Two screening tests

Douglas R. Stinson
Waterloo ON

August 6, 2024

We discuss question TYI 55 that was posed on Gil Kalai's blog [1] on July 25, 2024. The question is attributed to Michele Piccione and Ariel Rubinstein and it is introduced in their paper [2].

Suppose we have a population X of size N (of course the value of N is irrelevant). We are told that a subset $Y \subseteq X$ possesses a certain trait, where $|Y| = N/100$. There are two screening tests for this trait, denoted by A and B. There are no false negatives, but there are false positives. Let $\mathbf{T}$ be a random variable denoting the presence of the trait, and let $\mathbf{A}$ and $\mathbf{B}$ respectively denote random variables indicating a positive outcome to the two tests.

We are told that

$$\Pr[\mathbf{T}|\mathbf{A}] = 0.7 \tag{1}$$

and

$$\Pr[\mathbf{T}|\mathbf{B}] = 0.2. \tag{2}$$

We are also told that the two screening tests are conditionally independent. The task is to compute $\Pr[\mathbf{T}|\mathbf{A} \wedge \mathbf{B}]$.

We use an elementary counting approach to solve the problem. Consider the following three subsets of X:

$$\begin{aligned} X_{1,0} &= \{x \in X \setminus Y : A \text{ has a positive outcome and } B \text{ has a negative outcome}\} \\ X_{1,1} &= \{x \in X \setminus Y : A \text{ and } B \text{ both have positive outcomes}\} \\ X_{0,1} &= \{x \in X \setminus Y : A \text{ has a negative outcome and } B \text{ has a positive outcome}\}. \end{aligned}$$

Denote $a = |X_{1,0}|$, $b = |X_{1,1}|$ and $c = |X_{0,1}|$.

From (1), we have

$$\frac{\frac{N}{100}}{\frac{N}{100} + a + b} = 0.7,$$

so

$$a + b = \frac{.3N}{70}. \tag{3}$$

Similarly, from (2), we have

$$\frac{\frac{N}{100}}{\frac{N}{100} + b + c} = 0.2,$$

so

$$b + c = \frac{.8N}{20}. \tag{4}$$

It is also straightforward to see that the value we want to compute is

$$\Pr[\mathbf{T}|\mathbf{A} \wedge \mathbf{B}] = \frac{\frac{N}{100}}{\frac{N}{100} + b}.$$

Since the two screening tests are independent, we have

$$\Pr[\mathbf{A}|\overline{\mathbf{T}}] \times \Pr[\mathbf{B}|\overline{\mathbf{T}}] = \Pr[\mathbf{A} \wedge \mathbf{B}|\overline{\mathbf{T}}].$$

Therefore,

$$\frac{a+b}{\frac{99N}{100}} \times \frac{b+c}{\frac{99N}{100}} = \frac{b}{\frac{99N}{100}}.$$

Substituting (3) and (4) into this equation, we have

$$\frac{\frac{.3N}{70}}{\frac{99N}{100}} \times \frac{\frac{.8N}{20}}{\frac{99N}{100}} = \frac{b}{\frac{99N}{100}},$$

so

$$\frac{.24}{1400} = \frac{99b}{100N}$$

and hence

$$b = \frac{.24 \times 100N}{1400 \times 99} = \frac{N}{5775}.$$

Finally,

$$\Pr[\mathbf{T}|\mathbf{A} \wedge \mathbf{B}] = \frac{\frac{N}{100}}{\frac{N}{100} + \frac{N}{5775}} = \frac{\frac{1}{100}}{\frac{1}{100} + \frac{1}{5775}} = 0.9829787237.$$

The result of this computation agrees with [2].

Perhaps it would be instructive to redo our computation in a more general setting, as is done in [2]. Suppose we have a population X of size N and a subset $Y \subseteq X$ possesses a certain trait, where $|Y| = sN$ $(0 < s < 1)$. Suppose also that

$$\Pr[\mathbf{T}|\mathbf{A}] = \phi_1 \quad \text{and} \quad \Pr[\mathbf{T}|\mathbf{B}] = \phi_2.$$

Then, using the same reasoning as before, we have

$$a + b = \frac{sN(1-\phi_1)}{\phi_1} \quad \text{and } b + c = \frac{sN(1-\phi_2)}{\phi_2}.$$

Computing as before, we have

$$b = \frac{s^2N(1-\phi_1)(1-\phi_2)}{(1-s)\phi_1\phi_2}.$$

Finally,

$$\begin{aligned}
\Pr[\mathbf{T}|\mathbf{A} \wedge \mathbf{B}] &= \frac{sN}{sN + b} \\
&= \frac{sN}{sN + \frac{s^2 N(1-\phi_1)(1-\phi_2)}{(1-s)\phi_1\phi_2}} \\
&= \frac{(1-s)\phi_1\phi_2}{\phi_1\phi_2 + s(1-\phi_1-\phi_1)}.
\end{aligned}$$

This is easily seen to be equivalent to the formula given in [2], which is stated as follows:

$$\frac{\Pr[\mathbf{T}|\mathbf{A} \wedge \mathbf{B}]}{1 - \Pr[\mathbf{T}|\mathbf{A} \wedge \mathbf{B}]} = \frac{(1-s)\phi_1\phi_2}{s(1-\phi_1)(1-\phi_2)}.$$

References

[1] G. Kalai. Test Your Intuition 55: The Case of Two Screening Tests. `https://gilkalai.wordpress.com/2024/07/25/test-your-intuition-55-the-case-of-two-screening-tests/`.

[2] M. Piccione and A. Rubinstein. Failing to correctly aggregate signals.