

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

Claudio Struchiner & Luiz Max de Carvalho
Escola de Matemática Aplicada (EMAp), Getúlio Vargas Foundation.

8 de Maio de 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

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

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

$$\Pr(r = + | y_i = 0) =: u, \quad (2)$$

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

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

$$\Pr(r = + | \theta, u, v) =: \theta(1 - v) + (1 - \theta)u, \quad (4)$$

e podemos reescrever a probabilidade em (1):

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

Parâmetros para COVID-19: $u = 1 - \text{especificidade} = 1 - 0.97 = 0.03$ e $v = 1 - \text{sensibilidade} = 1 - 0.75 = 0.25$.

Problema(s)

- Escolha e justifique uma distribuição *a priori* para θ – lembre-se que neste exercício u e v são fixos;
- Derive $\Pr(\theta | y, n, u, v)$;
- Suponha que $y = 4$ e $n = 5000$. Qual a média *a posteriori* de θ ? Produza intervalos de credibilidade de 80, 90 e 95% para θ .
- Bônus.** Que melhorias você faria neste modelo? Que outras fontes de incerteza estão sendo ignoradas?

Dicas

- Lembrem-se de justificar **todas** as suas respostas, tanto matematica quanto estatisticamente. Isto inclui a escolha de métodos numéricos, se estes forem necessários;

¹Porquê?

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

³Exercício bônus: mostre porquê.

- O tópico abordado aqui é bastante conhecido e existe farta literatura. As palavras-chave são: diagnostic tests, sensitivity, specificity, Bayesian estimation. Por enquanto, não vamos recomendar os artigos pertinentes para que o problema não perca a graça uma vez que vocês vejam a solução. Quando formos discutir as abordagens para o problema, revelaremos alguns dos artigos que podem ser consultados;
- Vocês podem consultar os capítulos 7.2 e 7.3 de [DeGroot and Schervish \(2012\)](#) para o item a) e também o capítulo 3 de [McElreath \(2020\)](#) para os outros itens.
- Procurem se divertir com os problemas nas notas de rodapé; eles valem o esforço e envolvem apenas relembrar conceitos de probabilidade.

Referências

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and statistics*. Pearson Education.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.