August 9, 2021

Let's celebrate "A Gibbs sampler for a class of random convex polytopes", by Pierre E. Jacob, Ruobin Gong, Paul T. Edlefsen, and Arthur P. Dempster, while not forgetting the larger Dempster-Shafer theory.

By Glenn Shafer

Congratulations, Pierre, Robin, Paul, and Art. I lift my glass. Fifty-five years after Art introduced his simplicial method, you have implemented it for parametric models of reasonable complexity.

This is a testament to the steadfastness with which Art has pursued his vision. And a testament to your ingenuity.

I lift my glass to my fellow discussants, who are exploring how much farther this achievement can be pushed.

- More complex parametric models?
- Continuous ones?
- Alternatives to the simplex?
- Dependent observations?

There is much to do here.

Let's look even a little farther over the horizon.

The theory Art launched in the 1960s has had applications far beyond any statistical model.

- Far beyond models in which data is anticipated.
- Far beyond situations in which we know in advance the form of our data.
- Far beyond the essentially game-theoretic picture in which we begin with a protocol or "filtration" specifying step-by-step what we may learn in the future.

When I was working on Dempster-Shafer in the 1970s, I was most interested in problems where we evidence without a model. Evidence from various sources. Evidence that might nevertheless be quantified (if only on a Likert scale), perhaps discounted, perhaps combined by Dempster's rule.

This is the kind of problem that attracted attention in AI in the 1980s and led to the moniker "Dempster-Shafer".

Why did Dempster-Shafer attract more attention in AI, than in math stat? Because AI was not saddled with the concept of probability that has been math stat's birth-right and birth-curse.

Since Laplace, math stat has meant probability distributions,

- perhaps known (the Bayesian case),
- perhaps partially known (the parametric case),

- perhaps less known.

The options leave room for endless and repetitive argument about what probabilities mean, how to learn about them, what to do with them, but the probabilities have always been there. In the 1980s, at least, artificial intelligence was not saddled with this heritage.

For me, Dempster-Shafer belief functions have always been about problems where we have evidence but no numerical probabilities, known or unknown.

In my old-fashioned opinion, problems where deliberate about fact are of this type:

- Evidence but no probabilities.
- No known probabilities.
- No probabilities we can agree on.
- No unknown probabilities either.

To my recollection, even statisticians were comfortable with admitting this in the early 1970s. Today I'm not so sure.

A lot of things have changed in the past 50 years. For one thing, we're better at making probability predictions. We're better at making **good** probability predictions.

Consider a sequence of situations where you will give probabilities for a specified event based on data of a specified form:

- Every day, the probability of rain the next day.
- Before every ball game, the probability the home team will win.
- For every anesthesia, the probability the patient will survive.

Here you can do a good job.

Here is what I mean. Consider all probability predictors that are functions of data of the specified form. Then asymptotically, you can make probability predictions as good as the best of these *a posteriori*. This is a machine-learning type of statement; it is demonstrated, for example, in Chapter 12 of my 2019 book with Vladimir Vovk, *Game-Theoretic Foundations for Probability and Finance*.

But in problems where we deliberate, we seldom have this sequence of past similar problems. We have a unique case. Is Gracchus the murderer? The prosecutor may have a sequence of past cases. The defense has a different one.

What do we do? We make arguments. Weigh them. Take each for what it's worth. Combine them. No numbers.

- Like the great 16th-century Jesuit philosopher Luis de Molina did.
- Like courts still do.
- Like financial auditors do.
- Like you and I do.

The great 17th century mathematician Jacob Bernoulli proposed improving this accepted methodology by putting in a few numbers. He prized his law of large numbers not because it promised probabilities for the unique case, but because it might give probabilities for different arguments.

- Each argument relies on different experience.
- Estimate a probability for each argument,
- Weigh their relevance,
- Discount as appropriate,
- Combine by Dempster's rule.

This is what "Dempster-Shafer" means outside math stat. People have built general inference engines for it.

Can the ideas of this paper help us build or improve Dempster-Shafer computation at this level of generality?