LEGAL PROBABILISM

MARCELLO DI BELLO – ASU

HANDOUT #2 – AUGUST 29, 2024

1. BAYES' THEOREM

Recall that confusing $P(A|B)$ and $P(B|A)$ is known as the *inversion fallacy* or *prosecutor fallacy*. Bayes' theorem shows how the two probabilities are related, as follows:

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) = \frac{P(B|A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}P(A).$$

Bayes' theorem allows us to calculate the *conditional* probability of A given B from:
(i) the probability $P(A)$ regardless of B;
(ii) the probability $P(B)$, where $P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$;
(iii) the *likelihood* $P(B|A)$, i.e. the probability of B given A.

2. COLLINS AND BAYES' THEOREM

Let us stipulate that
(a) the guilty couple, in fact, fits the description D (blond, ponytail, mustache, etc.);
(b) the Collins match description D; and
(c) D has a frequency of 1 in 12,000,000.

Let M stand for *the Collins match the description D* and let G stand for *the Collins are guilty*.
Bayes' theorem tells us that

$$P(G|M) = \frac{P(M|G)}{P(M)}P(G) = \frac{P(M|G)}{P(M|G)P(G) + P(M|\neg G)P(\neg G)}P(G).$$

We can assume—simplifying a bit!—that
$P(G) = \frac{1}{n}$, with n the population of, say, Los Angeles and vicinities (maybe 6 million people?);
$P(M|G) = 1$; and
$P(M|\neg G) = \frac{1}{12{,}000{,}000}$.

So we have

$$P(G|M) = \frac{1}{\frac{1}{n} + \frac{1}{12{,}000{,}000} \times \frac{n-1}{n}} \times \frac{1}{n} = \frac{1}{1 + \frac{1}{12{,}000{,}000} \times (n-1)}.$$

With $n = 6,000,000$, we get

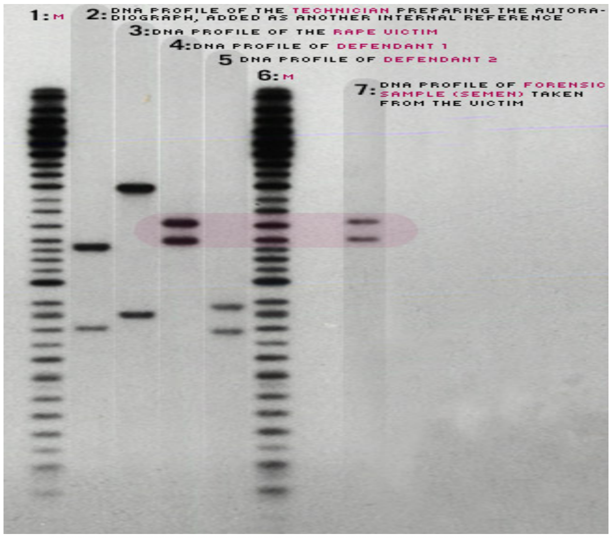
$$P(G|M) \approx \frac{1}{1+\frac{1}{2}} = \frac{1}{\frac{3}{2}} = \frac{2}{3}.$$

