Probabilidade e Inferência

**2016**

Jorge LS Leao

UFRJ / COPPE / PEE

26/12/2016



Leão, Jorge Lopes de Souza

Probabilidade e inferência (Rio de Janeiro) 2016

III , 27 p, 29,7cm (COPPE/UFRJ)

Inclui referências bibliográficas.

I.COPPE/UFRJ II.Título

1. Probabilidade

2. Inferência estatística

3. Matemática

 2016, by Jorge L. de Souza Leão

Creative Commons

Attribution-NonCommercial-NoDerivatives

4.0 International (CC BY-NC-ND 4.0)

Prefácio

Estas notas foram escritas para servirem como resumo teórico de probabilidades para uma disciplina sobre inferência Bayesiana em grafos nos cursos de mestrado e doutorado em engenharia elétrica.

Buscamos nestas notas apresentar os elementos da teoria da probabilidade de uma forma autocontida, quase introdutória, porém com uma visão matemática. Fazemos aqui uma analogia com a apresentação do cálculo infinitesimal: primeiro apresentamos os fundamentos matemáticos do cálculo, como sequências, limites, derivadas e integrais, para somente depois, e bastante depois, apresentar aplicações como a dinâmica da mecânica, o eletromagnetismo, etc.

Obviamente, a teoria da probabilidade é útil, sobretudo, para modelar experimentos aleatórios, seja lá o que isto for, já que ainda não os definimos. Entretanto, procuramos aqui, inicialmente, não falar nada sobre experimentos, assim como não se fala nada sobre as equações de Maxwell quando se introduz o cálculo.

Além disto, o público alvo destas notas já teve uma primeira apresentação à teoria da probabilidade no ensino médio e pelo menos uma disciplina, verdadeiramente introdutória, sobre probabilidade e estatística no ciclo básico de engenharia e de algumas outras carreiras.

Por isto, achamos que não é necessário iniciar o texto dizendo novamente que o lançamento de um dado não viciado proporciona uma probabilidade de 1/6 para a obtenção de qualquer uma de suas faces.

Supomos contudo, que o leitor já tenha um conhecimento básico de lógica, teoria dos conjuntos e cálculo.

O material apresentado baseia-se em parte nas referências (C. Geiss, 2009) e (James, 2013). Se por um lado não podemos dizer que o conteúdo destas notas seja original, acreditamos que o roteiro da apresentação tenha aspectos originais.

Como referência mais avançada para probabilidades, indicamos (Billingsley, 1995).

JLSL

Rio de Janeiro, dezembro de 2016.

Sumário

[1. Sigma-álgebras e mensurabilidade 5](#_Toc470652803)

[2. Medida de probabilidade 9](#_Toc470652804)

[3. Variáveis aleatórias 12](#_Toc470652805)

[4. Independência e probabilidade condicional 14](#_Toc470652806)

[5. Algumas propriedades básicas 18](#_Toc470652807)

[6. Distribuições e densidade de probabilidade 19](#_Toc470652808)

[7. Exemplos de distribuições contínuas e discretas 20](#_Toc470652809)

[8. Funções de variáveis aleatórias 21](#_Toc470652810)

[9. Valor esperado e integração 22](#_Toc470652811)

[10. Teoremas do limite 23](#_Toc470652812)

[11. Atribuições de probabilidades 24](#_Toc470652813)

[12. Estimação de parâmetros 25](#_Toc470652814)

[13. Teste de hipóteses 26](#_Toc470652815)

[Bibliografia 27](#_Toc470652816)

# Sigma-álgebras e mensurabilidade

*At a purely formal level, one could call probability theory the study of measure spaces with total measure one,*

*but that would be like calling number theory the study of strings of digits which terminate.*

*-Terence Tao*

*O objetivo deste capítulo é dar um embasamento para, no segundo capítulo, definir probabilidade como uma medida.*

**Definição 1.1:**[ÁLGEBRA]

Seja Ω um conjunto não vazio. Um sistema de subconjuntos de Ω é uma álgebra

sse

1. ∅ e Ω ∈ ,
2. Se A ∈ então ,
3. Se A e B ∈ então .

Ω é chamado de espaço amostral. Os elementos ω ∈ Ω são chamados de resultados elementares ou estados.

***fim def.***

*Na literatura, o complemento de um subconjunto A é representado às vezes por , mas também por .*

**Exemplo:** Isomorfismo entre a álgebra dos subconjuntos e a álgebra da lógica sentencial.

**Exemplo:** A álgebra booleana de dois elementos, 0 e 1, usada em circuitos lógicos, dada por:

<{∅,{∅}},∩,∪,‾,∅,{∅}> fazendo-se as correspondências naturais com os símbolos usuais.

**Definição 1.2:**[SIGMA-ÁLGEBRA]

Seja Ω um conjunto não vazio. Um sistema de subconjuntos de Ω é uma σ-álgebra

sse

1. ∅ e Ω ∈ ,
2. Se então ,
3. Se A1, A2, A3, ... ∈ então .

Aqui interpreta-se .

O par (Ω , ), onde é uma σ-álgebra sobre Ω , é chamado de espaço mensurável.

Os conjuntos , que formam a σ-álgebra, no contexto de um espaço de probabilidades que será definido adiante, são chamados eventos (eventos aleatórios ou eventos admissíveis ou subconjuntos observáveis).

Obs: Dado Ω, em geral, a escolha de uma σ-álgebra sobre Ω não é única!

***fim def.***

**Exemplo:** é a menor σ-álgebra sobre Ω.

**Exemplo:** é a maior σ-álgebra sobre Ω.

**Exemplo:** , onde A ⊆ Ω , é uma σ-álgebra sobre Ω.

É fácil verificar que se Ω é finito (#Ω ∈ ) então qualquer álgebra sobre Ω é uma σ-álgebra sobre Ω .

**Exemplo:** Uma álgebra que não é uma σ-álgebra.

Suponha e .

Pode-se verificar que é uma álgebra, pois:

1. ,
2. Se então também é uma união finita de intervalos em ,
3. Se e então também é uma união finita de intervalos em .

Os conjuntos e o evento elementar (evento que só possui um resultado elementar), com , são interpretados como intervalos de comprimento zero e portanto, são elementos de .

Mas não é uma σ-álgebra, pois não contém toda união enumerável de intervalos em . Por exemplo, o evento:

=

é uma união enumerável de intervalos em , mas não é uma união finita e então .

**Proposição 1.1**: [INTERCESSÃO DE SIGMA-ÁLGEBRAS É UMA SIGMA-ÁLGEBRA]

Seja Ω um conjunto não vazio, seja J≠∅ um conjunto índice e , com j∈J, uma família de σ-álgebras sobre Ω.

Então, é uma σ-álgebra também.

***fim prop.***

Demonstração:

..................

***fim dem.***

*A proposição 1.1 é importante para justificar a construção das menores σ-álgebras e o conceito de σ-álgebra gerada por um conjunto de subconjuntos de um espaço amostral.*

**Proposição 1.2**: [MENOR SIGMA-ÁLGEBRA QUE CONTEM UM SISTEMA DE CONJUNTOS]

Seja Ω um conjunto não vazio e um sistema qualquer de subconjuntos de .

Então, existe uma σ-álgebra mínima sobre Ω, denotada por σ, tal que .

Diz-se que σ é a σ-álgebra gerada por .

***fim prop.***

Demonstração:

......................

***fim dem.***

**Definição 1.3:** [MEDIDA]

Seja (Ω , ) um espaço mensurável.

Uma função é chamada uma medida

sse

e

para todo , e

para todo , , tem-se

.

A trinca é chamada um espaço de medida.

