How to Decide How Many Contributors in a Mixture

Charles Brenner, PhD

Abstract of presentation at 2019 AAFS meeting, Baltimore Md.

After attending this presentation, attendees will appreciate that defining "contributor" is not obvious, but once the word is understood in a sensible and practical way it is possible to decide how many contributors in a DNA mixture.

This presentation will impact the forensic science community by promoting usefully clear thinking about difficult and critical problems in evaluating DNA evidence.

All actual mixture programs calculate in a way that utilizes a "number of contributors" parameter. If one assumes an inappropriate number, the resulting final computation of evidential strength is wrong, perhaps by a lot. A number of published studies suggest various ideas about estimating the number, but none of them even claim to be nearly always right.

There is a way. If the mixture program is designed to calculate likelihood ratios via explicitly calculating both numerator and denominator (prosecution and defense) likelihoods L_p and L_d, and can calculate each of those likelihoods for a range of numbers of contributors -- $L_p(n)$ and $L_d(n)$ for n=1, 2, 3, ... contributors -- then examination of individual likelihoods says everything there is to know about the appropriate number of contributors. How to use the information may be debated but a simple answer is to choose the numbers n and m, for each of prosecution and defense independently, that maximize the respective likelihoods $L_p(n)$ and $L_d(m)$.

Note that this method subsumes the possibility of different numbers for prosecution and for defense. That's a possibility that co-authors and I discussed in early papers [1, 2] that has since been given lip service as theory but in practice it has not only been overlooked, it is precluded by some popular advanced mixture programs.

A futile body of literature and software has appeared for guessing the number of contributors in a DNA mixture. Ideas range from merely counting observed alleles per locus, to black box pattern recognition learning programs. All have a substantial error rate. None of them explore what it means to be a "contributor", hence all implicitly assume that the definition is obvious.

Obvious? Does "contributor" obviously mean someone who is the ultimate source of even a single cell? That's a bright-line definition (and the definition implied by allele-counting guesses, which assume a binary mixture model) but not a useful one: an inconsequential contribution is irrelevant to a mixture program. Methods tested against real but lab-concocted data embody a "ground truth" concept of contributor. That's appealing for developing and testing the algorithm but for real casework "ground truth" is not merely elusive, it's undefinable. One confounding concern among several is that low-level "contributors" exist or not according to our choice of detection threshold. Another interesting and subtle complication is the possibility that the mixture looks like a different number of contributors depending whether the suspect is assumed present or not. That means that, as discussed above, for prosecution and defense there are two *different* numbers, a situation that no analysis looking at the mixture data alone can possibly get right.

In summary, the essential first step in attacking the "number of contributors" problem is to define our terms in the relevant and practical sense. Since the purpose of the answer is to feed it into a computer program, by "mixture" we don't mean a collection of chemical DNA; we mean a collection of numbers, the measurements that are fed into the program. By "contributor" and the number thereof, the appropriate operational definitions are based on the operation of an adequate mixture model. That is, they are whatever a good continuous-model mixture program implements them as. Sensible definitions as a foundation guide thinking onto the right track.

[1] Charles Brenner, Rolf Fimmers, Max Baur, "*Likelihood ratios for Mixed Stains When the Number of Donors Cannot be Agreed*," *Int J Legal Med* (1996) 109(4):218-219

[2] Peter Gill, Charles Brenner, John Buckleton, *et al*., "*DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures,*" *For Sci International* (2006)160(2-3), 90-101