Statistical Theory for Likelihood Ratios in Forensic Analysis

Daniel Rabinowitz

Department of Statistics
Columbia University

Notation

Individual(s)

D	**D**efendant
S	**S**ource
A	Perpetrator of the alleged **A**ct

Their characteristic(s)

$$\theta_D, \; \theta_S \text{ distributed according to } \pi(\theta), \; \theta_A$$

Data

Y_1	evidence about θ_S modeled via $f^{\theta_S}(y)$
Y	all the rest of the evidence

Facts of Consequence to Culpability

$$\theta_S = \theta_D \longrightarrow D = S \longrightarrow D = A$$

State's Presentation of Y_1

Likelihood Ratio

- State $D = S$
  - operationally: $Y_1 \sim f^{\theta_D}(y)$
- Defense $D \neq S$
  - operationally: $Y_1 \sim \int \pi(\theta) f^{\theta}(y) d\theta$
  - Not $\theta_D \neq \theta_S$ because want to compute LRs?
  - or because point null embedded in composite alternative?
  - Turing's rule and "protection" on average.

State's Witness Reports

$$LR = \frac{f^{\theta_D}(Y_1)}{\int \pi(t) f^{t}(Y_1) dt} = \pi(\theta_D | Y_1)/\pi(\theta) \text{ is large,}$$

"providing strong evidence" for . . . the State's hypothesis.

Strength of Evidence for State's Hypothesis

Probability depends on LR and target, prior

$$\frac{P\{\text{target}\} \times LR}{P\{\text{target}\} \times LR \;+\; (1 - P\{\text{target}\})}$$

$P\{\theta_S = \theta_D \mid Y_1\}$

$$\frac{\pi(\theta_D) \times LR}{\pi(\theta_D) \times LR \;+\; (1 - \pi(\theta_D))} \quad \text{(miniscule)}$$

or $P\{S = D \mid Y_1,\, Y\}$

$$\frac{P\{S = D \mid Y\} \times LR}{P\{S = D \mid Y\} \times LR \;+\; (1 - P\{S = D \mid Y\})}$$

"Providing strong evidence for the prosecution's hypothesis" a *de facto* (and unusual) instruction about the standard for culpability? Also, implicitly, testifying beyond personal knowledge.

State's Finder-of-Fact's Calculation

$P\{S = D \mid Y_1,\ Y\}$

$$P\{S = D \mid Y_1,\ Y\} \;=\; \frac{P\{S = D \mid Y\}LR}{P\{S = D \mid Y\}LR + (1 - P\{S = D \mid Y\})}$$

$$P\{S = D \mid Y\} \;=\; (a_D + b_D) \Bigg/ \sum_i (a_i + b_i)$$

followed by $P\{A = D \mid Y_1,\ Y\}$

$$\begin{aligned} P\{A = D \mid Y_1,\ Y\} \;&=\; P\{S = D \mid Y_1,\ Y\} P\{A = D \mid S = D,\ Y\} \\ &\quad + (1 - P\{S = D \mid Y_1,\ Y\}) \frac{a_D}{(a_D + b_D)} \end{aligned}$$

How helpful for the purposes of FRE 702(a) is it to report the LR without explaining the subsequent calculations?

As Applied Challenges to

"providing strong evidence for" the hypothesis

$$\sum_i (a_i + b_i) \text{ is large,}$$

$$\text{or most } (a_i + b_i) \text{ are larger}$$

the Model for θ_S and Y_1

$$\text{(Posterior probability given } Y_1 \text{ that) } \theta_S \text{ is consistent with } \pi(\theta)$$

$$Y_1 \text{ is consistent with } \int f^{\theta}(y)\pi(\theta)d\theta$$

Validation

- Leverage statistics measuare "edge" not "hole."

Issues

- Power (sample size, State's quantile and the Confrontation Clause)
- Burden Shifting and Multiple testing
- Statistical versus practical significance

Reporting Uncertainty

Whose burden?

- Standard in *the discipline*: burden on analyst to report uncertainty ... *including* uncertainty in assessments of uncertainty.

Uncertainty in LR

- Propagate standard errors.
- Sensitivity analyses for assumptions. Subjective priors are judicial admissions of uncertainty?
    - random, known distribution
    - random, uncertain distribution
    - not random, subjective beliefs about distribution
    - randomness as a deliberate fiction because
        - only* Bayes' rules are admissible (as statistics term of art),
        - need Bayes factor to have an LR.

  "approximations" and "restrictions" are assumptions, too.
- Validation of π and f^{θ} to complement "black box" studies.