***fim def.***

*O que fica evidente para esta definição de medida é que:*

1. *Ela é definida para medir os conjuntos que formam a σ-álgebra ,*
2. *O conjunto vazio tem medida zero,*
3. *A medida é sempre não negativa,*
4. *A medida de uma união é a soma das medidas das partes, inclusive para uniões infinitas, o que é caracterizado pela σ-álgebra subjacente.*
5. *A medida de um conjunto não precisa ser necessariamente finita.*

*Intuitivamente, é isto que se espera de uma medida.*

*Ainda de uma maneira intuitiva, os pontos isolados do espaço amostral, os resultados elementares, não são tão importantes para os efeitos da medida, porque a medida é definida para os conjuntos da σ-álgebra, os eventos. Como será visto mais adiante, é nestes conjuntos da σ-álgebra que existe a variabilidade, isto é, a eventual pertinência de vários resultados elementares. A medida desta variabilidade é o verdadeiro objetivo da teoria da probabilidade.*

*Nestas notas, está-se interessado sobretudo em medidas de probabilidade definidas sobre subconjuntos de . Para isto, definem-se as álgebras de Borel no .*

**Definição 1.4:**[SUBCONJUNTOS ABERTOS E FECHADOS DE ]

Um subconjunto é dito aberto

sse

para todo ,

para algum ,

.

Um subconjunto é dito fechado

sse

é aberto.

Dados ,

(a,b) designa um intervalo aberto de a até b,

[a,b] designa um intervalo fechado de a até b.

Postula-se que o conjunto vazio é aberto e fechado.

***fim def.***

**Definição 1.5:** [ÁLGEBRA DE BOREL NO]

As σ-álgebras geradas pelos conjuntos dados abaixo são chamadas σ-álgebras de Borel e são escritas como B() :

σ() , onde é o sistema de todos os subconjuntos abertos de ,

σ() , onde é o sistema de todos os subconjuntos fechados de ,

σ() , onde é o sistema de todos os intervalos ,

σ() , onde é o sistema de todos os intervalos ,

σ() , onde é o sistema de todos os intervalos ,

σ() , onde é o sistema de todos os intervalos .

***fim def.***

**Proposição 1.3:** [SIGMA ÁLGEBRAS DE BOREL GERADAS NO ]

As seguintes σ-álgebras geradas pelos conjuntos dados acima são iguais:

σ() = σ() = σ() = σ() = σ() = σ() = B()

***fim prop.***

Demonstração:

..................

***fim dem.***

*Também se pode definir σ-álgebras de Borel em espaços métricos outros além do .*

...

(pag. 13 Geiss)

***fim dem.***

**Exemplo:** Comparando a σ-álgebra de Borel.

Seja B a σ-álgebra gerada por todos os intervalos semiabertos em , isto é, gerada por conjuntos da forma (a,b], onde a<b e onde a e/ou b podem ser ∞. Todo subconjunto razoável de é um conjunto de Borel. Contudo, comparando a σ-álgebra de Borel em com outros conjuntos tem-se que:

- B é muito maior que o conjunto de todos os intervalos semiabertos em e

- B é muito menor que o conjunto potência .

(ver Billingsley, 1995).

- Produto de sigma-álgebras

# Medida de probabilidade

*Neste capítulo, define-se probabilidade como uma medida para conjuntos de resultados chamados eventos. As definições baseiam-se no conceito de σ-álgebra, que permite definir probabilidades, de uma maneira precisa, para espaços razoavelmente complexos.*

**Definição 2.1:** [ESPAÇO DE MEDIDA SIGMA-FINITO]

Um espaço de medida , ou a medida μ, é chamada σ-finita, ou simplesmente finita,

sse

para alguma família , com k=1,2,... , e

para todo k=1, 2, ... , , e

para todo , , tem-se

1. e
2. .

***fim def.***

**Definição 2.2:** [MEDIDA DE PROBABILIDADE]

Seja (Ω , ) um espaço mensurável.

A função é chamada uma medida de probabilidade

sse

e

para todo , e

para todo , , tem-se

.

