Inferência Estatística com Abordagem Bayesiana

Rosangela Helena Loschi [1]

[1] Departamento de Estatística
Universidade Federal de Minas Gerais

14 de outubro de 2021

Tópicos a serem abordados

- Paradigmas Clássico e Bayesiano: os fundamentos
- Modelagem dentro da ótica Bayesiana
- Métodos de construção de dist. *a priori*
  - distribuições conjugadas
  - distribuições não informativas
- Inferência e Decisão
  - funções de perda e princípios para tomada de decisão.
  - Estimações pontual e intervalar
  - testes de Hipóteses: Fator de Bayes e Teste de Significância Bayesiano completo.
- Métodos Computacionais
  - SIR e método da Rejeição
  - MCMC: amostrador de Gibbs e Metropolis-Hastings

Referências Bibliográficas

- Statistical Inference: An Integrated Approach (2015) H. S. Migon, D. Gamerman e F. Louzada, 2a. edição, CRC Press.
- Estatística Bayesiana (2018), B. Murteira, C. D. Paulino, M. A. Amaral Turkman, G. L. Silva, 2ª edição.
- Bayesian Statistics: An Introduction.(2012) Peter M. Lee, 4ª Edição, Wiley.

Introdução

- Estatística é a ciência que resolve problemas onde a aleatoriedade está presente.
- Faz uso de **modelos probabilísticos** para resumir a **informação** relevante.
- **Problema fundamental abordado pela Estatística:** De posse de dados observados, deseja-se fazer afirmações e inferências sobre o sistema que gerou os dados.
- Derivar a inferência sobre parâmetros e modelos a partir dos dados é um **problema de indução**.

"Pode o passado informar sobre o futuro?"(David Hume)

- O problema de indução é um probema controveso. Mesmo na Filosofia, existem várias propostas de solução para ele.
- **Inferência Estatística**: é uma resposta científica para o problema de indução.

Introdução

- Existem várias Escolas de pensamentos sobre inferência Estatística: Fiducial, Verossimilhancista, Clássica ou frequencista (Fisher, Pearson, Neyman), Bayesiana (De Finetti, Savage).
- Cada escola de pensamento tem seus próprios princípios e métodos para fazer Inferência Estatística (Berger, 1984).
- Vamos discutir os princípios de duas destas escolas:
  - Clássica ou frequencista
  - Bayesiana

Paradigma Clássico

Objetivo da inferência se formula:

- que afirmações ou generalizações podem ser feitas sobre a população com base na informação trazida pela amostra colhida nesta mesma população?
- Em geral, sobre parâmetros θ que são resumos de alguma característica populacional.
- Os parâmetros θ: São **fixos** e **desconhecidos**

Informação (amostra):

- São dados estatísticos, resultantes de experimentos que são realizados para obtermos informação sobre o parâmetro.
- Os experimentos podem ser repetidos e, se repetidos em condições idênticas, podem fornecer resultados diferentes.
- Os dados observados $\boldsymbol{X}$ são **apenas um de muitos** - possivelmente infinitos - conjuntos de dados que poderiam ter sido observados nas mesmas circunstâncias.

Paradigma Clássico

- Aspectos importantes da inferência clássica:
  - **reconhecer a variabilidade que se verifica de amostra para amostra.**
- A interpretação dos resultados/inferência depende:
  - Dos dados observados;
  - E tão importante quanto: **Das premissas ou hipóteses feitas sobre o processo que gera os dados observáveis.**
- Os dados são tratados como uma realização de uma variável aleatória ou um vetor aleatório $\boldsymbol{X}$ com uma distribuição F_θ, que não é totalmente conhecida.
- F_θ é membro de uma família de distribuições $\mathcal{F}$ que é o modelo estatístico para $\boldsymbol{X}$.
- A especificação do modelo F_θ é uma parte essencial do processo de inferência.

Paradigma Clássico

A especificação do modelo F_θ é feita com uma síntese de vários fatores:

- considerações teóricas sobre a natureza do fenômeno estudo (EX: Características sobre o comportamento do tempo de vida do ser humano.)
- considerações teóricas sobre as técnicas experimentais (EX: lanço uma moeda, independentemente, um número fixo de vezes e conto o número de sucessos observados.)
- evidência nos dados amostrais;
- busca de parcimônia.
    - Escolher modelos mais simples com parâmetros interpretáveis para viabilizar o processo inferencial.

