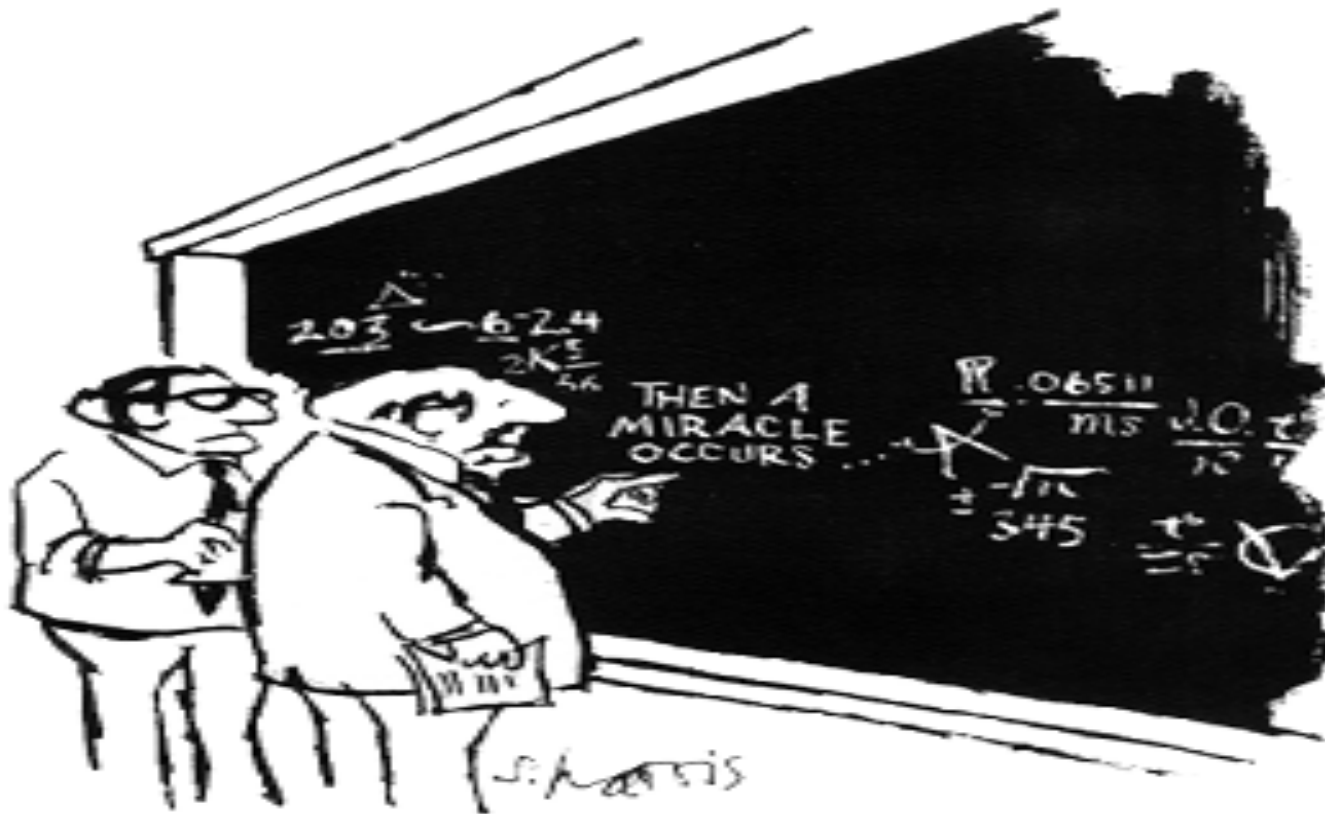# Panel on Similarity-Based Likelihood Ratio

Hal Stern
Department of Statistics
University of California, Irvine
sternh@uci.edu
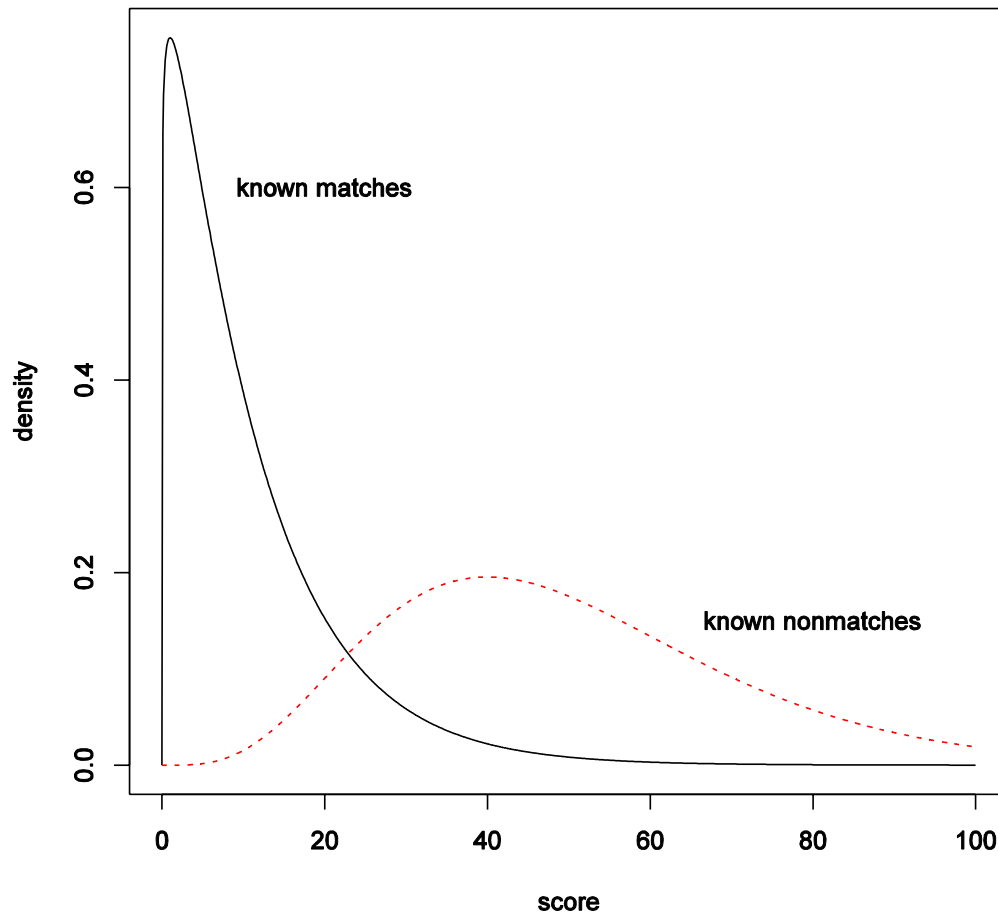
Likelihood ratios for forensic evidence

