Forensic Evidence Interpretation Using Likelihood Ratios

A Study on Prior Probabilities and LR Distributions for DNA Donors

R.M. Hallema

TU Delft Delft University of Technology

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Forensic Evidence Interpretation Using Likelihood Ratios

A Study on Prior Probabilities and LR Distributions for DNA Donors

by

R.M. Hallema

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday August 27, 2025 at 1:00 PM.

Thesis committee:	Dr. J. Söhl,	TU Delft, responsible supervisor
	Dr. R.C. Kraaij,	TU Delft
	Dr. ir. R.F.J. Ypma,	NFI, external supervisor

Cover: DNA illustration, source: `https://www.britannica.com/science/DNA`.
An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Preface

I've always had a strong drive to understand how things work and, admittedly, to win. As a child, I was fascinated by board games, often going as far as calculating the probabilities behind different strategies to improve my chances. This playful curiosity gradually evolved into a deeper interest in probability and decision-making under uncertainty. That made the stochastics specialization a natural fit, offering both intellectual depth and real-world relevance. In this thesis, I had the opportunity to apply these ideas within a forensic context, an area where statistical reasoning can have tangible impact.

I would like to express my sincere gratitude to my supervisors, Rolf Ypma and Jakob Söhl, for their insightful guidance and thoughtful feedback throughout this project. Our regular meetings were invaluable in shaping both the direction and clarity of the work. I also wish to thank Klaas Slooten for the helpful discussions during our three-weekly meeting and for providing a very convenient Excel sheet that streamlined likelihood ratio and posterior probability calculations.

Secondly, I am grateful to the Netherlands Forensic Institute (NFI) for hosting me, providing a workspace, and offering various opportunities to present my work. During the course of the project, I gave several presentations at intern demos, presented to the biological traces department, and created a poster for the NFI Science Fair. These occasions helped me gain new perspectives and significantly enriched the project.

I would also like to thank Richard Kraaij for serving on the graduation committee.

A special thanks goes out to everyone I've rowed with and those who have coached me at the student rowing association D.S.R. Proteus-Eretes. The training sessions, competitions, and camps have left me with unforgettable memories. More than that, the discipline of following a rigorous training schedule provided structure and balance during my student years.

Lastly, I want to thank my family and my girlfriend for their unwavering support, not only during this thesis, but throughout the entire journey of my studies.

The code that was used to perform the simulations and analyses presented in this thesis is available at: `https://github.com/RinzeHallema/MasterProject`

R.M. Hallema
Delft, August 2025

Abstract

This thesis investigates the interpretation of forensic evidence through the use of likelihood ratios (LRs), with a particular focus on the role of prior probabilities and LR distributions in forensic DNA analysis. In forensic science, LRs are commonly used to quantify the strength of evidence in favor of one hypothesis over another. However, challenges arise in practice due to the complexity of DNA mixtures and the necessity of integrating prior information in certain scenarios. The first part of this work explores when and how prior probabilities must be incorporated into LR calculations, demonstrating through theoretical exposition and case studies that neglecting priors or assuming equal priors can lead to misleading conclusions.

Two detailed case studies illustrate the impact of introducing new persons of interest (PoIs) and how prior knowledge about associations between individuals can alter posterior probabilities. A comparison is also drawn between categorical and probabilistic approaches in body fluid analysis, with the latter offering a more nuanced interpretation of mRNA profiling data.

In the second part, the thesis introduces methods to estimate LR distributions for DNA contributors. These include threshold-based and genotype sampling techniques, which are tested across synthetic mixtures with varying contributor ratios. Furthermore, the behavior of LRs is studied for relatives of the true donor.

The findings underscore the importance of transparently reporting assumptions about priors and the value of presenting LR tables to facilitate Bayesian reasoning by decision makers. Overall, the thesis contributes to a more robust and interpretable application of statistical reasoning in forensic science.

Contents

Chapter 1

Introduction and Overview

A key responsibility of forensic experts is to evaluate whether the collected evidence supports one hypothesis over another. For example, whether it is more likely that a suspect contributed to a DNA trace or that the DNA originated from an unknown individual. To quantify this, forensic scientists often use the likelihood ratio (LR). The LR expresses how much more likely the observed evidence is under one hypothesis than under the alternative hypothesis. Also, this LR serves as the bridge between prior belief and posterior belief. The prior belief means the belief before the forensic investigation about the hypotheses, and the posterior belief means the belief after the forensic investigation. The posterior belief can be interpreted as final conclusion about which hypothesis is more plausible. The mathematical background of this reasoning is introduced in Chapter 2.

While the theoretical foundation of LR-based reasoning is well-established, its practical application in forensic contexts introduces several challenges. One challenge is how to determine the prior probability of a hypothesis. For instance, how to determine what the probability is that a suspect contributed to a DNA trace. This prior probability is needed to compute the posterior probability and can significantly change the value of the posterior probability.

A second important challenge lies in interpreting the magnitude of the LR in a practical way. While the LR tells how much more likely the evidence is under one hypothesis compared to another, it does not tell whether the result is convincing enough to act upon. In this research, the LRs for donors of DNA mixtures are investigated under the hypothesis the suspect contributed to the DNA trace against the hypothesis an unknown individual contributed to the DNA trace.

In practice, forensic institutes use often thresholds as cutoff to either report or not report. For instance, a hit from a DNA database is only reported if the LR is higher than a certain LR. This motivates to study the expected LR for a donor of a DNA mixture.
This research has two main goals:

1. **Clarify how Bayesian reasoning can be applied in forensic science,** with a focus on the role of prior probabilities. The aim is to identify when prior probabilities are unavoidable in forensic research and how they can substantially affect the posterior probability.
2. **Develop and evaluate methods to derive LR distributions for contributors of DNA mixtures.** The order of magnitude of the LR depends on various variables as the quality of the DNA material, the uniqueness of the DNA material, the number of contributors to the DNA trace, etc. Therefore, the magnitude of the LR for the donors varies from case to case. With an LR distribution an indication can be made on the order of magnitude that is expected for the donors.

To address these goals, different types of forensic evidence are examined and several case studies are developed.

The structure of the thesis is as follows:

- **Chapter 2** introduces the basic concepts of forensic statistics, including hypotheses, likelihood ratios and Bayesian inference.
- **Chapter 3** explains DNA and genetic concepts that are essential for understanding forensic DNA analysis.

- **Chapter 4** focuses on the role of prior probabilities: when they are needed and how they influence the evidential value.
- **Chapter 5** introduces forensic DNA identification in detail and presents two case studies, including one where the introduction of a second suspect drastically changes the evidential value for the first suspect.
- **Chapter 6** covers body fluid analysis using mRNA markers, here the main question is what body fluids are present in the DNA trace. Different approaches are compared to evaluate this evidence.
- **Chapter 7** compares the frameworks used for DNA identification analysis and body fluid analysis. It focuses on the differences between the methods with the hope of gaining new insights.
- **Chapter 8** introduces two methods to obtain distributions to estimate LR distributions for DNA donors.
- **Chapter 9** simulates close relatives of the donors to investigate whether they can be distinguished from the actual contributors based on their LR values.
- **Chapter 10** concludes with a discussion of the main findings and suggestions for further research.

The core part of this thesis is divided into two parts. Part I (Chapters 4-7) focuses on understanding the current LR framework in forensic science and the practical challenges in applying Bayesian reasoning. Part II (Chapters 8-9) focuses on LR distributions: how they can be achieved and how to use them.

Chapter 2

Introduction to Forensic Statistics

Forensic scientists play a crucial role in the justice system by evaluating evidence and providing an assessment about a pair of hypotheses. Examples of possible hypotheses are: person x contributed DNA to a trace, sperm is present in the trace, or person x wrote these texts. The goal of forensic analysis is to evaluate these hypotheses such that the decision maker is able to draw conclusions based on the evidence. For evaluating the evidence a metric is needed to quantify the strength of the evidence. A widely used metric in forensic analysis is the Likelihood Ratio (LR). The LR provides a numerical representation of how strong the evidence supports a hypothesis compared to another hypothesis. The LR (in forensic setting) is simply the probability that we find the evidence E under the prosecution's hypothesis H_p divided by the probability that we find the evidence under the defendant's hypothesis H_d:

$$LR = \frac{P(E \mid H_p)}{P(E \mid H_d)}.$$

Interpreting which of the two hypotheses is more likely to be true according to the evidence is straightforward using the LR. An LR > 1 means that the evidence supports the prosecution's hypothesis, indicating that the evidence is more likely found when the prosecution's hypothesis is true rather than the defendant's hypothesis is true. When the LR $= 1$ the evidence is neutral, meaning that the evidence does not favor one of the two hypotheses. Lastly, when the LR < 1 the evidence supports the defendant's hypothesis, meaning that the evidence is more likely found under the defendant's hypothesis.

The magnitude of an LR can be difficult to interpret. Say that we found an LR of 100, then the evidence is more likely to be found under the prosecution's hypothesis than under the defendant's hypothesis. Whether this LR is labeled as 'strong evidence' or 'weak evidence' depends on the person assigning the label and the context. The label is irrelevant to the degree of certainty: what always matters is the value of the LR [9]. Consider the following two settings. In the medical setting where H_p is that a person recovers due to medicine A and H_d is that the recovery is due to medicine B, an LR of 100 in favor of H_p might be considered as 'strong evidence'. But on the other hand, if H_p is that person x contributed to a DNA trace and H_d is that an unknown person contributed to a trace, it might be labeled as 'weak evidence', since a higher LR might be expected if person x were indeed the true contributor.

The role of the forensic experts is to investigate the evidence objectively using scientific methods and after that give an unbiased result based on that. Therefore, the forensic expert is typically not given additional background information about the case. After this investigation, a report will be sent to a decision maker, such as the court. The forensic expert does not determine a conclusive result, but only provides an LR as an quantitative measure of the strength of the evidence. The decision maker combines the LR with other available information of the case to arrive at a final judgment. This is also the idea of using Bayesian reasoning in forensics. The Bayesian approach allows the forensic expert to determine the evidential value, which is reported as an LR. It is up to the decision maker to assign the prior probability. The prior probabilities are defined as $P(H_p)$ and $P(H_d)$. These prior probabilities are the beliefs of the hypotheses before observing the evidence. The posterior probability in this setting means the probability that the hypothesis is true given that we have found the evidence, thus $P(H_p \mid E)$ and $P(H_d \mid E)$. Using the Bayes' theorem we find the following expressions for the posterior probabilities

$$P(H_p \mid E) = \frac{P(E \mid H_p)P(H_p)}{P(E)} \quad \text{and} \quad P(H_d \mid E) = \frac{P(E \mid H_d)P(H_d)}{P(E)}.$$

Dividing the posterior probability of H_p by the posterior probability of H_d, we find the posterior odds. This posterior odds expresses which hypothesis is more likely after observing the evidence. The posterior odds can be obtained by the product of the LR and the prior odds. The prior odds expresses which hypothesis is more likely before observing the evidence.

$$\underbrace{\frac{P(H_p \mid E)}{P(H_d \mid E)}}_{\text{posterior odds}} = \underbrace{\frac{P(E \mid H_p)}{P(E \mid H_d)}}_{\text{likelihood ratio}} \times \underbrace{\frac{P(H_p)}{P(H_d)}}_{\text{prior odds}}$$

This Bayesian approach should result in an unbiased presentation of the evidence and provides a method for the decision maker to combine the evidence and draw a conclusion.

Chapter 3

Introduction to DNA and Genetic Concepts

Deoxyribonucleic acid (DNA) is hereditary material found in nearly all living organisms. DNA has two primary purposes [2]: (1) to make copies of itself so cells can divide and carry on the same information; and (2) to carry instructions on how to make proteins so cells can build the machinery of life. The genetic information encoded in DNA is passed from generation to generation. Half of a person's DNA is inherited from their mother, and the other half from their father.

The information in the DNA is encoded in sequences of four nucleobases: adenine (A), thymine (T), cytosine (C), and guanine (G). The variation in the sequences of the nucleobases can be used to identify individuals. There are specific regions, called loci, that contain short sequences of DNA that are repeated a variable number of times. The number of repeats at a locus is referred to as an allele, and each person inherits one allele from each parent at every locus. The alleles are represented by numbers that reflect the number of repeats of the DNA sequence.

When a DNA trace is recovered from a crime scene, the detected alleles can be compared to the DNA profiles of potential suspects. Statistical methods can be used to quantify whether it is likely that an individual contributed to the trace. Importantly, the strength of a match also depends on how common or rare the alleles are in the general population. Matches involving rare alleles provide stronger support than matches with common ones.

Because DNA is inherited, it also contains information about biological relationships. For example, a father and son share at least one allele at each locus, since the son inherits one allele from the father. For brothers, allele sharing is more variable. Each allele in one brother's genotype has a 50% chance of also being found in the genotype of the other brother. This is because each parent randomly passes on one of their two alleles to each child. These patterns of allele sharing allow forensic scientists to evaluate the biological relationships between individuals.

Figure 3.1 presents an example of these mixtures. This is a two-person mixture with proportions 300:150 in picograms (1 picogram = 10^{-12} gram). The major donor thus contributed twice the amount of DNA material compared to the minor donor. This mixture is classified as having high allele sharing, indicating that the donors of the mixture share more alleles than would typically be expected between unrelated individuals.

In Figure 3.1, the first column shows the markers (or loci). At each locus, every individual has two alleles. If both alleles at a locus are the same, the individual is said to be homozygous at that locus. If the two alleles are different, the individual is heterozygous. A DNA mixture of two persons can thus have at most four alleles per locus. Four alleles are observed when both donors are heterozygous at the locus, and none of their alleles overlap. The marker AMEL can be used to determine the sex of an individual, for example males have alleles X and Y and females have twice X.

In the figure, each allele is associated with a peak height, which is a measure of how much DNA material of this allele has been found in the mixture. For example, consider the third marker, counted from above, D1S1656. We note that there have been four alleles found for this marker in the mixture. This indicates that both donors are heterozygous for this marker and do not share an allele. By inspecting the peak heights, we

can make an informed guess about which alleles originate from the major donor and which from the minor donor. In this case, the alleles 13 and 15.3 show peak heights more than twice as high as 16 and 18.3, suggesting that alleles 13 and 15.3 originate from the major donor and the alleles 16 and 18.3 originate from the minor donor.

The informed guess can be confirmed as the genotypes of the donors are known. In Figure 3.2 the genotypes of the donors can be found. As expected, the genotype at D1S1656 is (13, 15.3) for the major donor and (16, 18.3) for the minor donor.

While the last example is straightforward, in practice it is often more complex. For instance, software needs to estimate how many persons contributed to the trace and what are the proportions of the contributed DNA material without prior knowledge. If contributors have proportions that are closer to each other than in the last example it becomes more difficult to assign the alleles to the correct contributor based on the peak heights. Furthermore, when alleles are shared between donors, their peak heights may be elevated due to combined contributors, complicating interpretation.

Marke	Allele	Height	Allele	Height	Allele	Height	Allele	Height
AMEL	X	11148	Y	10495				
D3S1358	14	3950	15	8780	17	6486		
D1S1656	13	8682	15.3	6469	16	2751	18.3	3051
D2S441	11	16330						
D10S1248	13	15720	15	1885	16	1455		
D13S317	11	6363	12	7488				
Penta E	7	13284	12	4910				
D16S539	11	3643	12	9117	13	7988		
D18S51	13	8407	15	8959	16	2423		
D2S1338	17	2180	18	7284	20	5258		
CSF1PO	10	5724	11	10454				
Penta D	9	11837	10	3565				
TH01	6	4147	7	1823	9.3	7340		
vWA	16	8994	17	4650	18	1938		
D21S11	27	2022	30	8631	31	5890		
D7S820	8	7550	11	6220				
D5S818	11	6539	12	8548				
TPOX	8	10564	11	1182				
D8S1179	13	4874	14	5530	15	4196		
D12S391	19.3	8051	21	8364	24	2893		
D19S433	13	6794	14	4217	15	4249		
SE33	15	2278	20	3086	28.2	2339	29.2	5763
D22S1045	11	2187	15	4679	16	6262		
DYS391	10	9092						
FGA	21	1580	22	6183	23	2965	24	7445
DYS576	16	5903	18	3251				
DYS570	17	1152	21	4402				

Figure 3.1: Alleles and peak heights of a two-person mixture where the proportions are 300:150.

	Donor 1	Donor 1	Donor 2	Donor 2
Marker	Allele 1	Allele 2	Allele 1	Allele 2
D1S1656	13	15.3	16	18.3
TPOX	8	8	8	11
D2S441	11	11	11	11
D2S1338	18	20	17	18
D3S1358	15	17	14	15
FGA	22	24	21	23
D5S818	11	12	11	12
CSF1PO	10	11	11	11
SE33	20	29.2	15	28.2
D7S820	8	11	8	11
D8S1179	14	15	13	13
D10S1248	13	13	15	16
TH01	6	9.3	7	9.3
vWA	16	17	16	18
D12S391	19.3	21	21	24
D13S317	11	12	11	12
Penta E	7	12	7	7
D16S539	12	13	11	13
D18S51	13	15	13	16
D19S433	13	15	14	14
Penta D	9	9	10	10
D21S11	30	31	27	30
D22S1045	15	16	11	16

Figure 3.2: Genotypes of the two individuals who contributed to the mixture shown in Figure 3.1. Donor 1 is the major donor and donor 2 is the minor donor.

In addition to identifying individuals through DNA, forensic scientists can also investigate which body fluids, such as blood, saliva, or semen, are present at the crime scene. This distinction can be crucial in certain cases, such as sexual assault. To determine the type of body fluid, the presence of body fluid specific markers is used. For example, hemoglobin is typically found in blood. Each body fluid has a set of characteristic markers. However, the analysis is complicated by the fact that some markers are shared across fluids, and even body fluid specific markers are not always consistently detected.

Part I

The Role of Prior Probabilities in Forensic Statistics

Chapter 4

Prior Probabilities

In forensic practice, the forensic expert typically determines the LR, while the decision maker combines this LR with prior information to reach a conclusion about the competing hypotheses. For instance, consider the hypothesis that there is blood in a trace versus the hypothesis that there is no blood. In such a case, the forensic expert can calculate an LR for these hypotheses without needing any prior information. The calculation of an LR in body fluid analysis is discussed in detail in Chapter 6.

However, there are cases in which the forensic expert cannot compute an LR without relying on prior probabilities. One case where prior information is needed is when the prosecution's hypothesis can be subdivided into several scenarios, and the forensic expert can only determine the likelihood for each scenario individually. In these cases, prior probabilities are needed to combine the LRs into a single LR for the prosecution's hypothesis.

Section 4.1 presents an example where prior probabilities are required to compute an LR. Section 4.2 describes the mathematical framework for incorporating prior into such calculations. Section 4.3 describes the alternative option to report the subdivided LRs.

4.1 Example: subdivision of a hypothesis

Suppose a DNA trace is found, which is a mixture of two individuals. There are two suspects and the forensic expert is asked to calculate for both of the suspects an LR for the contribution to the DNA trace. The forensic expert usually uses 'unknown individual' profiles in the calculation of the LR. A DNA profile of an unknown individual is based on population allele frequencies. More details on this process can be found in Chapter 5.

The complexity arises in calculating the LR for the contribution of either suspect. The forensic expert cannot directly compare the hypothesis that suspect 1 contributed with the hypothesis that suspect 1 did not contribute, as the latter cannot be modeled directly. Instead, the forensic expert evaluates the four hypotheses:

- both suspects contributed,
- suspect 1 and an unknown individual contributed,
- suspect 2 and an unknown individual contributed,
- two unknown individuals contributed

against the hypothesis that two unknown individuals contributed. These resulting four LRs can then be combined using prior probabilities to compute the LR for a separate suspect. Without prior probabilities, the forensic expert is unable to combine these LRs into a single LR for a suspect.

4.2 Mathematical framework: the role of the prior probability

Assume that there are two different scenarios in which the prosecution's hypothesis is true. For instance, the hypothesis that suspect 1 is a donor to a DNA trace is true if either suspect 1 and suspect 2 are the donors, or if suspect 1 and an unknown individual are the donors. Consider that there are two different scenarios H_{p_1}, H_{p_2}, if one of either is true, then H_p is true. If the scenarios H_{p_1}, H_{p_2} are mutually exclusive, i.e. they cannot both be true, we get using the law of total probability:

$$P(E \mid H_p) = P(E \mid H_{p_1})P(H_{p_1} \mid H_p) + P(E \mid H_{p_2})P(H_{p_2} \mid H_p).$$

In words, this means that the likelihood of finding the evidence under the prosecution's hypothesis is equal to the likelihood of finding the evidence under H_{p_1} times the prior probability of H_{p_1} given that H_p is true plus the likelihood of finding the evidence under the prosecution's is equal to the likelihood of finding the evidence under H_{p_2} times the prior probability of H_{p_2} given that H_p is true.

If the two scenarios are not mutually exclusive the likelihood of finding the evidence under the prosecution's hypothesis is as follows:

$$P(E \mid H_p) = P(E \mid H_{p_1})P(H_{p_1} \mid H_p) + P(E \mid H_{p_2})P(H_{p_2} \mid H_p) - P(E \mid H_{p_1} \cap H_{p_2})P(H_{p_1} \cap H_{p_2} \mid H_p).$$

When H_{p_1} and H_{p_2} are mutually exclusive $P(H_1 \cap H_2 \mid H_p) = 0$ and the formula reduces to the same formula that we found for mutually exclusive scenarios.

4.2.1 Generalization to multiple scenarios

The prosecution's hypothesis may in some cases be divided into more than two scenarios. Assume that there are $n \in \mathbb{N}$ mutually exclusive scenarios that can explain the prosecution's hypothesis, then we find similarly using the law of total probability that

$$P(E \mid H_p) = \sum_{i=1}^{n} P(E \mid H_{p_i})P(H_{p_i} \mid H_p).$$

In cases where the scenarios are not mutually exclusive a formula can be derived using the inclusion-exclusion principle.

4.3 Reporting without priors

A practical alternative is to report the LRs for each sub-hypothesis individually. In most cases where one of the hypotheses is clearly most likely this suffices. In cases where there is more doubt about the best fitting hypothesis more work is needed. The decision maker now has to combine the LRs with prior probabilities to an LR for the prosecution's hypothesis. This process may be challenging as statistical expertise in needed.

Chapter 5

Forensic DNA Analysis

In Chapter 4, we discussed how prior probabilities can be essential for evaluating forensic evidence. A field where this issue frequently arises is forensic DNA analysis. From crime scenes often DNA traces can be retrieved. These DNA traces contain useful information about the individuals whose body fluids were found on the crime scene. Often these DNA traces are mixtures of DNA material of several persons, which introduces complexity in interpreting the evidence.

Section 5.1 introduces the mathematical framework for calculating LRs and posterior probabilities in forensic DNA analysis. Section 5.2 outlines current guidelines on communicating the results. This chapter then presents two case studies. In Section 5.3 a case study is explored where the evidential value of a suspect completely diminishes after the introduction of a new suspect. In Section 5.4 a case study is considered where the suspects are associates. This affects the prior probabilities, which affects the posterior probabilities.

5.1 Calculating LRs and posterior probabilities in forensic DNA analysis

To assess whether a person contributed to a DNA trace, population allele frequencies are used to quantify how common or rare the observed alleles are. These frequencies allow us to compute the probability of observing a given genotype in the population. When both the mixture and the genotype of the person of interest (PoI) share rare alleles, this provides stronger support for the hypothesis that the PoI contributed to the trace than if only common alleles are shared.

Most often, DNA traces are mixtures of DNA material from several individuals. To calculate LRs in such cases, the statistical software requires hypotheses that fully explain the trace. For example, hypotheses for a two-person mixture need to include two individuals, i.e. suspect 1 and suspect 2 created the DNA trace, or suspect 1 and an unknown individual created the trace. As defendant's hypothesis, the statistical software uses the hypothesis that two unknown individuals created the trace.

Consider that a DNA trace is found and it is determined that it is a mixture of two contributors. The police has two suspects. It does not suffice to calculate only separate LRs for the suspects. In example, it is possible that the two hypotheses: suspect 1 and an unknown, and suspect two and an unknown are significantly more likely than two unknown individuals created the trace, and thus achieve both a high LR. But the hypothesis that suspect 1 and suspect 2 created together the trace is less likely than the alternative hypothesis. This suggests that while both suspects' profiles match portions of the trace, they do not explain together all the DNA observed. It is thus unlikely that both the suspects are donors of the trace, but it is likely that one of the two suspects is a donor of the trace.
To properly evaluate the evidence, we consider four mutually exclusive hypotheses:

- H_{12} : PoI 1 and PoI 2 contributed to the trace,
- H_{1u} : PoI 1 and another unknown person, that is unrelated to both PoIs contributed to the trace,
- H_{2u} : PoI 2 and another unknown person, that is unrelated to both PoIs contributed to the trace,
- H_{uu} : Two unknown unrelated persons contributed to the trace.

For all these four hypotheses, the defendant's hypothesis H_{uu} can be used to calculate an LR. Clearly, for the hypothesis H_{uu} this would result in an LR of 1, as the prosecution's is the same as the defendant's hypothesis. These LRs can then be used to assess which hypothesis is most likely to support the evidence with respect to the hypothesis H_{uu}. Let E be the event that we find the DNA trace and let S_1 be the event that PoI 1 contributes to the trace. Let S_1^c be the complement of S_1, thus PoI 1 did not contribute to the crime stain. By definition, S_1 is true if either PoI 1 and an unknown contributed, or PoI 1 and PoI 2 both contributed. The LR for the contribution of PoI 1 is given by the likelihood of finding the trace given that PoI 1 contributed divided by the likelihood of finding the trace given that PoI 1 did not contribute.
In the sequel the likelihood is defined as L and prior probabilities as π. In the subscripts is specified, which hypothesis is considered. For example, L_{2u} is the likelihood of H_{2u}, thus $P(E \mid H_{2u})$. The hypotheses are clearly mutually exclusive as they cannot co-occur. Thus, the law of total probability can be used to derive the likelihood that we find the trace under the hypothesis that PoI 1 is a donor. We find

$$\begin{aligned} P(E \mid S_1) &= P(E \mid H_{1u})P(H_{1u} \mid S_1) + P(E \mid H_{12})P(H_{12} \mid S_1) \\ &= L_{1u}\frac{\pi_{1u}}{\pi_{1u}+\pi_{12}} + L_{12}\frac{\pi_{12}}{\pi_{1u}+\pi_{12}} = \frac{L_{1u}\pi_{1u} + L_{12}\pi_{12}}{\pi_{1u}+\pi_{12}}. \end{aligned} \tag{5.1}$$

Here we used that

$$P(H_{1u} \mid S_1) = \frac{P(H_{1u})}{P(S_1)} = \frac{P(H_{1u})}{P(H_{1u}) + P(H_{12})} = \frac{\pi_{1u}}{\pi_{1u}+\pi_{12}}.$$

Similarly, we can calculate the defendant's likelihood $P(E \mid S_1^c)$.

$$\begin{aligned} P(E \mid S_1^c) &= P(E \mid H_{2u})P(H_{2u} \mid S_1^c) + P(E \mid H_{uu})P(H_{uu} \mid S_1^c) \\ &= L_{2u}\frac{\pi_{2u}}{\pi_{2u}+\pi_{uu}} + L_{uu}\frac{\pi_{uu}}{\pi_{2u}+\pi_{uu}} = \frac{L_{2u}\pi_{2u} + L_{uu}\pi_{uu}}{\pi_{2u}+\pi_{uu}}. \end{aligned} \tag{5.2}$$

Now, the LR of PoI 1 contributing to the DNA mixture against not contributing can be calculated. This LR is denoted using the subscript $1/\bar{1}$, thus meaning the LR for the hypothesis that PoI 1 contributed against the hypothesis that PoI 1 did not contribute. Using (5.1) and (5.2) follows that

$$LR_{1/\bar{1}} = \frac{P(E \mid S_1)}{P(E \mid S_1^c)} = \frac{L_{1u}\pi_{1u} + L_{12}\pi_{12}}{L_{2u}\pi_{2u} + L_{uu}\pi_{uu}} \times \frac{\pi_{2u}+\pi_{uu}}{\pi_{1u}+\pi_{12}}.$$

Statistical software can calculate the LRs of one specific explanation of the trace against the likelihood that two unrelated unknowns contributed to the trace. To obtain a formula with LRs, we can simply divide the right hand side by L_{uu}, then follows

$$LR_{1/\bar{1}} = \frac{LR_{1u,uu}\pi_{1u} + LR_{12,uu}\pi_{12}}{LR_{2u,uu}\pi_{2u} + \pi_{uu}} \times \frac{\pi_{2u}+\pi_{uu}}{\pi_{1u}+\pi_{12}}. \tag{5.3}$$

Note that prior probabilities are thus needed to calculate $LR_{1/\bar{1}}$. This formula can easily be extended for DNA mixtures that have more than two donors. It can also be easily extended for cases where there are more PoIs. When prior probabilities are available it is more logical to report posterior probabilities than LRs. Posterior probabilities are easier to interpret. The posterior probability can be calculated using the Bayes' theorem and the law of total probability as follows:

$$\begin{aligned} P(H_{1u} \mid E) &= \frac{P(E \mid H_{1u})P(H_{1u})}{P(E)} \\ &= \frac{P(E \mid H_{1u})P(H_{1u})}{P(E \mid H_{1u})P(H_{1u}) + P(E \mid H_{2u})P(H_{2u}) + P(E \mid H_{12})P(H_{12}) + P(E \mid H_{uu})P(H_{uu})}. \end{aligned} \tag{5.4}$$

Similarly, the posterior probabilities for the other hypotheses can be calculated. The posterior probability for a separate PoI can also be calculated, since the hypotheses are mutually exclusive follows that

$$P(S_1 \mid E) = P(H_{1u} \mid E) + P(H_{12} \mid E).$$

5.1.1 Example LR calculation using prior probabilities

To illustrate the use of equation (5.3), consider again a two-person DNA mixture with two PoIs. Statistical software has calculated LRs for the four possible hypotheses, which can be found in Table 5.1 presented on a logarithmic scale (base 10). From the table we note that the hypothesis that both PoIs contributed (H_{12}) receives the strongest support from the evidence. Also, the hypothesis H_{1u} has quite strong support and H_{2u} has weak support. Now, if prior probabilities are available, the LRs and priors can be used to calculate $LR_{1/\bar{1}}$ and $LR_{2/\bar{2}}$. Assume that there is moderate prior belief in individual contributors and weak belief in joint contribution, reflecting that the PoIs are not known to be associated. Take the prior probabilities $\pi_{1u} = \pi_{2u} = 0.2$ and $\pi_{12} = 0.04$. Lastly, take $\pi_{uu} = 1 - \pi_{12} - \pi_{1u} - \pi_{2u} = 0.56$. The prior probabilities sum up to 1 as one of the four hypotheses needs to be true. Then using equation (5.3) follows that

$$LR_{1/\bar{1}} = \frac{10^6 \cdot 0.2 + 10^{12} \cdot 0.04}{10^3 \cdot 0.2 + 1 \cdot 0.56} \times \frac{0.2 + 0.56}{0.2 + 0.04} \approx 6.3 \times 10^8$$

and

$$LR_{2/\bar{2}} = \frac{10^3 \cdot 0.2 + 10^{12} \cdot 0.04}{10^6 \cdot 0.2 + 1 \cdot 0.56} \times \frac{0.2 + 0.56}{0.2 + 0.04} \approx 6.3 \times 10^5.$$

Then using equation (5.4) follows the posterior probabilities

- $P(H_{1u} \mid E) \approx 10^{-9}$,
- $P(H_{2u} \mid E) \approx 10^{-6}$,
- $P(H_{12} \mid E) \approx 0.999995$,
- $P(H_{uu} \mid E) \approx 10^{-11}$.

Now, it can be concluded that the prior probabilities in combination with the LRs from Table 5.1 provides overwhelming support for the hypothesis that both PoIs contributed to the trace. The posterior probability for this hypothesis is approximately 99.999995%.

Hypothesis H	$\log_{10}(LR_{H,H_{uu}})$
H_{12}	12
H_{1u}	6
H_{2u}	3
H_{uu}	0

Table 5.1: LRs for four competing hypothesis for the example in Section 5.1.1.

5.1.2 Equal priors

By setting all prior probabilities equal, the LR formula for the contribution for PoI 1 simplifies to a formula without prior probabilities:

$$LR_{1/\bar{1}} = \frac{LR_{1u,uu} + LR_{12,uu}}{LR_{2u,uu} + 1}.$$

At first glance, this seems like a reasonable assumption, since this assumption does not favor one of the hypotheses. However, this can be a problematic assumption that leads to wrong conclusions. To illustrate, consider the hypotheses and their corresponding LRs in logarithmic scale from Table 5.2.

Hypothesis H	$\log_{10}(LR_{H,H_{uu}})$
H_{12}	-3
H_{1u}	6
H_{2u}	3
H_{uu}	0

Table 5.2: LRs for four competing hypothesis for the example in Section 5.1.2.

