# Likelihood Ratios for Mixtures: Binary Approach 

## Acknowledgement

I thank Simone Gittelson, Bruce Weir and John Buckleton for their helpful discussions.

## Disclaimer

Points of view in this presentation are mine and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.


## Likelihood Ratio





Cedar Crest College
Forensic Science Training Institute

## Standpoints of the prosecution and the defense



EPG of the crime stain

$H_{p}$ : The DNA came from the POI and an unknown contributor.
$H_{d}$ : The DNA came from two unknown contributors.


## Likelihood Ratio (LR)

The probability of observing the DNA typing results of the crime stain given the POI's genotype and that the DNA came from the POI and one unknown contributor

$$
L R=\frac{\operatorname{Pr}\left(G_{C S} \mid G_{P O I}, H_{p}\right)}{\operatorname{Pr}\left(G_{C S} \mid G_{P O I}, H_{d}\right)} \leftarrow \underbrace{\text { divided by }}_{\text {numerator }} \text { denominator }
$$

the probability of observing the DNA typing results of the criem stain given the POI's genotype and that the DNA came from two unknown contributors.


Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

## Likelihood Ratio (LR)

D18S51
$p_{14}=0.134$
$p_{16}=0.147$
$p_{18}=0.078$
$p_{20}=0.018$


Numerator:
What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype \{14,20\}?

$$
\begin{array}{|c|c|}
\hline \text { Major } & \text { Minor } \\
& \operatorname{Pr}(16,18) \times \operatorname{Pr}(14,20) \\
\hline 16,18 & 14,20 \\
& =2 p_{16} p_{18} \times 1 \\
& =2 p_{16} p_{18}
\end{array}
$$

## Likelihood Ratio (LR)

D18S51
$p_{14}=0.134$
$p_{16}=0.147$
$p_{18}=0.078$
$p_{20}=0.018$


Denominator:
What is the probability of obtaining these DNA typing results for the crime stain if the POI is not a contributor?

| Major |  |
| :---: | :---: |
| 16,18 | Minor |

$$
\begin{aligned}
& \operatorname{Pr}(16,18) \times \operatorname{Pr}(14,20) \\
& =2 p_{16} p_{18} \times 2 p_{14} p_{20} \\
& =4 p_{14} p_{16} p_{18} p_{20}
\end{aligned}
$$

## Likelihood Ratio (LR)

D18S51
$p_{14}=0.134$
$p_{16}=0.147$
$p_{18}=0.078$
$p_{20}=0.018$
$L R=\frac{2 p_{16} p_{18}}{4 p_{14} p_{16} p_{18} p_{20}}$
$=\frac{1}{2 p_{14} p_{20}}$
$=207.30$

## Likelihood Ratio (LR)

D18S51
$p_{14}=0.134$
$p_{16}=0.147$
$p_{18}=0.078$
$p_{20}=0.018$


The DNA typing results are 207 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

## Unrestricted LR - Peak Heights are ignored

## Likelihood Ratio (LR)

D18S51
$p_{14}=0.134$
$p_{16}=0.147$
$p_{18}=0.078$
$p_{20}=0.018$


Numerator:
What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype \{14,20\}?

$$
\begin{aligned}
& \operatorname{Pr}(16,18) \times \operatorname{Pr}(14,20) \\
& =2 p_{16} p_{18} \times 1 \\
& =2 p_{16} p_{18}
\end{aligned}
$$

## Likelihood Ratio (LR)

D18S51
$p_{14}=0.134$
$p_{16}=0.147$
$p_{18}=0.078$
$p_{20}=0.018$


Denominator:
What is the probability of obtaining these DNA typing results for the crime stain if the POI is not a contributor?

| Likelihood Ratio |  |  |  |
| :---: | :---: | :---: | :---: |
|  | Contributor 1 | Contributor 2 |  |
|  | 14,16 | 18,20 | $2 p_{14} p_{16} \times 2 p_{18} p_{20}$ |
|  | 14,18 | 16,20 | $2 p_{14} p_{18} \times 2 p_{16} p_{20}$ |
|  | 14,20 | 16,18 | $\left.2 p_{14} p_{20}\right\rangle 2 p_{16} p_{18}$ |
|  | 18,20 | 14,16 | $2 p_{18} p_{20} \times 2 p_{14} p_{20}$ |
| 14161820 | 16,20 | 14,18 | $2 p_{16} p_{20} \times 2 p_{14} p_{18}$ |
|  | 16,18 | 14,20 | $2 p_{16} p_{18} \times 2 p_{14} p_{20}$ |
| $=24 p_{14} p$ | $p_{18} p_{20}$ | 2) What is the probability of obtaining these DNA typing results if the PO is not a contributor? |  |



Cedar Crest College
Forensic Science Training Institute

## Likelihood Ratio (LR)

CSF1PO
$p_{10}=0.220$
$p_{11}=0.309$
$p_{12}=0.360$

lele at 12 is above the stochastic threshold
Numerator:
:
What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype \{12,12\}?

| Major | Minor |
| :---: | :---: |
| 10,11 | 12,12 |
| 10,11 | 10,12 |
| 10,11 | 11,12 |

$$
\begin{aligned}
& \operatorname{Pr}(10,11) \times \operatorname{Pr}(12,12) \\
& =2 p_{10} p_{11} \times 1 \\
& =2 p_{10} p_{11}
\end{aligned}
$$

## Likelihood Ratio (LR)

CSF1PO
$p_{10}=0.220$
$p_{11}=0.309$
$p_{12}=0.360$

allele at 12 is above the
Denominator:
stochastic threshold
What is the probability of obtaining these DNA typing results for the crime stain if the POI is not a contributor?

| Major | Minor |
| :---: | :---: |
| 10,11 | 12,12 |
| 10,11 | 10,12 |
| 10,11 | 11,12 |

## Likelihood Ratio (LR)

> CSF1PO
> $p_{10}=0.220$
> $p_{11}=0.309$
> $p_{12}=0.360$


Denominator:
allele at 12 is above the stochastic threshold

$$
\operatorname{Pr}(10,11) \times \operatorname{Pr}(12,12)+\operatorname{Pr}(10,11) \times \operatorname{Pr}(10,12)+\operatorname{Pr}(10,11) \times \operatorname{Pr}(11,12)
$$

| Major | Minor |
| :---: | :---: |
| 10,11 | 12,12 |
| 10,11 | 10,12 |
| 10,11 | 11,12 |

$$
\begin{aligned}
= & 2 p_{10} p_{11} \times p_{12}^{2}+2 p_{10} p_{11} \times 2 p_{10} p_{12} \\
& +2 p_{10} p_{11} \times 2 p_{11} p_{12} \\
= & 2 p_{10} p_{11} p_{12}\left(p_{12}+2 p_{10}+2 p_{11}\right)
\end{aligned}
$$

## Likelihood Ratio (LR)

CSF1PO
$p_{10}=0.220$
$p_{11}=0.309$
$p_{12}=0.360$

allele at 12 is above the
stochastic threshold

$$
\begin{aligned}
L R= & \frac{2 p_{10} p_{11}}{2 p_{10} p_{11} p_{12}\left(p_{12}+2 p_{10}+2 p_{11}\right)} \\
& =\frac{1}{p_{12}\left(p_{12}+2 p_{10}+2 p_{11}\right)} \\
& =1.96
\end{aligned}
$$

## Likelihood Ratio (LR)

CSF1PO
$p_{10}=0.220$
$p_{11}=0.309$
$p_{12}=0.360$

allele at 12 is above the stochastic threshold

The DNA typing results are about 2 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.



Allele 28 is below the
stochastic threshold

## Different Thresholds Used with CE Data



## Likelihood Ratio (LR)

$p_{30}=0.283$
$p_{32.2}=0.090$

Numerator:


Allele 28 is below the stochastic threshold

What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype \{28,28\}?

$$
\begin{aligned}
& \text { Major Minor } \\
& \text { 30,32.2 } 28, F \\
& \operatorname{Pr}(30,32.2) \times \operatorname{Pr}(28, F) \\
& =2 p_{30} p_{32.2} \times 1 \\
& =2 p_{30} p_{32.2}
\end{aligned}
$$

## Likelihood Ratio (LR)

D21S11
$p_{28}=0.159$
$p_{30}=0.283$
$p_{32.2}=0.090$


Allele 28 is below the stochastic threshold
What is the probability of obtaining these DNA typing results for the crime stain if the PO is not a contributor?

| Major | Minor |
| :---: | :---: |
| $30,32.2$ | $28, F$ |

$$
\begin{aligned}
& \operatorname{Pr}(30,32.2) \times \operatorname{Pr}(28, F) \\
& =2 p_{30} p_{32.2} \times\left[2 p_{28}\left(1-p_{28}\right)+p_{28}^{2}\right] \\
& =2 p_{30} p_{32.2}\left(2 p_{28}-p_{28}^{2}\right)
\end{aligned}
$$

```
"2p" or p}\mp@subsup{}{}{2}+2p(1-p
```


## Derivation of the $2 p$ Rule

- Two ways to think of it... (easy)

$$
2 p q \longrightarrow 2 p q \longrightarrow 2 p
$$

$$
\text { " } 2 p \text { " or } p^{2}+2 p(1-p)
$$

## Derivation of the $2 p$ Rule

- Two ways to think of it... (mathematical)


## 5 allele system - P Q R S T



## 5 allele system - P Q R S T


Prob $=p^{2}+2 p q+2 p r+2 p s+2 p t$
Prob $=p^{2}+2 p(q+r+s+t)$
Prob $=p^{2}+2 p(1-p)$
Prob $=p^{2}+2 p-2 p^{2}$
Prob $=2 p-p^{2}$

## Likelihood Ratio (LR)

D21S11
$p_{28}=0.159$
$p_{30}=0.283$
$p_{32.2}=0.090$


Allele 28 is below the stochastic threshold
$L R=\frac{2 p_{30} p_{32.2}}{2 p_{30} p_{32.2}\left(2 p_{28}-p_{28}^{2}\right)}$
$=\frac{1}{\left(2 p_{28}-p_{28}^{2}\right)}$
$=3.42$

## What does a $L \boldsymbol{R} \approx 3$ mean?

A. The person of interest committed the crime.
B. A total of 3 peaks were observed at this locus.
C. It is about 3 times more probable that the DNA came from the person of interest and an unknown contributor than that the DNA came from two unknown contributors.
D. There are 3 contributors to this DNA mixture.
E. The DNA typing results are about 3 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.


Response Counter


## Likelihood Ratio (LR)

TPOX
$p_{8}=0.525$
$p_{11}=0.252$


The peak at 11 is above

$$
G_{P O I}=\{11,11\}
$$

the stochastic threshold.
What is the probability of obtaining these DNA typing results for the crime stain if the POI is a contributor and the POI has genotype \{11,11\}?
Major Minor

$$
\operatorname{Pr}(8,8) \times \operatorname{Pr}(11,11)
$$

$$
=\cdots
$$

## Likelihood Ratio (LR)

TPOX
$p_{8}=0.525$
$p_{11}=0.252$


The peak at 11 is above
Denominator: the stochastic threshold.
What is the probability of obtaining these DNA typing results for the crime stain if the POI is not a contributor?

| Major | Minor | $\operatorname{Pr}(8,8) \times \operatorname{Pr}(11,11)+\operatorname{Pr}(8,8) \times \operatorname{Pr}(8,11)$ |
| :---: | :---: | :--- | :--- |
| 8,8 | 11,11 |  |
| 8,8 | 8,11 |  |

## What is the likelihood ratio?

A. $\frac{p_{8}^{2}}{p_{8}^{2}\left(p_{11}^{2}+2 p_{8} p_{11}\right)}=\frac{1}{p_{11}\left(p_{11}+2 p_{8}\right)}$
B. $\frac{1}{p_{11}+2 p_{8}}$
C. 1
D. $\frac{1}{2 p_{8} p_{11}}$
E. $\frac{1}{p_{11}^{2}}$
F. infinity
G. ???

## Likelihood Ratio (LR)

$p_{8}=0.525$
$p_{11}=0.252$

$=3.05$

## Likelihood Ratio (LR)

TPOX
$p_{8}=0.525$
$p_{11}=0.252$


The DNA typing results are about 3 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

## Likelihood Ratio (LR) for all loci

$H_{p}$ : The DNA came from the POI and an unknown contributor.
$H_{d}$ : The DNA came from two unknown contributors.

If $H_{p}$ is true, is the POI the major contributor or the minor contributor?

If $H_{p}$ is true, the POI could be either the major contributor or the minor contributor. Let us consider these possibilities to be equally probable. So if $H_{p}$ is true, there is a probability of $\frac{1}{2}$ that the POI is the major contributor and a probability of $\frac{1}{2}$ that the POI is the minor contributor.

We can only observe these DNA typing results if the POI is the minor contributor.

D18S51: Major Minor

CSF1PO: Major Minor
$G_{P O I}=\{12,12\}$

D21S11:
Major Minor
30,32.2 28,F
$G_{P O I}=\{28,28\}$
TPOX: Major Minor

| 8,8 | 11,11 |
| :---: | :---: |
| 8,8 | 8,11 |

$G_{P O I}=\{11,11\}$

## Likelihood Ratio (LR) for all loci

$H_{p}$ : The DNA came from the POI and an unknown contributor.
$H_{d}$ : The DNA came from two unknown contributors.

Numerator:
Because these DNA typing results are only possible when the POI is the minor contributor, and the POI is the minor contributor with a probability of $\frac{1}{2}$, we multiply the numerator of the likelihood ratio for the entire profile by $\frac{1}{2}$.

#  

| D21S11 | 3.42 |
| :--- | :--- |
| D7S820 | 3.74 |
| CSF1PO | 1.96 |
| D3S1358 | 2.39 |
| TH01 | 1.75 |
| D13S317 | 4.58 |
| D16S539 | 1.89 |
| D2S1338 | 5.03 |
| D19S433 | 1.29 |
| vWA | 1 |
| TPOX | 3.05 |
| D18S51 | 207.30 |
| D5S818 | 3.77 |
| FGA | 1 |

All Loci: $L R=2.5 \times 10^{7}$

## True or false?

A likelihood ratio of $\mathbf{2 . 5} \times \mathbf{1 0}^{\mathbf{7}}$ means that it is $\mathbf{2 . 5} \times$ $\mathbf{1 0}^{7}$ times more probable that the DNA came from the person of interest and an unknown contributor than that the DNA came from two unknown contributors.
A. True
B. False

## Response

Counter


## Likelihood Ratio (LR)

$$
L R=2.5 \times 10^{7}=25 \text { million }
$$

The DNA typing results are about 25 million times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

## Factor of 2

## Suppose...



What is the probability you will randomly pick a red marble?
$\operatorname{Pr}($ picking left $) \times \operatorname{Pr}($ left is red $)+\operatorname{Pr}($ picking right $) \times \operatorname{Pr}($ right is red $)$
$0.5 \times 1+0.5 \times 1$

$$
=1
$$

## Suppose...



What is the probability you will randomly pick a red marble?

$$
\begin{aligned}
& \operatorname{Pr}(\text { picking left }) \times \operatorname{Pr}(\text { left is red })+\operatorname{Pr}(\text { picking right }) \times \operatorname{Pr}(\text { right is red }) \\
& \qquad \begin{aligned}
0.5 & \times 1+0.5 \times 0 \\
= & 0.5
\end{aligned}
\end{aligned}
$$

## Now Suppose...

person of interest (POI)

$\{A, B\}$


A B C D
$\mathrm{Hp}: \mathrm{POI}$ and someone else
Hd: DNA came from 2 unknowns


Only 2 genotype combinations are possible for Hp


For the $H_{d}$ (recall this?)

|  | Contributor 1 | Contributor 2 |  |
| :---: | :---: | :---: | :---: |
|  | 7,8 | 9,10 | $2 p_{7} p_{8} \times 2 p_{9} p_{10}$ |
|  | 7,9 | 8,10 | $2 p_{7} p_{9} \times 2 p_{8} p_{10}$ |
|  | 7,10 | 8,9 | $2 p_{7} p_{10}>2 p_{8} p_{9}$ |
|  | 8,9 | 7,10 | $2 p_{8} p_{9} \times 2 p_{7} p_{10}$ |
| 78810 | 8,10 | 7,9 | $2 p_{8} p_{10} \times 2 p_{7} p_{9}$ |
|  | 9,10 | 7,8 | $2 p_{9} p_{10} \times 2 p_{7} p_{8}$ |

$=24 p_{7} p_{8} p_{9} p_{10}$
2) What is the probability of obtaining these DNA typing results if the POI is not a contributor?

## Likelihood Ratio

$$
\frac{2 p_{c} p_{d}}{24 p_{a} p_{b} p_{c} p_{d}}=\frac{1}{12 p_{a} p_{b}}
$$

This is like the box with 2 red marbles: $0.5+0.5=1$ in the numerator

Anytime you have a "mirror image" for the numerator, there is no factor of 0.5 (or factor of 2 in the denominator)

## Now Suppose...

person of interest (POI)

$\{A, B\}$

$\mathrm{Hp}: \mathrm{POI}$ and someone else
Hd: DNA came from 2 unknowns
person of interest (POI)

$$
\{A, B\}
$$

Hp: POI and someone else
Hd: DNA came from 2 unknowns


A B C D

|  | Contributor 1 <br> (MAJOR) | Contributor 2 <br> (MINOR) | probability of peaks in <br> crime stain EPG if $H_{p}$ is true | probability of peaks in crime <br> stain EPG if $H_{d}$ is true |
| :--- | :---: | :---: | :---: | :---: |
| 1 | AB | CD | $1 \times 2 p_{C} p_{D}$ | $2 p_{A} p_{B} \times 2 p_{C} p_{D}$ |
| 2 | AC | BD | 0 | 0 |
| 3 | AD | BC | 0 | 0 |
| 4 | BC | AD | 0 | 0 |
| 5 | BD | AC | 0 | 0 |
| 6 | CD | AB | 0 | 0 |


|  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | Contributor 1 <br> (MAJOR) | Contributor 2 <br> (MINOR) | probability of peaks in <br> crime stain EPG if $H_{p}$ is true | probability of peaks in crime <br> stain EPG if $H_{d}$ is true |
| 1 | AB | CD | $1 \times 2 p_{c} p_{D}$ | $2 p_{A} p_{B} \times 2 p_{C} p_{D}$ |
| 2 | AC | BD | 0 | 0 |
| 3 | AD | BC | 0 | 0 |
| 4 | BC | AD | 0 | 0 |
| 5 | BD | AC | 0 | 0 |
| 6 | CD | AB | 0 | 0 |

$$
\begin{aligned}
& \mathrm{Hp}=0.5 \times 1 \times 2 p_{c} p_{d}=0.5 \times 2 p_{c} p_{d} \\
& \mathrm{Hd}=2 p_{a} p_{b} \times 2 p_{c} p_{d}
\end{aligned}
$$

## Likelihood Ratio

$$
\frac{0.5 \times 2 p_{c} p_{d}}{4 p_{a} p_{b} p_{c} p_{d}}=\frac{0.5}{2 p_{a} p_{b}}=\frac{1}{4 p_{a} p_{b}}
$$

This is like the box with 1 red, 1 blue marble: 0.5 (left)+ 0.5 (right)
Therefore, a factor of 0.5 appears in the numerator (or factor of 2 in the denominator)

## So - why do we even need probabilistic genotyping?





## Whatever way uncertainty is approached, probability is the only sound way to think about it.


-Dennis Lindley

