

Probabilidade Bayesiana

Flávio Luiz Seixas¹

Instituto de Computação
fseixas@ic.uff.br, <http://www.ic.uff.br/~fseixas>

1 Paradigmas Frequentista e Bayesiano

O paradigma frequentista admite a probabilidade num contexto restrito a fenômenos que podem ser medidos por frequências relativas. O paradigma Bayesiano entende-se que a probabilidade é uma medida racional e condicional de incerteza. Uma medida do grau de plausibilidade de proposições quaisquer, as quais não precisam necessariamente estar associadas a fenômenos medidos por frequência relativa. Por exemplo, no paradigma Bayesiano admite-se falar da probabilidade de extinção de uma espécie, o que não seria admissível sob o paradigma frequentista.

A inferência estatística é o processo formal utilizado para fazer afirmações genéricas com base em informações parciais. Essas afirmações são probabilísticas pois se caracterizam por incluir componentes de incerteza.

Na perspectiva bayesiana, a inferência estatística sobre qualquer quantidade de interesse é descrita como a modificação que se processa nas incertezas à luz de novas evidências.

A conceituação frequentista admite falar em probabilidades somente no contexto de frequências relativas. Em contraste, na conceituação bayesiana, probabilidades quantificam as plausibilidades de proposições ou eventos. Ao atribuir plausibilidades diferenciadas a proposições, a formalização bayesiana de probabilidade estende a lógica dedutiva, restrita a classificar proposições em verdadeiras (probabilidade igual a 1) ou falsas (probabilidade igual a zero), para um conjunto de possibilidades entre estes dois extremos.

O rápido crescimento do uso do paradigma bayesiano em ciências aplicadas ao longo das últimas décadas foi facilitado pelo surgimento de vários programas para efetuar as computações estatísticas. Entre esses, destaca-se o R (programa de livre distribuições e de código aberto).

2 As Regras de Probabilidade

A probabilidade será um número real e função de dois argumentos: o evento incerto E e a premissa H . Utilizaremos o símbolo $Pr(E|H)$ lido como probabilidade de E dado que H é fato, ou a probabilidade de E condicionada ao fato H .

A Lei da convexidade

A probabilidade de um evento qualquer E , condicionado a H é um número real no intervalo $[0, 1]$

$$0 < Pr(E|H) < 1 \quad (1)$$

A Lei da adição

Se E_1 e E_2 são eventos exclusivos sob H , então a probabilidade da união lógica de $E_1 + E_2$ é igual a soma aritmética das suas probabilidades individuais condicionadas a H .

$$Pr(E_1 + E_2|H) = Pr(E_1|H) + Pr(E_2|H) \quad (2)$$

A Lei do produto

Se E_1 e E_2 são eventos quaisquer então a probabilidade do produto lógico $E_1 E_2$ condicionado a H é o produto da probabilidade de E_1 condicionado a H multiplicado pela probabilidade de E_2 condicionado a $E_1 H$

$$Pr(E_1 E_2|H) = Pr(E_1|H) \cdot Pr(E_2|E_1 H) \quad (3)$$

Nos casos em que estamos tratando de eventos independentes, a lei do produto pode ser reescrita como:

$$Pr(E_1 E_2|H) = Pr(E_1|H) \cdot Pr(E_2|H) \quad (4)$$

3 O Teorema de Bayes

Mutas propriedades do cálculo de probabilidades podem ser deduzidas a partir das três leis básicas indicadas na seção anterior. Depois teoremas adicionais merecem especial destaque, o Teorema da Probabilidade Total e o Teorema de Bayes.

Teorema da Probabilidade Total

Seja $E_j; j = 1, \dots, m$ um conjunto de m eventos exclusivos e exaustivos sob H , e seja A outro evento qualquer. Então $Pr(A|H)$ pode ser reescrito estendendo a conversa para a inclusão dos E_j .

