

Combining the evidential value of multiple partial DNA profiles using a lower bound for the Likelihood Ratio

Jacob de Zoete¹, Marjan Sjerps²

¹ Queen Mary University of London, ² Netherlands Forensic Institute
j.dezoete@qmul.ac.uk

Autosomal DNA profiles

Example #1 On the crime scene of a burglary, two crime stains containing biological material are secured. The stains provide two partial autosomal DNA profiles with partly overlapping loci. Both profiles match with the autosomal DNA profile of a suspect. The prosecution states (hypothesis H_p) that the suspect is the donor of both crime stains. The defense states (hypothesis H_d) that the suspect is not the donor of any of the crime stains.

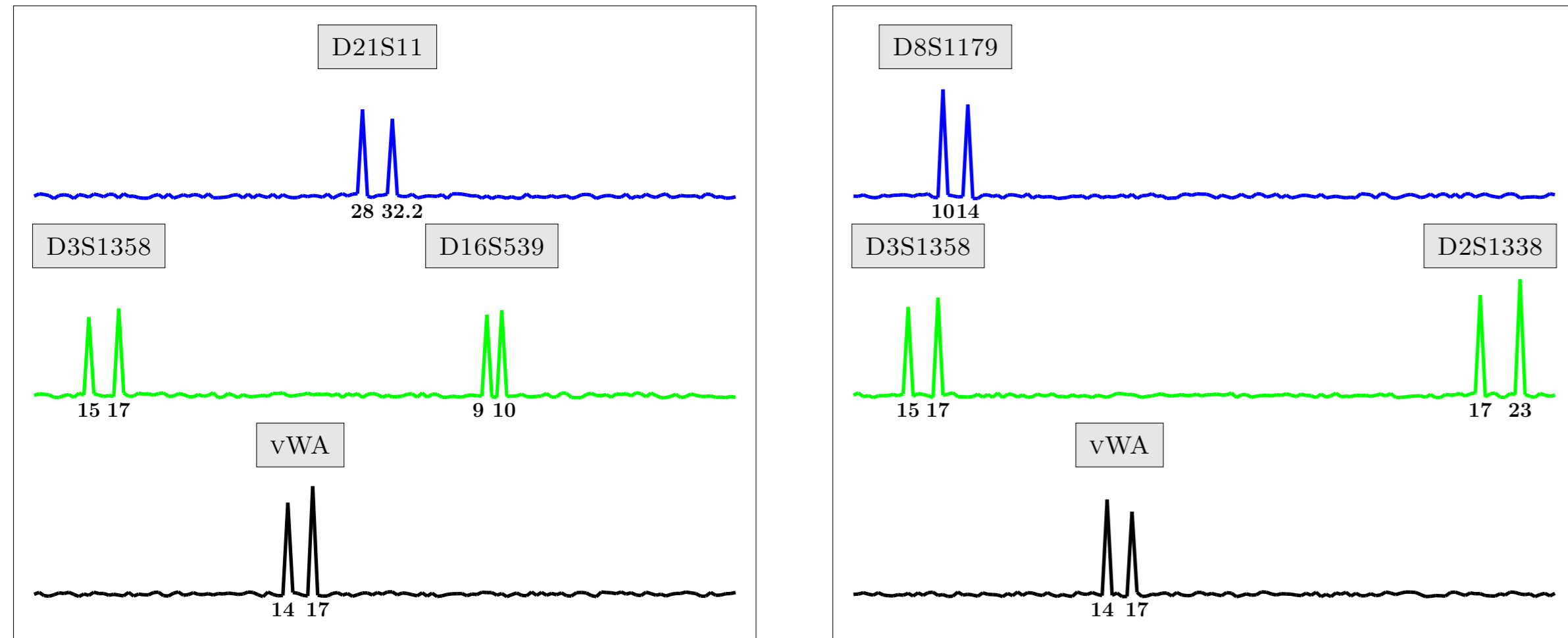


Figure 1: Partial autosomal DNA profiles - crime stain 1 and 2

locus	crime stain 1		crime stain 2		suspect	
	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
D3S1358	15	17	15	17	15	17
VWA	14	17	14	17	14	17
D16S539	9	10			9	10
D2S1338			17	23	17	23
D8S1179			10	14	10	14
D21S11	28	32.2			28	32.2

Table 1: Loci and alleles autosomal DNA profiles, Example #1.

Separate evaluation

The conditional probabilities corresponding with the DNA profiles obtained from crime stains 1 and 2 under H_d are computed using the R package *forensim*. The allele frequencies obtained from the Caucasian population data for the NGM STR loci from (Budowle et al., 2011) are used as background population data. The genotype of the suspect is denoted G_S .

$$\begin{aligned} \Pr(\text{crime stain profile 1 and 2} | H_p, G_S) &= 1 \\ \Pr(\text{crime stain profile 1} | H_d, G_S) &= 6.16 \cdot 10^{-7} \\ \Pr(\text{crime stain profile 2} | H_d, G_S) &= 2.91 \cdot 10^{-6} \end{aligned}$$

In situations like this, with multiple samples all matching the suspect (item), we have noticed that the police and legal decision makers can both overestimate and underestimate the evidential strength. For instance, one ignores the dependencies due to the overlapping loci and multiplies the individual random match probabilities reported. One may also concentrate on the smallest random match probability, ignoring the information in the non-overlapping loci of the other stain.

