The Mathematics of Mixtures
Wright State University
S. L. Zabell

August 30, 2003

There are two basic approaches to adjusting the DNA match probability whenever a mixture is present: one is the *random man not excluded* method, the other involves the use of the *likelihood ratio*.

1 Random man not excluded

Initial example: Consider the following common occurrence:

Four alleles detected in a mixture:

$$\begin{array}{rl} alleles & A, B, C, D \\ frequencies & a, b, d, d \end{array}$$

Possible consistent genotypes:

Four homozygotes: AA, BB, CC, DD.

Six heterozygotes: AB, AC, AD, BC, BD, CD.

Question: How does one allow for the multiple consistent genotypes?

Method 1: Brute force? Compute

Homozygotes frequencies: $a^2 + b^2 + c^2 + d^2$.

Heterozgyote frequencies: $2ab + 2ac + 2ad + 2bc + 2bd + 2cd$.

This equals: $(a + b + c + d)^2$.

Method 2: Insight! Think of A, B, C, D as one allele E.

The frequency of E is $a + b + c + d$.

The frequency of EE is $(a + b + c + d)^2$.

Consistent genotypes are homozygous EE.

Generalization to n alleles In general, if there are n alleles present, then there are

$$\underset{\textit{homozygotes}}{n} \quad + \quad \underset{\textit{heterozygotes}}{\frac{n(n-1)}{2}} \quad = \quad \underset{\textit{total}}{\frac{n(n+1)}{2}}$$

possible genotypes; and if these have frequencies $a_1, a_2, ..., a_n$, then the match probability is

$$(a_1 + a_2 + ... + a_n)^2.$$

2 The likelihood ratio

The scenario: Two persons are known to have contributed to a mixture.

The evidence:

E: Alleles present are A, B, C, D; defendant's profile is A, B.

There are two competing hypotheses:

H_0: two random individuals contributed to the mixture;

H_1: the suspect and a random individual contributed to the mixture.

If H_0, then the possible genotypes of the two contributors are:

$$AB, CD \qquad AC, BD \qquad AD, BC$$

The probabilities of these are (using Hardy-Weinberg twice) respectively

$$2(2ab)(2cd) \qquad 2(2ac)(2bd) \qquad 2(2ad)(2bc)$$

(Note these are all equal to $8abcd$.) These add up to $24abcd$.

The likelihood ratio is therefore:

$$\frac{P(E \mid H_0)}{P(E \mid H_1)} = \frac{24abcd}{2cd} = 12ab.$$

3 Some useful facts

The RMNE is not always conservative relative to LR To see this, suppose that c, d are very small. Then

$$P_{RMNE} \approx (a+b)^2 = 2ab + a^2 + b^2; \qquad LR = 12ab.$$

One need not have

$$a^2 + b^2 > 10ab.$$

For example, if $a = b = 0.1, c = d = 0.01$, then

$$P_{RMNE} = (0.22)^2 = 0.0484; \qquad LR = 0.12.$$

The RMNE can also be thought of as a likelihood ratio. Suppose the evidence is

F: the suspect type is consistent with A, B, C, D

but his genotype is not supplied. Then

$$\frac{P(F \mid H_0)}{P(F \mid H_1)} = \frac{(a+b+c+d)^2}{1} = (a+b+c+d)^2.$$

Note that the RMNE calculation assumes less than is usually known (since it ignores the genotype of the suspect) but the LR may assume more than is known (since it uses the exact number of contributors to the mixture is known).

4 The 70% rule

Suppose a mixture of DNA from two individuals is present in a sample; and the two individuals have the same genotype A, B. Let a_1, b_1 and a_2, b_2 denote the peak height contributions of the two individuals. If

$$\alpha < \frac{a_1}{b_1} < \beta \quad \text{and} \quad \alpha < \frac{a_2}{b_2} < \beta,$$

then

$$\alpha < \frac{a_1 + a_2}{b_1 + b_2} < \beta.$$

Proof. One has $\alpha b_1 < a_1$ and $\alpha b_2 < a_2$, hence $\alpha(b_1 + b_2) < a_1 + a_2$, hence

$$\alpha < \frac{a_1 + a_2}{b_1 + b_2}.$$

Similarly, one has $a_1 < \beta b_1$ and $a_2 < \beta b_2$, hence $a_1 + a_2 < \beta(b_1 + b_2)$, hence

$$\frac{a_1 + a_2}{b_1 + b_2} < \beta.$$

□

The point is that one cannot have a mixture of two A, B heterozygotes, each satisfying the 70% rule, and yet the mixture of the two not satisfying the rule.

References

[1] Clayton, T.M., Whitaker, J.P., Sparkes, R., and Gill, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International* **91**,55-70.

[2] Curran, J. M., Triggs, C. M., Buckleton, J. S., and Weir, B. S. (1999). Interpreting DNA mixtures in structured populations. *Journal of Forensic Sciences* **44**, 987-995.

[3] Evett, I. W., Buffery, C., Willott, G., and Stoney, D. (1991). A guide to interpreting single locus profiles of DNA mixtures in forensic case. *Journal of the Forensic Science Society* **31**, 341-47.

[4] Evett, I.W., Gill P.D., and Lambert, J.A. (1998). Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences* **43**, 62–69.

[5] Evett, I.W. and Weir, B. S. (1998). *Interpreting DNA Evidence*. Sinauer, Sunderland MA.

[6] Gill, P., Sparkes, R., Pinchin, R., Clayton, T., Whitaker, J., and Buckleton, J. (1998). Interpreting simple STR mixtures using allele peak areas. *Forensic Science International* **91**, 41-53.

[7] Perlin, M. W. and Szabady, B. (2001). Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Journal of Forensic Sciences* **46**, 1372–1377.

[8] Weir, B. S., Triggs, C. M., Starling, L., Stowell, L. I., Walsh, K. A. J., and Buckleton, J. (1997). Interpreting DNA mixtures. *Journal of Forensic Sciences* **42**, 213-222.

[9] Wang, T. (2002). Comment on Perlin and Szabady. *Journal of Forensic Sciences* **47**, 1176.