LIONMAIL @COLUMBIA

Michael H Andreae <ma2196@columbia.edu>

please comment on my revised research plan

Bob Carpenter <carp@alias-i.com> Thu, Mar 12, 2015 at 9:36 PM
To: Michael H Andreae <ma2196@columbia.edu>

Overall, I think this is very compelling. It looks very solid on both the stats and the clinical side, which is what I think NIH looks for (and why we can't write these things on our own!).

First off, Andrew hates hates hates photos of researchers. And that one of Bayes isn't even confirmed to be him :-)

He's also not a big fan of acronym soup.

Andrew suggested that we put the plan in the form of a scientific hypothesis we'll investigate (i.e., whether modeling these things will help with prediction). I see you get there later, but maybe you should start with the hypothesis.

You probably want to cite Dr. Gong's grant by directorate/number if this is going to NIH.

I wouldn't say "complex" as this sounds like a bad thing. I'd say "richer" or "more explanatory" or even "sophisticated" or "advanced" if you must (I'm not a big fan of these last two).

"groundbreaking" is one word (it must be impossible for you to tell, coming from German, because English is so damn random).

"envelop" -> "envelope"

I'd be surprised if there wasn't any previous work on Bayes and medical records.

"STAN" -> "Stan" --- it's not an acronym, it's named after Stanislaw Ulam.

I'd just drop "real-time".

It's nice you spell out the "unique innovation" for them. Grant reviewers love it when you make their life easy because they often have to fill out forms and suss this kind of stuff out themselves.

The major thing you have to do is separate out what *your* contribution is going to be above and beyond what Dr. Gong's already doing. And how you'll test it.

I think "ultra-fast" may be pitching Stan a bit high --- it's very fast for MCMC, but slow compared to other approaches.

I don't know about Andrew, but I'm not a big fan of AUC evals, because in practice they are noisy and in practice you wind up going with a single operating point based on a cost/benefit analysis, not operating along the whole ROC curve.

"live saving" -> "life saving". Voicing is also tough for Germans, like v/f and d/t! (Sorry, can't resist the linguistic commentary).

Andrew's also not a big fan of p-values or Excel-generated graphics, so I'm sure he'll have some commentary on Figure 1. He's especially unhappy about graphs with quantities in them that aren't explained, such as what those p-values are tests for.

Wow --- I didn't know "attenuate" had meaning other than the "make more peaked" one we use in signal processing. Yours is indeed a correct usage!

"A pragmatic trials" --- "a" is singular and "trials" is plural, so it should be either "a pragmatic trial" or just "pragmatic trials". You can read all about this in my second book on semantics, which has a model of plural syntax and semantics in English :-)

For Table 1, the same issue --- you need to describe what the numbers mean in the caption --- if you don't, Andrew will ask you to. Compare the figures in his books and papers. Also, why is there a question in the caption?

I like the section on "Big Data" --- that's exactly how we've been trying to explain things ourselves. (People always like it when you agree with them!)

I'm not a big fan of the "Krushcke" diagram as in figure 2 because it only shows one possible shape for the distributions. I'm going to take a wild guess that Andrew's not going to like the term "distrogram", but I've been wrong before.

Also, I think your particular diagram is confusing. I assume you want to give beta a fixed prior and make beta0 a random effect (varying intercept in Andrew's terminology). Then beta0[i] for group i gets a hierarchical prior normal(mu,sigma), with hyperpriors on mu and sigma. What is the outcome measurement y[n]? Is it binary (Bernoulli) or repeated binary (binomial)? Assuming it's binary, that'd produce a model like this:

```
for (k in 1:K)
  beta[k] ~ normal(0,5);

mu ~ normal(0,10);
sigma ~ cauchy(0,2.5);

for (i in 1:I)
  beta0[i] ~ normal(mu, sigma);

for (n in 1:N)
  y[n] ~ bernoulli(inv_logit(beta0[ii[n]] + x[n] * beta));
```

where x[n] is a vector of patient-level predictors, and ii[n] in 1:I is the patient group and y[n] is a boolean indicator of outcome.

You don't need quotes on "partial pooling" -- you only want to use quotes when you mention a phrase or when you are trying to scare people. If it's a technical term, italicize it if you must.

