PROBABILITY AND THE LAW

MARCELLO DI BELLO – STANFORD UNIVERSITY

HANDOUT #4 – JAN 31, 2014

0. TAXI CABS AND BASE RATES

Imagine that there are two taxi companies, Green Cabs Inc. and Blue Cabs Inc., whose vehicles are respectively painted green and blue. There are no other taxi companies around. On a misty day a cab hits and injures a passerby, but it drives off. A witness reports that it was a blue cab. The witness is right only 80 percent of the time. This means that his *reliability* equals 0.8 in the sense that he gets the color right 80 percent of the time. Given the witness report, what is the probability that the taxi cab involved in the accident was in fact blue?

1A. 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.

1B. 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 $P(M\neg G)$)?

2. How the frequency of a DNA profile is estimated

(STEP 1) frequency of each STR allele;

$\Rightarrow$ count the number of a given STR allele in a given database, yielding f_i.

(STEP 2) genotype frequency at each STR locus.

$\Rightarrow F_1 = 2 \times (f_i \times f_j)$

(STEP 3) frequency of the entire DNA profile (with 14 STR loci).

$\Rightarrow F = F_1 \times F_2 \times F_3 \times F_4 \times ... \times F_{14}$

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?

3. 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)

4. 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.