"I think you should be more explicit here in step two."

Likelihood ratios for forensic evidence

- The score-based approach

Question 1 - Subjectivity

- Probability is personal???
  - Nice reviews of probability and statistical inference yesterday
  - On probability: It can be very difficult (impossible?) to argue with a subjective Bayesian
  - But when it comes to assessing the performance of an expert one can argue that frequentist assessment matters
  - For example … calibrated Bayes (Little, 2006)

Question 1 - Subjectivity

- Likelihood ratio is personal???
  - LR (actually Bayes Factor) can be developed through a purely subjective Bayesian approach
    - "my prior"
    - "my likelihood"
    - For the expert or for the juror/judge
  - Why has LR been so successful for DNA?
    - Probability model that everyone can agree on
    - Information about allele frequencies
    - Published literature

Question 1 - Subjectivity

- OK to have different LRs?
  - Different LRs are in fact inevitable ("sensitivity analysis"?)
    - Different parametric assumptions
    - Parametric vs non-parametric models (see, e.g., Lucy and Aitken, 2004)
    - Different estimation approaches for unknown parameters

Question 2 – Role of Calibration

- Recent issue of Multivariate Behavioral Research features an article by Hoitjink et al., "Why Bayesian Psychologists Should Change the Way They Use the Bayes Factor", which advocates calibrating BF to achieve desired frequency of "misleading evidence" or "to identify appropriate thresholds"

- Published with rebuttal and a couple of complementary articles

- Rebuttal makes the key point that one can calibrate the "Bayes" out of the BF

- In the specific context of Hoitjink et al. calibration reproduced traditional t-test

- My view is that it is important to study frequentist properties of Bayesian (and other procedures) but not desirable to let those frequentist properties drive the Bayesian procedure

Question 3 – The Way Forward

- This is really a key question with no unique answer
- John Butler's 2x2 table of "fundamental understanding" and "practical use" makes a key point about bridging the basic/applied research gap
- Need to continue developing full probability models for high-dimensional complex data types (e.g., Neumann et al., 2015 for latent prints) but we are very far away in many disciplines
- Score-based likelihood ratios can be useful

Question 3 – The Way Forward

- Score-based likelihood ratios

    $LR = Pr(E=(E_x, E_y) \mid H_p) \ / \ Pr(E = (E_x, E_y) \mid H_d)$

    $SLR = Pr(\ S(E_x, E_y) \mid H_p) \ / \ Pr(\ S(E_x, E_y) \mid H_d)$

    – Numerator is the distribution of scores in "known" matches
    – It seems that this can often be well-characterized (e.g., DFIQI talk)
    – Denominator is potentially more challenging. Why?
    – Which non-matches?

Question 3 – The Way Forward

- Denominator of score-based likelihood ratio (see, e.g., Hepler et al., 2012)
  - One approach that seems intellectually reasonable is to estimate denominator using distribution of $S(E_x, E_z)$ over all z in database
    - A potentially large computational burden in every case
    - Different reference distribution for each case
    - Effectiveness will depend on size and representativeness of database
  - A second approach would be to estimate the denominator using the distribution of $S(E_w, E_z)$ for non-matching w, z pairs
    - This suggests a common reference distribution independent of properties of the crime scene evidence
    - Perhaps closest non-matches (e.g., DFIQI talk)
    - Can't be correct to have a single reference distribution
  - Middle ground estimates the denominator as a function of key characteristics of the evidence
    - Example – DFIQI presentation showed non-match distribution as a function of number of features identified in the crime scene sample

Question 3 – The Way Forward

- Another key element of the way forward is not related to LRs

- Reliability and validity studies
  - Ulery et al. 2012 PNAS "black box" study was very useful as our followup "white box" studies
  - This can be done more generally
  - Reliability as a function of evidence characteristics (e.g., quality)
  - Accuracy as a function of evidence characeteristics