Paradigma Clássico

Construído o modelo F_θ, são desenvolvidos métodos para respoder peguntas tais como:

- Os dados são compatíveis com a família $\mathcal{F}$?
- Dado que a especificação de $\mathcal{F}$ é correta, que conclusões podemos tirar sobre o parâmetro θ_0 que indexa a distribuição F_θ que "apropriadamente" descreve o fenomeno em estudo?

Paradigma Clássico

Os métodos clássicos são avaliados sob o princípio da amostragem repetida. Isto é,

- As inferências, decisões e avaliações são feitas com base em procedimentos que pressupõem **infinitas repetições hipotéticas do experimento realizado em condições idênticas**.
- Não apenas a amostra de fato observada é considerada.

Um dos aspectos importantes deste princípio:

- frequências são usadas como medida de incertezas.
- Daí, nesta escola, a **probabilidade** de ocorrência de um evento é interpretada como a **frequência relativa** de sua ocorrência.

Paradigma Clássico

Exemplo: Suponha que queiramos estimar a média populacional μ com base em uma amostra de tamanho n, $X_1, \ldots, X_n$.

- Seja $\bar{X}_n = (X_1 + \cdots + X_n)/n$ o estimador escolhido.
- Se $n = 3$ e a amostra é $x_1 = 61, x_2 = 60$ e $x_3 = 62$, a estimativa pontual é $\bar{x}_3 = 61 = \hat{\mu}$
- **a estimativa pontual depende apenas da amostra observada $\Rightarrow$ Princípio da repetibilidade amostral não é usado.**

Paradigma Clássico

- No entanto, se desejamos verificar as qualidades do estimador $\bar{X}_n$, por ex., se queremos avaliar se é não-viciado, temos:
  - Determinar $E(\bar{X}_n) = \int z f_{\bar{X}_n}(z) dz$;
  - Precisamos encontrar a Distribuição de $\bar{X}_n$ (captar a variabilidade em $\bar{X}_n$ induzida por todas as amostras possíveis);
  - **Assim, utilizamos o princípio da repetibilidade amostral.**

- Se F_θ é uma $N(\mu, \sigma^2)$ e se $X_1, \ldots, X_n$ é uma amostra independente e identicamente distribuída (iid)

$$\bar{X}_n \sim N(\mu, \sigma^2/n) \Rightarrow \bar{X}_n \text{ é não-viciado}$$

Paradigma Clássico

Exemplo: Suponha que queiramos estimar a média populacional μ com base em uma amostra de tamanho n, $X_1, \ldots, X_n$ e com $\gamma 100\%$ de confiança.

- Usamos o **Princípio da Repetibilidade Amostral** para **construir** o Intervalo de Confiança (IC).
- $\bar{X}_n$ é a estatística pivotal.
- Se F_θ é uma $N(\mu, \sigma^2)$ e se $X_1, \ldots, X_n$ é uma amostra iid

  $\Rightarrow \bar{X}_n \sim N(\mu, \sigma^2/n) \leftarrow$ fruto do princípio da repetibilidade amostral

- Pré-experimentalmente, o IC é:

$$P\left(\bar{X}_n - \ Z_\gamma \sigma/\sqrt{(}n) < \mu < \bar{X}_n + \ Z_\gamma \sigma/\sqrt{(}n)\right) = \gamma. \qquad (1)$$

- Os limites do IC são aleatórios.
- γ é a probabilidade de cobertura do valor μ pelo IC.

Paradigma Clássico

- Pós-experimentalmente, a probabilidade em (1) perde o sentido pois os limites do IC são números:

$$[\bar{x}_n - \ Z_\gamma \sigma / \sqrt(n); \bar{x}_n + \ Z_\gamma \sigma / \sqrt(n)]$$

- O **Princípio da Repetibilidade Amostral** é também usado para **interpretar** Intervalo de Confiança (IC).
- se μ fosse conhecido faria sentido falar na seguinte

$$P\left(\bar{x}_n - \ Z_\gamma \sigma / \sqrt(n) < \mu < \bar{x}_n + \ Z_\gamma \sigma / \sqrt(n)\right) = 1$$

  se $\mu \in IC$ e é zero, caso contrário.
- Como μ é desconhecido, γ passa a ter o seguinte sigunificado: Repete-se a seleção de amostras de tamanho n infinitas vezes, calcula-se o intervalo com confiança $\gamma 100\%$ com cada uma delas. $\gamma 100\%$ dos IC contruídos conterão o verdadeiro valor de μ.

