

Estimando a prevalência de uma doença a partir de um teste diagnóstico

Lucas Emanuel Resck Domingues
Lucas Machado Moschen
Vitor Bitarães

Escola de Matemática Aplicada (EMAp)
Fundação Getúlio Vargas

29/04/2020

Introdução

Suponha que desejamos estimar a proporção $\theta \in (0, 1)$ de indivíduos infectados com um determinado patógeno em uma população. Suponha ainda que dispomos de um teste laboratorial, que produz o resultados $r = \{-, +\}$ indicando se o indivíduo (y_i) é livre (0) ou infectado (1). Se o teste fosse perfeito, poderíamos escrever a probabilidade de observar $y = \sum_{i=1}^n y_i$ testes positivos em n testes realizados como¹

$$\Pr(y \mid \theta, n) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}. \quad (1)$$

Infelizmente, o teste não é perfeito, acertando o diagnóstico com probabilidades fixas da seguinte forma²

$$\Pr(r = + \mid y_i = 0) := 1 - u, \quad (2)$$

$$\Pr(r = - \mid y_i = 1) := 1 - v, \quad (3)$$

de modo que agora, assumindo $u + v > 1$, escrevemos³

$$\begin{aligned} \Pr(r = + \mid \theta, u, v) &= \Pr(r = + \mid y_i = 0) \Pr(y_i = 0) + \Pr(r = + \mid y_i = 1) \Pr(y_i = 1) \\ &= \Pr(r = + \mid y_i = 0) \Pr(y_i = 0) + (1 - \Pr(r = - \mid y_i = 1)) \Pr(y_i = 1) \\ &= (1 - u)(1 - \theta) + (1 - (1 - v))\theta \\ &= 1 - u + \theta(u + v - 1) \end{aligned} \quad (4)$$

e podemos reescrever a probabilidade em~(1):

$$\Pr(y \mid \theta, n, u, v) = \binom{n}{y} [1 - u + \theta(u + v - 1)]^y [u - \theta(u + v - 1)]^{n-y}. \quad (5)$$

¹Porquê?

²Naturalmente, $u, v \in (0, 1)$, levando em conta a restrição $u + v > 1$.

³Exercício bônus: mostre porquê.

Problemas

a) Escolha e justifique uma distribuição *a priori* para θ – lembre-se que neste exercício u e v são fixos;

Resposta:

Uma distribuição de probabilidade plausível para o parâmetro θ é dada pela distribuição beta, uma distribuição flexível, que pode assumir várias formas, dependendo da escolha de valores de a e b . Nesse caso, a priori é $\xi(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}$. Claro que temos que conhecer os valores a e b . Segundo este artigo, os hiperparâmetros podem ser escolhidos em um painel de experts, usando resultados de estudos anteriores.

Nesse caso, a ideia seria aproximar a moda da distribuição *Beta* para a prevalência acreditada por alguns epidemiologistas e, também, aproximar o desvio padrão como uma parte, um quarto, por exemplo, dos limites inferior e superior, acreditados pelo epidemiologista.

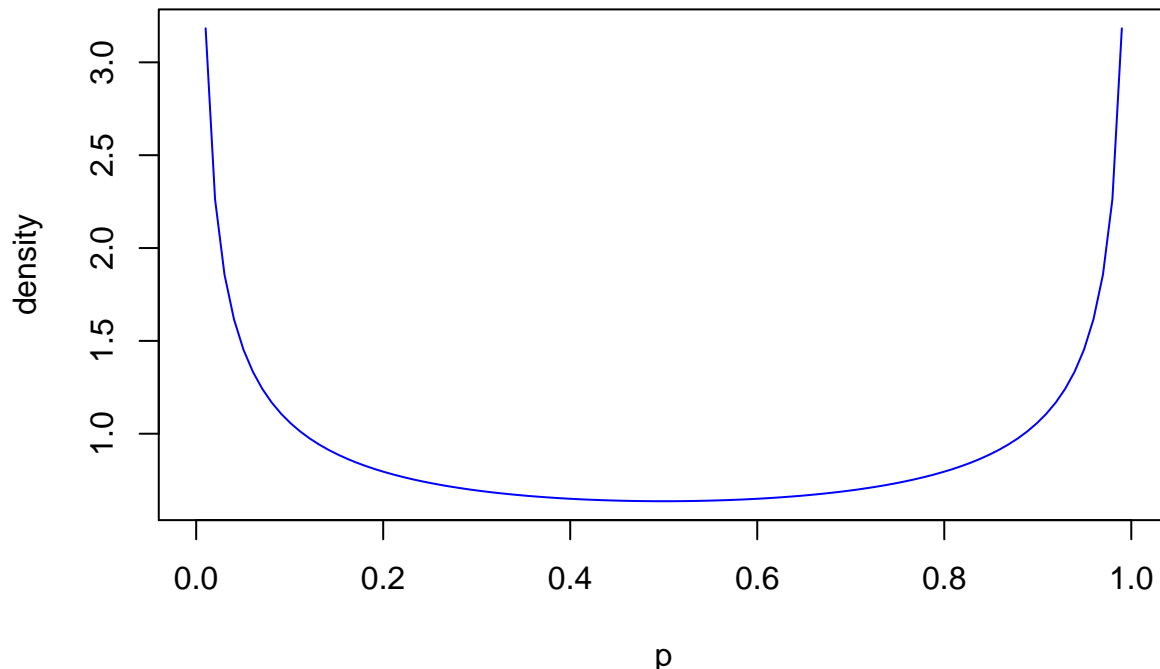
Veja, entretanto, que a escolha dos hiperparâmetros vai influenciar bastante o resultado, já que podemos estar, por exemplo, quase impossibilitando a probabilidade de algumas prevalências.

Podemos contruir uma priori não informativa, nesse caso que não temos informação. No caso da distribuição Beta, poderíamos atribuir $a = b = 1$, para termos uma distribuição uniforme no intervalo $[0, 1]$. Entretanto, existem técnicas para construir essas distribuições. Intuitivamente, a ideia é minimizar o impacto da priori na nossa posteriori, e, nesse caso, a distribuição dela vai se aproximar à estimativa da verossimilhança máxima. Neste artigo de Jeffrey, ele comenta um método para essa construção, baseado na informação de Fisher.

Usando os cálculos obtidos pelo artigo, obteríamos que essa priori seria de fato uma distribuição beta, mas $a = b = 0.5$. Isso, claro, se nossa verossimilhança for uma Binomial.

Vejam os formato de uma *Beta*(0.5,0.5)

```
p = seq(0,1, length=100)
plot(p, dbeta(p, 0.5, 0.5), ylab="density", type="l", col=4)
```



O que acham de usar a distribuição normal truncada?

Distribuição Normal Truncada

b) Derive $\Pr(\theta | y, n, u, v)$;

Resposta:

- c) Suponha que $y = 4$ e $n = 5000$. Qual a média $\sim a posteriori$ de θ ? Produza intervalos de credibilidade de 80, 90 e 95% para θ .

Resposta:

- d) **Bônus.** Que melhorias você faria neste modelo? Que outras fontes de incerteza estão sendo ignoradas?

Resposta: