Recent methodological advances in Bayes factors for use in forensic analysis and reporting

Dan J. Spitzner

University of Virginia
Department of Statistics

June 14, 2023

Background

Two recent publications:

(Both appeared in 2023)

Part I

The American Statistician

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

A Statistical Basis for Reporting Strength of Evidence as Pool Reduction

Dan J. Spitzner

Probability

Two misunderstandings of probability:

- **Prosecutor's fallacy:** "confuses the probability of finding the evidence on an innocent person with the probability that a person on whom the evidence is found is innocent"
  (ICAA-RSS, 2017, p. 26)
- **Defense fallacy:** "uses the probability value for an evidence match. . . to argue that for a large enough population. . . their client is only one of many people that could be guilty, and is thus innocent due to reasonable doubt" (p. 28)

Each insufficiently acknowledges the relevance of **prior probabilities**

Prosecutor's fallacy: hypothetical example

One-in-a-million: (Thompson, 2018, sec. III.C)

- DNA profile "found in only one person in 1 million in the general population"
- Is "one-in-a-million" rare enough to conclude that the victim is the source of the material?
- **Consider:** "in a nation as large as the United States there are likely to be over 300 people who share the one-in-a-million DNA profile."

To reduce the pool further—down to the victim—requires "other evidence in the case."

Prosecutor's fallacy: courtroom example

Doheny case: England & Wales Court of Appeals, 1996

Involves a semen stain found on the clothing of a woman who had been raped

> *"There are probably only four or five white males in the United Kingdom from whom that semen could have come" and "the defendant is one of them." The jury's task is to decide, based on "all the evidence," whether "it was the defendant who left that stain or whether it is possible that it was one of the other small group of men who share the same DNA characteristics."*
>
> (See, Redmayne, 2002)

The pool reduction strategy

Pool reduction: State the extent to which forensic material reduces an initial pool of plausible sources to a smaller pool, but not to a particular source

- Avoids controversies around the theory of **individualization**, where "a person or thing is specifically distinguished from all other persons or things of the same kind"
  (Biedermann, Bozza, and Taroni, 2008)
- A refinement of **classification**, wherein "we set out with a goal to individualize," but "fail to narrow the source item to a category of one" (Inman and Rudin, 2001)

Present purpose: Establish a statistical foundation for pool reduction *via* reinterpretion as a Bayes factor

Other reporting strategies

Name	Example
Random match probability (RMP)	Among a certain population 1 in 10 million would match to a DNA sample
Likelihood ratio (LR)	The probability of this evidence is 100 times higher under hypothesis H than A
Likelihood ratio verbal equivalent (VE)	The LR indicates 'moderately strong' evidence of hypothesis H over A
Random match equivalent (RME)	The LR indicates strength-of-evidence equivalent to 1 in 100 RMP

(Thompson and Newman, 2015; Thompson 2012)

Other reporting strategies (continued)

Perception experiment: (Thompson and Newman, 2015)

- RMP, LR, VE only
- Expected effect for DNA, but not for shoeprints, unless RMP

Speculative explanation:

- RMP perceived as scientific/discriminating
- LR or VE seem "a conclusion without evidence"
- DNA is already perceived as highly scientific

Additional speculations: (Thompson 2012)

- RME may seem artificial, but perceived as discriminating

Present speculation:

- Pool reduction may be perceived as discriminating
- Less artificial through connection to Bayes factors

The p-value debate

Recent timeline:

- 2016 ASA statement on p-values
  (Wasserstein & Lazar, 2016)
- 2019 special issue in *The American Statistician* on p-values, statistical significance, reproducibility, and related issues
  (see, Wasserstein, Schirm, & Lazar, 2019)
- 2021 ASA statement clarifying the 2016 statement
  (Benjamini *et al.*, 2021)

Present purpose: Strengthen pool reduction as a strategy for **forensic** evidence-reporting by also establishing it as a strategy for evidence-reporting **in general**

Source attribution

Example 1:

The remains of a murder victim is found in a forest. A suspect is found to have blood stains on his shirt. How strong is the evidence that the blood on the suspect's shirt is that of the victim?

Hypothetical blood types:

Blood type:	O	A	B	AB
Frequency:	41.5%	29.3%	23.0%	6.2%

Evidence report: Based on a comparison of forensic material, a pool of 1000 plausible sources would be reduced to 62, leaving 61 that remain to possibly exclude by other evidence.

Bayesian interpretation

Setup:

- $\Theta = \{\theta\}$ initial pool of 1000 plausible sources, including the victim
- $\theta^* =$ victim

Relevant assertions:

- M_p: Assertion of the **prosecutor**
  $\Rightarrow$ Victim is the source, $\Theta_p = \{\theta^*\}$
- M_d: Assertion of the **defense**
  $\Rightarrow$ Victim is not the source, $\Theta_d = \{\theta \in \Theta : \theta \neq \theta^*\}$
- M_j: Assertion of **jurispridence**
  $\Rightarrow$ One or the other, $\Theta_j = \Theta_p \cup \Theta_d$

Bayesian interpretation (continued)

Setup: M_p *vs* M_j

- **Prior:** equal-probability sampling from Θ_j or Θ_p
- **Data-generation:** $\pi(y|\theta) = 1$ if y is the blood-type of θ
- **Data:** $y =$ blood-type AB

Bayes factor:

$$BF_{pj}(y) = \frac{\pi_p(\theta^*)/\pi_j(\theta^*)}{\pi_p(\theta^*|y)/\pi_j(\theta^*|y)} = \frac{1000/1}{62/1}$$

Bayesian interpretation (continued)

Bayes factor for M_p *vs* M_j, nested $\Theta_u \subset \Theta_v$:

$$BF_{uv}(y) = \frac{\pi_u(y)}{\pi_v(y)} = \frac{P[M_u|y]/P[M_u|y]}{P[M_u]/P[M_v]} = \frac{\pi_u(\theta^*)/\pi_v(\theta^*)}{\pi_u(\theta^*|y)/\pi_v(\theta^*|y)}$$

Interpretations:

1. **Likelihood ratio:** integrated likelihood, $\pi_u(y) = \int_{\Theta_u} \pi_u(y|\theta)\pi_u(\theta)d\theta$
2. **Bayesian updating:** Multiplicative factor transforming prior odds to posterior odds
3. **Pool reduction:** Ratio of relative initial pool-size, $\pi_u(\theta^*)/\pi_v(\theta^*)$ to relative final pool-size, $\pi_u(\theta^*|y)/\pi_v(\theta^*|y)$

Note: inverse-probability is pool-size

Source attribution (continued)

Example 2:

Two blood stains are found at a crime scene where it is clear there has been a struggle. Were two people injured in the struggle, or just one?

Evidence report: **For each** plausible source of one stain, forensic evidence reduces a pool of 1000 sources of the other stain to 62

Observe:

- This is a **relative** reduction in pool-size
- The **absolute** reduction of pool-size is from $1,000^2$ to 62^2

Initial pool-size

Concrete population:

- 300 million in U.S. population reduces to 300

Versions of concrete populations:

- 100,000 county residents reduce to 6,200
- 500,000 residents of five-county area reduce to 31,000

Conventions:

- **Initial pool-size of 1000:** Recalls percentage values carried to one decimal point
- **Final pool-size of 2:** For drastic reductions, avoids reference to fraction of a source

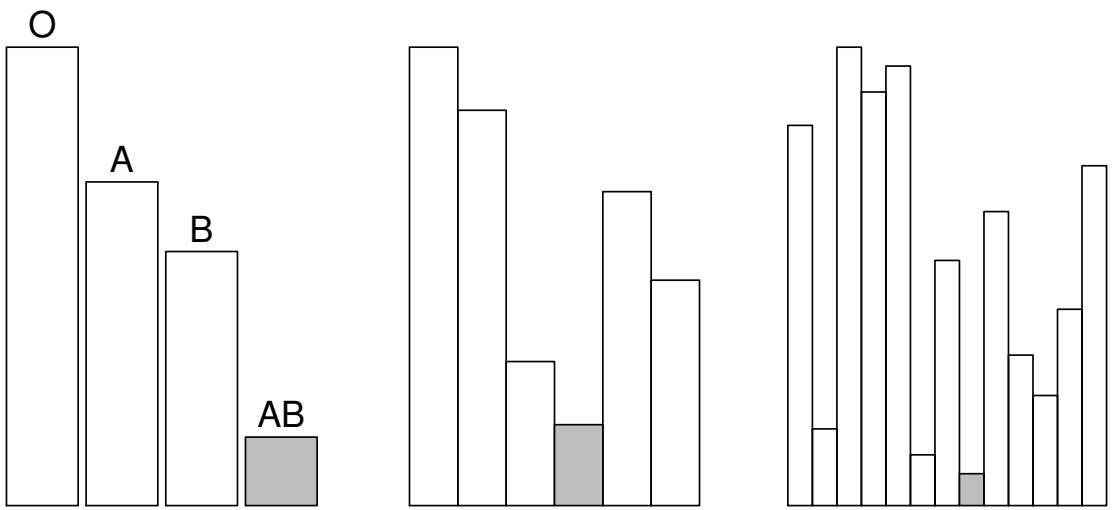
Note: Just the attempt to identify a concrete population can be a worthwhile exercise

