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



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



The most frustrating aspect of working in computational probability/statistics is that it's basically impossible to construct algorithms that actually return well-defined probabilistic quantities and this results in no end of chaos.

At best algorithms can return approximations with quantifiable error, but to understand when approximations are useful you have to learn enough math to understand the exact result and how the algorithmic approximation relates to that exact result. Many do not do this.

More commonly programmers project the heuristics that prove successful in other computing problems -- pattern matching, type consistency, unit testing, relying on compiler errors, etc -- but these test only the algorithm and not the relevance of the algorithm to a stats problem.

Unaware of these subtleties many end up conceptually replacing the algorithm for the output being approximated, assuming that algorithmic properties are inherent and well-defined features of the underlying probabilistic/statistical system.

Needless to say this generalizes...poorly. Even worse: without formal knowledge of what is being approximated the poor generalization performance itself is easy to ignore and naive applications drift ever so steadily away from any well-defined mathematical objective.

At some point the algorithms drift too far and it becomes impossible to make any formal critique of the emergent heuristics. How can you criticize an algorithm when it's doing everything people believe it's supposed to be doing?

Of course statistical procedures and methodology in general follow the same pattern. Many methods that are abused today were at one point grounded in mathematical validation, only for those foundations to be gradually lost as the methods were taught less and less carefully.

All of this is to say that most math people going around critiquing heuristic methods and advocating for learning more irritating, burdensome math are not being exclusionary cynics: we're just trying to help ensure that contributions are, and will continue to be, constructive.

Probability and statistics is fundamentally difficult. Compromising the math by replacing subtle concepts with vaguely overlapping algorithms and heuristics leads only to superficial inclusion that typically does more harm than good.

In order to responsibly expand our communities we need recognize this challenge, respecting the math and consistently evaluating our current understandings while also working like hell to guide those starting their journey towards a meaningful destination.