From the table, it is evident that both PoI 1 and PoI 2, each together with an unknown contributor, are more likely than the hypothesis with two unknowns contributors. However, the hypothesis of PoI 1 and PoI 2 together is not more likely than the hypothesis of two unknowns. This can occur, for instance, when both profiles separately match parts of the DNA mixture, but the two profiles together fail to match DNA material that is present in the DNA mixture. Meaning that the two profiles are not likely to have created the trace together. Using equal priors, we find

$$LR_{1/\bar{1}} = \frac{10^6 + 10^{-3}}{10^3 + 1} \approx 10^3 \quad \text{and} \quad LR_{2/\bar{2}} = \frac{10^3 + 10^{-3}}{10^6 + 1} \approx 10^{-3}.$$

Thus, the likelihood of observing the DNA trace under the hypothesis that PoI 1 contributed is 1000 times greater than under the hypothesis that PoI 1 did not contribute. The likelihood of observing the trace under the hypothesis that PoI 2 contributed is 1000 times smaller than under the hypothesis that PoI 2 did not contribute. These result are difficult to interpret on their own, as they do not directly reflect the probability of a hypothesis being true. A more informative approach is to calculate the posterior probabilities.

- $P(H_{1u} \mid E) \approx 0.9999$,
- $P(H_{2u} \mid E) \approx 0.0001$,
- $P(H_{12} \mid E) \approx 10^{-9}$,
- $P(H_{uu} \mid E) \approx 10^{-6}$.

The posterior probabilities suggest a very strong evidence in favor of hypothesis H_{1u}.
Now suppose that police information strongly supports the involvement of PoI 2 in the trace, while the genotype of PoI 1 has been found in a database. We can reflect this belief in the prior probabilities. Take $(\pi_{1u}, \pi_{2u}, \pi_{12}, \pi_{uu}) = (10^{-9}, 0.5, 10^{-9}, 0.5)$. Then follows that

$$LR_{1/\bar{1}} = \frac{10^6 \cdot 10^{-9} + 10^{-3} \cdot 10^{-9}}{10^3 \cdot 0.5 + 1 \cdot 0.5} \times \frac{0.5 + 0.5}{10^{-9} + 10^{-9}} \approx 1000$$

and

$$LR_{2/\bar{2}} = \frac{10^3 \cdot 0.5 + 10^{-3} \cdot 10^{-9}}{10^6 \cdot 10^{-9} + 1 \cdot 0.5} \times \frac{0.5 + 10^{-9}}{0.5 + 10^{-9}} \approx 1000.$$

Under the new priors, both PoIs appear roughly 1000 times more likely to have contributed than not. This is a significant change for PoI 2, for which under equal prior we found $LR_{2/\bar{2}} \approx 10^{-3}$.
Now consider the posterior probabilities under the updated prior. We find

- $P(H_{1u} \mid E) \approx 2 \cdot 10^{-6}$,
- $P(H_{2u} \mid E) \approx 0.9999$,
- $P(H_{12} \mid E) \approx 2 \cdot 10^{-15}$,
- $P(H_{uu} \mid E) \approx 0.0001$.

The posterior probabilities here are strongly in favor of hypothesis H_{2u}.

This example thus clearly illustrates how the assumption of equal priors, although seemingly a neutral assumption, can lead to wrong conclusions. Under equal priors, the analysis misleadingly points to PoI 1 as contributor and not to PoI 2. However, when the prior probabilities are updated to reflect reasonable contextual knowledge, the analysis points to PoI 2 as contributor and not to PoI 1.

5.2 Guidelines for reporting LRs

In cases with a single PoI and an n-person DNA mixture, two hypotheses are typically considered. Firstly, the hypothesis that the PoI created the trace with $n-1$ other unknown persons and secondly the hypothesis that the trace is created by n unknowns unrelated to the PoI. Statistical software can calculate an LR for the first hypothesis against the second hypothesis. In cases where there are more PoIs it is more complex. There are more possible hypotheses and the question arises which hypotheses are relevant to consider. However, the LR for a specific PoI cannot be obtained without the use of prior probabilities as explained in Section 5.1.

Slooten (2022) advocates that the most natural way to retain the overview on all hypotheses is by reporting a table of LRs as the output of the forensic laboratory. This table allows the decision maker to move from prior odds to posterior odds. In some cases, evidence supports the proposition that all PoIs contributed. This was also illustrated in Section 5.1.1. In that case the simplest summary is that within the considered hypotheses the trace is 10^6 (in the example) times more likely if both PoI 1 and PoI 2 contributed.

Other articles suggest to report an LR per person. Thus, using the hypothesis that the PoI contributed to the trace against the hypothesis that the PoI did not contribute to the trace. However, as explained in Section 5.1 this is not possible without invoking prior probabilities. In Hicks et al. (2021) an LR for a certain PoI having contributed is obtained by setting all prior probabilities equal, in their article this is denoted as *Assumption 1.* It is written that "in general this would seem reasonable but it would be best if this was disclosed in some way." As discussed in Section 5.1.2, this assumption may lead to incorrect conclusions. For this reason, we recommend avoiding this assumption.

Another alternative approach is to explore how different prior assumptions influence the evidential value. Instead of reporting single LRs, the forensic expert could calculate several posterior probabilities using a set of clearly defined prior probability scenarios. This allows the decision maker to see how these assumptions influence the posterior probabilities. However, this approach requires additional effort from the forensic expert and may not always be necessary, particularly in cases where there is a really dominant hypothesis. This additional effort could also be made by the decision maker, since the decision maker is ultimately responsible for determining the prior probabilities.

5.3 Case study 1: one PoI weakens the evidential value of the other

This section presents a real-world case example. Initially, a PoI had a relatively high evidential value. However, a few years later, a second PoI emerged whose presence significantly altered the interpretation of the first PoI's evidential value. We explore different ways this could be reported to a decision maker and discuss how certain reporting choices might lead to a misinterpretation.

5.3.1 Case background

The DNA trace is a two-person mixture. Initially, only one PoI was under consideration. Based on the prosecution's hypothesis that PoI 1 and an unknown individual contributed to the trace versus the defendant's hypothesis that two unknown individuals contributed to the trace the software calculates an LR of 60,000.

While an LR of 60.000 appears convincing, it raises the question of whether this is sufficient to support a conviction. Such a high LR may result from coincidence or from the possibility that the PoI is a relative of the true contributor, thereby sharing many genetic characteristics.

A few years later, a second PoI was identified. For this individual (PoI 2), the LR for the hypothesis that PoI 2 and an unknown individual contributed to the trace versus two unknown individuals was calculated to be 60.000.000. This significantly diminished the evidential value previously assigned to PoI 1. Also, the LR for the hypothesis that both PoI 1 and PoI 2 contributed against the hypothesis that two unknown individuals contributed to the trace is calculated, this LR is 0.001.

From this LRs it seems that either PoI 1 or PoI 2 contributed to the trace, but not both. Also, it seems that PoI 2 is more likely to have contributed to the trace than PoI 1. The low LR for contribution of both PoIs can be explained as follows: there may be DNA material present in the DNA trace that is not present in the DNA material from PoI 1 and PoI 2, implying that PoI 1 and PoI 2 can not have created the trace together. Separately, both PoIs have DNA material that is also found in the mixture, suggesting that each of the PoIs together with an unknown they may have created the trace, or at least are more likely to create the trace in comparison to two unknown individuals.

5.3.2 Equal priors

When assuming a priori that all hypotheses are equally likely, the LRs for the contribution of the PoIs can be calculated as follows:

$$LR_{1/\bar{1}} = \frac{LR_{12,uu} + LR_{1u,uu}}{LR_{2u,uu} + 1} \quad \text{and} \quad LR_{2/\bar{2}} = \frac{LR_{12,uu} + LR_{2u,uu}}{LR_{1u,uu} + 1}$$

The complete derivation of this formula can be found in the Section 5.1.
In this specific example $LR_{12,uu} \ll LR_{1u,uu}$ and $1 = LR_{uu,uu} \ll LR_{2u,uu}$. The LR for PoI 1 then reduces approximately to

$$LR_{1/\bar{1}} = \frac{LR_{1u,uu}}{LR_{2u,uu}} = \frac{60.000}{60.000.000} = \frac{1}{1000}.$$

For PoI 2, also $LR_{12,uu} \ll LR_{2u,uu}$, and we find

$$LR_{2/\bar{2}} = \frac{LR_{2u,uu}}{LR_{1u,uu}} = \frac{60.000.000}{60.000} = 1000.$$

The calculations above are based on the assumption that all hypotheses are a priori equally likely. This implies that the probability that the trace is from PoI 1 together with an unrelated unknown individual (π_{1u}) is 0.25, thus 25%. Similarly, $\pi_{12} = \pi_{2u} = \pi_{uu} = 0.25$. This assumption may be reasonable if PoI 1 was identified based on external information (e.g., police investigation), indicating a non-negligible belief in their involvement. However, if PoI 1 is not identified by investigative leads but instead results from a database search, assuming a prior probability of 0.25 is no longer justifiable. As of September 2024, the Dutch DNA databank contains over 400.000 profiles. Approximately 50% of crime scene DNA traces in the Netherlands result in a match with a profile in this databank. Assuming that all persons in the databank are equally likely to contribute, the prior probability for a randomly selected PoI from the database contributing to the trace is approximately $\frac{1}{400,000} \times \frac{1}{2} \approx 10^{-6}$. Note that this is a rough estimate.

5.3.3 Unequal priors

We now consider the case where both PoIs originate from a DNA database. We assign prior probabilities $\pi_{1u} = \pi_{2u} = 10^{-6}$. This reflects the assumption that both PoI 1 and PoI 2 have a prior probability 10^{-6} of contributing to the trace. Assuming PoI 1 and PoI 2 have no relation to each other, it is reasonable to treat the prior probabilities as independent, yielding $\pi_{12} = \pi_{1u} \cdot \pi_{2u} = 10^{-12}$. This independence reflects that the probability for PoI 1 and PoI 2 to both contribute to the trace is much smaller than the contribution of only one of the two PoI. The remaining prior probability is then assigned to $\pi_{uu} = 1 - 2 \cdot 10^{-6} - 10^{-12}$, ensuring that the total probability sums to 1.

With unequal priors, the formula for the LR of contribution for PoI 1 is derived in 5.1 and is given by

$$LR_{1/\bar{1}} = \frac{LR_{12,uu}\pi_{12} + LR_{1u,uu}\pi_{1u}}{LR_{2u,uu}\pi_{2u} + LR_{uu,uu}\pi_{uu}} \times \frac{\pi_{2u} + \pi_{uu}}{\pi_{12} + \pi_{1u}},$$

and for PoI 2

$$LR_{2/\bar{2}} = \frac{LR_{12,uu}\pi_{12} + LR_{2u,uu}\pi_{2u}}{LR_{1u,uu}\pi_{1u} + LR_{uu,uu}\pi_{uu}} \times \frac{\pi_{1u} + \pi_{uu}}{\pi_{12} + \pi_{2u}}.$$

We find that the LR of contribution for PoI 1 is now 9.84×10^2 and the LR of contribution for PoI 2 is now 5.66×10^7. Interestingly, these LRs are considerably larger than LRs calculated under equal priors (10^{-3} and 10^3). These LR values are significant because most laboratories apply reporting thresholds to databank hits. If such a threshold lies between 10^3 and 10^7, this example illustrates that PoI 2 would be reported under unequal priors, but not under equal priors.

Forensic experts often avoid specifying priors, as these cannot always be chosen objectively. Nevertheless, the role of prior probabilities should not be overlooked. A PoI with a high prior probability requires less evidential support (i.e., a lower LR) to reach the same posterior level as one with a low prior. In other words, "extraordinary claims require extraordinary evidence". Often, the exact relationship between the PoIs is unknown, making it difficult to assign appropriate priors. The priors may not be agreed upon by all parties. Therefore, most forensic institutes chose to report a table of likelihoods to the decision maker.

5.3.4 Detailed explanation of the change in evidential value in terms of the posterior

To combine likelihoods and the prior probability the posterior probability can be calculated. We begin by revisiting the case involving only PoI 1. The two hypotheses we consider are H_{1u} : PoI 1 contributed to the trace together with an unknown, that is unrelated to PoI 1 and H_{uu} : two unknown individuals contributed to the trace, that are unrelated to PoI 1 and each other. The posterior for PoI 1 is calculated as follows

$$\text{Posterior } 1 = \frac{LR_{1u,uu}\pi_{1u}}{LR_{1u,uu}\pi_{1u} + \pi_{uu}}.$$

With equal priors, i.e., $\pi_{1u} = \pi_{uu} = 0.5$, the resulting posterior is 0.99998. For priors $\pi_{1u} = 0.01$ and $\pi_{uu} = 0.99$ this gives a posterior of 0.99835. Lastly, the 'database prior' $\pi_{1u} = 10^{-6}$ and $\pi_{uu} = 1 - 10^{-6}$ gives a posterior of 0.05660. These results show that when there is prior support for PoI 1, the posterior strongly favors their contribution. However, when PoI 1 originates from a databank search, the posterior drops to approximately 5.7%, indicating weak support for contribution.

Next, consider the case with two PoIs, under the four competing hypotheses: $H_{12}, H_{1u}, H_{2u}, H_{uu}$. The posterior for PoI 1 is then calculated as follows

$$\text{Posterior } 1 = \frac{LR_{12,uu}\pi_{12} + LR_{1u,uu}\pi_{1u}}{LR_{12,uu}\pi_{12} + LR_{1u,uu}\pi_{1u} + LR_{2u,uu}\pi_{2u} + LR_{uu,uu}\pi_{uu}}.$$

With equal priors, the posterior probability for PoI 1 is 0.00099 and for PoI 2 is 0.999. Assuming both PoIs originate from the Dutch DNA databank, the posteriors for PoI 1 is 0.000856 and 0.856 for PoI 2. Thus, in both cases the posterior is in strong favor for PoI 2. The posterior can also be calculated for the contribution of both PoI 1 and PoI 2. The posterior probability for H_{12} is 1.67×10^{-11} under equal priors, and 1.43×10^{-17} under databank priors, providing extremely strong evidence against this hypothesis.

5.3.5 Conclusion of the case study

This case study illustrates several challenges in the interpretation of forensic DNA evidence. First, it is important to note that a LR alone is, in most cases, insufficient to justify a conviction.

Second, this example underscores that the prior probabilities should be chosen with caution. Priors can significantly change the LR. The seemingly innocuous assumption of equal priors can result in a substantial underestimation of evidential strength, which in this case may lead to no reporting. If forensic experts chose to use prior probabilities, or assume they are equal, such assumptions should be clearly communicated to decision makers.

The recommendation by Slooten (2022) to report a table of LRs for all relevant hypotheses proves useful in this example. This is demonstrated in Table 5.3.

Hypothesis H	$LR_{H,H_{uu}}$
H_{12}	0.001
H_{1u}	60.000
H_{2u}	60.000.000
H_{uu}	1

Table 5.3: LRs of the compared hypotheses.

The table clearly shows that, for non-extreme prior probabilities, there is strong support for the hypothesis that one of the PoIs contributed to the trace, but not both. Moreover, it is evident that π_{2u} would need to be significantly smaller than π_{1u} for PoI 1 to be more likely than PoI 2 to have contributed.

When only PoI 1 is considered, the posterior probability shows that the evidential value depends heavily on the prior. If PoI 1 was identified through a databank search, the posterior remains low. However, if there is already case-based suspicion, the posterior supports PoI 1's contribution strongly. When both PoIs are considered, the posterior probabilities indicate that, under equal priors, the evidence strongly favors PoI 2. However, if there is already case-based suspicion, the posterior supports PoI 1's contribution strongly.

5.4 Case study 2: PoIs are associates, how do the priors change and influence results?

This case study examines how prior probabilities affect results depending on whether the PoIs are unrelated or known associates. When PoIs are associates, the presence of one individual's DNA may increase the probability of the presence of the other. This information is considered as background information and is not for the forensic expert to investigate. It is the decision maker's role to consider this information when assigning prior probabilities. We consider here a three person DNA mixture where there are three PoIs. LRs for all eight hypotheses were computed using statistical software. These LRs are presented in logarithmic scale in Table 5.4.

Hypothesis H	$\mathbf{Log_{10}(LR}_{H,H_{uuu}}\mathbf{)}$
H_{123}	33
H_{12u}	27
H_{13u}	29.4
H_{23u}	12.2
H_{1uu}	17.5
H_{2uu}	12.1
H_{3uu}	7.9
H_{uuu}	0.0

Table 5.4: LRs for the considered hypotheses

Table 5.4 shows that all hypotheses (other than H_{uuu}) are substantially more likely to explain the observed DNA trace than H_{uuu}. The hypothesis H_{123} has the highest LR and is at least $10^{3.6}$ times more likely to explain the trace than any other considered hypothesis.

5.4.1 Overview of the considered cases

In this Section we will consider three different cases with different prior probabilities.

- **Case 1:** Equal priors $\implies$ all eight priors for the hypotheses are 0.125.
- **Case 2:** PoI 1 is suggested by the police and PoI 2 & PoI 3 come from the DNA databank. The corresponding priors for hypothesis involving a PoI with two unknowns contributors are respectively $(0.3, 10^{-6}, 10^{-6})$. We treat the priors independently, meaning that for example $\pi_{12} = \pi_{1u}\pi_{2u}$.
- **Case 3:** Again PoI 1 is suggested by the police and PoI 2 & PoI 3 come from the DNA databank. Now assume that PoI 1 and PoI 2 are associates. Then we should no longer assume independence for PoI 1 and PoI 2 anymore. Finding DNA for PoI 1 increases the probability of finding DNA for PoI 2 and also the other way around. This should be reflected in the prior probabilities. We use the prior probabilities $(\pi_{1uu}, \pi_{2uu}, \pi_{3uu}, \pi_{12u}, \pi_{13u}, \pi_{23u}, \pi_{123}) = (0.1, 10^{-6}, 10^{-6}, 0.2, 10^{-6}, 10^{-12}, 2 \times 10^{-7})$.

5.4.2 Intuition

In case 1 the prior probabilities are uniform, meaning that there is no background information that one of the hypotheses is more likely than the others. In this case we would expect that the posterior reflects that H_{123} is most likely, due to the fact that this has an LR that is at least $10^{3.6}$ times higher than all other hypotheses and the prior does not change this information.
Case 2 is interesting as the dominant term now shifts using the priors. While the dominant term in terms of LRs is for hypothesis H_{123}, the dominant in terms of LR $\times$ prior is for the hypothesis H_{13u}. The prior $\pi_{123} = 0.3 \times 10^{-6} \times 10^{-6} = 0.3 \times 10^{-12}$ and the prior $\pi_{13u} = 0.3 \times 10^{-6} \times (1 - 0.3 - 2 \times 10^{-6}) \approx 0.3 \times 10^{-6}$. This difference in priors is greater than the difference in LR, thus indeed the posterior term for H_{13u} is greatest. We thus expect the posterior for PoI 1 and for PoI 3 to be large and for PoI 2 to be small.
In case 3 it will be interesting to see how the different prior probabilities will shift the posterior.

5.4.3 Extension formulas to a three person DNA mixture with three PoIs

For a three-person DNA mixture involving three PoIs, we extend the two-person formula 5.3 as follows:

$$
\begin{aligned}
LR_{2/\bar{2}} &= \frac{LR_{H_{123,uuu}}\pi_{123} + LR_{H_{12u,uuu}}\pi_{12u} + LR_{H_{23u,uuu}}\pi_{23u} + LR_{H_{2uu,uuu}}\pi_{2uu}}{LR_{H_{13u,uuu}}\pi_{13u} + LR_{H_{1uu,uuu}}\pi_{1uu} + LR_{H_{3uu,uuu}}\pi_{3uu} + \pi_{uuu}} \\
&\times \frac{\pi_{13u} + \pi_{1uu} + \pi_{3uu} + \pi_{uuu}}{\pi_{123} + \pi_{12u} + \pi_{23u} + \pi_{2uu}} \\
&= \frac{10^{33}\pi_{123} + 10^{27}\pi_{12u} + 10^{12.2}\pi_{23u} + 10^{12.1}\pi_{2uu}}{10^{29.4}\pi_{13u} + 10^{17.5}\pi_{1uu} + 10^{7.9}\pi_{3uu} + \pi_{uuu}} \times \frac{\pi_{13u} + \pi_{1uu} + \pi_{3uu} + \pi_{uuu}}{\pi_{123} + \pi_{12u} + \pi_{23u} + \pi_{2uu}}.
\end{aligned}
$$

Recall that we find S_2 as the event that PoI 2 is a donor of the trace. Then the Posterior for PoI 2 can be calculated as follows:

$$
P(S_2 \mid E) = \frac{\pi_{123}LR_{123,uuu} + \pi_{12u}LR_{12u,uuu} + \pi_{23u}LR_{23u,uuu} + \pi_{2uu}LR_{2uu,uuu}}{\sum_{H\in\mathcal{H}} \pi_H LR_{H,uuu}},
$$

where in the summation is the summand over all eight possible hypotheses. The corresponding formulas for PoI 1 and PoI 3 can be calculated similarly.

5.4.4 Results and conclusion case study

	Case 1	Case 2	Case 3
$LR_{1/\bar{1}}$	3.52e+20	1.41e+17	7.42e+20
$LR_{2/\bar{2}}$	3.98e+03	6.12e+03	6.37e+03
$LR_{3/\bar{3}}$	1.00e+06	1.94e+08	4.55e+05
$P(S_1 \mid E)$	1	1	1
$P(S_2 \mid E)$	0.999	7.90e-3	0.999
$P(S_3 \mid E)$	0.999	0.996	0.500

Table 5.5: LRs of contribution and posterior probabilities for all three PoIs and the three cases

In Table 5.5 we note that the LRs for PoI 1 are really large, also the posterior for PoI 1 is in all three cases 1. Strictly speaking the posteriors should all three be a bit smaller than 1, but they are rounded up in excel. In case 1 the posterior probability for PoI is 0.9999... where the 16th decimal is the first decimal unequal to 9. For PoI 1 the evidential value for contribution is really strong all three cases. In case 1, as expected, for all three PoIs the posterior is almost 1. Meaning that contribution of all PoIs is likely.
In case 2 the posterior for PoI 2 is really low, but for PoI 3 it is almost 1.
In case 3 the posterior for PoI 2 is almost one and for PoI 3 a half. The half is due to the fact the the posteriors for H_{123} and H_{12u} are equally likely and the two dominant posteriors.
The shift in posterior probability for PoI 2 between case 2 and case 3 is due to the assumed association between PoI 1 and PoI 2 in case 3, which increases the prior probability of their joint contribution.

This case study highlights the critical importance of carefully selecting the prior probabilities, as they can significantly alter the conclusions. In case two the conclusion would be that PoI 2 did not contribute to the trace and case three the conclusion is that PoI 1 did contribute to the trace. This change of result is caused by the change of prior probabilities and thus highlights the proper use of prior probabilities.

A sound strategy for the forensic expert here is to report a comprehensive table of LRs rather than relying solely on the highest LR. Reporting only the highest LR and stating that the highest LR is at least $10^{3.6}$ times more likely to explain the trace than all other considered hypotheses is insufficient, as such a margin may be overridden by differences in prior probabilities, thereby altering which hypothesis is most supported.

Chapter 6

Body Fluid Analysis

Evidentiary biological traces provide not only information about the contributors to the trace, but also about the types of body fluids present. This distinction is particularly important in case such as sexual assaults, where identification of specific body fluids can substantially influence the interpretation of the evidence in court. For judicial interpretation, it is important to distinguish whether only skin cells were present at the crime scene or whether additional fluids such as vaginal mucosa or menstrual secretion were also identified.
Several analytical methods exist for identifying specific body fluids. For instance, seminal fluid can be identified using RSID (Rapid Stain Identification Series) semen test or the PSA (Prostate Specific Antigen) test and for saliva there is the RSID saliva test. However, not all body fluids have corresponding detection methods, and most current methods are only suitable to identifying a single body fluid type per analysis. Moreover, conducting these tests consumes part of the evidentiary sample, thereby reducing the material available for subsequent DNA analysis.

6.1 mRNA profiling: categorical method

An alternative method for body fluid identification involves the use of messenger RNA (mRNA) profiling. Messenger RNA molecules reflect the activity of specific markers within a cell. Body fluids can, to some extend be distinguished based on their markers. For instance, the marker hemoglobin is typically detected in blood samples. However, hemoglobin may also be present in other body fluids. While hemoglobin is consistently present in blood, some other markers appear only intermittently across samples. For example, the marker CD93 is detected in approximately 58% of the blood samples. Figure 6.1 presents the detection rate of 15 relevant markers across different body fluids. Body fluid-specific markers are indicated in purple. If all body fluid-specific markers for a particular fluid are detected, it is likely that the corresponding fluid is present in the mixture. The detection rates represent the percentage of samples in which a given marker is present. The detection rates were derived from a dataset consisting of single-source body fluid samples [1].

This dataset, compiled by the NFI, provides insight into marker overlap across body fluids. It includes the peak heights for 15 relevant markers from single samples are available expressed in relative fluorescence units (rfu). Additionally, there are two housekeeping markers used to assess the mRNA profile quality, as well as two sex determination markers. It is important to note that these samples may not be entirely pure; for instance, saliva can easily spread and may be detected on samples primarily containing other fluids such as skin.
Because many markers are shared among body fluids, and even fluid-specific markers are not consistently present, distinguishing between fluids can be challenging. To illustrate marker occurrence across fluids, a heatmap in Figure 6.2 displays the mean peak heights per marker for each body fluid. The body fluids may be distinguished based on the peak heights, but the difficulty is what happens when the quantity/quality of the sample is not so good. For example, do the peak heights for the same marker sum up in a mixture or is it the maximum peak height. The peak heights for different markers can differ substantially due to numerous variables, such as for example physical condition of the donor [3].

[1] The dataset is available from an online repository `https://github.com/NetherlandsForensicInstitute/body_fluids_mRNA/blob/master/Datasets/Dataset_NFI_rv.xlsx`

		HBB	ALAS2	CD93	HTN3	STATH	BPIFA1	MUC4	MYOZ1	CYP2B7P1	MMP10	MMP7	MMP11	SEMG1	KLK3	PRM1
1	Blood	1	0.960	0.579	0	0	0	0	0	0	0	0	0.032	0	0	0
2	Menstrual secretion	1	0.496	0.451	0	0.009	0	0.566	0.531	0.31	0.319	0.381	0.558	0	0	0
3	Nasal mucosa	0.008	0	0.432	0.008	0.976	0.504	0.616	0.016	0.016	0	0.008	0.024	0.024	0	0
4	Saliva	0.159	0.009	0.028	0.907	0.907	0.019	0.009	0.019	0.009	0	0.009	0	0	0	0
5	Semen fertile	0.011	0.011	0	0	0	0.011	0.011	0	0	0	0	0	0.832	0.789	0.958
6	Semen sterile	0	0	0	0	0	0	0	0	0	0	0	0.031	0.875	0.656	0
7	Skin	0.264	0.014	0.111	0	0.083	0.028	0.194	0.056	0	0	0.0278	0	0	0	0
8	Vaginal mucosa	0.009	0	0.157	0	0	0	0.922	0.722	0.557	0	0.043	0.009	0	0	0
9	Skin penile	0.146	0	0.042	0	0	0	0.333	0.021	0	0.021	0.021	0.042	0	0	0.104

Figure 6.1: Detection rate markers per body fluid. The colored squares highlight markers that are body fluid specific markers to a specfic body fluid. Taken from [7]

6.1.1 n/2 method

A straightforward method for body fluid identification is the n/2 method, introduced by Lindenbergh et al. For this method all body fluids have a couple of body fluid specific markers allocated, which are highlighted in Figure 6.1. The method provides a categorical verdict based on the presence of these body fluid specific markers in samples obtained from the crime stain. More precisely, let x denote the number of observations of specific markers associated with a given body fluid in the collected samples, and let n represent the maximum number of such markers that could theoretically be detected. For example, with two samples and three body fluid specific markers, the total number of theoretical detections would be $n = 2 \times 3 = 6$. If $x \geq \frac{n}{2}$, the method reports that there is an indication for presence of the body fluid. If $0 < x < \frac{n}{2}$, it is reported that no reliable statement is possible for the body fluid, and lastly if $x = 0$ it is reported that there is no indication of presence for the body fluid.

Example

Consider that we have taken two samples from a crime scene. In the first sample the markers HBB, ALAS2, CD93, MUC4, SEMG1 and KLK3 are present and in the second sample are the markers HBB, ALAS2, CD93 present. Based on this, the n/2 method concludes that there is an indication of the presence of blood and sterile semen. For saliva and fertile semen, the method yields no reliable conclusion. For the other body fluids there is no indication of presence.

Difficulties for the n/2 method

A drawback of this categorical method is the 'fall-off-the-cliff-effect', which arises from the hard cut off between the different verdicts. For instance, the method yields the same verdict whether 1 out of the 12 markers is detected or 5 out of the 12 are detected, yet this does not correspond to a proportional increase in evidential support. However, if 6 out of the 12 markers are detected, the verdict changes abruptly. The categorical approach does not reflect the gradual increase in support as more markers are found.

By contrast, probabilistic models provide a probabilistic statement about the presence of certain cell types, rather than issuing a fixed categorical verdict. A multi-label classification framework is particularly suitable for identifying the simultaneous presence of multiple body fluid types withing a sample. In the probabilistic model we can set up an LR system, such that the background information of the case can be used as prior information after which the posterior can be calculated. The n/2 method does not make use of the possibly available background information in the case.

Interestingly, for menstrual secretion, the average detection rate over the corresponding specific markers is lower than 0.5, as could be seen in the Figure 6.1. This implies that for single-source menstrual secretion samples, the n/2 method will most likely yield a 'no reliable conclusion' outcome, despite the fluid being present. The method therefore fails to incorporate the differences in the amplification rates for different markers and cell types.

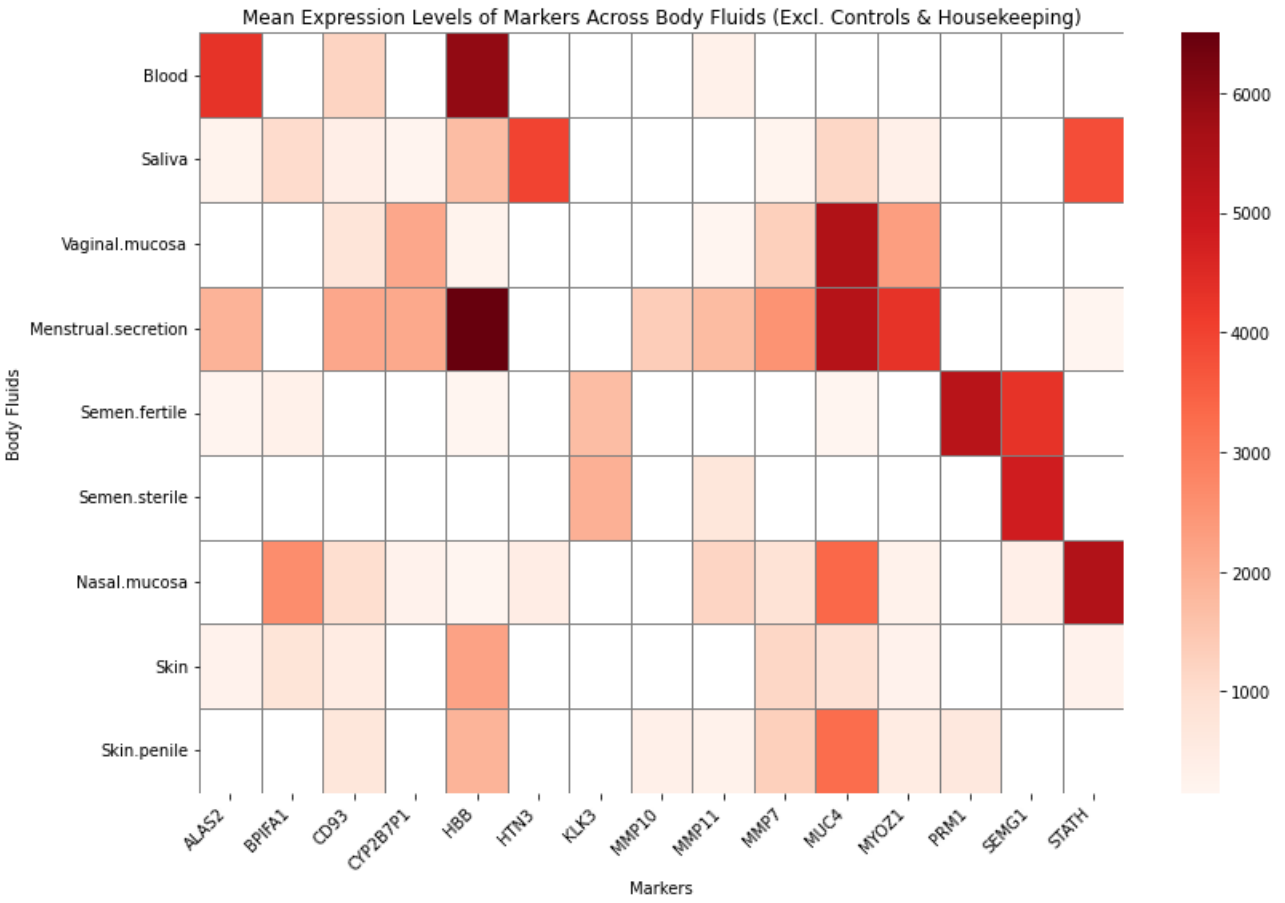


Figure 6.2: Heatmap showing the mean relative fluorescence unit (rfu) values of 15 mRNA markers across nine body fluid types. Higher rfu values indicate stronger marker expressions.

Additionally, the method performs poorly when distinguishing between nasal mucosa and saliva in mixed samples. Both body fluids are characterized by only two specific markers, one of which they share. Since this shared marker has a high detection rate (higher than 0.9) in both fluids, the presence of either body fluid often causes the n/2 threshold to be exceeded for both, resulting possibly in a misclassification.

