Forensics II: Further interpretation of the likelihood ratio and exclusion power

Thore Egeland

Norwegian University of Life Sciences

thore.egeland@nmbu.no

Contents

- Formulation of hypotheses using IBD parameters.
- Testing in forensic genetics vs classical approaches
- Exclusion power.
- Bayesian approach: Including prior, non-DNA information. Controversial also in forensics.
- Decision theory: Justify thresholds used for conclusions.
- Further discussion of LR.

Alternative formulations of hypotheses. p-values?

- H_1: AF biological father of CH.
- H_2: AF and CH unrelated.

Parametric reformulation:

- H_1: $\kappa = (0, 1, 0)$
- H_2: $\kappa = (1, 0, 0)$

Generalisation: consider all (non-inbred) alternatives:

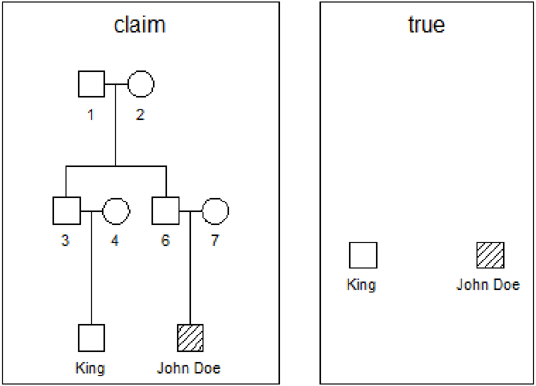
- H_1: $\kappa = (0, 1, 0)$
- H_2: $\kappa \neq (0, 1, 0)$

Forensic genetics: I have never seen latter formulation and classical p-value based testing outside academia.[Kaur, PhD, NMBU, 2016]

Power

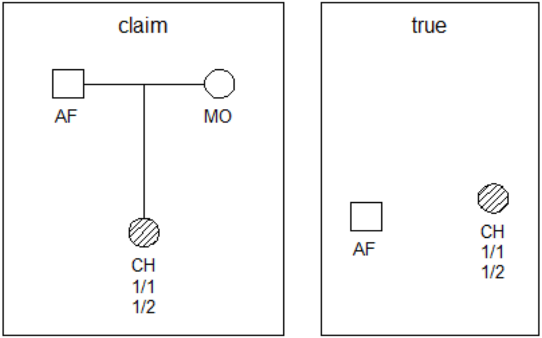
- Generally:
  - Power calculations can be used to determine *sample size*
- Forensic genetics:
  - How many and who should we genotype?
  - How many, which markers should be used?
  - ...

Generic example...



- What data do we need to exclude John Doe as the first cousin of the King given that he is unrelated?

Exclusion Power (EP). Two equifrequent SNPs



$$EP = P(\text{"claim" incompatible with genotypes} \mid \text{"true"})$$

$$EP_1 = P(g_{AF} = 2/2) = 0.5^2 = 0.25, \;\; EP_2 = 0$$

$$EP = 1 - (1 - EP_1) \cdot (1 - EP_2) = 0.25 \text{ for both markers}$$

- `forrel::exclusionPower`

Two approaches to paternity testing: EP versus LR

- Method 1 (used, not recommended): Assume AF is not excluded. Calculate EP *not* using genotype data for AF. If EP is close to 1, report strong evidence in favour of paternity versus unrelated
- Method 2 (recommended): compute the LR as before.

Differences between these approaches

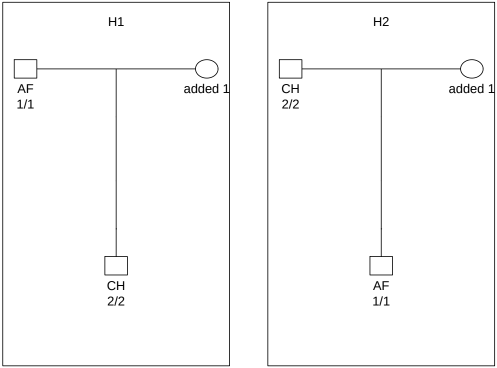
EP

- Does not use the genotype of the alleged father, only that of the child
- Can be computed prior to having any alleged father
- E.g., to judge whether to do a database search (how many possible fathers to expect)
- $EP = P(LR = 0 \mid H_D)$

LR

- Uses all available genetic information on both individuals
- Is therefore better informed than EP