Paradigma Clássico

Exemplo: Suponha que queiramos testar se a média populacional μ é maior ou igual a um valor conhecido μ_0 com base em uma amostra de tamanho n, $X_1, \ldots, X_n$ e com nível de significância γ.

- A estatística de teste é $\bar{X}_n$.
- O Valor P, dado por

$$P - Value = P(\bar{X}_n > \bar{x}_{obs} \mid \mu = \mu_0).$$

- O Valor P se baseia no **Princípio da Repetibilidade Amostral** pois é dado pela proporção de amostras que, se obtidas em condições identicas, fornecem valores para a estatística de teste tão extremos quanto o observado, se a média populacional real for μ_0.

Paradigma Clássico

Moral da História:

A amostra observada é utilizada e é importante no processo de inferência

MAS não apenas ela.

Amostras que poderiam ter sido observadas mas não foram também são importantes neste processo.

Perguntas

Qual é a probabilidade de que a altura média da população brasileira seja superior a 185cm?

Perguntas

Um Estatístico Clássico deveria responder:

- não faz sentido falar em probabilidade neste caso pois a média populacional é um parâmetro;
- A média populacional é um objeto fixo, não há uma variabilidade intrinsica.
- Não tenho como usar o princípio da repetibilidade amostral!

Perguntas

No entanto, damos uma resposta probabilística a esta pergunta. Por que?

Perguntas

R: Por que é natural traduzirmos através de uma probabilidade nossa incerteza sobre objetos desconhecidos.

Perguntas

Um Estatístico Bayesiano responderia:

- a média populacional é um parâmetro;
- é fixo e desconhecido para mim.
- Baseado no que sei sobre a propulação, $P(\mu > 185) < 0,0025$.

Perguntas

Quando falamos a frase: "A probabilidade de dar cara no lançamento de uma moeda é 1/2."De quem é esta probabilidade? É minha ou é da moeda?

Em qualquer dos mundos, PROBABILIDADE é uma medida de incerteza.

Paradigma Bayesiano

- **Objetivo da Inferência:** usar a informação disponível para reduzir a incerteza sobre um objeto em estudo (por ex., um parâmetro θ).
- A Escola Bayesiana segue a **Postura Subjetivista**.
- **Postura Subjetivista**: É uma maneira de pensar inferência estatística onde se assume que a incerteza sobre TUDO o que é **desconhecido** deve ser traduzida por uma medida de probabilidade.
- **Aleatoriedade** (o que deve ser probabilizado) está relacionado ao **Desconhecimento** e não à variabilidade inerente ao objeto.
- Por isto faz sentido falar de probabilidade para um parâmetro.
- **O tratamento dado a um parâmetro é o mesmo dado a uma variável aleatória não-observável.**

Paradigma Bayesiano

Assim, com um modelo probabilístico, se pode descrever a sua incerteza sobre

- $\boldsymbol{X}$ que são variáveis ou vetores aleatórios: pois não sabemos antecipar seu resultado antes de a observarmos e seu resultado varia de objeto para objeto.
- θ é um objeto desconhecido. É um objeto fixo MAS não sabemos antecipar seu valor. Não há varibilidade em θ.

Paradigma Bayesiano

A interpretação de Probabilidade no mundo Bayesiano

- A probabilidade é **Subjetiva**.
- É uma opinião pessoal, não um atributo físico do objeto.
- A probabilidade de um evento descreve a SUA incerteza sobre a ocorrência de tal evento.
- A probabilidade não é uma propriedade do objeto em estudo.
- "**Probabilidade não existe.**´´(De Finetti): no sentido de que não é um atributo físico, uma característica de um objeto.
- É uma medida da incerteza individual construída com base no conhecimento prévio que se tem sobre o fenomeno em estudo.

Paradigma Bayesiano

Exemplo: Se desejamos determinar a probabilidade de $A =$ ocorrer cara no lançamento de uma moeda.

