabilidade, então

1. $P(A^c) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. $P(A) \leq 1$.

Prova: Parte 1, segue do fato que $\Omega = A \cup A^c$, K3, e K4, pois

$$1 = P(\Omega) = P(A) + P(A^c).$$

Parte 2, segue da Parte 1, do fato que $\Omega^c = \emptyset$, e K3, pois

$$P(\emptyset) = 1 - P(\Omega) = 0.$$

Parte 3, segue do fato que $1 = P(\Omega) = P(A) + P(A^c) \geq P(A)$, já que $P(A^c) \geq 0$ por K2. ∎

Propriedades de uma Medida de Probabilidade

Teorema

Monotonicidade. *Se* $A \subseteq B$, *então* $P(A) \leq P(B)$.

Prova: Note que $B = A \cup (B - A)$, onde A e $B - A$ são disjuntos. Então K4 implica que $P(B) = P(A) + P(B - A)$. O resultado segue do fato que $P(B - A) \geq 0$. ∎

Corolário

$P(A \cup B) \geq \max(P(A), P(B)) \geq \min(P(A), P(B)) \geq P(A \cap B)$.

Teorema

Uma expressão exata para a probabilidade de uma união não-disjunta é dada por

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

Prova: Como $A \cup B = A \cup (B - A)$, e A e $B - A$ são disjuntos, K4 implica que $P(A \cup B) = P(A) + P(B - A)$. E como $B = (A \cap B) \cup (B - A)$, $A \cap B$ e $B - A$ são disjuntos, K4 implica que $P(B) = P(A \cap B) + P(B - A)$. Logo,

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

Propriedades de uma Medida de Probabilidade

Teorema

Probabilidade de Partições. *Se $\{A_i\}$ é uma partição enumerável de Ω feita de conjuntos em $\mathcal{A}$, então para todo $B \in \mathcal{A}$*

$$P(B) = \sum_i P(B \cap A_i).$$

Prova: Como $\{A_i\}$ é uma partição, segue que

$$B = B \cap \Omega = B \cap (\cup_i A_i) = \cup_i (B \cap A_i).$$

O resultado segue então por K5$'$. ∎

Propriedades de uma Medida de Probabilidade

Teorema

Desigualdade de Boole. *Para n eventos arbitrários* $\{A_1, \ldots, A_n\}$, *a desigualdade de Boole é*

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

Prova: Provaremos por indução matemática em n. A desigualdade é trivialmente verdadeira para $n = 1$ e verdadeira para $n = 2$, pois é uma consequência imediata do Teorema 4.13. Assuma que a desigualdade é válida para $n = k$ e vamos provar que ela é válida para $n = k + 1$. Para ver isto, escrevemos $\cup_{i=1}^{k+1} A_i = A_{k+1} \cup \cup_{i=1}^{k} A_i$.
Pela desigualdade para $n = 2$,

$$P(\cup_{i=1}^{k+1} A_i) \leq P(A_{k+1}) + P(\cup_{i=1}^{k} A_i).$$

Pela hipótese do passo indutivo, para $n = k$,

$$P(\cup_{i=1}^{k} A_i) \leq \sum_{i=1}^{k} P(A_i),$$

portanto, a desigualdade de Boole é verdadeira. ∎

Propriedades de uma Medida de Probabilidade

Corolário

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

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

Prova: 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,

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

∎

Propriedades de uma Medida de 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

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.

No caso particular de $n = 3$, o princípio de inclusão-exclusão afirma que

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) \\ &- P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Prova: A prova é por indução matemática em n e será omitida. ∎

Exercícios

Exemplo

Professor Leônidas está tentando calcular a probabilidade $p = P(A)$ do evento A, e determinou que ela é uma raiz do seguinte polinômio de grau cinco:

$$(p-3)(p-3\sqrt{-1})(p+3\sqrt{-1})(p+0.3)(p-0.3) = 0.$$

Baseado nesta fato, qual é o valor de p?

Exemplo

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 $\mathcal{A}$.

Exemplo

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?

Exercícios

Exemplo

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?

Exercícios

Solução: Para determinar esta probabilidade, vamos utilizar a probabilidade clássica. 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 $0,97$.

Exercícios

Exemplo

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?

Exercícios

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

Exercícios

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:

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

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:

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

Exercícios

Exemplo

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

Exercícios

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. 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:

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

Exercícios

Exemplo

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?

Exercícios

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. 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

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

Exercícios

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 $1 - \frac{1}{e}$.

Exercícios

Exemplo

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

Exercícios

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