Extensions for evidence-reporting in general

- Unequal probability sampling
- Non-nested models

  Note: M_p *vs* M_j, not M_p *vs* M_d, benefits the defense
- Pool expansion

  $\Rightarrow$ Similar to logic that is misapplied within the defense fallacy
- Supplemental descriptive graphics



Part II

Statistical Methods & Applications
https://doi.org/10.1007/s10260-023-00683-4

ORIGINAL PAPER

Calibrated Bayes factors under flexible priors

Dan J. Spitzner[1]

Bayes factors in forensics, bookended across 40 years

Lindley (1977):

- Proposes a Bayes factor for forensic source attribution
- Mathematical core is two-sample testing with Gaussian priors

Lund & Iyer (2017):

- Argue against the normative use of Bayes factors in forensics
- Main criticism: **oversensitivity** to the prior distribution
- Nonparametric/flexible priors have become relevant (See, Aitken and Lucy, 2004; Aitken and Taroni, 2004)
- Frequent reference to subjective Bayesian foundations
- Refuted by Morrison (2017); Aitken *et al.* (2018); Gittelson *et al.* (2018)

Prior oversensitivity

Robustness in model selection: (See, Liseo, 2000)

- **Sensitivity analysis:** examine results across a range of priors (see Berger, 1985, sec. 4.7; Berger and Sellke, 1987)
- **Default priors:** replace the prior with one analytically derived (Jeffreys, 1961; Berger and Pericchi,1996; Bayarri *et al.*, 2012)
- **Approximations:** focus on asymptotically stable elements (Schwarz, 1978; Kass and Wasserman, 1995)
- **Robustifying procedures:** modify or reinterpret techniques to reduce sensitivity (O'Hagan, 1995; Spitzner, 2019)

Flexible priors

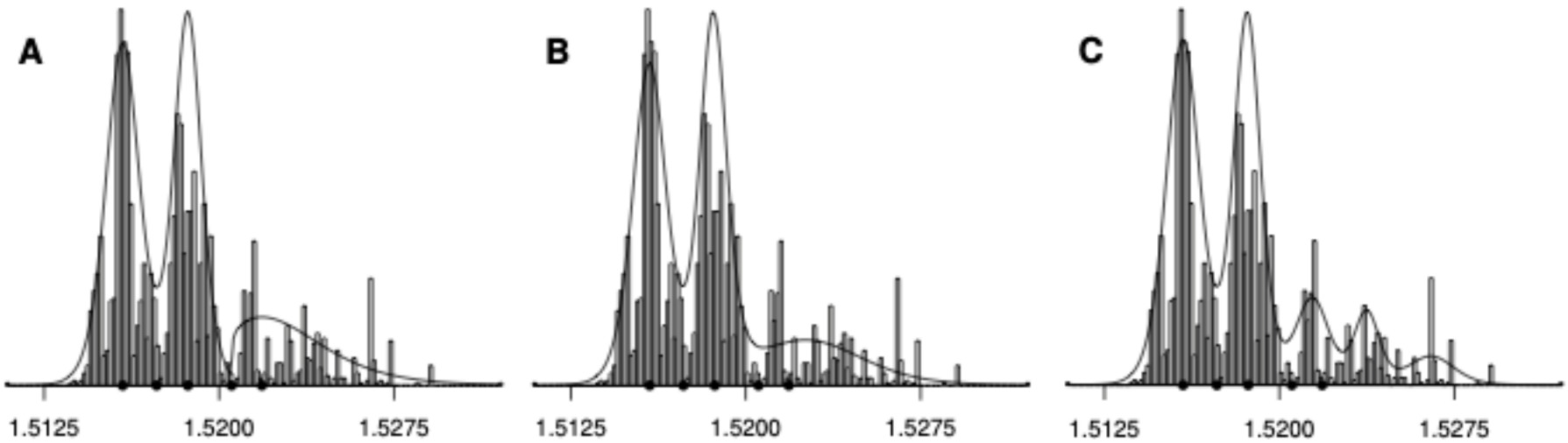
Present purpose: Demonstrate a **robustifying procedure** for Lindley's problem that is consistent with **subjective Bayesian** perspectives and accommodating to **flexible priors**

