

Bayesian, Frequentist, and why I do not care

Ben Klemens

19 January 2013

Really intelligent people have made attempts to formalize the intuitive definition of odds/chance-of/probability/likelihood for a few centuries now. We have no new data: what real-world evidence on the link between our intuition about odds and the real world do we have that Pascal didn't have in the 1600s? [The entropy angle is interesting, because we've only had the concept for, I dunno, 150 years, and information entropy for not even a century.]

When there's a major problem that thousands of people possessing full information have written about in great detail, and it's still unresolved, that's when we conclude that it's a question of philosophy. It can be resolved in the sense that philosophical questions can be resolved, wherein distinct parties come up with internally consistent and personally satisfactory storylines and there is no objective test to determine which storyline is objectively correct.

Ad hominem Articles that talk about Bayesians and Frequentists often remind me of those comedians with bits about how *Black people, they do it like this, But then White people, they do it like this:* I end up wondering if the people being made fun of really exist here in the real world.

A lot of attacks on Frequentists are really attacks on oversimplified undergrad stats textbooks, or seat-of-the-pants statisticians who do things that seem to work but have no serious theoretical basis. Here's a review of a book by Nate Silver¹, written by a non-statistician, who got the impression from Silver's writing that Frequentists do nothing but produce inferior research. Using his commonsense prior knowledge, he concludes that Silver is beating up a straw man that is unlikely to be representative of real Frequentists.

Characterizing a class of people based on the subset that is least paying attention is fun for the feelings of superiority, but is otherwise a waste of time. Here's an essay² (PDF) where Jayne says that he's done with Frequentists, pointing to certain people who had trouble understanding how likelihood functions relate to states of nature and calling them "handicapped". Some day, Bayesian methods will be common enough that there are oversimplifying textbooks that provide recipes for users to misapply, and software that advertises itself as being so easy that even an undergrad who knows nothing of statistics can get Bayesianly correct estimates without knowing a lick of statistics. At which point we'll have essays about how Bayesians don't have a deep understanding of statistics.

¹<http://www.openlettersmonthly.com/too-much-signal/>

²<http://bayes.wustl.edu/etj/articles/prob.as.logic.pdf>

On models A model intermediates between data and parameters. We typically express a model in terms of a probability/likelihood function, which I will write as $P(d, \beta)$, where d is a data element (typically scalar or matrix, typically known), β is a set of parameters (typically a vector, typically unknown), and the output to $P(d, \beta)$ is a probability/likelihood, a single nonnegative scalar.

At this point, you might want to go and read this entry on probability versus likelihood (Entry #024), which explains that the function $P(d, \beta)$ expresses both subjective and objective components in a single joint distribution, and why I have to finesse the question of whether this is a probability or a likelihood—it's both. I think that if you are comfortable with how $P(d|\beta)$ is verifiable and (relatively) objective, while $P(\beta|d)$ is subjective and (in most cases) can not be verified, then you have a handle on the crux of the debate.

Having gotten the caveats about stereotyping out of the way, my working definition is that Bayesians see $P(d, \beta)$ as a fundamentally subjective expression, but strive to rely as much as possible on an objective probabilistic framework; Frequentists see $P(d, \beta)$ as fundamentally objective, but acknowledge and strive to accommodate subjective elements where needed.

Defining the model Isn't the distinction between Bayesians and everybody else right in the title—that one team uses Bayes's rule and everybody else doesn't? Bayes's rule has a one-line proof, and your typical stats/probability textbook covers it by page twenty. To not believe in Bayes's rule is equivalent to not believing basic standard probability theory. Models that don't explicitly use Bayes's rule at some point over the course of their typical description can frequently still be restated to make the underlying rule more apparent; explicit use of Bayes's rule is often computationally expensive, so Bayesian models are frequently reshaped to a version with no mention of Bayes's rule.

I found this article by David Draper³ (PDF, 95 slides) to be a nice summary of the descriptive-side situation. The first 25pp give an overview of the efforts to formalize the problem of defining a model and separating the objective from the subjective, listing the many conditions that are required to really make for a consistent theory. It points out that Bayes's rule is essential to consistency, but then turns around and seems to imply that only models that explicitly use Bayes's rule are going to be consistent, which is clearly not true (and I'm guessing/hoping that DD would acknowledge this).

[Grammar tip: possessives in English end in an apostrophe-s combination. As an exception, this 's is omitted for plurals ending in *s*, such as *the pedestrians' crossing*. The exception about plurals does not apply to words that are not plural but happen to end in *s*. See rule #1 in Strunk and White's *Elements of Style*⁴ for details.]

