F11   DNA Confronts Bayes and There's Trouble

Charles H. Brenner, PhD*, DNA-VIEW, Oakland, CA 94611-1336

Learning Overview: After attending this presentation, attendees will have learned that mathematics applied to the real world has its limitations, which we are beginning to come up against with some of the advanced problems that recent DNA methods try to deal with.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by suggesting humility, modesty, and some disappointment.

DNA evidence enters legal decision-making via a Bayesian framework. The framework has two components: an objective calculation of the evidentiary strength of the DNA, and a step in which the decision-maker (the court, judge, or jury) incorporates that information to sway their subjective opinion. Whether the swaying occurs according Bayes Theorem, or whether that's a convenient fiction and a vain hope, is irrelevant to the present discussion. The only relevant point here is that the mathematical calculation is an objective process. Everything subjective or otherwise outside the knowledge of the DNA analyst is relegated to the domain of the court where it sensibly belongs.

This framework has served well historically. The evidential value of DNA data is a Likelihood Ratio (LR) computed by mathematical calculation based on a model. "Model" means a mathematical model of reality—a quantitative theory as simplified as reasonable but as complicated as necessary—that describes how the DNA data comes about. A model that deals only with measurements—allele sizes, allele population frequencies—lies within the DNA analyst's proper purview.

A dilemma arises when the DNA analysis model requires a prior probability assumption about something than can't be thought of probabilistically, or which current scientific knowledge cannot provide.

Two examples arise mostly from advanced (i.e., continuous model) mixture analysis. The analysis requires assumptions about the number of contributors, and there is no mathematically sound way to choose. Deciding how to weigh among different numbers of contributors requires assuming for example that a three-person mixture is *à priori* (i.e., without even looking) more probable than a four-person, but it is hard to imagine data that allows such an inference. All one can do is hope that the choice does not matter much. Another vexing problem arises when adjusting a "degradation" parameter in the model in order that the calculation be correct, assuming degraded contribution from the suspect. If that substantially increases the LR, one has created a more satisfying explanation of the mixture—but that end doesn't justify the means. Even if the suspect happens to be connected to the mixture, that doesn't justify framing him to prove it.

Parentage attribution can be problematic. Suppose a child shares alleles with an alleged father at all but two or three Short Tandem Repeat (STR) loci. Three hypotheses stand out: (1) Hp=man is the father and there are a few mutations; (2) Ha=man is the uncle ("a" for "avuncular"); or (3) H0=man is unrelated to the child.

When Ha isn't under consideration, a reasonable DNA analysis reports the LR supporting Hp over H0, properly leaving the question of the prior probabilities for Hp versus H0 to the court. With two different versions of non-paternity—Ha and H0—similar could be done only by making an assumption as to the relative prior probabilities of Ha versus H0, but, at least for the DNA laboratory, that is not possible.

Further examples can be listed.

The conclusion from these problematic situations is not that we need better mathematics; the traditional Bayesian framework is the principled mathematical approach. Some mathematical-seeming problems simply don't have a clean solution.

Bayes, DNA Evidence, Limitations