Probability v Likelihood

Ben Klemens

2 July 2009

Here's the question: should we distinguish probabilities from likelihoods?

In case you didn't even know there was a distinction, here are the definitions. First, let the word *odds* take our intuitive meaning. A *probability* gives the odds of an event, given any parameters. Given that the mean is zero and the variance one, what are the odds that the draw will be between 1.1 and 1.2? A *likelihood* gives the odds of parameters given data. We drew a 1.3 from the distribution; what are the odds that the mean is zero?

Now, the probability can be verified. We can make a million draws from the distribution, and then count up what percentage are between 1.1 and 1.2, and call that the odds. Likelihood can't be verified. We have only one distribution to draw from, so we have no story about re-drawing from millions of different distributions and developing a confidence that the data comes from one or another.

Some folks take this as settling the question, concluding that there's a distinction to be made. The odds of data is relatively concrete; the odds of a parameter is at best a metaphor to probability. The most famous person who stopped here is Mr R A Fisher.

Fisher is an interesting character, in that his techniques are indisputably the baseline for modern statistics, but his larger worldview didn't survive. Do you know the definition of a fiducial distribution? Well, it was central to Fisher's overall system. Fisher was UCL's Galton Professor of Eugenics, and the reputation of that field hasn't been very good ever since that one holocaust. [I do *not* claim that Fischer endorsed the Holocaust. However, he was clear in his endorsement of selective breeding and other such Eugenic principles that we now consider to be politically incorrect and/or evil.]

Fisher was vehement in distinguishing between probability and likelihood—in fact, he coined the term likelihood to make that distinction. [There's a comment on the matter from him on p 329 of *Modeling with Data*.]

But things are a little more blurred than that. We'll start with the probability side, and the story above about making a million draws of an event. This is problematic for more cases than it is easy. What are the odds of rain tomorrow? We only have one tomorrow to live. We could look at comparable prior days, but how similar must a situation be before it's properly comparable?

What percentage of cars passing an intersection are SUVs? If we observe over the course of a day, aren't we confounding the rush hour rate with the mid-day rate and the midnight rate? If we take full days, will Monday have the same rate as Sunday, and does a Monday in January have the same rate as a Monday in April?

1

The frequentist interpretation of probability assumes an infinite stream of data that is identical in all manners but the single variable we care about. This is obviously a fiction, but we don't mind because there are enough cases where it approximately works. Our weatherfolk and surveyors worked out something serviceable and run with it. [We can think of regression analysis as an attempt to patch over the 'all else equal' assumption.]

In the probability v likelihood context, the distinction starts to blur. We only have one distribution, so the likelihood is a human-invented fiction. We only have one tomorrow, so the probability of rain is also a human-invented fiction.

Now let's go the other way, and consider how imaginary or subjective these parameters are. Maybe we've observed a few manufacturers and we know that contaminants per million is Normally distributed with a different, observable-by-history mean for each manufacturer. Now we have a situation where the parameters of the Normal are not taken from an imaginary distribution, but are just as much observed as the data is. The odds that a given data draw is taken from one Normal distribution or another can be calculated from the records on hand.

The joint distribution   Me, I spend a lot of my time writing code. What would a probability function look like? It would take in some data and parameters, and put out a nonnegative number: $p(\text{data}, \text{params})$. What would a likelihood function look like? Well, it would take in some data and parameters, and put out a nonnegative number: $l(\text{data}, \text{params})$. And in fact, for a given model, like a Normal distribution, the two functions are identical.

So how's that for a trip: by our traditional interpretation, a function viewed one way (with the parameters fixed) is an objective and verifiable fact of nature; the very same function viewed another way (with the data fixed) is subjective and human-invented. Let the data be x and the parameter be β, then $P(x|\beta)$ is objective and $P(\beta|x)$ is subjective.

In a couple of ways, $P(x, \beta)$ is a combination of the objective and subjective. A full, unconditional distribution of data, $P(x)$, would certainly count as objective in our classification scheme (assuming away the practical problem of gathering it), and $P(\beta|x)$ is the subjective likelihood, and we can combine the two to produce the joint, not-conditional likelihood $P(\beta|x) \cdot P(x) = P(x, \beta)$. Of course, you can also do it the other way, again combining an objective and subjective component to get $P(x|\beta) \cdot P(\beta) = P(x, \beta)$.

So is $P(x, \beta)$ objective or subjective? What can we do with such a function?

I have no idea what Fisher would say [though I suppose this is easy enough to research; I invite comments from anybody who has done so]. He seemed to be resistant to allowing the two conditional probabilities to be merged, and took pains to distinguish between a function of the data given the parameters (such as a probability) and a function of the parameters given data (of which the fiducial distribution is one); I'm not sure where he'd class the joint distribution from which both sides can be derived.

The post-Bayesian modernists are fine with accepting the entire thing as subjective. We don't know what probability is, and we don't know what likelihood is. It's all subjective from top to bottom.

Here, I'm taking something of a more moderate position, because I have no idea

whether the fundamental philosophy questions are even answerable. I can tell you that for any sufficiently well-specified situation, I can give you a function $P(x, \beta)$, from which we can derive slices that are functions only of the data or of the parameters. Sometimes, this is a mix of the data, the model, and subjective beliefs; sometimes it's just a table of observed data.

So that's why I don't distinguish between probability and likelihood. The philosophy issues are hard to untangle, and it's easy to find equations that are objective, subjective, or a mix of both depending on context and opinion. We may make distinctions between parameters and data, but the probability/likelihood formula $P(x, \beta)$ doesn't care about the distinction. It's just a function where one input is a Roman character and the other Greek.