We all want our models to lean to the objective side; how do we make that happen? Here we get to the rift between the different schools. The Frequentists lean heavily on the CLT and other such objectively-proven and subjectively-applied theorems. Sometimes the CLT is preeminently defensible (as alluded below, typically in the world of controlled experiments on non-living things); sometimes it's an irritating refusal to

³<http://users.soe.ucsc.edu/~draper/draper-census-bureau-workshop-2011.pdf>

⁴<http://www.bartleby.com/141/strunk>

really accept the ambiguity of the situation.

Central limit theorems, when applicable, work. If elements really have a fixed probability of leaving the set at any given time, then lifespans really do follow a Poisson distribution. Inquiries in the physical sciences typically have a derivable distribution like these built in to the controlled experiment, while in the social sciences we typically have no idea what we're doing, and can't control anything in the data. In both the physical and social sciences, the model is a human-imposed structure on the data, but we can already see that in some situations people will be more inclined to label the final statements about parameters as objective and in some will be inclined to call them subjective. The reader will note that Bayesians tend to hang out in the social sciences.

The Bayesians point out that there is still subjective information to be had, and maybe we can codify it all into a prior distribution. OK, great, you've concentrated all your subjectivity into one place. This forces a certain format on the model, and makes you no more or less objective than you were before, but does make it easy to incorporate correctly-codified additional information.

The typical Bayesian setup involves using a prior distribution, which is allegedly subjective auxiliary information, and a likelihood distribution, which is often described as the real model. This core-plus-auxiliary terminology creates a lot of trouble, and life is easier when we take the entire pipeline, prior plus likelihood, as the unit that we call a model. [This is how Apopenia does it: the `apop.update` function returns a single model estimated using the data and setup you provided.] Now we're back where we were with the Frequentists: a single model that expresses what information we have and what beliefs we have about the structure of the model, including some portion of its inner workings that is clearly subjective and some portion that is more clearly objective. Any claim that we could somehow componentize the objective and subjective is chimerical anyway.

[There are people who take the principle of maximum entropy to be an objective rule about the universe, which has been observed to hold many times over. I have a coworker who gets really pissed off when anybody mentions the principle of maximum entropy, and thinks of any work using it as subjective bunk. The maxEntropists are making another attempt to confront the problem that we have too little information and have to fill it in with some sort of assumption, and having used a different principle, wound up with a different final structure to the model.]

The descriptive models of the different schools embody subjective information in different ways. But in the end, they all have the same form of an assumed structure $P(d, \beta)$.

The inferential step To this point, I've been talking about descriptive statistics: the art of combining objective facts and subjective modeling decisions into a single final model, then estimating its parameters from the data. Bayesians, Frequentists, maxEntropists tend toward different models, that each feels is on the better side of the objective-subjective scale. Then we have the inferential step, where we take the model's final distribution describing the parameters, $P(\beta|d)$, and try to say something interesting about β . A lot of ink is spilled at this point, about how much one should rely on certain intervals and whether to call them *confidence intervals* or *credible intervals*, and how much one can use a point on the distribution, like $P(\beta = 3|d)$ and

what you would name such a thing. [I began writing this thanks to a blog post by a coworker⁵ on the different schools; his post focuses on the interpretation of the inference step.] Also, we should again remember to distinguish between human Frequentists and crappy undergrad stats textbooks.

I do like to make one quibble about language: when an author says something like ‘given the data, we fail to reject the null hypothesis with 91% certainty’, a key element is left out; this should read ‘given the data *and the model*, we fail to reject the null hypothesis with 91% certainty’. We developed a model, and every statement we make from that point on will rely on that model and its subjective and objective foundations. Careful authors tend to get this right, undergraduate stats textbooks never do.

Quibbling after that point is largely a restatement of the same philosophy of science question from the descriptive step: the probability distribution of a parameter is fundamentally unobservable in most of the situations we’re dealing with, so how do we express that we’ve imposed a model and are stating probabilities using this invention of ours, and how do those invented probabilities relate to reality?

Summary sentence: people have trouble dealing with things that are not easily categorized, and $P(d, \beta)$ is fundamentally a combination of observable, unobservable, subjective, and objective. This is conceptually difficult—I think I’ve used the term *mind-blowing* before. Among the reasonable people, some (which I have called Bayesians) class models as subjective and then from that point reach toward saying something objective; some (what I’ve called Frequentists) class models as fundamentally objective, but acknowledge and strive to accommodate subjective aspects. They eventually meet in the middle, though their models will have different forms expressing the differences in perspective.

⁵<http://civilstat.com/?p=1076>