Flexible parametric form:

$$\pi(\theta_i|\phi_i) = \sum_g \omega_g p_g(\theta_i|\phi_i)$$

- θ_i = target parameter of sample i
- ϕ_i = nuisance parameter of sample i
- ω_g = weight of g'th mixture component
- $p_g(\theta_i|\phi_i)$ = symmetric or skewed mixture component

Example: Refractive index of glass



(Aitken and Taroni, 2004; Curran, 2011; Lund and Iyer, 2017)

Calibrated Bayes factors

Idea: Use a default prior not for substitution, but calibration

For M_u *vs* M_v...

Step 1. Solve for **neutral data**, $\tilde{y}$, under a default prior

$$BF^D_{uv}(\tilde{y}) = 1$$

Step 2. Form a **neutral-data comparison** by substituting $\tilde{y}$ in the elicited prior

$$NDC_{uv}(y) = \frac{BF_{uv}(y)}{BF_{uv}(\tilde{y})}$$

Other aspects

Direct calculation:

- Formulas for $BF_{uv}(\tilde{y})$ are available in certain broad scenarios

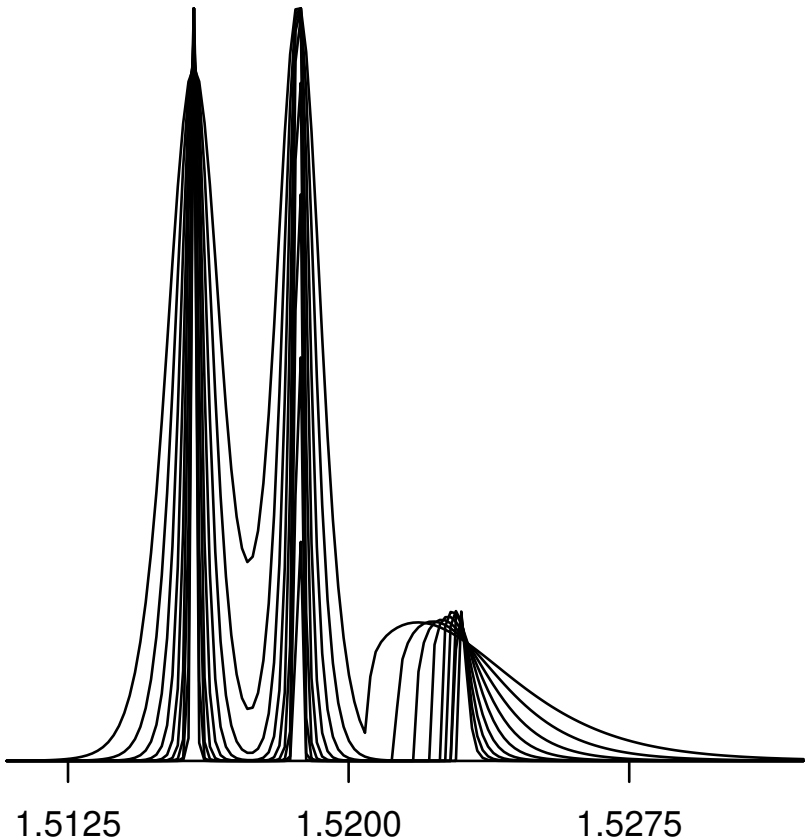
  $\Rightarrow$ Typically, neutral data, $\tilde{y}$, need not actually be calculated