$$Pr(A|H) = \sum_{j=1}^m Pr(A|E_j H) \cdot Pr(E_j|H) \quad (5)$$

Teorema de Bayes

Sejam E e F dois eventos quaisquer e $Pr(E|H) > 0$, então:

$$Pr(F|EH) = \frac{Pr(E|FH) \cdot Pr(F|H)}{Pr(E|H)} \quad (6)$$

3.1 Exemplo

Um estudo de uma mamografia no diagnóstico de câncer é apresentado na Tabela 1. Os dados foram obtidos experimentalmente sobre a efetividade do exame na detecção de um tumor de mama maligno ou benigno. Por exemplo, se um tumor é maligno Ca , a probabilidade de que o exame resulte positivo é $Pr(Pos|Ca) = 0,792$, ou seja, 79,2%. De forma similar temos $Pr(Neg|Ca') = 0,904$ como a probabilidade de que o exame resulte negativo se o tumor não é maligno (Ca'). Os percentuais para falsos positivos e falsos negativos são 9,6% e 20,8%, respectivamente.

Table 1. Resultados dos testes de câncer de mamas

Resultado do teste	Realidade	
	Ca (Tumor maligno)	Ca' (Tumor benigno)
Pos (Positivo)	0,792	0,096
Neg (Negativo)	0,208	0,904

Com essa tabela, fez-se a seguinte pergunta: "Suponha que uma paciente pertença a uma população (mesmo grupo etário, hábito alimentar, etc.) na qual a incidência geral de câncer de mama é de 1%. Detectado um nódulo no seio desta paciente, pede-se uma mamografia para avaliar a possibilidade de que se trate de um tumor maligno; o resultado é positivo. De posse deste conjunto de informações, qual é, em sua opinião, a probabilidade de tratar-se de um tumor maligno?"

Pelo Teorema de Bayes:

$$\begin{aligned}
 Pr(Ca|Pos) &= \frac{Pr(Pos|Ca) \cdot Pr(Ca)}{Pr(Pos)} \\
 &= \frac{Pr(Pos|Ca) \cdot Pr(Ca)}{Pr(Pos|Ca) \cdot Pr(Ca) + Pr(Pos|Ca') \cdot Pr(Ca')} \quad (7) \\
 &= \frac{0,792 \cdot 0,01}{0,792 \cdot 0,01 + 0,096 \cdot 0,99} = 0,077
 \end{aligned}$$

Observe que a acurácia retrospectiva do exame $Pr(Pos|Ca)$ é diferente de acurácia preditiva $Pr(Ca|Pos)$. Na prática, a importância atribuída à alta probabilidade de um resultado positivo quando o tumor é de fato maligno,

$Pr(Pos|Ca) = 0,792$ foi excessiva com relação a baixa probabilidade de incidência desse tipo de câncer $Pr(Ca) = 0,01$. No contexto mais geral de investigação científica, o Teorema de Bayes expressa o mecanismo pelo qual hipóteses científicas e evidências empíricas são integrados.

Seja F uma hipótese científica cuja probabilidade corrente é $Pr(F|H)$. Seja E a evidência contida nos dados, experimentais ou observacionais, e cuja probabilidade sob a premissa F é dada por $Pr(E|F, H)$. Então, se efetivamente foi observado E , o Teorema de Bayes permite calcular a probabilidade atualizada $Pr(F|E, H)$ da hipótese F . Portanto, a probabilidade atualizada de F é composta pela sua probabilidade a priori, modificada pela acresção das novas evidências E presentes nos dados.

4 Variáveis aleatórias

As variáveis aleatórias são de dois tipos. As discretas associadas a alguma contagem, e as contínuas que envolvem medições ou mesmo razões. A distinção é necessária para que as regras do cálculo das probabilidades sejam corretamente aplicadas a cada caso. Segue um resumo das propriedades e distinções dos principais modelos de probabilidades para variáveis aleatórias discretas e contínuas.

4.1 Variáveis Aleatórias Discretas

Uma variável aleatória discreta X é uma função que associa as proposições de interesse a um conjunto enumerável (ou categórico, que pode ser nominal ou ordinal), não necessariamente finito, de valores. Se, por exemplo, X caracteriza o número de amostras de sangue avaliadas antes do aparecimento de uma amostra que contém um vírus de interesse, então os possíveis valores para X que caracterizam um conjunto infinito $0, 1, 2, \dots$ já que não existe um limite superior para delimitá-lo.