I think you want to cite BDA 3, and for pooling, maybe the original Efron and Morris work on the Stein estimator using empirical Bayes. I don't know of any theoretical proofs. There's also more on pooling in Gelman and Hill than in BDA. You migth also want to site Efron's big data in bio book; it uses empirical Bayes (i.e., max marginal likelihood, which uses the data twice and is only approximate), but we should be able to do it with full Bayes using Stan. If not, we can

do MML. Efron, one of the most influential statisticians in the late 20th century wrote a nice paper on Bayes:

http://statweb.stanford.edu/~ckirby/brad/papers/2005NEWModernScience.pdf

And here's his book:

http://www.cambridge.org/us/academic/subjects/statistics-probability/statistical-theory-and-methods/large-scale-inference-empirical-bayes-methods-estimation-testing-and-prediction

The original Efron and Morris paper is this:

http://www.stat.cmu.edu/~acthomas/724/Efron-Morris.pdf

The discussion of institutional learning and bias is great. More people should be aware of this effect, where the experiment affects its own measurements.

One answer to how all these effects can be integrated is with Gaussian processes, as in Andrew and Aki's pet birthday problem (cover of BDA 3).

The contrast isn't Bayes/non-Bayes for hierarchical model, but more of full Bayes vs. max marginal likelihood. There aren't maximum likelihood estimates for hierarchical models.

I'm not sure the neon sign from the Wikipedia is doing you any favors, but then I don't know how doctors react to things like this. And speaking of figure 5, the likelihood is P(B|A). But this isn't really the likelihood, as you're confusing event probabilities and densities. The likelihood is a density function. Of course, in discrete cases, the event probability and the likelihood are the same function, even though they are conceptually different.

New information is that event B occurred, not p(B).

You don't show any integrals. They come in on the denominator when working with densities. I dont think this is right:

"Instead of analyzing the odds, we simulate throwing the dice repeatedly."

I think you mean "analytically deriving the odds". I'm not keen on dice or coin examples (and neither is Andrew --- you'll note there's not a single game of chance in BDA).

If we throw dice repeatedly, that's probably giving us a simulation (i.e., Monte Carlo) estimate of the odds.

"allow to push" ---> "push"

"STAN" -> "Stan".

Andrew didn't develop Stan, Matt and I did, but whatever, it's winner-take-all in academia and that may sound better for the grant.

"Marcov" -> "Markov"

Step size itself is tuned, as well as step size * number of steps. Too low a step size wastes computing power wheras too large loses efficiency.

Matt developed NUTS. Gelman was just on the paper. But hey, it's still winner take all and Andrew gets the credit. He's like the last author of a bio or

medical paper, rather than the first. He should get the credit for hiring me and Matt, at least :-)

You need some help with the rest of the formulas and model, but I'm sure Andrew can do that as your mentor :-)

Why do you think we won't be able to run x-validation? If we can fit the model, we can just throw x-validation at a cluster and be done with it. Easy-peasy.

I don't see how you safeguard against "overfitting" by throwing in more data. You need to calibrate on out-of-sample data.

Natural language in this domain is a complete mess and insanely noisy. I know, having worked on it. What kind of NLP techniques do you plan to use? The proposal doesn't say.

OK, I can't help but tell you that it's not inverse-log (that'd be exp), but inverse logit (i.e., 1/(1 + exp(-x)), the logistic sigmoid function).

Timeline: "mentors" -> "mentor's"

But I don't see a timeline for you.

Shouldn't that be Drs. Gong, Gelman, and Hall? Or is it Dr. Gong and Mr. Gelman to doctors?

"STAN" --> "Stan"!!!

I hate calling places "prestigous". If they don't know its prestige, telling them won't help.

"specially" -> "specialty"

Good luck!

- Bob

P.S. You guys should submit a paper on ShinyStan to JStatSoft.

[Quoted text hidden]

> <K01_Research_Plan_Michael_Andreae.tex><K01_Research_Plan_Michael_Andreae.pdf><K01_bibliography_24Feb15.bib><nihunsrt.bst><Timeline.pdf><BayesTheorem.jpg><Distrogram.pdf><posteriorpredictivecheck.png><Hamiltonian.png><Bayesian_imputation.pdf><SOFA_fig.png><Alert_trigger_fig.png><Thomas_Bayes.pdf>