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

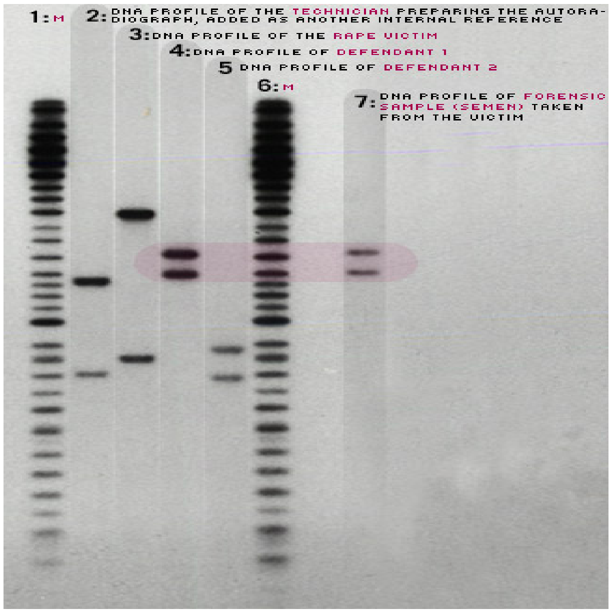
$$P(G|M) \approx \frac{1}{1+\frac{1}{2}} = \frac{1}{\frac{3}{2}} = \frac{2}{3}.$$

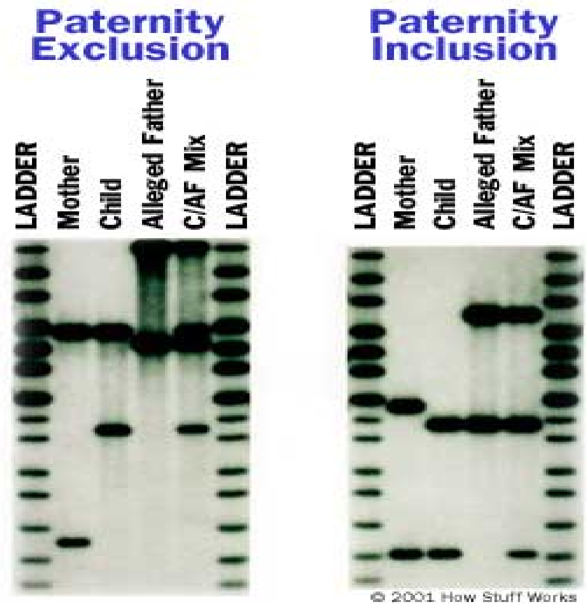
3. DNA EVIDENCE BASICS

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





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

QUESTION: Which one of the two approaches is better?

5. Estimating the frequency of a DNA profile

S1: determining frequency of each (STR) allele;

⇒ counting how many times an allele occurs in the database, yielding f_i.

S2: determining frequency of the STR genotype (from both parents).

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

S3: determining frequency of the entire DNA profile (usually, 12 or 13 STR genotypes).

⇒ $F = F_1 \times F_2 \times F_3 \times F_4 \times ... \times F_{13}$
[this value can be **astronomically small**; also called Random Match Probability]

QUESTIONS:
Regarding step (S1), are databases good indicators of an allele's frequency?
Steps (S2) and (S3) rely on **genetic models**; can we apply the *product rule* here?

6. The cold hit controversy

Argument that a DNA match in a *cold hit* case is less significant than in a *standard case*:

> There is a very high chance 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*).

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.

7. What can DNA evidence establish?

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

declared match $\rightarrow$ factual match $\rightarrow$ source $\rightarrow$ presence $\rightarrow$ involvement $\rightarrow$ guilt

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

8. Are DNA profiles 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 (Superior Court, Los Angeles County), 1995 WL 672671 (Sep. 26, 1995)

9. Can a DNA profile frequency be trusted?

"we recognize that we are considering an extreme extrapolation using these models. We are not operating near the center of their prediction range where they are more testable and tested. The models have been extensively tested in this central range and there is some considerable reason to believe that they are robust there, but they are still models and the probabilities produced by them are still untestable." Buckleton (2005), section 3.4.5.

10. What does a DNA profile frequency mean?

> "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 [*or a frequency, in the terminology I have been using*]. That is a personal, subjective probability, which incorporates a set of beliefs with regard to the reliability/robustness of the underlying model. ... Thus, I would not say ...that the probabilities [*i.e. frequency estimates*] are untestable estimates. I would ask—Is it rational for me to assign such a small match probability [*i.e. small DNA profile frequency*]?"
> Ian Everett, quoted by Buckleton (2005), section 3.1.

11. Bayes' theorem—the odds formulation

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.

12. BAYES' THEOREM (IN TERMS OF ODDS) AND DNA EVIDENCE

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 of the DNA profile in question. We want to know the probability of G given M and the probability of S given M. Bayes' theorem can be used as follows:

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

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

QUESTIONS:

- What value to give to $P(S)$ or $P(G)$?
- Is it correct to put $P(M|S) = 1$ or $P(M|G) = 1$?
- Is it correct to put $P(M|\neg S) = f$ or $P(M\neg G)$)?

13. HOW TO ARRIVE AT THE PROBABILITY OF GUILT GIVEN DNA EVIDENCE?

Possible answer: Bayes' Theorem, though some think we should not calculate the source or guilt probability.

14. FALLACIES IN REASONING WITH PROBABILITIES (ESP. WITH DNA EVIDENCE)

Inversion fallacy
Base rate fallacy
Uniqueness fallacy
Database fallacy

15. CAN WE DO AWAY WITH NUMBERS/STATISTICS/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*.

16. CAN WE PRESENT DNA EVIDENCE 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).