***fim def.***

*Como é fácil verificar que , tem-se que P é uma medida.*

*Pode-se dizer, coloquialmente, que uma medida de probabilidade é uma medida finita com medida total igual a 1, ou normalizada para a qual .*

**Definição 2.3:** [ESPAÇO DE PROBABILIDADE]

Um espaço de probabilidade é uma trinca (Ω , , P), onde:

(1) Ω é um conjunto não vazio, chamado de espaço amostral.

(2) é uma σ-álgebra de subconjuntos Ai de Ω.

(3) P é uma medida de probabilidade em :

P : [0,1]

que dá uma probabilidade para todos , isto é, P(A) ∈ [0,1].

- Os elementos de são chamados eventos, eventos aleatórios ou subconjuntos observáveis.

- Dado um resultado elementar e um evento , diz-se que o evento acontece (para o resultado ) se . Se , diz-se que não acontece (para o resultado ).

- Novamente, dados Ω e , a escolha de uma medida de probabilidade P não é única.

***fim def.***

**Proposição 2.1:** [PROPRIEDADES ELEMENTARES]

Seja um espaço de probabilidade.

Então, tem-se:

1. Se , então

.

1. Se , então

.

1. Se , então

(eventos possivelmente não disjuntos).

***fim prop.***

Demonstração:

..................

***fim dem.***

*Deve-se agora detalhar melhor a nomenclatura que será usada com a medida de probabilidade.*

Seja um espaço de probabilidade.

Seja a função uma medida de probabilidade deste espaço.

*Quando se escreve P(A)=p, com p∈[0,1], o argumento A é um subconjunto de Ω que pertence à σ-álgebra .*

*Pode-se então utilizar expressões com subconjuntos de Ω e os operadores da teoria dos conjuntos, como a união, a intercessão, o complemento e a diferença.*

*Além disto, utiliza-se também a notação da lógica das proposições referindo-se à pertinência a conjuntos, como apresentado a seguir.*

Sejam A e B ∈ .

P(AB) P(A∧B) P(AB),

P(A,B) P(A∨B) P(AB),

P() P() P(),

P(A/B) P(A­B) .

*É fácil verificar que:*

*P(AB) = P(BA) ,*

*P(A,B) = P(B,A) ,*

*mas em geral, P(A/B) ≠ P(B/A) .*

**Proposição 2.2:** [CONTINUIDADE POR BAIXO]

Dados ,

se , então

.

***fim prop.***

Demonstração:

***fim dem.***

**Proposição 2.3:** [CONTINUIDADE POR CIMA]

Dados ,

se , então

.

***fim prop.***

Demonstração:

***fim dem.***

# Variáveis aleatórias

**Definição 4.1:** [FUNÇÃO DEGRÁU MENSURÁVEL]

Seja um espaço de probabilidade.

A função é chamada uma função degrau mensurável

sse

dados

e

,

,

onde

(impulso unitário de pertinência ao conjunto )

***fim def.***

Exemplos:

.

.

.

.

**Definição 4.2:** [VARIÁVEL ALEATÓRIA]

Seja um espaço de probabilidade.

A função é chamada uma variável aleatória

sse

para alguma sequência de funções degrau mensuráveis ,

para todo

.

***fim def.***

**Proposição 4.1:** [INVERSA DE UMA VARIÁVEL ALEATÓRIA]

Seja um espaço de probabilidade.

A função é uma variável aleatória

sse

para todos ,

.

Isto é, a imagem inversa de intervalos (abertos) em são eventos de .

***fim prop.***

Demonstração:

...

***fim dem.***

*Isto é, a imagem de uma variável aleatória é uma sigma-álgebra (de Borel) tal que a imagem inversa de intervalos abertos (eventos da álgebra de Borel) são eventos de (mas não todos?).*

*Pode-se ver que nem toda variável aleatória mantém a estrutura de .*

- Tipos de variáveis aleatórias:

* - contínuas,
* - discretas,
* - mistas. (ver https://www.probabilitycourse.com/chapter4/4\_3\_2\_delta\_function.php)

- PDF: Probability Density Function

- PMF: Probability Mass Function -> When discrete: Probability Distribution

- CDF: Cumulative Distribution Function: contínuas, discretas ou mistas!