De modo geral, o valor $Pr(X = x_j|H)$ denota a massa de probabilidade no ponto x_j . Sobre o conjunto de todos os pontos plausíveis, ou seja, pontos com massa de probabilidade maior que zero, a Lei da adição garante que:

$$\sum_j Pr(X = x_j|H) = 1 \quad (8)$$

A média ou valor esperado de X , $E(X)$, e a sua variância, $V(X)$, são definidas por:

$$\begin{aligned} E(X) &= \sum_j x_j Pr(X = x_j|H) \\ V(X) &= \sum_j (x_j - E(X))^2 Pr(X = x_j|H) \end{aligned} \quad (9)$$

4.2 Variáveis Aleatórias Contínuas

Quando os possíveis valores de X foram um subconjunto que compreende pelo menos um intervalo da escala de números reais, a variável aleatória é denominada contínua. Variáveis aleatórias contínuas são caracterizadas pela sua função distribuição cumulativa $F(x)$ que é definida para todo o número real $x \in (-\infty, \infty)$.

$$F(x) = Pr(X \leq x|H) \quad (10)$$

A função densidade de probabilidade é definida por $f(x) = \frac{F(x)}{dx}$. A integração da densidade de probabilidade sobre os números reais é igual a 1, ou seja: $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$.

A média e a variância de X também são expressos em integrais:

$$\begin{aligned} E(X) &= \int x \cdot f(x) dx \\ V(X) &= \int (x - E(X))^2 \cdot f(x) dx \end{aligned} \quad (11)$$

5 Distribuições de Probabilidade

Denomina-se de distribuição de probabilidade de alguma variável aleatória a regra geral que define a função de massa de probabilidade (variável discreta) ou de densidade de probabilidade (variável contínua) para a variável de interesse.

5.1 Distribuições Discretas

Distribuição Uniforme Discreta $Ud(a, b)$

$$p(x) = \frac{1}{b - a + 1} \quad (12)$$

Distribuição Binomial $Bin(n, \theta)$

$$p(x) = \binom{n}{x} \cdot \theta^x \cdot (1 - \theta)^{n-x} \quad (13)$$

Distribuição Hipergeométrica $Hip(M, N, n)$

$$p(x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad (14)$$

Distribuição de Poisson $\text{Poi}(\mu)$

$$p(x) = \frac{e^\mu \mu^x}{x!} \quad (15)$$

Distribuição Binomial Negativa $\text{BinN}(\mathbf{a}, \theta)$

$$p(x) = \frac{(a+x-1)!}{x!(a-1)!} \cdot \theta^a (1-\theta)^x \quad (16)$$

5.2 Distribuição Contínua**Distribuição Uniforme $\text{U}(\mathbf{c}, \mathbf{d})$**

$$p(x) = \frac{1}{d-c} \quad (17)$$

Distribuição Beta (α, β)

$$p(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} \quad (18)$$

Distribuição Beta Gerenalizada**Distribuição Exponencial****Distribuição Gama****Distribuição Gama Inversa****Distribuição Normal****Distribuição Student****5.3 Exemplo de códigos gerados no R**

```

1 # binomial
2 x <- 0:20
3 plot(x, dbinom(x, size=20, prob=.3), type='h')
4
5 # Hipergeometrica
6 m <- 10; n <- 7;
7 x <- 0:9
8 plot(x, dhyper(x, m, n, k), type='h')
9
10 # Poisson
11 m <- 1.8
12 x <- 0:9
13 plot(x, dpois(x,m), type='h')
14
15 # Binomial invertida
16 a <- 5
17 x <- 0:40
18 plot(x, dnbinom(x,a,prob=0.3), type='h')

```

6 Análise Bayesiana de Dados

A inferência bayesiana consiste na construção de uma distribuição de probabilidade posterior via o Teorema de Bayes. Essa distribuição resulta da combinação de informações prévias, sumarizadas em uma distribuição denominada priori, com dados estatísticos descritos por algum modelo probabilístico e resumidos na função de verossimilhança.

A distribuição posterior é a forma mais completa de expressar o estado do conhecimento sobre o fenômeno investigado. Toda pergunta específica é respondida a partir da análise da distribuição posterior. Ela contém toda a informação necessária para a inferência.

Além disso, o processo é dinâmico. A distribuição posterior de hoje pode se transformar na priori em estudos futuros, caracterizando o elemento cumulativo de aquisição de informações.