6.2 mRNA profiling: probabilistic methods

6.2.1 LR framework for mRNA analysis

Computing an LR for mRNA based body fluid analysis is more complex than for DNA based identification analysis. In forensic DNA analysis, the alternative hypothesis can be constructed involving unknown contributors based on the allele frequencies in the population. In contrast, constructing such an alternative hypothesis for mRNA based body fluid identification is not straightforward. Ideally, one would require information about which body fluids, and consequently, which markers, would be expected in the absence of the alleged crime. In practice, this is intractable, as the necessary contextual information is typically case-specific and often unavailable.

Consequently, a different strategy must be employed for mRNA based inference. One such approach involves considering all possible combinations of body fluids. Assuming the nine body fluids shown in Figure 6.1 are of interest, there are 2^9 possibilities of possible mixtures (including single-fluid and no-fluid scenarios, which are also referred to as mixtures for simplicity).

All 2^9 mixtures can be generated in silico by computationally combining marker data from single-source body fluid samples. Each generated mixture is then evaluated for its similarity to the crime scene sample using a probabilistic model.

To compute an LR for the body fluids under investigation, referred to here as the specified body fluids, we define the following hypotheses:

- H_p : The sample contains the specified body fluid(s), and possibly others.
- H_d : The sample does not contain all of the specified body fluids.

The first step is to compute the probability for each of the 2^9 created mixtures that include the specified body fluid(s), that the mixture matches the observed mixture of the crime scene sample. These probabilities are estimated using a probabilistic model (e.g., using logistic regression). Next, we sum over all probabilities in which the body fluids from the prosecution's hypothesis are present to obtain the probability that the specified body fluid(s) are present in the crime scene mixture. An LR for the competing hypotheses is then computed as follows:

$$LR = \frac{\sum_{m \in H_p} P(m)}{1 - \sum_{m \in H_p} P(m)}.$$

Here, $P(m)$ represents the estimated probability that mixture m corresponds to the true composition of the crime scene sample. The summation is over all mixtures that contain at least the specified body fluid(s) from H_p.

6.2.2 Prior probabilities incorporated in training data

In probabilistic methods, prior probabilities are often implicitly embedded within the training data used for classification. However, this approach conflicts with the principles of forensic inference. The consensus is that the forensic expert calculates an LR, while the decision maker assigns prior probabilities based on contextual case information. The classification outcomes of probabilistic methods often depend on the relative frequencies of body fluids in the training data. For example, if a particular body fluid is present in 50% of the training data, then the model implicitly treat the prior probability for that body fluid as 50%.

Ypma et al. conducted a sensitivity analysis to evaluate this effect. They trained probabilistic models under varying background levels. With the background levels we mean the presence of the body fluids. They compared uniform background levels, thus all body fluids are as many times present, against adjusted background levels. First, a uniform background level of 50% was assigned to all body fluids. After that, different dataset splits are tested where background levels for selected fluids are set to 90%. If large deviations of LRs are noted, this indicates that the model's output is sensitive to the choice of the prior.

6.2.3 Best probabilistic methods for mRNA profiling

Zoete et al. present two probabilistic methods for making probabilistic inferences about the presence of specific body fluids in a sample. The first method is a Bayesian Network (BN) with a naïve Bayes assumption, where the detection rate of each marker is considered conditionally independent given the body fluid type. This model generates LRs, which can subsequently be converted into a posterior probability. The second method is a multinomial logistic regression (MLR) model, which directly estimates posterior probabilities. The prior probabilities are incorporated in the training process, based on the relative frequencies of different body fluids in the training data.

The performance of both probabilistic methods, along with the categorical n/2 method, is tested in Zoete et al. For the BN and MLR the majority rule is used, which means that the output of the models is instead of a probabilistic statement the mixture that is most likely. While this simplification results in a loss of information, it allows a direct comparison with the categorical n/2 method. Even under this majority rule simplification both probabilistic models outperform the n/2 method. However, all three the methods show limitations when applied to samples containing saliva.

Although both probabilistic models perform similarly the paper concludes that the BN with naïve Bayes assumption method is preferable as this method is easier to adapt to complex situations. Furthermore, it does not assume a prior distribution and it is easier to implement/understand.

Ypma et al. explored five classification models: (multinomial) logistic regression, multi-layer perceptron, support vector machine, and random forest. After training, each model outputs a score for a new sample. This score is mapped with a calibration step into an LR. The authors conclude that multinomial logistic regression method is the most suitable method and note: "Its performance is comparable to other models, its robustness is similar or better and its interpretability is much higher. "

6.3 Case study: Evaluating a body fluid mixture using the n/2 method and multinomial logistic regression.

In this Section, we analyze a trace recovered from a crime scene, with the objective of determining whether vaginal mucosa and semen are present in the sample. Table 6.1 presents the detection results for each mRNA marker across three replicates.

	Vaginal mucosa			Semen			Other markers								
Markers	MUC4	MYOZ1	CYP2B7P1	SEMG1	KLK3	PRM1	HBB	ALAS2	CD93	HTN3	STATH	BPIFA1	MMP10	MMP7	MMP11
Samples	2/3	1/3	3/3	2/3	1/3	0/3	3/3	3/3	2/3	0/3	0/3	0/3	0/3	0/3	1/3

Table 6.1: Detection results for three replicates of the sample.

6.3.1 Evaluation based on detection rates

Before applying the n/2 method and multinomial logistic regression, we first interpret the mixture based on the detection rates from Figure 6.1. An initial examination reveals that six out of nine body fluid specific markers for vaginal mucosa are present. For semen, the number is slightly lower: three out of nine. It should be noted that PRM1 is a specific marker for fertile semen only, and not for sterile semen, thus for sterile semen three out of six specific markers are present. Additionally, the three blood specific markers were detected eight out of nine possible observations.

The marker MMP11 was detected in one of the three replicates. This detection could originate from either blood or semen sterile, as they have a low detection rate for this marker. Furthermore, KLK3 is a useful marker. In Figure 6.1, we find that this marker is only detected for semen sterile and semen fertile, supporting the inference that one of these body fluids is likely present.

6.3.2 n/2 method

Using the n/2 method we find the following results:

- For vaginal mucosa, six out of nine specific markers are detected, which exceeds the n/2 threshold. The method concludes that there is an indication of presence for vaginal mucosa.
- For semen fertile, three out of nine specific markers are detected, which falls below the n/2 threshold, but greater than zero. The method concludes that there is no reliable statement possible for semen fertile.
- For semen sterile, three out of six specific markers are detected, which meets the n/2 threshold. The method concludes that there is an indication of presence for semen sterile.

6.3.3 Multinomial logistic regression

Although an MLR model has not been trained in this study, we outline the general approach for illustrative purposes. In MLR, the log LR for the presence of a specific body fluid is computed as: $\log \text{LR} = \beta_0 + \sum_{i=1}^{p} \beta_i r_i$. Here, the β's are parameters estimated from the training data. The r_i's are the detection results (e.g., 2/3 for MUC4) and p is the number of markers. If no markers are detected $r_i = 0$ for all i and thus the log LR will be equal to the intercept β_0. The coefficient β_i reflects how much the detection of marker i increases or decreases our belief in the presence of the target body fluid. For instance, when modeling the likelihood for blood, we would expect the β_i corresponding to hemoglobin to be positive, as hemoglobin is a specific body fluid for blood. To compute log LRs for specific hypotheses we must first train the MLR model on labeled data containing the relevant body fluid classes. Once trained, the model can evaluate new samples and output log likelihood ratios for each body fluid of interest.

6.3.4 Conclusion

To conclude, the n/2 method is simple, quick to apply, and easy to interpret. It offers a clear verdict based on the presence or absence of specific markers. This makes the n/2 method a suitable starting point for forensic research. However, its simplicity limits the evidential strength it can provide.

The MLR method is more complex and requires model training and prior probabilities. By contrast, the probabilistic framework allows for a more nuanced and quantitative interpretation of the evidence.

Chapter 7

Comparison Forensic DNA Analysis and mRNA Analysis

Forensic DNA analysis and mRNA analysis represent two distinct but complementary domains within forensic science. While DNA profiling seeks to identify individuals who may have contributed to a trace, mRNa profiling is primarily concerned with determining which body fluids are present in a sample. This chapter provides a comparative analysis of the two domains with the aim of understanding their respective methodological frameworks, statistical structures, and reporting practices.

7.1 Comparison of the state space

In DNA profiling, it is often possible to estimate the number contributors to a DNA trace based on the number of observed alleles, especially when the number of contributors low. For mRNA analysis determining the number of body fluids in a mixture is more difficult. This is because many markers are shared across different body fluids. The presence of all markers for a specific body fluid does not always guarantee that the body fluid is actually present.

One notable advantage in mRNA profiling is that the set of potential contributors, the body fluids, is finite and predefined. This constrained state space enables a complete, tractable enumeration of all possible combinations (2^9 in total), facilitating systematic hypothesis testing. In contrast, forensic DNA profiling involves an open set of potential contributors, making exhaustive exploration of the state space computationally infeasible.

Another key difference lies in the construction of the alternative hypothesis. In DNA analysis, the allele frequencies of the population are used for the alternative hypothesis. This enables the comparison between a PoI and a weighted average over the allele frequencies. The alternative hypothesis for mRNA profiling is more difficult. Ideally, we would want to know what body fluids would be present if the crime did not take place, but this heavily depends on the context of the scene. For example, background body fluids will differ between a bathroom, bedroom or a park.

7.2 Comparison on reporting

In DNA analysis, it was suggested by Slooten et al. to report a table with LRs. This allows the decision maker to incorporate their prior beliefs and move from prior odds to posterior odds. For mRNA profiling making such a table is more complicated. There are 2^9 possible mixtures, which makes the table enormous and difficult to use. Moreover, to derive an LR for the presence blood in the trace, the decision maker would need to assign priors for all hypotheses. This can be done using equation (5.3), where the PoIs are replaced by body fluids. One possible solution is to report only those mixtures that have an LR above a certain threshold (e.g., greater than 1). Here the LR refers to the prosecution's hypothesis that a specific body fluid (or several) is present in the mixture, possibly with other body fluids, against the defendant's hypothesis that not all specified body fluids are present. Note that applying MLR as described in Section 6.2.3 a considerable amount of effort is required, as the model needs to be trained based on the hypotheses.

Reporting only a table of LRs per body fluid is not sufficient, it does not reflect the dependencies between

the body fluids in the mixture. For example, it is unclear for the decision maker how the LR for menstrual secretion change depending on whether blood is assumed to be present or absent in the trace. The trace is often a mixture of different body fluids this dependency is often significant and should not be ignored in interpretation.

If background information clearly indicates that a specific body fluid is present, the MLR can be trained on this information. All mixtures of the training data should then contain this body fluid. If it is clear from the background information that one specific body fluid is not present in the mixture, then this body fluid should be excluded from the training data. In contrast, if one contributor of a DNA trace is already known, the this contributor can be conditioned on during the analysis, which simplifies the modeling of the remaining contributors.

7.3 Conclusion

In conclusion, forensic DNA analysis and mNRA analysis differ substantially in both statistical framework and reporting complexity. DNA analysis benefits from the available allele frequencies in the population, enabling the construction of likelihood ratios based on the alleles. However, it requires the availability of genotype profiles from the PoIs, which can limit its applicability. By contrast, mRNA analysis does not rely on individual profiles but instead assumes a fixed set of potential body fluids. The main complexity in mRNA analysis is that reliable marker frequencies that would be expected if the crime did not take place are not available. As a result, defining a meaningful alternative hypothesis is more difficult in mRNA analysis.

In DNA analysis, LRs can often be clearly presented in a table covering a manageable set of hypotheses. In mRNA analysis, however, dependencies between body fluids and the implicit incorporation of prior probabilities through model training make interpretation more complex. Consequently, summarizing the evidential value in a transparent and interpretable way requires greater care and methodological nuance.

Part II

LR Distributions for the Donors of DNA Mixtures

Chapter 8

LR Distribution

In this chapter, we explore methods to estimate the distribution of LRs in advance of a forensic investigation. Specifically, we focus on the LR obtained by comparing the prosecution's hypothesis that a PoI a contributor of the trace together with an unknown (or several unknowns) against the defendant's hypothesis that two unknowns created the trace (or more unknowns). Having access to this LR distribution in advance allows the forensic expert to check whether the LR observed for a suspect aligns with what would be expected if the suspect were indeed a true contributor. We use a dataset of laboratory-generated DNA mixtures[1]. The dataset covers a broad range of scenarios, including mixtures with varying DNA proportions per donor, differing levels of allele sharing, and mixtures containing between two and five contributors. Each mixture is produced three times to account for technical variability in DNA processing, such as pipetting inconsistencies. These technical replicates help ensure that observed differences in LRs are due to the biological and statistical features of the mixtures, rather than random experimental noise.

8.1 Method 1: Posterior probability threshold

In this method, we calculate the posterior probabilities for all possible genotypes at each locus, representing the potential allele combinations that a given donor may have. The posterior probability is defined as the probability that donor i has genotype (a_1, a_2) at locus j, given the observed mixture E. At each locus, all alleles not present in the mixture are grouped together into a placeholder called allele Q. This allows us to account for possible allele drop-in and drop-out, while significantly reducing the number of genotypes for which posterior probabilities must be calculated. For each locus, we generate all possible genotypes that can be formed using the observed alleles in the mixtures, along with allele Q. For each of these genotypes, we calculate the corresponding LRs and posterior probabilities.

8.1.1 Threshold method in detail

For each locus, DNAStatistx calculates the unnormalized posterior probabilities $P(E|D_j^i = (a_1, a_2))P((a_1, a_2))$ for all generated allele pairs (a_1, a_2). This represents the probability of observing the evidence (i.e., the DNA mixture), assuming that donor i has genotype (a_1, a_2) at locus j, multiplied by the prior probability (i.e., the population frequency of the genotype). The likelihoods $P(E|D_j^i = (a_1, a_2))$ are marginal likelihoods as the likelihoods are marginalized over the donors. In the sequel, the unnormalized posterior probabilities will be abbreviated to $P(E \mid (a_1, a_2))P((a_1, a_2))$.

Suppose that at locus j, alleles 13 and 14 are observed in the DNA mixture. We restrict genotype generation to combinations involving only these observed alleles, and group all other alleles into a placeholder denoted as Q. Thus, Q means that there is a dropout, i.e. an allele present in the genotype of the donor is not present in the mixture. Therefore, the set of considered allele pairs at this locus is: $(13, 13), (13, 14), (13, Q), (14, 14), (14, Q)$, and (Q, Q).

[1] All generated mixtures and genotypes of the donors are available in the online repository `https://github.com/NetherlandsForensicInstitute/DNANet/tree/main/resources/data`

The posterior probability for the allele combinations at a specific locus can be calculated as follows

$$P((a_1, a_2) \mid E) = \frac{P(E \mid (a_1, a_2))P((a_1, a_2))}{P(E)} = \frac{P(E \mid (a_1, a_2))P((a_1, a_2))}{\sum_{(b_1,b_2)} P(E \mid (b_1, b_2))P((b_1, b_2))}.$$

In the first step, Bayes' theorem is applied and in the second step the law of total probability is used. For each locus, the possible allele combinations are ordered on posterior probability. We construct a dictionary G that, for each locus, stores the most probable allele combinations whose cumulative posterior probability exceeds a threshold t (e.g., $t = 0.5$). This ensures that, with at least probability t, the true allele combination for the donor is included in G. In most cases, the total posterior probability of selected combinations exceeds the threshold, since the last included combination will typically push the cumulative sum above t.

Next, we use dictionary G to construct a set S containing full genotypes that can be formed by combining the selected allele pairs across loci. Assuming independence across loci, the posterior probability that the true genotype is included in S is at least $t^{\text{Number of loci}}$. The size of S grows exponentially in the number of allele combinations possible at each locus. The number of genotypes in S is the product over all considered loci of the number of allele combinations possible at the locus.

8.1.2 Disadvantages of the threshold method

The main limitation of this method is that it relies on setting a threshold to determine which allele combinations are included. When the contributors to the mixture cannot be clearly distinguished, the posterior probabilities for the best-fitting allele combinations tend to be lower. This means that more allele combinations must be stored in order to exceed the threshold. The number of possible genotypes generated from the dictionary grows exponentially with the number of allele combinations, which can make the method computationally expensive or even infeasible. As a result, there is a trade-off between computational efficiency and accuracy, defined here as the probability that the correct genotype is included in the generated set. A higher accuracy thus means that the threshold needs to be placed higher and this implies that more genotypes are generated, giving more computation time.

For several of the mixtures analyzed, the threshold could not be set higher than $t = 0.2$. Then when considering 23 loci, there is only a $0.2^{23} \approx 10^{-16}$ probability of having the correct genotype in the genotype set. This will not lead to a good expectation of the LR that is expected for the true donor and the posterior probabilities of the genotypes obtained will be low.

