

The model asserts the likelihood

Ben Klemens

12 July 2013

[Part of a series of posts that started here (entry #146)]

Recall from an earlier post (entry #147) that some authors define a model to be equivalent to a parameterized likelihood distribution. That is, there is a one-to-one mapping between models and likelihood functions: if you give me a likelihood, then that is the core of my model, from which I can derive the CDF, RNG, and many implications. Conversely—and I'll discuss this in detail next time—if you give me a well-defined model, even one defined in non-probabilistic terms, then there exists a likelihood function giving the probability of a given combination of parameters and data.

For this post, I'll stick to some very typical likelihood functions from Stats 101: the Normal distribution, which, given μ and σ , gives the probability that a given observation will occur (herein, $L_N(x, \mu, \sigma)$); and ordinary least squares, which, given β s and the variance of the error term σ , gives the probability that an observation (Y, X) will occur as $L_N(Y, \beta X, \sigma)$.

This entry will be about social norms. The short version is that there is no truly objective measure of a model, because most measures use the model's likelihood to make statements about the likelihood of modeled events; instead we have certain established customs.

Reading numbers off of the map The Normal distribution is the limit of many very plausible physical processes, in which several independent and identically distributed elements are averaged together, so it is often very plausible that repeated observations from a real-world process follow a Normal distribution. The model 'X is Normally distributed' is really the model 'X is the mean of several iid subelements' after a few theorems are applied to derive the implications of the iid model. There are many situations where that model is easily plausible, and in those cases it is easy to confound the map and the terrain. That's the joy of a good map.

For a linear regression, there is rarely a hard-and-fundamental reason for the dependent variable to be a linear function of independent inputs [maybe (number of eyes in the population) = $\beta_0 + \beta_1(\text{number of noses}) + \epsilon?$]. In common usage, it is nothing but a plausible approximation, and we can't break it down to a simpler, more plausible model the way the Normal model is implied by the iid subelements model.

And even if the overall picture is plausible, it is easy to push the model to the point of reading features from the map that may or may not be on the actual terrain. The stats student runs a regression, finds that the model assigns probability .96 to the claim

that $\beta \neq 0$, and muses how that was derived from or is reflected in facts about the real world. It isn't: it's a calculation from the model likelihood.

The claim that we reject the null with $p = 0.04$ is calculated using a data set and the model likelihood, and we colloquially call it a test of the model. That is, we test the model using the model. To make this preeminently clear, here's another valid model, letting $\hat{\beta} = (X'X)^{-1}X'Y$, (i.e., the standard OLS coefficients); then for our model, let the coefficient be:

$$\beta = \begin{cases} \hat{\beta}, & |\hat{\beta}| > 1 \\ 1, & |\hat{\beta}| \leq 1 \end{cases}$$

This is a useful model: over many arbitrary data sets, we know that β will correlate to the slope of the X - Y line, albeit a little less than $\hat{\beta}$ does. But it has the massive advantage over OLS that we reject the null hypothesis that $\beta = 0$ with 100.00% certainty. Just as with OLS, the model as stated is not inconsistent with the data—any data—given this model's assumptions. Now that you have an internally consistent model where the parameters are always significantly different from zero, you can publish anywhere, right?

Objective functions The other way to describe the β s for the ordinary least squares model is, as per the name, to find the line that minimizes squared error $(Y_i - X_i\beta)^2$. Your favorite undergrad stats textbook should show you a diagram with a line of best fit through a scatter of data points, and vertical dotted lines from each point to the line; it is those squared lengths that we are minimizing. Why not the shortest distance from datum point to the line, which is typically a diagonal line, or the horizontal line from datum to line of best fit? The only honest answer is that the math is easier with the vertical distance.

So you could go to any seminar where the speaker fit a regression line and ask *why did you minimize the squared vertical length instead of the Euclidian distance between points and the line?* or better still, *if you used Euclidian distance, would your parameters still be significant?* The speaker might respond that it's implausible that there's a serious difference in results given the two objective functions (because the odds are low that the speaker really tried any alternative objective functions), or might cite computational convenience. Or the speaker might not have a chance to answer, as the other seminar attendees cut you off for being an obstruction to the progress of the seminar. Linear regression has been around long enough that all of its epistemology has been worked through, and discussing it again is largely rehash.

In a much-recommended paper by William Thurston¹, he explains how he got over a similar block regarding the very objective world of mathematics:

When I started as a graduate student at Berkeley, I had trouble imagining how I could “prove” a new and interesting mathematical theorem. I didn't really understand what a “proof” was.

By going to seminars, reading papers, and talking to other graduate students, I gradually began to catch on. Within any field, there are certain

¹<http://arxiv.org/abs/math.HO/9404236>

theorems and certain techniques that are generally known and generally accepted. [...] Many of the things that are generally known are things for which there may be no known written source. As long as people in the field are comfortable that the idea works, it doesn't need to have a formal written source.

Linear regression has been widely accepted, so authors don't have to discuss the fundamentals when presenting regression-based papers. But it is an arbitrary set of assumptions that are salient for their historical place and human plausibility. Setting aside its social position, it is on the same footing as any other internally consistent model, notably in that claims about the model's consistency with data are made using the likelihood that the model itself defines and are therefore plausible iff the model itself is plausible.