Exercises in Bayesian reasoning: Proceedings of the Church of Human Potential

Joachim Vandekerckhove

The sinister serologist

The sinister serologist

In the research labs of the Church of Human Potential, Reverend William Stryker oversees the Mutant Serology Department. One project is to develop a field test for the “X gene” that distinguishes mutants from baseline humans.

The sinister serologist

In the research labs of the Church of Human Potential, Reverend William Stryker oversees the Mutant Serology Department. One project is to develop a field test for the “X gene” that distinguishes mutants from baseline humans.

The sinister serologist

Reverend Stryker's new field test can accurately detect the X mutation 99% of the time, but will also falsely indicate that a baseline human is a mutant 2% of the time.

The sinister serologist

Reverend Stryker's new field test can accurately detect the X mutation 99% of the time, but will also falsely indicate that a baseline human is a mutant 2% of the time.

When Stryker presents his results to the Pentagon, Senator Robert Kelly asks two questions:

The sinister serologist

Reverend Stryker's new field test can accurately detect the X mutation 99% of the time, but will also falsely indicate that a baseline human is a mutant 2% of the time.

When Stryker presents his results to the Pentagon, Senator Robert Kelly asks two questions:

- What is the probability that a subject is a mutant, when your field test says that it is mutant?
- What is the probability that a subject is a mutant, when your field test says that it is baseline?

The sinister serologist

Call the event that a specific subject is a mutant $\mathcal{M}$, and that it is baseline $\neg\mathcal{M}$.

Call the event that Stryker's field test diagnoses a subject as a mutant D, and that it diagnoses it baseline $\neg D$.

Senator Kelly's interest is in the probability the subject is indeed a mutant given it has been diagnosed as a mutant, or $P(\mathcal{M}|D)$, and the probability the subject is a mutant given it has been diagnosed as baseline, or $P(\mathcal{M}|\neg D)$.

The sinister serologist

Senator Kelly, who is a trained statistician, has all the relevant information to apply Bayes' Rule to find these probabilities:

- He knows the prior probability that a subject is a mutant is $P(\mathcal{M}) = .001$, and thus the prior probability that a subject is baseline is $P(\neg\mathcal{M}) = 1 - P(\mathcal{M}) = .999$.
- The probability of a correct mutant diagnosis given the subject is a mutant is $P(D|\mathcal{M}) = .99$, and the probability of an erroneous baseline diagnosis given the subject is a mutant is thus $P(\neg D|\mathcal{M}) = 1 - P(D|\mathcal{M}) = .01$.
- When the subject is baseline, the field test incorrectly diagnoses it as a mutant with probability $P(D|\neg\mathcal{M}) = .02$, and correctly diagnoses the subject as baseline with probability $P(\neg D|\neg\mathcal{M}) = 1 - P(D|\neg\mathcal{M}) = .98$.

The sinister serologist

When Stryker's field test gives a mutant diagnosis, the posterior probability that the subject is really a mutant is given by:

$$P(\mathcal{M}|D) = \frac{P(\mathcal{M})P(D|\mathcal{M})}{P(\mathcal{M})P(D|\mathcal{M}) + P(\neg\mathcal{M})P(D|\neg\mathcal{M})}.$$

Senator Kelly can now plug in the values to find that the posterior probability the subject is a mutant given a mutant diagnosis is:

$$P(\mathcal{M}|D) = \frac{.001 \times .99}{.001 \times .99 + .999 \times .02} \approx .047.$$

The sinister serologist

A mutant diagnosis from Stryker's field test raises the probability the subject is a mutant from $.001$ to roughly $.047$. This means that when a subject is diagnosed as a mutant, the posterior probability the subject is *not* a mutant is $P(\neg\mathcal{M}|D) = 1 - .047 \approx .953$.

The sinister serologist

A mutant diagnosis from Stryker's field test raises the probability the subject is a mutant from $.001$ to roughly $.047$. This means that when a subject is diagnosed as a mutant, the posterior probability the subject is *not* a mutant is $P(\neg\mathcal{M}|D) = 1 - .047 \approx .953$.

The low prior probability that a subject is a mutant means that, even with the field test having 99% accuracy to correctly diagnose a mutant subject as such, a subject diagnosed as a mutant is still probably baseline.

The sinister serologist

Analogous calculations show that the posterior probability that a subject is a mutant, given it is diagnosed as baseline, is:

$$P(\mathcal{M}|\neg D) = \frac{.001 \times .01}{.001 \times .01 + .999 \times .98} \approx .000010.$$

The sinister serologist

Analogous calculations show that the posterior probability that a subject is a mutant, given it is diagnosed as baseline, is:

$$P(\mathcal{M}|\neg D) = \frac{.001 \times .01}{.001 \times .01 + .999 \times .98} \approx .000010.$$

The posterior probability that a subject is a mutant despite being diagnosed as baseline is almost vanishingly small, so Senator Kelly can be relatively confident if Stryker's field test returns a baseline diagnosis.

The sinister serologist

An advantage of using Bayes' Rule in this way is that it gracefully extends to more complex scenarios.

The sinister serologist

An advantage of using Bayes' Rule in this way is that it gracefully extends to more complex scenarios.

Consider the perhaps disappointing value of $P(\mathcal{M}|D)$: a mutant diagnosis only brings the posterior probability up to 5%. Suppose, however, that Stryker knows that his diagnosis is statistically independent from the diagnosis of his research associate Didier Raoult, and suppose that both Stryker and Raoult return the mutant diagnosis. Due to the independence of the diagnoses, we know that

$$\begin{aligned} P(D_S, D_L|\mathcal{M}) &= P(D_S|\mathcal{M}) \times P(D_L|\mathcal{M}) = .99 \times .99 = .9801 \\ P(D_S, D_L|\neg\mathcal{M}) &= P(D_S|\neg\mathcal{M}) \times P(D_L|\neg\mathcal{M}) = .02 \times .02 = .0004 \end{aligned}$$

The sinister serologist

Applying Bayes' Rule once more (and considering that the joint event (D_S, D_L) can be treated like any other event), the posterior probability that the subject is really a mutant is now given by:

$$\begin{aligned} P(\mathcal{M}|D_S, D_L) &= \frac{P(\mathcal{M})P(D_S, D_L|\mathcal{M})}{P(\mathcal{M})P(D_S, D_L|\mathcal{M}) + P(\neg\mathcal{M})P(D_S, D_L|\neg\mathcal{M})} \\ &= \frac{.001 \times .99 \times .99}{.001 \times .99 \times .99 + .999 \times .02 \times .02} \approx .71, \end{aligned}$$

which illustrates the value of multiple independent sources of evidence: a subject that has been independently diagnosed as a mutant is quite likely to be one. A third independent diagnosis would put the posterior probability over 99%.