On the other hand, the expert also misses information needed to properly combine the findings. This is because when multiple pieces of evidence are found, one needs to take the prior probability that they originated from different sources into account. In **Example #1** above the defense hypothesis does not specify whether the two stains originate from one or two sources. Hence, for a correct evaluation, both should be considered when evaluating the combined evidential value, and their prior probabilities become part of the likelihood ratio. The problem is that these prior probabilities are generally beyond the scope of expertise of the forensic specialist and are usually left to the trier of fact to decide. Hence, such a combined evaluation can only be reported in the form of a formula or table of the prior probabilities when the forensic specialist refrains from assigning them. Such reporting forms may be too difficult to understand.

In summary, in a situation in which multiple pieces of evidence 'match' with a single suspect (item), neither the expert nor the police/legal decision maker has all information needed to properly combine the findings at the source level. This poses the interesting problem of how to deal with this in a practically feasible way. We propose a solution to this problem in the form of a lower bound of the likelihood ratio that does not require the prior probabilities mentioned above.

Combined evaluation

Derivation

$$\begin{aligned} \Pr(\text{crime stain profile 1 and 2} | H_d, 1 \text{ donor}, G_S) &= 5.71 \cdot 10^{-10} \\ \Pr(\text{crime stain profile 1 and 2} | H_d, 2 \text{ donors}, G_S) &= 6.16 \cdot 10^{-7} \cdot 2.91 \cdot 10^{-6} = 1.79 \cdot 10^{-12} \end{aligned}$$

Hence,

$$\begin{aligned} LR_{\text{comb}} &= \frac{\Pr(\text{profile 1} | \text{profile 2}, H_p, G_S) \cdot \Pr(\text{profile 2} | H_p, G_S)}{\Pr(\text{profile 1} | \text{profile 2}, H_{d,1} \text{ donor}, G_S) \cdot \Pr(\text{profile 2} | H_{d,1} \text{ donor}, G_S) \cdot \Pr(H_{d,1} \text{ donor}) + \Pr(\text{profile 1} | \text{profile 2}, H_{d,2} \text{ donors}, G_S) \cdot \Pr(\text{profile 2} | H_{d,2} \text{ donors}, G_S) \cdot \Pr(H_{d,2} \text{ donors})} \\ &\geq \frac{\Pr(\text{profile 1, profile 2} | H_p, G_S)}{\max_{i \in \{1,2\}} \Pr(\text{profile 1, profile 2} | H_{d,i} \text{ donors}, G_S)} = \frac{1}{5.71 \cdot 10^{-10}} \approx 1.75 \cdot 10^9 \end{aligned}$$

Report

The probability of observing the evidence when the suspect left both traces is *at least* $1.75 \cdot 10^9$ times larger than when he did not leave any trace.

Y-chromosomal DNA profiles

Example #2 On the crime scene of a burglary, two crime stains containing biological material are secured. The stains provide two partial Y-chromosomal DNA profiles. Both profiles match with the Y-chromosomal DNA profile of a suspect. The prosecution states (H_p) that the suspect is the donor of both crime stains. The defense hypothesis (H_d) claims that the suspect is not the donor of any of the crime stains.

