Introdução à Inferência Baysiana

James D Santos

2023-10-03

Table of contents

Preface

Este material foi criado para a disciplina Introdução à Inferência Bayesiana, do curso de bacharelado em Estatística, da Universidade Federal do Amazonas.

1

2 Aula 1 - Introdução

Os objetivos desta aula são:

1. Apresentar a notação
2. Explicar sobre as fontes de informação
3. Apresentar as inferências básicas
4. Discutir como se dá o processo de elicitação de prioris

2.1 Notação

Variáveis aleatórias cujo valor podem ser observados serão denotadas por letras maiúsculas. Exemplos:

- X é número de acidentes diários na Avenida Torquato Tapajós
- Y é o nível máximo diário do Rio Negro

Valores observados de variáveis aleatórias serão denotados pela respectiva letra minúscula.

Parâmetros serão considerados aleatórios, mas serão representados por letras gregas minúsculas, como θ, λ, etc.

Vetores aleatórios serão representados por letras em negrito. Exemplos:

- $\mathbf{X} = \{X_1, \dots, X_n\}$ é um vetor de variáveis aleatórias.
- $\mathbf{x} = \{x_1, \dots, x_n\}$ é um vetor observado de variáveis aleatórias.
- $\theta = \{\alpha, \beta\}$ é um vetor de parâmetros.

Definition 2.1. O suporte de uma variável aleatória é o conjunto de todos os seus possíveis valores. Quando necessário, o suporte de variáveis aleatórias são representada pela versão caligráfica de sua letra correspondente.

Exemplos: o suporte de X é $\mathcal{X}$; o suporte de Y é $\mathcal{Y}$; o suporte de Z é $\mathcal{Z}$.

Definition 2.2. O espaço paramétrico é o conjunto de todos os possíveis valores do parâmetro. Eles são representados pela versão maiúscula da letra grega utilizada para seu respectivo parâmetro.

Exemplo: o espaço paramétrico do parâmetro θ é representado por Θ.

Tanto a função de densidade quanto a de probabilidade serão denotadas por funções começando com letras minúsculas. Por exemplo,

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

onde $x, \lambda > 0$ é a densidade da distribuição exponencial, enquanto que

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

com $x \in \mathbb{N}$ e $\lambda > 0$ é a é a função de probabilidade da distribuição Poisson.

2.2 Fontes de informação

2.2.1 A função de verossimilhança

Seja $\mathbf{x} = \{x_1, \dots, x_n\}$ uma amostra observada. Supomos que $\mathbf{x}$ é uma das possíveis amostras das variáveis aleatórias $\mathbf{X} = \{X_1, \dots, X_n\}$. Supomos ainda que $X \sim F(.|\theta)$. Assim, condicionada ao conhecimento de θ, a distribuição da amostra está completamente especificada.

::: {#def-Funcao de verossimilhanca} Para $\mathbf{x}$ fixado, a função

$$L : \Theta \Rightarrow [0, \infty)$$

é denominada verossimilhança. :::

Sua interpretação é a seguinte: para $\theta_1, \theta_2 \in \Theta$, se

$$L(\theta_1) > L(\theta_2),$$

dizemos que θ_1 é mais verossímil que θ_2. Isto porque a probabilidade de observar uma amostra na vizinhança de $\mathbf{x}$ é maior se considerarmos que θ_1 é o valor do parâmetro. A verossimilhança é uma das fontes de informação utilizada na inferência bayesiana (e a única fonte da inferência frequentista).

Example 2.1. Considere uma amostra de 30 variáveis aleatórias independentes com distribuição Poisson(θ). Neste caso, a função de verossimilhança será

$$L(\theta) = \frac{e^{-30\theta}\theta^{\sum_{i=1}^{30} x_i}}{\prod_{i=1}^{30} x_i!}.$$

Vamos simular uma amostra de tamanho 30 deste modelo (o valor de θ será omisso de propósito)

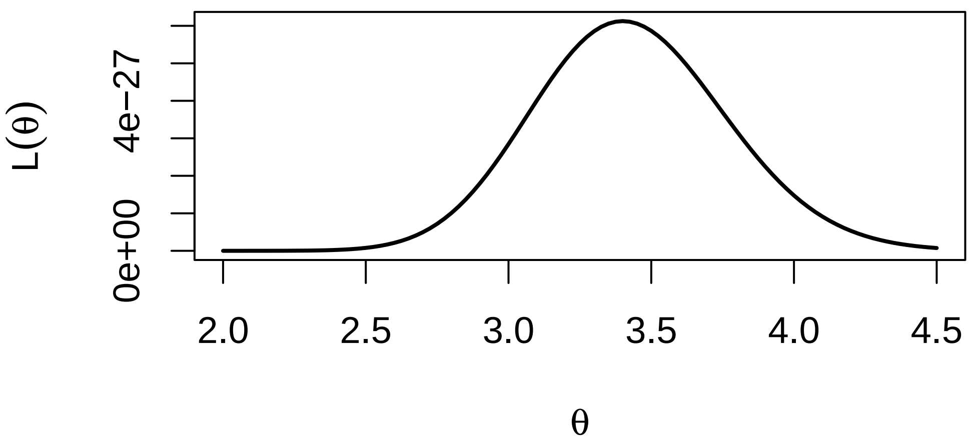
```
x <- rpois(30,theta)
x
```

```
[1] 2 4 2 5 6 0 3 5 3 3 6 3 4 3 1 5 2 0 2 6 5 4 3 8 4 4 3 3 2 1
```