- Vetores de variáveis aleatórias, probabilidade conjunta

- Produto de espaços de probabilidade

- Álgebras de Borel no Rn

# Independência e probabilidade condicional

**Definição 3.1:** [EVENTOS INDEPENDENTES]

Seja um espaço de probabilidade.

Seja uma família indexada de subconjuntos da σ-álgebra , com , onde I é um conjunto índice arbitrário, não vazio.

Os eventos Ai são independentes

sse

para todo i1, ... , in ∈ I,

Observação: A família indexada de subconjuntos pode ser infinita.

***fim def.***

*É importante observar que não se pode definir independência de eventos dizendo que se exige simplesmente que a probabilidade da intercessão de eventos independentes é o produto das probabilidades destes eventos. É preciso exigir esta igualdade para todos os valores de n.*

*Por exemplo, dados A e B que não sejam independentes, tem-se:*

*.*

*Considerando-se mais um evento C=∅, ter-se-ia:*

*,*

*o que não serviria para garantir a independência dos três eventos.*

*Resumindo, é preciso exigir a regra do produto dois-a-dois (n=2), três-a-três (n=3), etc.*

**Definição 3.2:** [PROBABILIDADE CONDICIONAL]

Seja um espaço de probabilidade.

Suponha A, B , com P(A)>0.

Diz-se que P(B|A) é a probabilidade condicional de B dado A

sse

***fim def.***

*Primeiramente, pode-se verificar que a probabilidade condicional é de fato uma medida de probabilidade.*

*Pode-se também dizer, de maneira a permitir uma compreensão mais intuitiva, que P(B|A) é a probabilidade condicional de B dado A como espaço amostral de B.*

**Proposição 3.1:** [INDEPENDÊNCIA DE DOIS EVENTOS]

A e B são independentes

sse

P(B|A) = P(B)

***fim prop.***

Demonstração:

( Temos que ,

mas se A e B são independentes, ,

então = = P(B).

(Temos que P(B|A) = P(B), então = P(B) ,

daí ,

e pela definição, A e B são independentes.

***fim dem.***

**Proposição 3.2:** [PROBABILIDADE TOTAL]

Suponha que

1. A , com P(A)>0

(2) Bk , com P(Bk)>0, para k=1, 2, ... , n

(3) com i e j = 1, 2, ... , n

para todo i ≠ j ,

(mutuamente exclusivos)

(4) (exaustivos)

Então

Pode-se observar que os eventos Bk formam uma partição de eventos mutuamente exclusivos de Ω.

***fim prop.***

Demonstração:

..............

***fim dem.***

**Proposição 3.3:** [PROBABILIDADE TOTAL 2]

Suponha que

1. A, C , com P(A) e P(C)>0

(2) Bk , com P(Bk)>0, para k=1, 2, ... , n

(3) com i e j = 1, 2, ... , n

para todo i ≠ j ,

(mutuamente exclusivos)

(4) (exaustivos)

Então

Pode-se observar que os eventos Bk formam uma partição de eventos mutuamente exclusivos de Ω.

***fim prop.***

Demonstração:

..............

***fim dem.***

**Definição 3.3:** [INDEPENDÊNCIA CONDICIONAL]

sse

P(A|B,C)=P(A|C)

***fim def.***

**Proposição 3.4:** [EQUAÇÃO DE BAYES]

Suponha que

1. A , com P(A)>0

(2) Bk , com P(Bk)>0, para k=1, 2, ... , n

(3) com i e j = 1, 2, ... , n

para todo i ≠ j ,

(4) .

Então

Os eventos Bj são chamados hipóteses, as probabilidades P(Bj) são chamadas as probabilidades *a priori* de Bj e as probabilidades P(Bj|A) são chamadas as probabilidades *a posteriori* de Bj.

***fim prop.***

Demonstração:

.............. Geiss deixa a demonstração como um exercício...

