

Sherlock Holmes and the DNA Likelihood Ratio

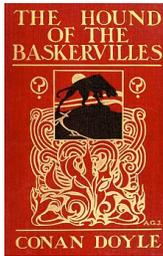
American Academy of Forensic Sciences
February, 2011
Chicago, IL

Mark W Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA



Cybergenetics

Cybergenetics © 2003-2011



as you value your life or your reason
keep away from the moor



Sir Henry Baskerville
Northumberland Hotel

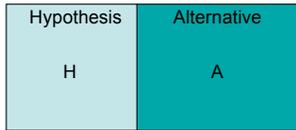
Dr. Mortimer: "We are coming now rather into the region of guesswork."

Sherlock Holmes: "Say, rather, into the region where we **balance probabilities** and choose the most likely. It is the scientific use of the imagination, but we have always some material basis on which to start our speculation. Now, you would call it a guess, no doubt, but I am almost certain that this **address has been written in a hotel.**"

Dr. Mortimer: "How in the world can you say that?"

Balance of Probability

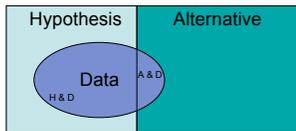
Belief before observing data



$$\text{Odds}(\text{Hypothesis}) = \text{Prob}(\text{H}) / \text{Prob}(\text{A})$$

Observe Data

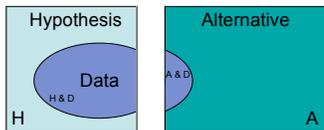
Belief after observing data



$$\text{Odds}(\text{Hypothesis} \mid \text{Data}) = \text{Prob}(\text{H} \ \& \ \text{D}) / \text{Prob}(\text{A} \ \& \ \text{D})$$

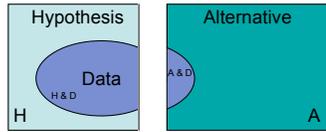
Likelihood Ratio

Information gained by observing data



$$\text{LR} = \frac{\text{Odds}(\text{Hypothesis} \mid \text{Data})}{\text{Odds}(\text{Hypothesis})} = \frac{\text{Prob}(\text{H} \ \& \ \text{D}) / \text{Prob}(\text{A} \ \& \ \text{D})}{\text{Prob}(\text{H}) / \text{Prob}(\text{A})}$$

Likelihood Ratio



$$LR = \frac{\text{Prob(Data | Hypothesis)}}{\text{Prob(Data | Alternative)}} = \frac{\text{Prob(H \& D)}}{\text{Prob(H)}} \bigg/ \frac{\text{Prob(A \& D)}}{\text{Prob(A)}}$$

Sherlock Holmes: "If you examine it carefully you will see that both the pen and the ink have given the writer trouble. The **pen has spluttered twice in a single word** and has **run dry three times in a short address**, showing that there was very little ink in the bottle. Now, a **private pen or ink-bottle is seldom allowed to be in such a state**, and the **combination of the two** must be quite rare. But you know the **hotel ink and the hotel pen, where it is rare to get anything else**. Yes, I have very little hesitation in saying that could we **examine the waste-paper baskets of the hotels** around Charing Cross until we found the remains of the mutilated Times leader we could lay our hands straight upon the person who sent this singular message."



- Hypothesis** letter written in a hotel
- Alternative** not written in a hotel
- Data** pen splutter, ink ran dry
- Action** check waste baskets?

The pen data & likelihoods

- Hypothesis** written in hotel
- Alternative** privately written
- Data** pen spluttered (2x in one word)

Prob(pen splutter data | hotel hypothesis)
"rare to get anything else" ~ 50%

Prob(pen splutter data | private alternative)
"seldom allowed to be in such a state" ~ 10%

Sir Henry Baskerville
Northumberland Hotel

The pen likelihood ratio

Hypothesis written in hotel
Alternative privately written
Data pen spluttered (2x in one word)

Prob(pen splutter data | hotel hypothesis)
"rare to get anything else" ~ 50%

Prob(pen splutter data | private alternative)
"seldom allowed to be in such a state" ~ 10%

$$LR = \frac{\text{Prob}(\text{splutter} | \text{hotel})}{\text{Prob}(\text{splutter} | \text{private})} = \frac{50\%}{10\%} = 5$$

The ink data & likelihoods

Hypothesis written in hotel
Alternative privately written
Data ink ran dry (3x in short address)

Prob(ink ran dry data | hotel hypothesis)
"rare to get anything else" ~ 50%

Prob(ink ran dry data | private alternative)
"seldom allowed to be in such a state" ~ 10%

*Sir Henry Baskerville
Northumberland Hotel*

The ink likelihood ratio

Hypothesis written in hotel
Alternative privately written
Data ink ran dry (3x in short address)

Prob(ink ran dry data | hotel hypothesis)
"rare to get anything else" ~ 50%

Prob(ink ran dry data | private alternative)
"seldom allowed to be in such a state" ~ 10%

$$LR = \frac{\text{Prob}(\text{ran dry} | \text{hotel})}{\text{Prob}(\text{ran dry} | \text{private})} = \frac{50\%}{10\%} = 5$$

Joint likelihood ratio

Combine independent events by multiplication

$$LR_S = \frac{\text{Prob}(\text{splutter} | \text{hotel})}{\text{Prob}(\text{splutter} | \text{private})} = 5$$

$$LR_R = \frac{\text{Prob}(\text{ran dry} | \text{hotel})}{\text{Prob}(\text{ran dry} | \text{private})} = 5$$

$$\begin{aligned} LR &= LR_S \times LR_S \times LR_R \times LR_R \times LR_R \\ &= 5 \times 5 \times 5 \times 5 \times 5 \\ &= 3,125 \end{aligned}$$

DNA mixture evidence

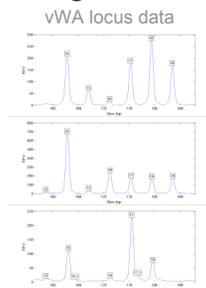
- **quantitative STR data**
peak height is proportional to DNA amount
- **likelihood**
explains data under alternative hypotheses
- **joint likelihood (within locus)**
permits statistical combination of evidence to infer more informative genotypes
- **likelihood ratio**
data support for suspect match hypothesis, relative to population alternative
- **joint likelihood ratio (between loci)**
combines the locus LR's into joint statistic

Regina v. Broughton

- low template mixture
- three DNA contributors
- triplicate amplification
- post-PCR enhancement

- no match score found
- computer interpretation

A match between suspect and evidence is **3,620,000** times more probable than coincidence.



Conclusions

- Victorian science used likelihood inference
- Sherlock Holmes relied on the LR
- LRs have long-standing general acceptance
- likelihood principle applies to DNA evidence
- likelihoods enable combining DNA data

M.W. Perlin, "Explaining the likelihood ratio in DNA mixture interpretation," *Promega's Twenty First International Symposium on Human Identification*, San Antonio, TX, 14-Oct-2010.



Cybergenetics

<http://www.cybgen.com/information/presentations.shtml>
<http://www.cybgen.com/information/publications.shtml>

perlin@cybgen.com
