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

Lucas Emanuel Resck Domingues, Lucas Machado Moschen, Vitor Bitarães

4/29/2020

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 como[1]

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

Infelizmente, o teste não é perfeito, acertando o diagnóstico com probabilidades fixas da seguinte forma[2]

$$\Pr(\ r = + \mid y_i = 0) := 1 - u, \tag{2}$$
$$\Pr(r = - \mid y_i = 1) := 1 - v, \tag{3}$$

de modo que agora, assumindo $u + v > 1$, escrevemos[3]

$$\Pr(r = + \mid \theta, u, v) := \theta(1-v) + (1-\theta)u, \tag{4}$$

e podemos reescrever a probabilidade em~(1):

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

Problemas

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

Resposta:

Uma primeira idea é a distribuição BETA que já conversamos em aula.

Procurar usadas na biografia.

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

Resposta:

[1] Porquê?

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

[3] Exercício bônus: mostre porquê.

c) 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.

Resposta:

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

Resposta: