

How to Think Like an Epidemiologist

Don't worry, a little Bayesian analysis won't hurt you.

By Siobhan Roberts

Aug. 4, 2020

Sign up for Science Times Get stories that capture the wonders of nature, the cosmos and the human body. [Get it sent to your inbox.](#)

There is a statistician's rejoinder — sometimes offered as wry criticism, sometimes as honest advice — that could hardly be a better motto for our times: “Update your priors!”

In stats lingo, “priors” are your prior knowledge and beliefs, inevitably fuzzy and uncertain, before seeing evidence. Evidence prompts an updating; and then more evidence prompts further updating, so forth and so on. This iterative process hones greater certainty and generates a coherent accumulation of knowledge.

In the early pandemic era, for instance, airborne transmission of Covid-19 was not considered likely, but in early July the World Health Organization, with mounting scientific evidence, conceded that it is a factor, especially indoors. The W.H.O. updated its priors, and changed its advice.

This is the heart of Bayesian analysis, named after Thomas Bayes, an 18th-century Presbyterian minister who did math on the side. It captures uncertainty in terms of probability: Bayes's theorem, or rule, is a device for rationally updating your prior beliefs and uncertainties based on observed evidence.

Reverend Bayes set out his ideas in “An Essay Toward Solving a Problem in the Doctrine of Chances,” published posthumously in 1763; it was refined by the preacher and mathematician Richard Price and included Bayes's theorem. A couple of centuries later, Bayesian frameworks and methods, powered by computation, are at the heart of various models in epidemiology and other scientific fields

As Marc Lipsitch, an infectious disease epidemiologist at Harvard, noted on Twitter, Bayesian reasoning comes awfully close to his working definition of rationality. “As we learn more, our beliefs should change,” Dr. Lipsitch said in an interview. “One extreme is to decide what you think and be impervious to new information. Another extreme is to over-privilege the last thing you learned. In rough terms, Bayesian reasoning is a principled way to integrate what you previously thought with what you have learned and come to a conclusion that incorporates them both, giving them appropriate weights.”

With a new disease like Covid-19 and all the uncertainties it brings, there is intense interest in nailing down the parameters for models: What is the basic reproduction number, the rate at which new cases arise? How deadly is it? What is the infection fatality rate, the proportion of people with the virus that it kills?

But there is little point in trying to establish fixed numbers, said Natalie Dean, an assistant professor of biostatistics at the University of Florida.

“We should be less focused on finding the single ‘truth’ and more focused on establishing a reasonable range, recognizing that the true value may vary across populations,” Dr. Dean said. “Bayesian analyses allow us to include this variability in a clear way, and then propagate this uncertainty through the model.”

A textbook application of Bayes's theorem is serology testing for Covid-19, which looks for the presence of antibodies to the virus. All tests are imperfect, and the accuracy of an antibody test turns on many factors including, critically, the rarity or prevalence of the disease.

The first SARS-CoV-2 antibody test approved by the F.D.A., in April, seemed to be wrong as often as it was right. With Bayes's theorem, you can calculate what you really want to know: the probability that the test result is correct. As one commenter on Twitter put it: “Understanding Bayes's theorem is a matter of life and death right now.”

The logic of uncertainty

Joseph Blitzstein, a statistician at Harvard, delves into the utility of Bayesian analysis in his popular course “Statistics 110: Probability.” For a primer, in lecture one, he says: “Math is the logic of certainty, and statistics is the logic of uncertainty. Everyone has uncertainty. If you have 100 percent certainty about everything, there is something wrong with you.”

By the end of lecture four, he arrives at Bayes's theorem — his favorite theorem because it is mathematically simple yet conceptually powerful.

“Literally, the proof is just one line of algebra,” Dr. Blitzstein said. The theorem essentially reduces to a fraction; it expresses the probability P of some event A happening given the occurrence of another event B.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

“Naïvely, you would think, How much could you get from that?” Dr. Blitzstein said. “It turns out to have incredibly deep consequences and to be applicable to just about every field of inquiry” — from finance and genetics to political science and historical studies. The Bayesian approach is applied in analyzing racial disparities in policing (in the assessment of officer decisions to search drivers during a traffic stop) and search-and-rescue operations (the search area narrows as new data is added). Cognitive scientists ask, ‘Is the brain Bayesian?’ Philosophers of science posit that science as a whole is a Bayesian process — as is common sense.

Take diagnostic testing. In this scenario, the setup of Bayes’s theorem might use events labeled “T” for a positive test result — and “C” for the presence of Covid-19 antibodies:

$$P(C|T) = \frac{P(T|C) \times P(C)}{P(T)}$$

Now suppose the prevalence of cases is 10 percent (that was so in New York City in the spring), and you have a positive result from a test with accuracy of 87.5 percent sensitivity and 97.5 percent specificity. Running numbers through the Bayesian gears, the probability that the result is correct, and that you do indeed have antibodies is 79.5%. Decent odds, all things considered. If you want more certainty, get a second opinion. And continue to be cautious.

The Coronavirus Pandemic: Latest Updates ›

Updated 39 minutes ago

- The Supreme Court blocks Biden’s mandate for large employers but allows the vaccine mandate for health care workers.
- Djokovic may be a divisive figure abroad, but he remains a hero in Serbia.
- W.H.O. officials see signs that Omicron may be peaking in Africa.

An international collaboration of researchers, doctors and developers created another Bayesian strategy, pairing the test result with a questionnaire to produce a better estimate of whether the result might be a false negative or a false positive. The tool, which has won two hackathons, collects contextual information: Did you go to work during lockdown? What did you do to avoid catching Covid-19? Has anyone in your household had Covid-19?

“It’s a little akin to having two ‘medical experts,’” said Claire Donnat, who recently finished her Ph.D. in statistics at Stanford and was part of the team. One expert has access to the patient’s symptoms and background, the other to the test; the two diagnoses are combined to produce a more precise score, and more reliable immunity estimates. The priors are updated with an aggregation of information.

“As new information comes in, we update our priors all the time,” said Susan Holmes, a Stanford statistician, via unstable internet from rural Portugal, where she unexpectedly pandemicked for 105 days, while visiting her mother.

The Coronavirus Pandemic: Key Things to Know

The latest Covid data in the U.S. As the Omicron surge causes case counts to reach record highs and hospitalizations to surpass last winter’s peak, here’s how to think about the data and what the early numbers suggest about the variant’s potential toll and possible peaking in parts of the United States.

< >

That was the base from which Dr. Holmes refined a preprint paper, co-authored with Dr. Donnat, that provides another example of Bayesian analysis, broadly speaking. Observing early research in March about how the pandemic might evolve, they noticed that classic epidemiological models tend to use fixed parameters, or constants, for the reproduction number — for instance, with an R0 of 2.0.

But in reality, the reproduction number depends on random, uncertain factors: viral loads and susceptibility, behavior and social networks, culture and socioeconomic class, weather, air conditioning and unknowns.

With a Bayesian perspective, the uncertainty is encoded into randomness. The researchers began by supposing that the reproductive number had various distributions (the priors). Then they modeled the uncertainty using a random variable that fluctuates, taking on a range of values as small as 0.6 and as large as 2.2 or 3.5. In something of a nesting process, the random variable itself has parameters that fluctuate randomly; and those parameters, too, have random parameters (hyper-parameters), etcetera. The effects accumulate into a “Bayesian hierarchy” — “turtles all the way down,” Dr. Holmes said.

The effects of all these up-and-down random fluctuations multiply, like compound interest. As a result, the study found that using random variables for reproductive numbers more realistically predicts the risky tail events, the rarer but more significant superspreader events.

Humans on their own, however, without a Bayesian model for a compass, are notoriously bad at fathoming individual risk.

“People, including very young children, can and do use Bayesian inference unconsciously,” said Alison Gopnik, a psychologist at the University of California, Berkeley. “But they need direct evidence about the frequency of events to do so.”

Much of the information that guides our behavior in the context of Covid-19 is probabilistic. For example, by some estimates, if you get infected with the coronavirus, there is a 1 percent chance you will die; but in reality an individual's odds can vary by a thousandfold or more, depending on age and other factors. “For something like an illness, most of the evidence is usually indirect, and people are very bad at dealing with explicit probabilistic information,” Dr. Gopnik said.

Modeling humility

Even with evidence, revising beliefs isn't easy. The scientific community struggled to update its priors about the asymptomatic transmission of Covid-19, even when evidence emerged that it is a factor and that masks are a helpful preventive measure. This arguably contributed to the world's sluggish response to the virus.

“The problems come when we don't update,” said David Spiegelhalter, a statistician and chair of the Winton Centre for Risk and Evidence Communication at the University of Cambridge. “You can interpret confirmation bias, and so many of the ways in which we react badly, by being too slow to revise our beliefs.”

There are techniques that compensate for Bayesian shortcomings. Dr. Spiegelhalter is fond of an approach called Cromwell's law. “It's heaven,” he said. In 1650, Oliver Cromwell, Lord Protector of the Commonwealth of England, wrote in a letter to the Church of Scotland: “I beseech you, in the bowels of Christ, think it possible you may be mistaken.”

In the Bayesian world, Cromwell's law means you should always “keep a bit back — with a little bit of probability, a little tiny bit — for the fact that you may be wrong,” Dr. Spiegelhalter said. “Then if new evidence comes along that totally contradicts your main prior belief, you can quickly ditch what you thought before and lurch over to that new way of thinking.”

“In other words, keep an open mind,” said Dr. Spiegelhalter. “That's a very powerful idea. And it doesn't necessarily have to be done technically or formally; it can just be in the back of your mind as an idea. Call it ‘modeling humility.’ You may be wrong.”

[Like the Science Times page on Facebook. | Sign up for the Science Times newsletter.]

A version of this article appears in print on , Section D, Page 8 of the New York edition with the headline: Thinking Like an Epidemiologist