A próxima figura mostra os valores da função de verossimilhança para vários valores de θ para a amostra observada.

```
# função de verossimilhança
vero <- function(q){
  sapply ( q, function(q) prod(dpois(x, q)) )
}

# gráfico da função de verossimilhança
oo <- par( cex = 1.2)
curve( vero(x),2,4.5, xlab = expression(theta), ylab = expression( L(theta)) , lwd = 2)
```




```
par(oo)
```

Podemos notar que o valores mais verossímeis para θ estão entre 2 e 4. Podemos ainda procurar o valor mais verossímil, denominado **estimativa de máxima verossimilhança (emv).** Pode-se mostrar, utilizando cálculo diferencial, que este valor é equivalente à média amostral. Contudo, com o objetivo de utilizar ao máximo o poder computacional que temos disponível, vamos encontrar esse valor utilizando a função `optimize`.

```
# menos o logaritmo da função de verossimilhança
lvero <- function(q) -log( vero(q))

# encontrando a emv:
optimise(lvero, c(2,4))
```

```
$minimum
[1] 3.399981

$objective
[1] 60.35731
```

O valor 2,6 é a estimativa de verossimilhança (este é o valor exato da média amotral). Sob o ponto de vista frequentista, esta seria a nossa estiamtiva para o valor de θ.

2.2.2 A distribuição a priori

Sob o ponto de vista bayesiano, a informação existente sobre θ antes da observação da amostra deve ser levada em consideração. Isto é feito traduzindo tal informação em termos de probabilidades.

Definition 2.3. A distribuição de θ é denominada **distribuição a priori**.

Os parâmetros da distribuição **a priori** são denominados hiper parâmetros.

As distribuições a priori : * agregam o conhecimento sobre parâmetro antes da observação da amostra (tal conhecimento pode ter sido gerado de uma amostra prévia). * podem ser muito ou pouco informativas, dependendo do grau de crença sobre os valores em particular do espaço paramétrico. Em geral isto é feito alterando a variância da distribuição:

$$\text{variância} = \frac{1}{\text{precisão}}$$

2.2.3 Reunindo as fontes de informação - distribuição a posteriori

Sejam $f(\theta)$ a densidade/função *a priori* para θ e $L(\theta)$ a função de verossimilhança.

Como θ é considerado aleatório, podemos analisar sua distribuição **após** observar a amostra x, ou seja

$$\theta|x.$$

Esta distribuição é denominada *posteriori*

::: {#thm-Teorema de Bayes}

Seja x uma amostra observada. Considere a priori $\theta \sim f(\theta)$ e a verossimilhança $L(\theta)$. Então a função de densidade (ou probabilidade) *a posteriori* de $\theta|x$ é dada por

$$f(\theta|x) = \frac{L(\theta)f(\theta)}{f(x)}.$$

O denominador é denominado distribuição preditiva, sendo igual a

$$f(x) = \sum_{\theta \in \Theta} L(\theta)f(\theta),$$

se θ é v.a. discreta ou

$$f(x) = \int_{\Theta} L(\theta)f(\theta)d\theta$$

se θ é v.a. contínua. :::

2.3 Inferência estatística

Denominamos por estatística qualquer função da amostra. Utilizamos estatísticas para fazer as seguintes inferências:

- *Estimação pontual*: trata-se de uma estatística com o objetivo de inferir o valor de θ. Tal estatística é denominada **estimador**.
- *Estimação por região*: trata-se de uma estatística, digamos $T(X)$, com o objetivo de cobrir o valor de θ, ou seja, fazer a inferência $\theta \in T(X)$. As estimações intervalares são as mais comuns, nas quais $T(X) = (L(X), U(X))$.
- *Testes de hipóteses*: são estatísticas construídas para decidir se aceitamos a afirmação (hipótese) $H : \theta \in \Theta_0$, onde Θ_0 é um subconjunto de Θ conhecido por hipótese.

Note que a distribuição **a posteriori** é função da amostra. Assim, toda função desta distribuição é uma estatística. Assim:

- *Estimação pontual*: em geral é uma medida que representa a região de alta densidade (ou probabilidade) da **posteriori**. A média da posteriori, assim como a mediana ou a moda são escolhas comuns.
- *Estimação por regiões*: em geral procuramos por uma região T da posteriori que satisfaça $P(\theta \in T(x)|x) = \gamma$, onde γ é denominado nível de credibilidade (não confundir com nível de confiança)
- *Testes de hipóteses*: em geral, aceitamos $H : \theta \in \Theta_0$ se $P(H|x)$ é elevada.