8.2 Method 2: Sampling genotypes

Instead of generating a set of genotypes using a posterior probability threshold, an alternative approach is to apply a sampling method. In this approach, genotypes are sampled according their posterior probabilities. In this way the genotypes are sampled proportional to their probability and also the number of times a specific genotype is sampled gives an indication of the probability that that is the correct genotype of the donor.

As before, DNAStatistx is used to compute unnormalized posterior probabilities. Thus, for each donor, it calculates for each locus all possible allele combinations (a_1, a_2)

$$P(E \mid (a_1, a_2))P((a_1, a_2)).$$

Using the Bayes' formula and the law of total probability the posterior probability can be calculated.

$$P((a_1, a_2) \mid E) = \frac{P(E \mid (a_1, a_2))P((a_1, a_2))}{P(E)} = \frac{P(E \mid (a_1, a_2))P((a_1, a_2))}{\sum_{(b_1,b_2)} P(E \mid (b_1, b_2))P((b_1, b_2))}.$$

These posterior probabilities are then used as sampling weights to draw allele combinations per locus. The sampling process works as follows: on each locus we sample an allele combination by using the corresponding

posterior probabilities as weights.

In this way, the genotypes that have highest probability of occurrence will be sampled more often. According to the law of large numbers, the proportion of times a genotype is sampled converges to its posterior probability of being the true donor genotype.

8.3 Calculate the LR and posterior probability for the generated genotypes.

Given the set of generated genotypes, the next step is to evaluate how well each genotype explains the observed DNA mixture. We consider the following hypotheses:

- H_1 : The allele combination for donor i on locus j is (a_1, a_2).
- H_2 : The allele combination for donor i on locus j is unknown.

To compute the likelihood of the evidence under H_2 we apply the law of total probability by computing a weighted average of the likelihoods over all possible allele combinations, using their genotype frequencies in the population.

$$LR(a_1, a_2) = \frac{P(E \mid (a_1, a_2))}{P(E \mid \text{unknown})} = \frac{P(E \mid (a_1, a_2))}{\sum_{b_1, b_2} P(E \mid (b_1, b_2))P((b_1, b_2))}.$$

The likelihood $P(E|(a_1, a_2))$ can be obtained by dividing the unnormalized posterior probability calculated by DNAStatistx by the population frequency of the allele combination.

$$\frac{P(E \mid (a_1, a_2))P(a_1, a_2)}{P(a_1, a_2)} = P(E|(a_1, a_2)).$$

As the allele pairs at each locus are independent, the LR for a full genotype g, comprising allele pairs across multiple loci, can be computed by taking the product of LRs at each locus. Thus,

$$LR(g) = \prod_{\text{locus}} LR(a_1, a_2).$$

The posterior probabilities for the genotypes can be calculated using Bayes' law and the law of total probability. Assuming independence between loci, the posterior probability of a full genotype can be obtained by taking the product of the posterior probabilities at each individual locus.

$$\begin{aligned} P(g \mid E) &= \prod_{\text{locus}} P((a_1, a_2) \mid E) \\ &= \prod_{\text{locus}} \frac{P(E \mid (a_1, a_2))P((a_1, a_2))}{P(E)} \\ &= \prod_{\text{locus}} \frac{P(E \mid (a_1, a_2))P((a_1, a_2))}{\sum_{(b_1, b_2)} P(E \mid (b_1, b_2))P((b_1, b_2))} \end{aligned}$$

8.4 Results two-person mixtures

8.4.1 Mixture with 300:150 DNA proportions

In this section, we compare the threshold method and the sampling method for generating a set of genotypes for the mixture shown in Figure 3.1. The DNA proportions in the mixture are 300:150 in picograms.

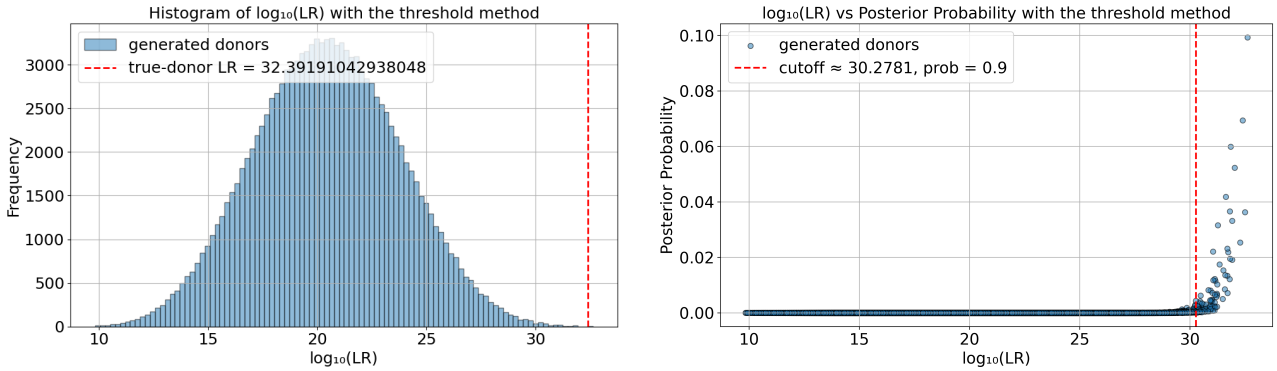
Threshold method

We begin by applying the threshold method to the major donor. Through testing, it was determined that a threshold of $t = 0.999$ could be applied. This high threshold is feasible in this case as the alleles from the major donor and the minor donor are distinguishable. Thus, for each locus, the posterior probability of having considered the correct allele combination is 99.9%.

Figure 8.1a shows the LRs of all generated genotypes. The distribution appears approximately a normal distribution. However, this histogram reflects only the number of genotypes that result in a given LR. It does not account for the posterior probabilities of these genotypes, and therefore does not accurately represent the expected LR for the donor. For example, we would expect that the LR of 10^{30} is much more likely to be the LR of the true donor than the LR 10^{10}. This relationship is not captured in the histogram. The issue with this histogram is that it does not plot LR against their posterior probabilities. The true-donor LR is for the major $10^{32.4}$ and is indicated by a red line in the figure. It seems from the histogram that the true-donor LR is even higher than the genotypes coming from the threshold method, but this is thus caused by the fact that LRs are not plotted against the probability of obtaining this LR.
On the other hand, for this mixture the histogram gives a couple of insights. One key insight is that, because the threshold is high we know that with a posterior probability of at least $0.999^{23} \approx 0.977$ the correct genotype is in the range of the LRs from the histogram. This probability is even higher if we use the achieved sum of posterior probabilities per locus instead of the threshold and then take the product over these values. By taking allele combinations until the threshold is achieved, the actual posterior probability is often higher than the threshold. The allele combinations do often not exactly sum up to the threshold. Taking the product over the achieved posterior probabilities, we find that there is a probability of 0.9966 that the genotype of the donor is in the set. This histogram therefore provides an indication of the range of LRs that might be expected for the true donor.

In Figure 8.1b the LRs of the genotypes are placed against the posterior probabilities of the genotypes. In this case, the genotype with the highest LR also has the highest posterior probability. This is not necessarily the case. The relative frequency (prior probability) of the genotype affects the posterior probability. The highest LR has a posterior probability close to 10%.

The scatterplot allows us to infer the minimum LR that is expected for the true donor. For example, the red line in the scatterplot indicates the posterior probabilities on the right hand side of the line have 90% of the posterior mass. This indicates that, with 90% posterior probability, the LR of the true donor is greater than $10^{30.2855}$. This provides useful information in forensic investigations. If an LR 10^{25} is observed, this method allows us to estimate the probability that the LR of the true donor exceeds 10^{25}.

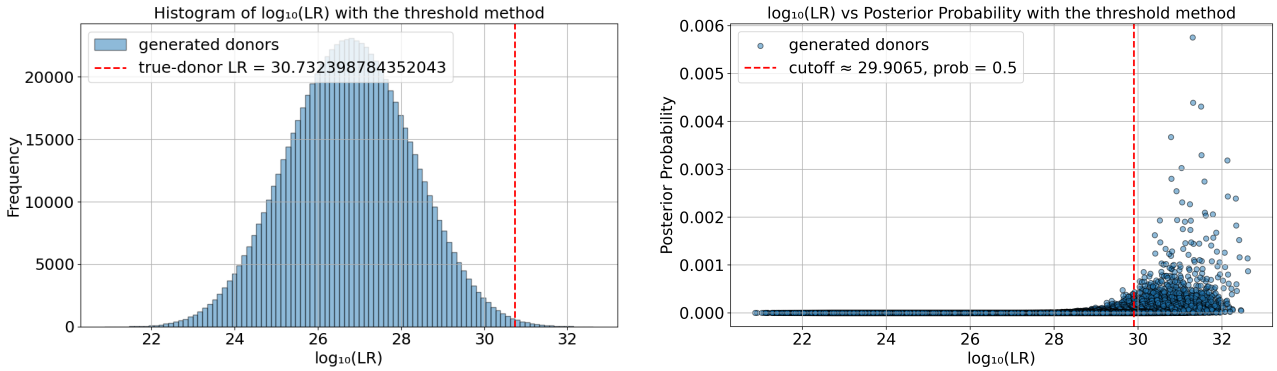


(a) Histogram

(b) Scatterplot

Figure 8.1: Histogram and scatterplot for the major donor in the mixture with 300:150 DNA proportions, using a threshold of $t = 0.999$ in the threshold method. In the histogram, the red line marks the LR of the true donor's genotype. In the scatterplot, the red line represents a posterior cutoff: the cumulative posterior probability of genotypes to the right of this line is at least 0.9.

Next, we apply the threshold method to the minor donor. Due to greater uncertainty about which alleles belong to the minor donor than for the major donor, a lower threshold of $t = 0.95$ was used.



(a) Histogram

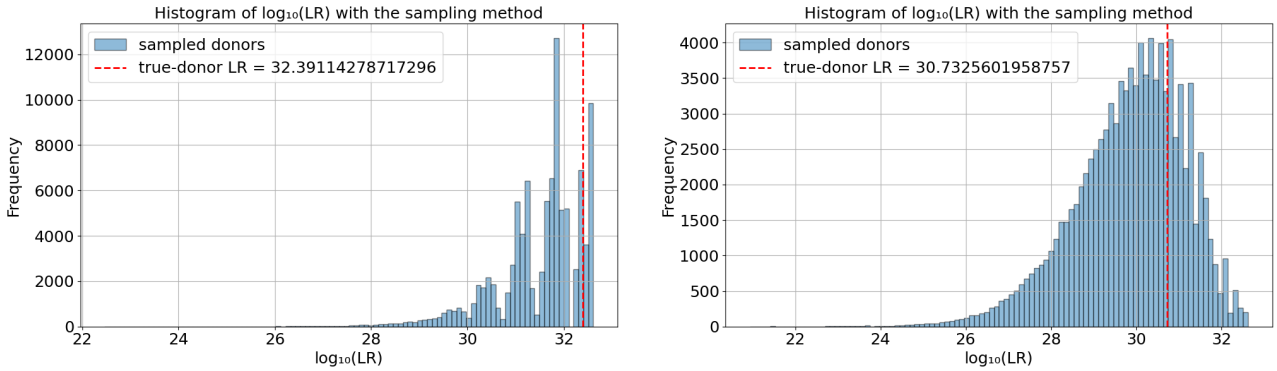
(b) Scatterplot

Figure 8.2: Histogram and scatterplot for the minor donor in the mixture with 300:150 DNA proportions, using a threshold of $t = 0.95$ in the threshold method. In the histogram, the red dashed line marks the LR of the true donor's genotype. In the scatterplot, the red dashed line represents a posterior cutoff: the cumulative posterior probability of genotypes to the right of this line is at least 0.5.

The probability that the actual genotype of the donor is in the LR range of the histogram is calculated to be 0.76. Again, the red line is on the far right side of the histogram in Figure 8.2a. Figure 8.2b shows that the LR of the true genotype is greater than $10^{29.9029}$ with 50% posterior probability.

Sampling method

Using the sampling method, we generate $n = 100.000$ genotypes for both the major donor as the minor donor.



(a) Major donor

(b) Minor donor

Figure 8.3: Histogram of the LR distribution for the major and minor donor in the 300:150 DNA mixture. The genotypes are generated using the sampling method. The red dashed line marks the LR of the true donor's genotype.

Figures 8.3a and 8.3b the LR distributions, with the true donor's LR indicated by a red dashed line. In contrast with the threshold method, these histograms approximate the actual distribution of the LRs, as the genotypes are sampled proportionally to their posterior probabilities. The rightmost bin in the major donor's histogram corresponds to a single genotype with an LR of approximately $10^{32.6}$. The genotype was sampled approximately 10.000 times, meaning it appeared in about 10% of all samples. This result is consistent with the posterior probability previously calculated using the threshold method. Figure 8.3a shows multiple peaks. This is due to the fact that relatively few different genotypes are sampled, which results from the alleles being easily distinguishable based on peak heights. The limited number of distinct genotypes are sampled repeatedly, leading to the visible spikes in the distribution.

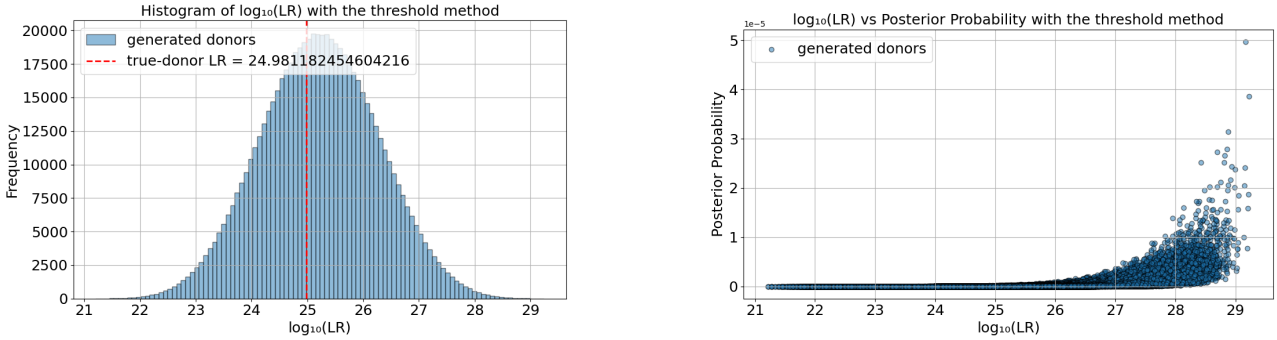
8.4.2 Mixture with 150:150 DNA proportions

In this section, a two person mixture with DNA proportions of 150:150 picograms is analyzed. The goal was to include equal amounts of DNA from both contributors. If that is the case, we would expect that the LRs

and posteriors for the major/minor donor are more or less the same, as there is no major/minor donor, but two profiles that donated as much DNA. Since there is no true major or minor donor in this case, both possibilities were explored. We evaluated both scenarios and assigned the genotype with the highest LR as the major donor and the one with the lowest LR as the minor donor. In one configuration, the major donor has an LR of approximately 10^{25} and the minor donor 10^{24}. In the other configuration, the respective LRs are 10^{7} and 10^{13}. Based on these results, we proceed with the first configuration.

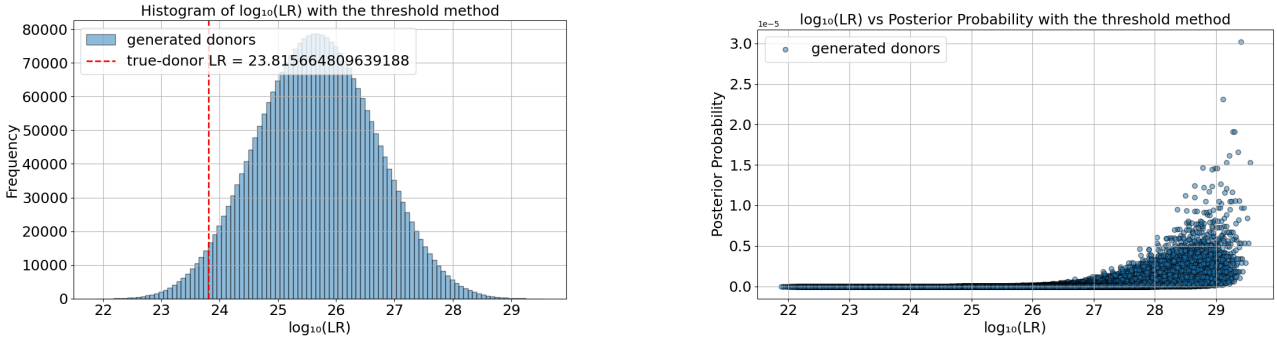
Threshold method

The threshold is set here at $t = 0.75$ for both donors. The histograms for both donors appear similar in shape and spread. More genotypes were sampled for the minor donor than for the major donor. In the scatterplots 8.4b and 8.5b, we note that the highest posterior probability for a sampled genotype is 5×10^{-5}. This difference low posterior probabilities can be attributed to the lower degree of allele distinguishability in the 150:150 mixture, as both donors contributed equal amounts of DNA. In contrast, the 300:150 mixture allows for better separation of contributors, leading to higher LRs and posterior probabilities for the most likely genotypes.