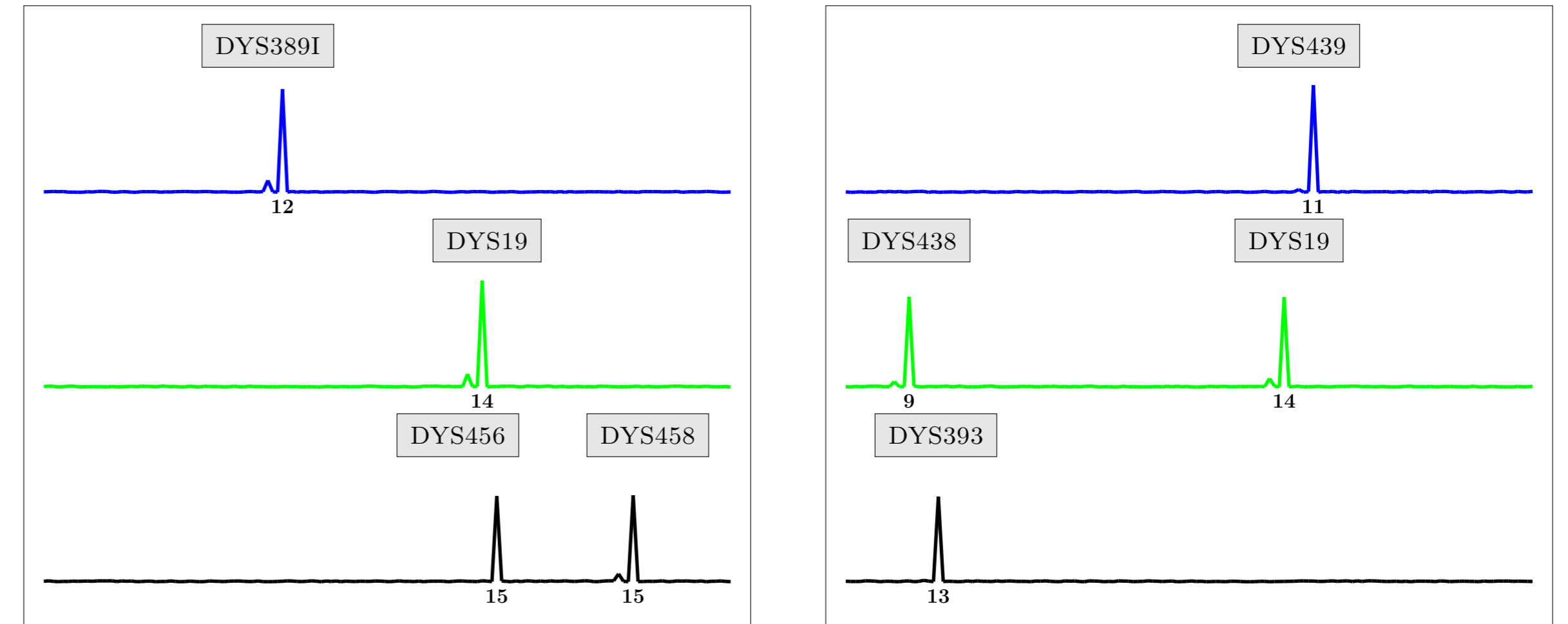


Figure 2: Partial Y-chromosomal DNA profiles - crime stain 1 and 2

locus	DYS456	DYS389I	DYS458	DYS19	DYS393	DYS439	DYS438
crime stain 1	15	12	15	14			
crime stain 2				14	13	11	9
suspect	15	12	15	14	13	11	9

Table 2: Loci and alleles Y-chromosomal DNA profiles, Example #2.

Separate evaluation

If the random match probabilities of the partial Y-chromosomal DNA profiles are based on the observed number of matches of these profiles in the *Y-Chromosome STR Haplotype Reference Database* (obtained from YHRD.org, accessed Feb 23, 2017), the conditional probabilities of observing the evidence given the different hypotheses are,

$$\begin{aligned} \Pr(\text{crime stain profile 1 and 2} | H_p, G_S) &= 1 \\ \Pr(\text{crime stain profile 1} | H_d, G_S) &= \frac{1}{186} \\ \Pr(\text{crime stain profile 2} | H_d, G_S) &= \frac{1}{302} \end{aligned}$$

Combined evaluation

Derivation

$$\begin{aligned} \Pr(\text{crime stain profile 1 and 2} | H_d, 1 \text{ donor}, G_S) &= \frac{1}{126409} \\ \Pr(\text{crime stain profile 1 and 2} | H_d, 2 \text{ donors}, G_S) &= \frac{1}{186} \cdot \frac{1}{302} = \frac{1}{56172} \end{aligned}$$

Report

The probability of observing the evidence given that suspect left both traces is *at least* **50 000** times more likely than when he did not leave any trace.

Discussion

Apart from the mathematical derivation, one could also approach the problem with a more intuitive reasoning. When for the evaluation of the evidence the presented hypotheses consist of several sub-hypotheses, the evidential strength is at least as large as the evidential strength obtained when evaluating the evidence in a situation that is most favorable for the defense and least favorable for the prosecution. This corresponds with, for the numerator, the sub-hypothesis that *minimizes* the conditional probability to observe the evidence for all prosecution scenarios and for the denominator, the sub-hypothesis that *maximizes* the conditional probability of observing the evidence for all defense scenarios.

A drawback of this method is that it can be too conservative. For example, for **Example #2**, the lower bound corresponds to a scenario where it is assumed that two unknown individuals are the sources of the material. Background information related to the crime stains can provide very strong evidence that the material originated from the same source. If the evidential value is reported based on the assumption that the evidence originates from one source, the LR will be more discriminative. On the other hand, this assumption could be criticized during trial clouding the issues and slowing down the judicial process. Another option is to assign prior probabilities to the number of distinct donors. However, similar to making an assumption about the number of distinct donors, the choice of prior distribution can be criticized. Furthermore, one could debate whether it makes sense to assign fixed probabilities for this uncertainty in the first place.

The benefit of applying this lower bound approach is that knowledge regarding the dependency relation between observation is not lost by reporting the LR's separately. Furthermore, it represents a more relevant evaluation; it better addresses the question of interest. In other words, when forensic scientists believe that their reports do not optimally represent their belief (for example because their knowledge on the dependency structure between pieces of evidence is lost), one is obliged to clarify this in the report. This lower bound likelihood ratio approach may be an effective and practical method to achieve this.