2.4 Elicitação de prioris

Note que o Teorema de Bayes deve unir as duas fontes de informação em uma nova fonte, sumarizada pela **posteriori**.

Não é incomum escolhermos certas prioris que trazem pouca informação, de modo que a moda **a posteriori** estará concentrada ao redor do estimador de máxima verossimilhança. Contudo, se há informações disponíveis, é recomendável gastar um tempo imaginando como traduzir estas informações em probabilidades. Nestes casos, diremos que a **priori** será elicitada.

O modo mais usual para a elicitação é encontrar os hiperparâmetros através de estatísticas da informação prévia.

Considere o problema de estimar o número médio de pessoas infectadas pelo vírus da AIDS em 2020 (os dados atuais são de 2018).

O nosso parâmetro de interesse é a média $\theta > 0$.

Abaixo, temos a informação do número de casos registrados nos últimos 10 anos:

```
library(knitr)
Ano <- 2009:2018
Total <-    c(40818,40409, 42355,42086 ,42934,41746,40506,38924,37999,37161)
kable(data.frame(Ano,Total))
```

Ano	Total
2009	40818
2010	40409
2011	42355
2012	42086
2013	42934
2014	41746
2015	40506
2016	38924
2017	37999
2018	37161

Podemos considerar a informação anterior para construir nossa **priori**. A média deve oscilar em torno do total de cada ano. Uma técnica simples e bastante útil é elicitar uma **priori** que tenha a mesma média e desvio padrão da informação disponível. Temos que

- Médias dos anos anteriores: 40.493,09
- Desvio padrão dos anos anteriores: 1.927,29

Como $\theta > 0$, podemos pensar em utilizar $\theta \sim \text{Gama}(a, b)$, onde

$$E(\theta) = \frac{a}{b} = 40.493,09$$

e

$$\sqrt{Var(\theta)} = \sqrt{\frac{a}{b^2}} = 1.927,29.$$

Um pouco de álgebra revela que

$$a = 441\ ,\ b = 0,011$$

```
plot.new()
plot.window( ylim=c(0,.00025), xlim = c(30000,50000) )
curve(dgamma(x, 441, .011), lwd = 2 , add = T)
axis(1)
title( xlab = expression(theta))
```

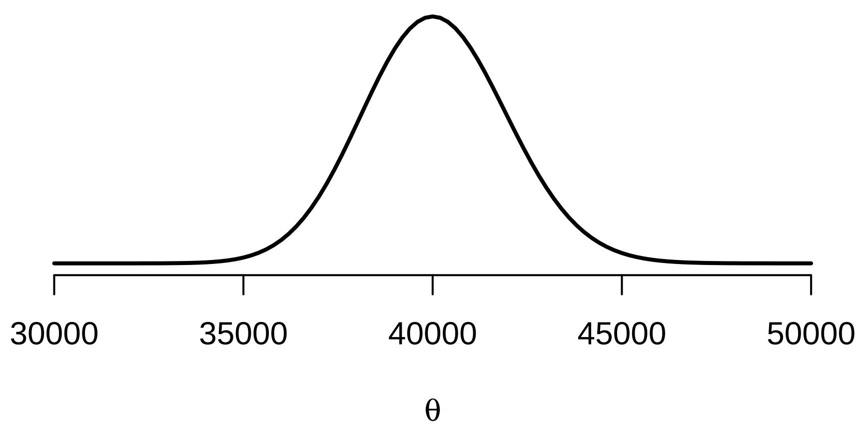


Figure 2.1: Densidade a priori elicitada a partir da média e do desvio padrão de informações passadas

É importante notar que a priori depende de um fator subjetivo. No exemplo anterior, você poderia ter utilizado a moda e o desvio padrão para fazer a sua elicitação, ou mesmo alguns quartis. Ou ainda, utilizado uma Weibull ou outra distribuição contínua. Isto teria resultado em prioris diferentes.

Naturalmente, isto nos conduziria a **posterioris** diferentes. Esta é a principal crítica ao método bayesiano: duas pesquisas com a mesma amostra podem ter resultados distintos, dependendo da priori.

Para evitar resultados discrepantes, temos que garantir que não haja conflitos entre as fontes de informação. Em geral, vamos querer que a **posteriori** esteja mas próxima da verossimilhança que da **priori**.

2.5 O problema da exposição

Uma exposição gratuita recebeu vários visitantes em um dia. Nesta exposição existe um livro de visitas, que o visitante pode optar por não assinar. Os organizadores da exposição dizem que

- Entre 60% e 80% dos visitantes assinam o livro.