Twitter Thread by \mathfrak{Michael "Shapes Dude" Betancourt}

\mathfrak{Michael "Shapes Dude" Betancourt}
@betanalpha

A short thread on what I think is particularly useful intuition for the application of computational statistics.

Deterministic methods, like variational Bayes, utilize _rigid_ approximations. Ultimately these methods try to find the best way to wedge an approximating solid into the desired solid. If the two shapes are close enough then you can get a good fit.

But this rigidity also means that the approximating solid can't always contort to fit into the desired solid. If the two shapes don't match up well enough then we will end up in lots of awkward configurations no matter how hard we push.

In particular it's hard to quantify how good the fit of the approximate solid into the desired solid might be without knowing the shape of the desired solid already, which isn't possible in practice. This is one reason why these methods have such limited empirical diagnostics.

Stochastic, methods, on the other hand, utilize more fluid approximations that are able to take the shape of their container given enough time. In this respect we can think of them like a liquid or gas filling the desired shape.

In some sense the nature of the stochastic approximation determines the viscosity of the fluid, and how quickly it's able to expand into the desired shape (especially when the shape has an awkward geometry).

All stochastic methods will fully expand into the desired shape _eventually_, but the rate of expansion may be too slow or erratic to be particularly useful in practice. That said exactly _how_ the fluid expands helps us understand how well the method is working.

These analogies are particularly useful when trying to compare methods. Fitting deterministic methods is often fast no matter how good of an approximation they provide; there's only so many ways to wedge incongruent shapes together.

On the other hand stochastic methods can quickly expand into simple shapes but they tend to slow down when encountering more complex shapes, for example in the case of statical inference when trying to fit degenerate posterior density functions.

All of this is to say that most analyses that argue "we had to use this deterministic approximation because Markov chain Monte Carlo was too slow" are bullshit. In almost all of these cases the disparity in speed is due to nasty posterior degeneracies and complex uncertainties.

In these cases the speed boost from using a deterministic method arises only by ignoring most of that uncertainty and providing a specious picture of the actual inferences.

Could that approximation actually be equivalent to accurate inferences for some implicit, more reasonably regularized model? Sure, but if you can't explicitly quantify that model how can you tell whether or not the implicit regularization actually is reasonable?

On the other hand improving the initial model with explicit regularization not only clearly communicates the modeling changes but also speeds up the stochastic approximations, too! Yes it's more work but work you should have been doing anyways.

When we say that Markov chain Monte Carlo _explores_ we really mean it. It will do it's damnedest to explore as much as it can, but it can only go so fast when exploring difficult terrain. That's often not a problem of the method so much as the terrain to which it's been sent!

Use that struggle to you advantage. Learn about the degeneracies in your inferences and motivate principled resolutions such as more careful prior models, auxiliary measurements, and the like. See for example https://t.co/vBLqlH7QdA.

And please, please stop assuming that your fits are slow only because Markov chain Monte Carlo is terrible and some magical new algorithm will come and make the scary chains go away. The pathologies have been calling from within your models the entire time!