***fim dem.***

**Corolário 3.5:** [EQUAÇÃO DE BAYES SIMPLIFICADA]

Suponha que A,B , com P(A)>0 e P(B)>0.

Então

***fim prop.***

Demonstração:

Esta proposição decorre diretamente das proposições 3.2 e 3.3 acima.

***fim dem.***

*Neste ponto, é importante observar que, como disseram* (W. von Linden, 2014)*, toda a probabilidade é uma probabilidade condicional. Neste sentido, poder-se-ia escrever P(A|I), onde I é o conjunto de todas as “informações subjacentes” (background information).*

**Proposição 3.6:** [REGRA DA CADEIA] Também chamada Regra da Multiplicação ou da Fatorização da Probabilidade Conjunta.

Demonstração:

...

***fim dem.***

Marginalização

# Algumas propriedades básicas

- Lema de Fatou

- Lema de Borel-Cantelli

- Teorema da extensão de Carathéodory

- Medida de Lebesgue

# Distribuições e densidade de probabilidade

- distribuições e densidades

- Sistema PI

?????? Colocar aonde?

**Definição 6.4:** [MEDIDA DE DIRAC]

Seja um espaço de medida, com e .

A medida

é chamada uma medida de Dirac.

***fim def.***

*A medida de Dirac é uma medida de probabilidade e será usada em espaços discretos.*

**Definição 6.6:** [MEDIDA DE CONTAGEM]

Seja e .

A medida

é chamada uma medida de contagem.

***fim def.***

*A medida de contagem não é uma medida de probabilidade, mas será usada em espaços discretos para calcular uma medida de probabilidade.*

# Exemplos de distribuições contínuas e discretas

**Experimentos aleatórios:**

Os conceitos de espaço amostral, resultados elementares, eventos aleatórios, medida de probabilidade e variável aleatória foram apresentados como conceitos matemáticos abstratos, embora seus nomes já evidenciem sua origem histórica e certamente sua principal aplicação.

Estes conceitos matemáticos podem ser usados para construir modelos para processos chamados experimentos aleatórios.

Estes experimentos podem ser de um tipo que admita um grande número de repetições e cujos resultados apresentem uma variabilidade para a qual não se tem um modelo determinístico conhecido, mas para o qual se supõe algum tipo de regularidade.

Os experimentos também podem ser de outro tipo, para o qual não é possível fazer nenhuma repetição, mas para o qual se pode justificar algum tipo de crença sobre o resultado, crença esta medida por uma função de probabilidade.

Como é usual nestes casos, modelar um processo do mundo real (ou virtual!) é uma arte, que usualmente representa um compromisso entre a exatidão, o detalhe e a economia de recursos técnico-matemáticos e computacionais.

Além disto, estes conceitos matemáticos podem ser usados em algoritmos para calcular resultados absolutamente determinísticos, como o cálculo de π com uma precisão especificada.

Neste texto, portanto, procura-se sempre deixar clara a separação entre os aspectos teóricos de probabilidades e a sua aplicação à modelagem de processos, experimentos e algoritmos.

# Funções de variáveis aleatórias

- funções geradoras de momentos,

- funções características

# Valor esperado e integração

# Teoremas do limite

# Atribuições de probabilidades

# Estimação de parâmetros

# Teste de hipóteses

# Bibliografia

**Billingsley, Patrick. 1995.** *Probability and measure.* Third edition. New York : John Wiley and Sons, Inc, 1995.

**C. Geiss, S. Geiss. 2009.** *An introduction to probability theory.* Jyvaskyla : Department of Mathematics and Statistics, University of Jyvaskyla, 2009.

**James, Barry R. 2013.** *Probabilidade: um curso em nível intermediário.* Terceira edição. Rio de Janeiro : IMPA (Coleção projeto Euclides), 2013.

**W. von Linden, V. Dose, U. von Toussaint. 2014.** *Bayesian Probability Theory: applications in the physical sciences.* Cambridge : Cambridge University Press, 2014.