- **No Mundo Clássico:** $p(A) = \lim_{n \to \infty} \frac{N.Caras}{n}$
- Faz uso do Princípio da Amostragem Repetida.
- Pressupõe: (i) experimento pode ser realizado infinitas vezes em condições identicas; (ii) experimentos são realizados independentemente. (O QUE É INDEPENDÊNCIA??)
- **A probabilidade é uma propriedade da moeda**

Paradigma Bayesiano

- **No Mundo Bayesiano:** $p(A)$ é um número que reflete o Seu conhecimento sobre a moeda.
- A única realização do experimento que importa é a que será feita.
- Não se admite a possibilidade de repetibilidade amostral
- Se $P(A) = 0,5$ isto pode refletir duas coisas:
  - (i) que, para você, a moeda é honesta;
  - (ii) que você nada sabe sobre a moeda.

IMPORTANTE: Apesar de ser uma escolha subjetiva, suas probabilidades devem seguir as Leis de Probabilidade.

“ Se você É COERENTE, sua medida de incerteza é uma medida de probabilidade.´´ (De Finetti, 1931, 1937).

Paradigma Bayesiano

Sobre a informação disponível para se realizar inferências:

- Informação tem um papel vital em inferência, melhor dizendo, em Estatística!
- Temos duas fontes de informação:
  - a amostral, que vem do experimento realizado.
  - a que que vem de sua experiência ou conhecimento prévios ou de outras fontes de informação.

Em Inferência:
Faz sentido utilizar toda a informação disponível, ou somente a informação amostral é relevante?

Paradigma Bayesiano

Se somente a informação amostral é relevante para a inferência e tomadas de decisão, por que é que quando apresentamos certos sintomas procuramos um médico especialista e não um recém formado?

- Ambos terão acesso aos mesmos dados amostrais (os sintomas, os resultados de exames, etc.) para expressarem seus diagnósticos!

Paradigma Bayesiano

Você deseja contratar uma pessoa que tenha grande capacidade de “acertar os resultados"de provas a que se submetem, independentemente do tema. Se apresentam para a competição

- um especialista em música que diz ser capaz de diferir entre as músicas de Haydn e Mozart (São músicos do mesmo período e é uma tarefa bem difícil distinguir entre um e outro).
- Uma senhora inglesa que diz ser capaz de acertar a ordem em que o seu chá é servido, antes ou após o leite ter sido servido.
- Um bebado que diz ser capaz de acertar os resultados no lançamento de uma moeda.

Qual é a probabilidade de Você contratar cada um destes candidatos? Antes de ver os resultados, quem voce contrataria?

Paradigma Bayesiano

Se ambos são submetidos a dez provas e acertam todas elas

- Sua inferência/decisão baseada apenas nos dados observados é a mesma.
- Esta é uma decisão razoável?

Paradigma Bayesiano

- Em Estatística Bayesiana **ambas** as fontes de informação - **a amostral e a vinda de experiências prévias**- são importantes para as inferências e decisões.
- A informação dos dados (amostral) é resumida pela **função de verosimilhança** $\leftarrow f(\boldsymbol{x} \mid \theta)$
- A informação inicial sobre θ, vinda de experiências prévias ou outras fontes de informação, é resumida pela **distribuição** *a priori* $\leftarrow \pi(\theta)$.
- A melhor e mais completa inferência sobre θ é a **distribuição** *a posteriori* $\leftarrow \pi(\theta \mid \boldsymbol{x})$.

Paradigma Bayesiano

- A distribuição *a posteriori* é uma **atualização** da distribuição *a priori*.
- Tal atualização é feita misturando-se a informação sobre θ trazida pela amostra $\boldsymbol{x}$ com sua informação *a priori* sobre θ.
- O **Teorema de Bayes** é a ferramenta mais comumente utilizada para realizar esta atualização

$$\pi(\theta \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{\int_{\Theta} f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta}$$

- A distribuição *a posteriori* é calculada após observarmos a amostra $\boldsymbol{x}$.
- **IMPORTANTE:** Relacionado ao experimento aleatório, **apenas a amostra observada** é relevante para o processo de inferência. Amostras que poderiam ter sido observadas mas não foram não são consideradas nos procedimentos inferenciais.

Paradigma Bayesiano *versus* Paradigma Clássico

Para um Bayesiano

- Probabilidade é uma opinião subjetiva;
- Tanto a informação amostral quanto a não amostral são relevantes para a inferência
- Apenas a amostra observada é utilizada no processo inferencial.

Para um Clássico

- Probabilidade é uma frequência relativa;
- A informação não-amostral não é relevante para a inferência;
- Tanto a amostra observada, quanto todas as amostras que poderiam ter sido observadas mas não foram são relevantes em inferência.