Peer Review for Yoni Brande

PHSX 815
Project 3

TRISTON RUISECO

April 8, 2021

"To Peer Reviewer: Am I framing this problem correctly? Is it sufficient to use a prebuilt minimization routine (like in Python or something) and use its error estimates to talk about how well the parameter can be estimated from the simulated data? Do I just do this for smaller or larger simulated datasets? Is my math on the Bernoulli -> likelihood -> minimization function right?"

Am I framing this problem correctly?

The way I understand this project, your approach definitely seems correct: generate data -> define function with output proportional to likelihood of a hypothesis given generated data -> maximize function (or minimize 1/f) to find most likely hypothesis. My only comment on anything to do with this is that the $\propto$ symbol indicates direct proportionality not inverse proportionality as suggested in your mock paper, but that's just semantics. Given that what you said is true the way that you meant it, you clearly understand what's going on.

Is it sufficient to use a prebuilt minimization routine (like in Python or something) and use its error estimates to talk about how well the parameter can be estimated from the simulated data?

This is certainly worth speaking to Professor Rogan about (I plan to discuss this with him myself as well), but my reasoning on this is that the error estimates on the minimization routine are the only error associated with the precision of the parameter associated with hypotheses, which to me is not necessarily the error associated with using maximum likelihood estimation to identify the "true" parameter of the system with some degree of confidence unless the routine also incorporates some statistical analysis. I believe the lecture material most relevant to this was in the March 23, 2021 lecture in the "Measurement" section, so maybe check that out, particularly the discussion starting around 57:45.
As for the sufficiency of using a prebuilt routine, I imagine that, just like with sorting, using pre-existing highly refined implementations is a vastly more efficient route than using custom implementations unless there is a lot to be gained by using a custom made routine that is constructed to account specifically for something special about the nature of your data.

Do I just do this for smaller or larger simulated datasets?

The requirement as indicated by the rubric is that we be able to generate data with many different true values of a parameter, then see how well our maximum likelihood estimation does at finding the correct value. To me this implies that the bare minimum case is generating data for several

different true values of a parameter, then see how the uncertainty of the estimated parameter scales according to one tweak, i.e. the true value of the parameter, the size of the dataset, etc.

Is my math on the Bernoulli -> likelihood -> minimization function right?

In your final formula you have $f(p) = -[\xi \log(p) + (n - \xi) \log(1 - p)]$ "where $\xi = \sum_{i=1}^{n}$.", and, if I'm not mistaken, I think here you meant to instead write "where $\xi = \sum_{i=1}^{n} x_i$." If so, I started from the same point with the same goal and obtained the same result myself. So, either we both took slightly different paths from the same starting point and arrived at the same wrong answer (seems unlikely), or you're good to go!