(a) Histogram (b) Scatterplot

Figure 8.4: Histogram and scatterplot for the major donor in the mixture with 150:150 DNA proportions, using a threshold of $t = 0.75$ in the threshold method. In the histogram, the red dashed line marks the LR of the true donor's genotype.

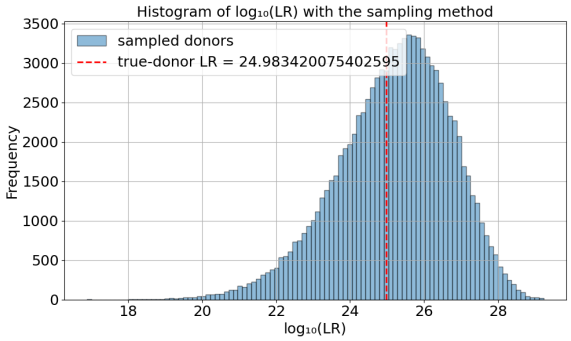


(a) Histogram (b) Scatterplot

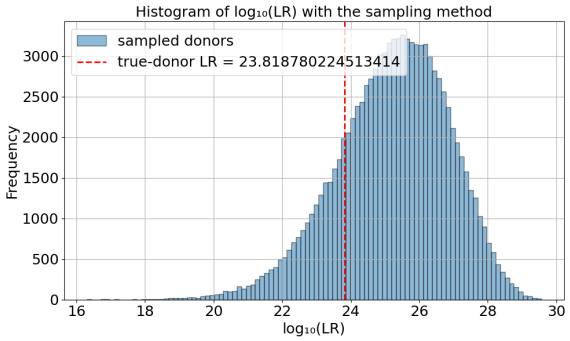
Figure 8.5: Histogram and scatterplot for the minor donor in the mixture with 150:150 DNA proportions, using a threshold of $t = 0.75$ in the threshold method. In the histogram, the red dashed line marks the LR of the true donor's genotype.

Sampling method

Figure 8.6b shows the LR distributions for both donors in the 150:150 mixture. The distributions are highly similar, as expected given the equal DNA contribution from both donors. Unlike the 300:150 mixture, where a few genotypes were sampled repeatedly, resulting in visible peaks in the LR distribution, no such peaks are observed here. This again reflects the reduced distinguishability between contributors and the more uniform spread of posterior probabilities across possible genotypes.



(a) Major donor



(b) Minor donor

Figure 8.6: Histogram of the LR distribution for the major and minor donor in the 150:150 DNA mixture. The genotypes are generated using the sampling method. The red dashed line marks the LR of the true donor's genotype.

8.4.3 Comparison threshold method and sampling method

Both methods provide valuable insights into the expected LRs. The threshold method, however, has several drawbacks. Its histograms reflect only the number of genotypes associated with a given LR, not the probability of obtaining that LR, which makes the interpretation less straightforward. Additionally, the threshold method requires manual selection of a threshold per donor, which must be adapted to each mixture depending on the posterior probabilities. This makes it difficult to apply the method in a standardized or automated way.

In contrast, the sampling methods does not require a threshold and directly approximates the probability distribution of the LR, as genotypes are sampled proportionally to their posterior probabilities. Based on these advantages, we proceed with the sampling method for the remainder of this study.

8.5 Validation of the sampling method

The sampling method produces a distribution of LRs based on genotypes drawn using posterior probabilities as weights. Since the true donor genotype is known for each mixture, we can validate the method by examining where the true-donor LR lies within the sampled LR distribution.

For each two-person mixture and each donor (major and minor), we compute the percentile of the true-donor LR within the LR distribution obtained through sampling. A percentile of x% means that x% of the sampled genotypes have a lower LR than the true genotype, and the remaining (100-x)% have a higher LR. Since genotypes are sampled proportionally to their posterior probabilities, we expect the true-donor LR to fall randomly within the distribution, leading to a uniform distribution of percentiles if the sampling method is well-calibrated.

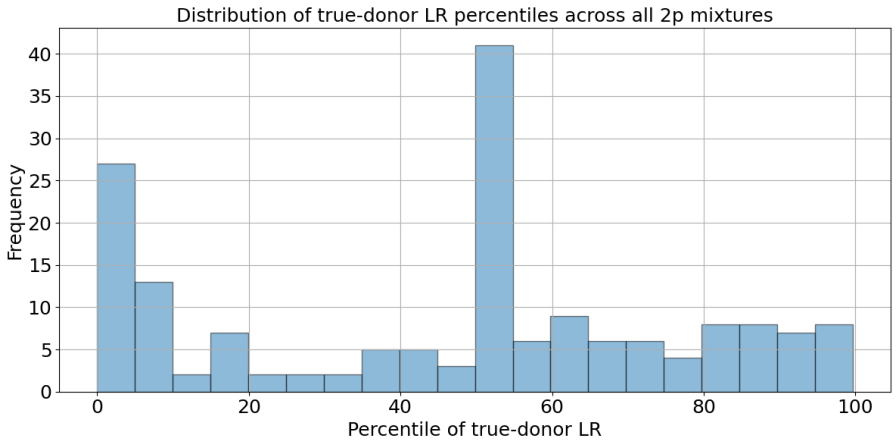
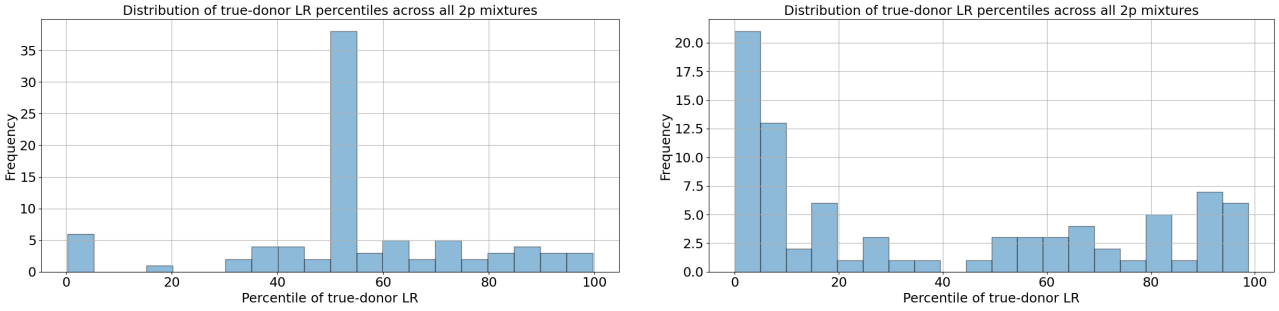


Figure 8.7: For each two-person mixture and each donor, the percentile of the true-donor LR is calculated within the LR distribution obtained using the sampling method. These percentiles are plotted as a histogram.

Figure 8.7 shows the percentiles for all two person mixtures and all donors (major/minor) in a histogram. The peak at 50% occurs when only one genotype is repeatedly sampled. In such cases, at each locus there is one allele combination whose posterior probability dominates over the other. As a result, at each locus, only one allele combination is sampled. Thus, only one genotype is sampled repeatedly. This genotype typically corresponds to the genotype of the true donor. It thus has the same LR as the donor. In this way the percentile is 50%.
In eight cases, the true-donor LR falls below all sampled genotypes, resulting in a percentile of 0%. No consistent pattern could be identified among these cases, and further investigation is required to understand the conditions under which the percentiles of 0% occur.

The histogram does not show a clear uniform distribution. A noticeable concentration of percentiles falls within the 0-10% range. Also,there seems to be greater density between 55-100% than between 10-45%. To better understand this behavior, the percentiles are split into two histograms: one for the major donor and one for the minor donor.



(a) Major donor

(b) Minor donor

Figure 8.8: Histograms of the percentiles for the true-donor LRs in the LR distribution splitted for the major donor and the minor donor.

A couple of differences could be observed in Figure 8.8. The peak at 50% appears only in the major donor histogram. This supports the idea that in some mixtures the major donor's genotype is so well supported by the evidence that it is sampled exclusively. The peak heights of the dominant donor are so strong that there is no debate possible for the genotype of the major donor. This leads to only one genotype being sampled over and over. The percentile of the true-donor LR is in this way 50%.
In contrast, the minor donor histogram does not show a peak at 50%. Through allele sharing it is more difficult to assign alleles to the minor donor, which increases the number of plausible genotypes. For example, suppose a locus has strong peaks at alleles A and B, and a much smaller peak at allele C. It may be clear that the major donor has genotype (A,B), but for the minor donor, multiple genotypes such as (A,C), (B,C) and (C,C) could be plausible. This uncertainty increases the diversity of sampled genotypes and dilutes the sampling frequency of the true one.

The sharp peak between 0-10% in the combined histogram is largely driven by the minor donor probabilities. Finally, we observe that for both donors the lack of percentiles between 10-45%, while a noticeable concentration is present between 55-100%. This asymmetric pattern deviates from the expected uniform distribution. Further investigation is needed to understand the underlying cause of this imbalance.

8.6 Comparison of DNA kits and their impact on LR estimation

DNA kits used for measuring the DNA have improved over time. In particular, the number of loci measured has increased substantially. This means modern measurements give much more information. However, the DNA databank still stores a large number of profiles measured on the older kits. The kit SGM includes 10 loci, NGM includes 15 loci and the newest kit, PPF6C PowerPlex Fusion, 23 loci. The number of loci directly affects the evidential strength of a DNA match, and thus the magnitude of the resulting LR. For instance, consider a two-person mixture where the DNA profile of the major donor is known. We would expect the LR based on 23 loci to be substantially higher than one based on only 10 loci, as more genetic information is incorporated into the LR calculation.

In practice, a DNA databank match is typically reported when the LR exceeds 10^5. This raises the ques-

tion of whether profiles that do not exceed the 10^5 threshold when only 10 or 15 loci are available might surpass this threshold when 23 loci are used.

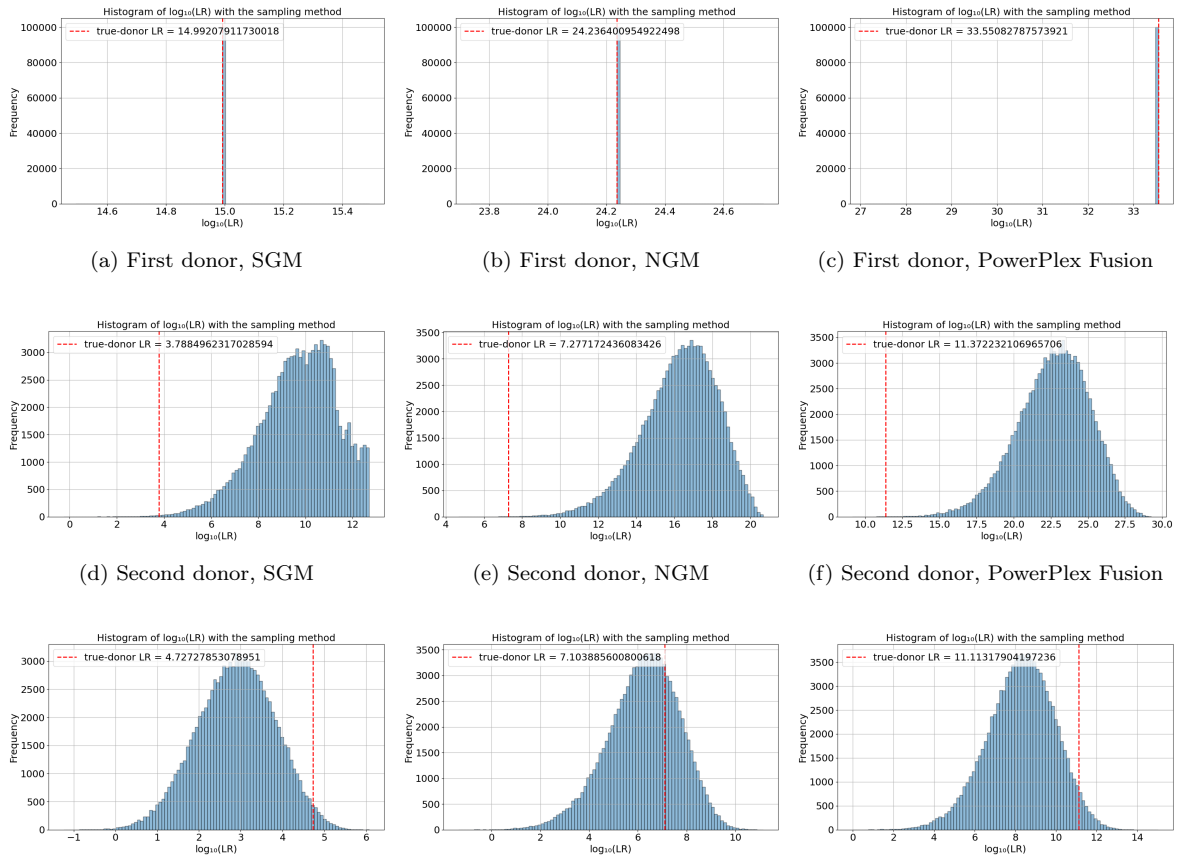
However, an increase in the number of loci does not guarantee a higher LR. For example, it could be the case that the first 10 loci match (relatively) good, but the additional loci show less similarity or mismatch. In such cases, including more loci may actually decrease the LR.

In addition to variation of the number of stored loci, the quality of the DNA mixture may also vary. For example, degraded DNA, caused by environmental exposure, time, or physical and chemical processes, can lead to missing alleles at certain loci, which reduces the number of loci available for LR calculation.

The number of loci available can therefore vary both across stored reference profiles and across mixtures. A flexible analysis method is required to accommodate this variability.

8.6.1 Mixture with 300:30:30 DNA proportions

In this section, we analyze a three-person mixture with proportions 300:30:30. The major donor thus contributed significantly more DNA to the mixture than the second and third donor. It can thus be expected that the alleles of the major donor are distinguishable from the alleles from second and third donor. The second and third donor contributed equal amounts of DNA, so their alleles are difficult to distinct.



(a) First donor, SGM (b) First donor, NGM (c) First donor, PowerPlex Fusion

(d) Second donor, SGM (e) Second donor, NGM (f) Second donor, PowerPlex Fusion

(g) Third donor, SGM (h) Third donor, NGM (i) Third donor, PowerPlex Fusion

Figure 8.9: LR distributions obtained using the sampling method for a three-person mixture (300:30:30) across three different DNA kits. Each row corresponds to one donor; each column represents a kit.

From Figure 8.9, we observe that the number of loci included in the analysis substantially affects the resulting LRs for all three donors. The LRs increase as more loci are included in the analysis.

As expected, the first donor, who contributed the largest amount of DNA material achieves the highest LR across all three kits. For all three kits, there is one genotype that is sampled almost all of the times. The LR of

this genotype increases from 10^{15} (SGM) to $10^{24.2}$ (NGM) to $10^{33.6}$ (PowerPlex Fusion). All three LRs exceed the threshold used for databank reporting.

For the second and third donor, we note that the true-donor LRs are similar. The LR distributions differ. The sampling method gives relatively higher LRs in the LR distribution for the second donor than for the third donor. This suggests that the model finds the second donor more distinguishable than the third donor. The LRs for the sampled donors for NGM and PowerPlex Fusion are almost all higher than the databank search threshold for the second donor. For the kit SGM, the left tail is lower than the databank search threshold and also the true-donor LR. It follows that, when the true profile of the second donor is stored in the databank using the kit SGM, it will not always be detected by a databank search. For the third donor with both kits NGM and PowerPlex Fusion a part of the distribution falls below the databank threshold. The true-donor LR for SGM is lower than the threshold. It is also visible that a part of the LRs from the sampled donor for PowerPlex Fusion do not exceed the databank search threshold. This indicates that the third donor's contribution is harder to resolve, making genotype matching more challenging, especially with kits that include fewer loci.

A summary of the true-donor LRs for the three kits and the three donors can be found in 8.1.

	SGM	**NGM**	**PowerPlex Fusion**
First Donor	15.0	24.2	33.6
Second Donor	3.8	7.8	11.4
Third Donor	4.7	7.1	11.1

Table 8.1: True-donor LRs per kit, shown on a logarithmic scale, for the 300:30:30 mixture.

Chapter 9

Identifying Relatives

Using the sampling method described in Chapter 8, we can estimate the expected order of magnitude of LRs for the contributors of a mixture. An important question is whether we can distinguish between an actual donor and a relative of the donor based on the LR.