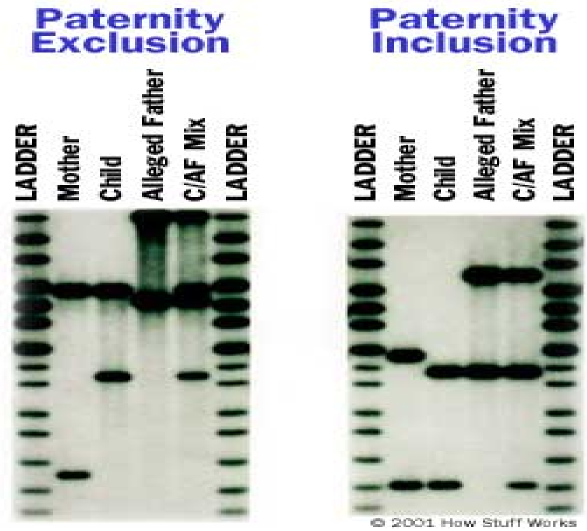
3. DNA EVIDENCE BASICS

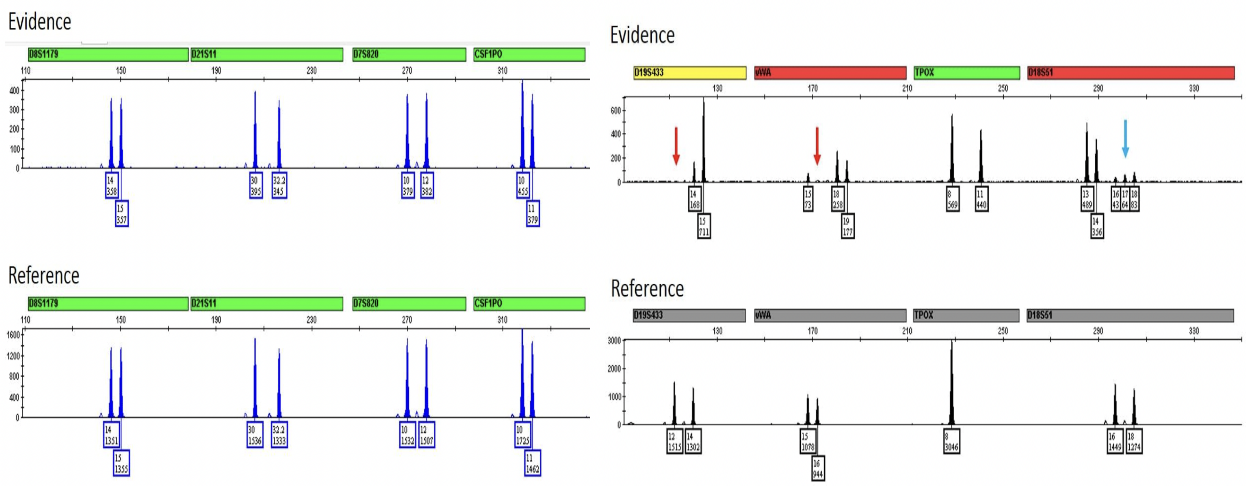
DNA evidence consists of two or three pieces of information:
(1) match between an individual's DNA profile and crime scene DNA profile;
(2) estimate of the DNA profile's frequency (also known as *Random Match Probability*);
(3) background information (e.g. shape, conditions, arrangement, location of the traces).

DNA evidence is used in criminal cases (e.g. rape) and civil cases (e.g. disputed paternity).

4. DECLARING A MATCH







A **tolerance window** can be used within which a match is declared [dichotomous statement]. *Alternatively*, we can use **degrees of congruence**.

5. ASSESSING THE FREQUENCY/PROBABILITY OF A DNA PROFILE

S1: frequency/probability of each STR (Short Tandem Repeat) allele;

$\Rightarrow$ counting how many times an allele occurs in the database, yielding f_i.

S2: frequency/probability of STR genotype (from both parents).

$\Rightarrow F_1 = 2 \times (f_i \times f_j)$ [why?]

S3: frequency/probability of entire DNA profile (currently 20 STR genotypes).

$\Rightarrow F = F_1 \times F_2 \times F_3 \times F_4 \times ... \times F_{20}$
[this value can be **astronomically small**; also called Random Match Probability (RMP)]

QUESTIONS: Are databases good data about allele's frequency? Can the *product rule* be applied?

6. THE MEANING OF THE DNA PROFILE PROBABILITY

"In the DNA context, I take some numbers (that are estimates of things like allele proportions ...) and stick them into a formula. Out comes a number and on the basis of that I assign ... a probability. That is a personal, subjective probability, which incorporates a set of beliefs with regard to the reliability/robustness of the underlying model. (Ian Everett, quoted in Buckleton, Triggs and Walsh (eds), Forensic DNA Evidence Interpretation, CRC Press, 2005, sec. 3.1.)

7. WHETHER DNA PROFILES ARE UNIQUE

"Ladies and gentlemen, his blood on the rear gate with that match, that makes him one in 57 billion people that could have left the blood . . . there is only five billion people on the planet. Ladies and Gentleman, that is an identification, okay, that proves it is his blood. Nobody else's on the planet; no one." *People v. Simpson*, Transcript (Sup. Ct.. LA County), 1995 WL 672671 (Sep. 26, 1995)

QUESTION: Is this right? Run the calculations. What probability would the formula $1-(1-p)^n$ give us, where n is the earth population and p the genetic profile frequency/probability? By plugging the numbers in the quotation, this probability is about 10%. What does that mean?

8. WHAT DNA EVIDENCE CAN ESTABLISH

Do not conflate *source*, *presence*, and *guilt*. Keep in mind the following inferential chain:

declared match → factual match → source → presence → involvement → guilt

QUESTIONS: Can DNA evidence **alone** establish guilt? Can it establish source?

9. BAYES' THEOREM—THE ODDS FORMULATION AND DNA EVIDENCE

Another formulation of Bayes' theorem, which makes calculations easier, is in terms of odds:

$$\frac{P(A|B)}{P(\neg A|B)} = \frac{P(B|A)}{P(B|\neg A)} \times \frac{P(A)}{P(\neg A)}.$$

In other words,

$$\textit{posterior odds} = \textit{likelihood ratio} \times \textit{base rate odds}.$$

The posterior probability $P(A|B)$ is usually given by $\frac{PO}{1+PO}$, were PO are the posterior odds.

Let G be the proposition that the defendant is guilty; let S be the proposition that the defendant is the source of the crime traces; let M be the proposition that the defendant and the traces match; let F represent the frequency/probability of the DNA profile in question. Using Bayes's theorem, we can arrive at the probability of G given M:

$$\frac{P(G|M)}{P(\neg G|M)} = \frac{P(M|G)}{P(M|\neg G)} \times \frac{P(G)}{P(\neg G)},$$

and the probability of S given M:

$$\frac{P(S|M)}{P(\neg S|M)} = \frac{P(M|S)}{P(M|\neg S)} \times \frac{P(S)}{P(\neg S)}.$$

QUESTIONS: What value to give to $P(S)$ or $P(G)$? Shall we put $P(M|S) = 1$ or $P(M|G) = 1$? What about $P(M|\neg S) = F$ or $P(M \neg G) = F$? Finally, how to arrive at the probability of guilt or source given DNA evidence?

10. Fallacies in reasoning with probabilities (esp. with DNA evidence)

Inversion fallacy
Base rate fallacy
Uniqueness fallacy

11. The cold hit controversy

In a cold-hit case, the accused is identified not via standard, old-school investigative methods but via a database search of DNA profiles, each associated with a person. Here is an argument that a DNA match in a *cold hit* case is less significant than in a *standard case*:

> There is a very high probability of getting, say, 10 consecutive heads if one makes a sufficient number of attempts at tossing a coin (*even if getting 10 consecutive heads is a very unlikely event*). Likewise, there is a very high chance of getting a match if the database is sufficiently large (*even if the profile in question is very rare*).

And here is a *counter-argument*:

> Relative to each attempt, the probability of getting 10 consecutive heads is the same. Likewise, the probability of finding a matching individual (given a low frequency DNA profile) is not affected by how many attempts one makes. Further, if one had a database containing all individuals on the planet, finding one match (only) would be a proof of the DNA profile's uniqueness. So, we could even say that the more profiles are searched, the more significant the match.

12. How to present DNA evidence with/without using numbers

Option 1: jurors are given a (declaration of the) DNA match, but no numbers.

Option 2: jurors are also given the frequency of the DNA profile or the RMP.

Option 3: jurors are also given the probability that the defendants is the source of the DNA traces at the crime scene (possibly through Bayes' theorem).

13. Should we do away with probabilities?

"How often have I said to you that when you have eliminated the impossible whatever remains, HOWEVER IMPROBABLE, must be the truth? We know that he did not come through the door, the window, or the chimney. We also know that he could not have been concealed in the room, as there is no concealment possible. Whence, then, did he come?" Conan Doyle, *The Sign of Four*.