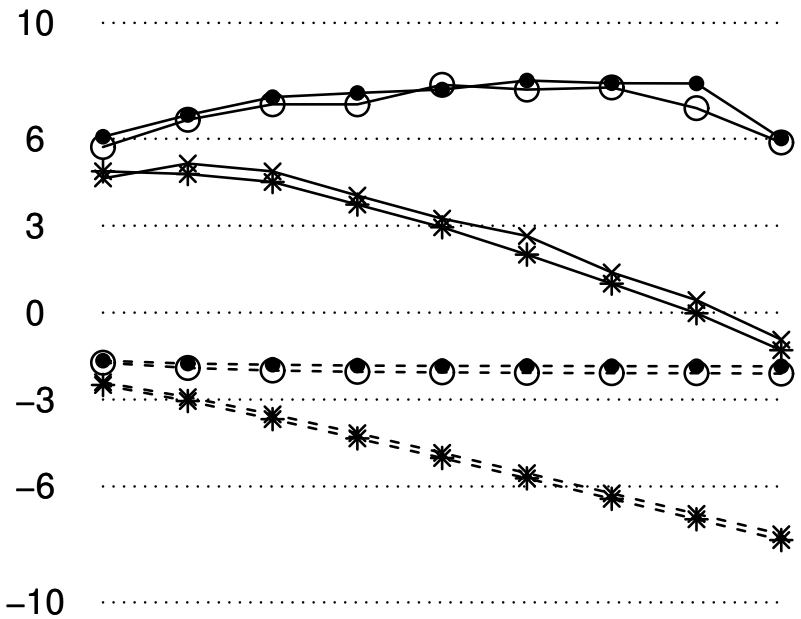
Conditioning and computation:

- Work conditionally in the presence of **nuisance parameters**
- Compatible with **MCMC calculation**

Sensitivity analysis

$-2 \log BF_{12}(y)$ and $-2 \log NDC_{12}(y)$ as functions of prior scale





(Modeled after Aitken and Taroni, 2004, sec. 10.4.2)

Calibrated prior odds

Posterior odds:

$$P[M_u|y]/P[M_v|y] = \rho_{uv} BF_{uv}(y) = \tilde{\rho}_{uv} NDC_{uv}(y)$$

Prior odds:

$$\begin{aligned} \text{Draft value:} \quad & \tilde{\rho}_{uv} = \tilde{P}[M_u]/\tilde{P}[M_v] \\ \text{Refined value:} \quad & \rho_{uv} = P[M_u]/P[M_v] \end{aligned}$$

$$\Rightarrow \rho_{uv} = \tilde{\rho}_{uv}/BF_{uv}(\tilde{y})$$

Possible interpretations:

- **Calibration** of prior odds *after* elicitation
- **Debiasing** of prior odds *within* elicitation

What have we seen?

Pool reduction:

- The pool-reduction strategy has a sound statistical basis

Calibrated Bayes factors:

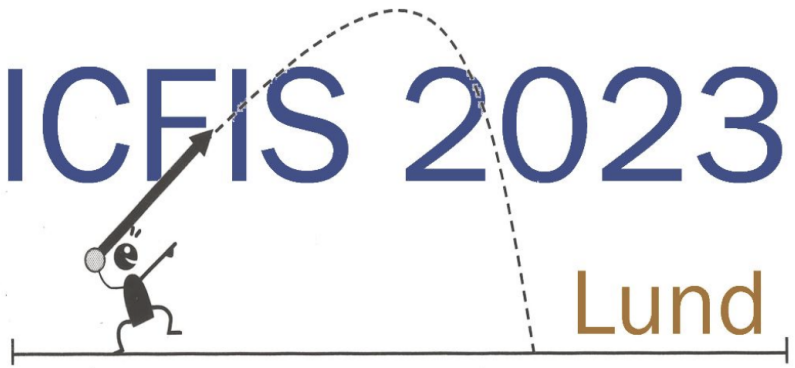
- Robust Bayes factor is compatible with elicited prior knowledge and flexible parametric priors
- Offers a concept by which to guide methodology toward complex situations in forensics

Support

This work was supported by the **Center for Statistics and Applications in Forensic Evidence (CSAFE)** through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and the University of Virginia.

Thanks

Thank You!!



spitzner@virginia.edu

https://profile.virginia.edu/djs4y