The mixtures examined in this study are generated in the laboratory, so the genotypes of the actual donors are known. Based on the known genotypes, we generate hypothetical sons and siblings. For these generated relatives we use the same statistical framework to calculate LRs as we do for the actual donors. By comparing the resulting LR distributions, we assess how well the LR can differentiate between the true donor and a relative.

9.1 Generating relatives

9.1.1 Generating sons

For simplicity, we use the term sons, though the same approach yields to generate daughters, fathers and mothers. To generate a son, one allele at each locus is randomly inherited from the donor, and the other is sampled from the population according to allele frequencies. This reflects the way the alleles are inherited from the parents, i.e. one allele from the father and one allele from the mother. We assume only one of the two parents is a donor of the mixture.

9.1.2 Generating siblings

To generate a sibling, each of the donor's alleles is inherited with 50% probability. If only one allele is inherited, the other allele is sampled with the allele frequencies of the population. If no allele is inherited, both alleles are sampled with the allele frequencies of the population. This approach reflects the genetic relatedness between full siblings, in this way siblings share on average 50% of the alleles.

9.1.3 Differences between the sons and the siblings

The hypothetical sons inherit exactly one allele per locus from the donor, meaning that one allele is always shared. The hypothetical siblings inherit at each locus two alleles from the donor with 25% probability, one allele with 50% probability and no alleles with 25% probability. Thus, on average, both the hypothetical sons and siblings inherit one allele at each locus from the donor. However, due to the possibility of inheriting either two or zero alleles, the variance in the LR distribution is greater for the hypothetical siblings than for the hypothetical sons. For example, some generated genotypes will have relatively higher LRs (inherited relatively many alleles from donor) and some generated genotypes will have relatively lower LRs (inherited relatively less alleles from the donor).

9.2 Mixture with 150:60:30 DNA proportions

9.2.1 The first donor

The first donor has relatively high LRs, which is expected given that the first donor contributed 2.5 times more DNA than the second donor. The peak heights of the alleles originating from the first donor are higher than

those of the other donors. For the first donor, one genotype is sampled in approximately 66.000 of the 100.000 iterations, corresponding to a posterior probability of 66%. This genotype is also the true genotype of the donor. The corresponding LR is around 10^{33}.

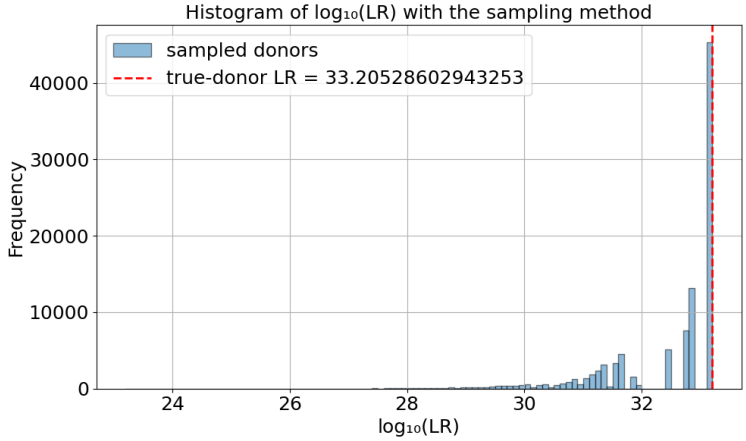


Figure 9.1: LR distribution for the major donor of the 150:60:30 mixture.

Figure 9.2 shows that both the hypothetical sons and the siblings yield extremely low LRs. This is expected, given the dominance of the first donor in the mixture. As a result, it follows that the probability that other alleles are from the major donor is very small. Since the hypothetical sons and siblings inherit approximately half of their alleles from the donor and the remaining alleles are drawn from the general population, many of their genotypes include alleles that are either not present in the mixture or appear with much lower peak heights than expected for the major contributor. These discrepancies lead to substantially lower likelihoods and, consequently, low LRs for the relatives.

Figure 9.2 also shows that the means of the LR distributions of the sons and siblings are approximately equal, but the variance is greater for the siblings. This is a consequence of how alleles are inherited as described in Section 9.1.3.

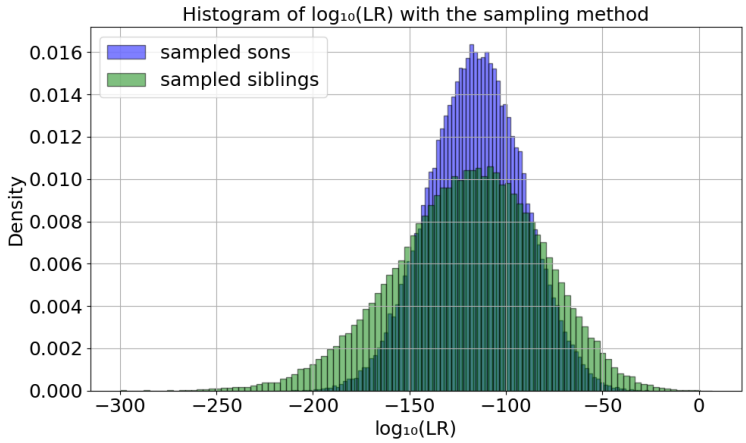


Figure 9.2: LR distributions for the hypothetical sons and siblings for the first donor of the 150:60:30 mixture

9.2.2 The second donor

The second donor contributes a smaller proportion to the mixture. Some alleles may be shared between the first and second donors, making it more difficult to attribute specific alleles to the second donor. As a result, the true-donor LR is lower than that of the first donor. The second donor has an LR of $10^{20.5}$ and the LR distribution of the sampled genotypes is around 10^{20}. As shown in Figure 9.3, there is some overlap between the LR distribution of the generated donors and that the distributions of the relatives.

To quantify the overlap between the LR distributions, we proceed as follows. From both the sets of sampled donors and sampled relatives take one genotype randomly. Now by doing this a number of times, we can estimate the probability that a randomly sampled relative yields a higher LR than the true donor. For the son this probability is 10^{-5} and for the sibling this probability is 0.00015. The maximum LR observed for a hypothetical son is 5.8×10^{15}, and for a sibling it is 2.9×10^{16}.

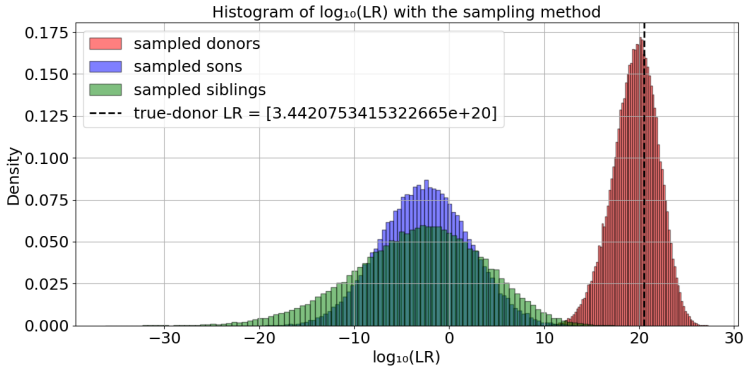


Figure 9.3: LR distributions for the hypothetical sons and siblings for the first donor of the 150:60:30 mixture

9.2.3 The third donor

Finally, the third donor has a true-donor LR of almost 10^{16}, which is lower than that of the first and second donors, as expected due to its smaller DNA contribution. Figure 9.4 shows a greater overlap between the LR distributions of the hypothetical sons and siblings and that of the sampled donor genotypes.

The probability that a hypothetical son has a higher LR than the true donor is 0.00864, and for a sibling this probability is 0.01549. The maximum LR observed for a hypothetical son is 1.1×10^{16}, and for a sibling it is 3.7×10^{20}.

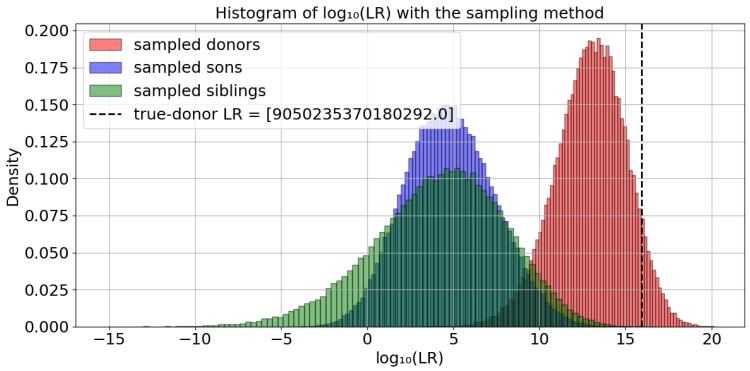


Figure 9.4: LR distributions for the hypothetical sons and siblings for the third donor of the 150:60:30 mixture

9.2.4 Conclusion LR distribution relatives

For the major donor the hypothetical relatives yielded extremely low LRs. This is the consequence of the fact that the major donor contributed significantly more DNA material to the mixture than the other donors. It is thus clear which alleles are in the genotype of the major donor. This suggests that, in general, the risk of a relative being mistakenly matched instead of the actual contributor is very small for dominant donors.

We saw that for the second and the third donor there was a small chance that a relative could be retrieved in a databank search. This possibility should be kept in mind in forensic investigations. In general we expect the overlap between the distributions to increase further for the smaller donors if the number of contributors increases. In this case it is more difficult to assign alleles to the smaller donors by the overlap of alleles and

thereby the probability of generating a hypothetical relative with alleles that are not observed in the mixture decreases. Also, for the same reasoning we expect the overlap to increase when the amount of contributed DNA of the donors lies closer to each other.

Chapter 10

Conclusion

This thesis is set out with two goals:

1. Clarify how Bayesian reasoning can be applied in forensic science, with a focus on the role of prior probabilities.
2. Develop and evaluate methods to derive LR distributions for contributors of DNA mixtures.

These goals were pursued through a combination of mathematical derivation, illustrative case studies, and evaluation of forensic practice. This concluding chapter reflects on the extent to which these goals were achieved and outlines the main findings that emerged along the way.

10.1 The role of prior probabilities in forensic statistics

While LRs are widely accepted as a way to quantify evidential value, this thesis has demonstrated that in practice, their computation and interpretation are not always inseparable from prior assumptions. The prosecution's hypothesis may consist of several mutually exclusive sub-scenarios. For example, the hypothesis that PoI 1 contributed may be divided into the scenarios in which PoI 1 and PoI 2 contributed, and the scenario in which PoI 1 and an unknown individual contributed. To compute an LR for the prosecution's hypothesis against the hypothesis that PoI 1 did not contribute, the LRs for the sub-scenarios against the hypothesis that two unknown individuals contributed to the trace need to be combined using prior probabilities.

It became clear that assuming equal priors is not a neutral act. Rather, it can systematically bias the resulting inference, particularly when PoIs originate from databases or have (familial) relationships. Case studies demonstrated how results can drastically change depending on the assumed prior probabilities.

By presenting the relevant LRs and deferring the choice of priors to the decision maker, the forensic expert respects the mathematical structure of Bayesian inference and leaves the choice of priors open for the decision maker, if this is needed in the case. In this way, the thesis contributes to a more careful and transparent use of statistical reasoning in forensic contexts.

10.2 LR distributions for DNA contributors and relatives

Obtaining LR distributions for DNA mixtures contributors is practically important, as forensic laboratories often use LR thresholds to decide whether to report a match. If true contributors fall below such thresholds, or if relatives can exceed them, these decisions risk being misleading.

Two methodological approaches were developed to obtain LR distributions. The first was a threshold-based approach, which adds for each locus allele combinations to a set until the sum of posterior probabilities of the allele combinations exceed the threshold. With these allele combinations all possible genotypes are generated. The second approach was a more flexible sampling-based approach in which genotype profiles are sampled to empirically estimate the LR distribution. This method proved especially valuable in estimating realistic LR distributions.

Examples showed that the LR values for true donors vary widely, depending on factors such as: the number of contributors, DNA proportions, number of loci. The LR distribution for relatives was also investigated. It was found out that, in some cases, there is overlap between the LR distribution of a relative and that of the true donor. Experiments comparing different DNA kits revealed that the number of loci considered significantly affects the LR distribution, or the order of magnitude of the LR. This can result in the LR exceeding a threshold for one kit but not for another.

By simulating such scenarios and plotting the resulting LR distributions, the thesis offers practical tools to interpret LRs found in forensic research.

10.3 Concluding remarks

Taken together, the results of this thesis form a coherent argument in favor of more transparent, context-aware forensic evaluation. Prior probabilities can only be used with the utmost caution and LR distributions shape the practical reliability of threshold-based reporting. Both aspects are essential for understanding what a reported LR actually means in context.

The next and final chapter will reflect more broadly on these findings, discuss their limitations, and identify directions for future research and practical improvement.

Chapter 11

Discussion and Suggestions for Further Research

11.1 Discussion

Throughout this thesis, DNAStatistx has been an important tool by computing unnormalized posterior probabilities. In this thesis, we calculated LRs for a PoI being the first donor, rather than simply being a donor. This approach differs from real-world casework. It is more useful to determine whether the PoI contributed, regardless of which donor they are. The methods discussed in this thesis can also be applied without marginalizing over the donors. For instance, prior probabilities are still needed to combine the LRs from sub-scenarios, that PoI 1 and PoI 2 contributed, and that PoI 1 and an unknown individual contributed, into the broader hypothesis that PoI 1 contributed to the trace. Also, LR distributions can be achieved when the donors are not marginalized.

It was expected that the percentiles of true-donor LRs within their LR distribution would follow a uniform distribution. In Section 8.5 we saw that the percentiles for all two-person mixtures did not give a uniform distribution. Most notably, there were eight true-donor LRs that fell below all sampled LRs, producing a percentile of 0%. This is not consistent with the assumed posterior model. Two other observations that are made from the percentile plot are: (1) a high number of percentiles fall between 0-10% for the minor donor, and (2) for both donors, there appear to be relatively more percentiles in the 55-100% range than in the 10-45% range. These observations indicate that, although the method is promising and practically applicable, it is not yet perfectly calibrated and warrants further investigation.

11.2 Suggestions for further research

11.2.1 Generating a DNA mixture containing both parents

An intriguing extension of the work in Chapter 9 on relative identification would be to examine the scenario in which both of an individual's parents contribute to a DNA mixture, but the individual themselves does not. In such a case, the child's genotype would entirely consist of alleles that also appear in the mixture, potentially yielding high LR values despite the child not being a contributor.

To explore this, one could generate parent pairs for existing donor genotypes and construct synthetic mixtures including the mother and the father (and possibly other contributors). From there, the LR for the child could be computed to determine how often the system mistakenly treats the child as a contributor. The simulation requires a different strategy from those used previously: for each locus, one allele from the child should be assigned to the father and the other to the mother, with the second allele in each profile drawn randomly using population allele frequencies as weights.

11.2.2 Reference profiles

Another area that merits further investigation is the effect of replicate mixtures on LR outcomes. The investigated mixtures are produced three times. This raises questions about the stability and robustness of LR estimates across replicates. For example, to what extent do the different samples yield consistent LR values for

the donors? Should these replicates be treated independently, or combined in a principled way to strengthen the inference?

Answering these questions would involve running the full LR pipeline for each replicate and analyzing the variation. If large discrepancies are observed, this might suggest that stochastic effects need to be better understood before reporting final LR values.

Bibliography

[1] David J Balding and John Buckleton. Interpreting low template DNA profiles. *Forensic Science International: Genetics*, 4(1):1–10, 2009.

[2] John M. Butler. *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. Elsevier Academic Press, Burlington, MA, 2nd edition, 2005.

[3] Jacob de Zoete, James Curran, and Marjan Sjerps. A probabilistic approach for the interpretation of RNA profiles as cell type evidence. *Forensic Science International: Genetics*, 20:30–44, 2016.

[4] Tacha Hicks, Zane Kerr, Simone Pugh, Jo-Anne Bright, James Curran, Duncan Taylor, and John Buckleton. Comparing multiple PoI to DNA mixtures. *Forensic Science International: Genetics*, 52:102481, 2021.

[5] Alexander Lindenbergh, Petra Maaskant, and Titia Sijen. Implementation of RNA profiling in forensic casework. *Forensic Science International: Genetics*, 7(1):159–166, 2013.

[6] K Slooten. The comparison of DNA mixture profiles with multiple persons of interest. *Forensic Science International: Genetics*, 56:102592, 2022.

[7] Naomi Voerman. A probabilistic approach to quantify the strength of evidence of presence of cell types from RNA data using a multi-label method. Master's thesis, Universiteit Leiden, 2019.

[8] RJF Ypma, PA Maaskant-van Wijk, R Gill, M Sjerps, and M Van den Berge. Calculating LRs for presence of body fluids from mRNA assay data in mixtures. *Forensic Science International: Genetics*, 52:102455, 2021.

[9] Rolf Ypma, Charles Berger, and Marjan Sjerps. Interpretatie van de bewijskracht; wanneer is bewijs 'sterk' of 'zwak'? *Expertise en Recht*, 2024:212–217, 12 2024.