Bayesian approach: Motivation

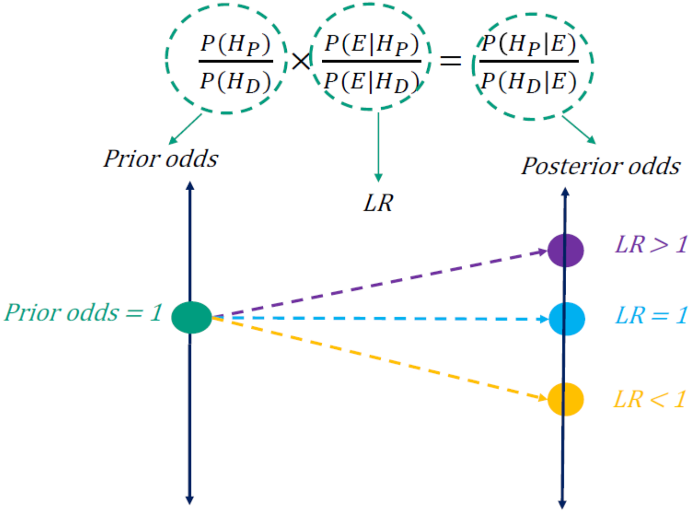


- H_1 more likely *apriori* than H_2 based on age information
- How do we include non-DNA information? Prior

Bayesian framework

- Specify $P(H_P), P(H_D)$, typically subjectively or
- Prior odds: $P(H_P)/P(H_D)$
- *Flat prior* $P(H_P) = P(H_D) = 0.5$ often used.
- I avoid using the common *uninformative prior* for flat prior.

Bayes theorem on odds form



Prior and posterior odds

Assume

- *prior odds* $\frac{P(H_1)}{P(H_2)} = 1000.$

Then

$$\text{prior odds} * \text{LR} = \text{posterior odds},$$
$$1000 * 0.66 = 666.$$

Interpretation: H_1 is 666 times more probable than H_2.

Posterior probability of paternity. Bayes theorem

$$P(H_1 \mid E) = \frac{P(E \mid H_1)P(H_1)}{P(E \mid H_1)P(H_1) + P(E \mid H_2)P(H_2)}$$
$$= \text{"Probability of } H_1 \text{ given evidence"}$$

Important special forensic case: $P(H_1) = P(H_2) = 0.5$.
The Essen-Möller index for paternity:

$$W = P(H_1 \mid E) = \frac{LR}{1 + LR}.$$

Allows inteligible statements like:
"The probability that he is the father is 99.73%".
Problem: the prior ...

Main practical problems in forensics

- Do we report LR, posterior probability or posterior odds?
- Or should we report on a verbal scale? Both numbers and verbal statements?
- How do we choose thresholds?

One Verbal Scale for LR

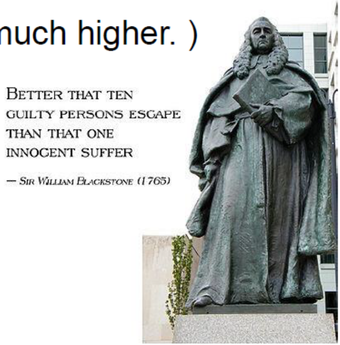
LR	Expert guidance*
1	… do not support one proposition over the other
2 - 10	weak support
10 - 100	moderate support
100 - 1000	moderately strong support
1000 - 10000	strong support
10000 - 1 million	very strong support
Over 1 million	extremely strong support

*ENFSI Guideline for Evaluative Reporting in Forensic Science

How do we specify thresholds?. Decision theory

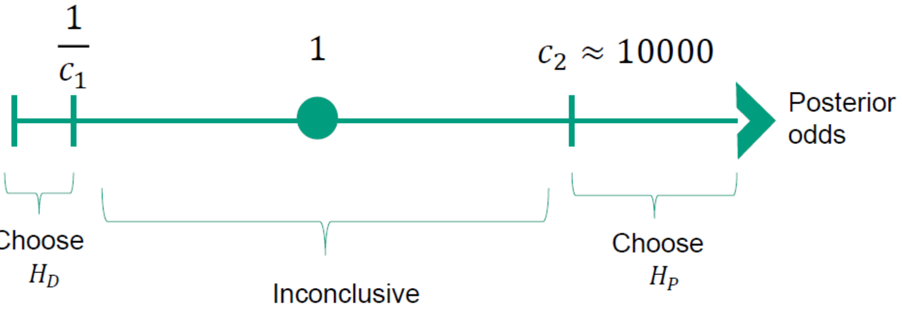
- Blackstone's ratio:
$(1 + c_2)/(1 + c_1) =$ 10 (in practice much higher.)

		TRUTH	
		Guilt H_P	Innocence H_D
VERDICT	Guilt H_P	0	$1 + c_2$
	Innocence H_D	$1 + c_1$	0



Make no decision: cost = 1

Optimal decision rule



If c_1 and c_2 are specified, an optimal decision rule can be determined.
See Tillmar and Mostad (2014) for an application

Adding evidence I

- If prior odds $= 0$ or $LR = 0$

  ➡ posterior odds $= 0$

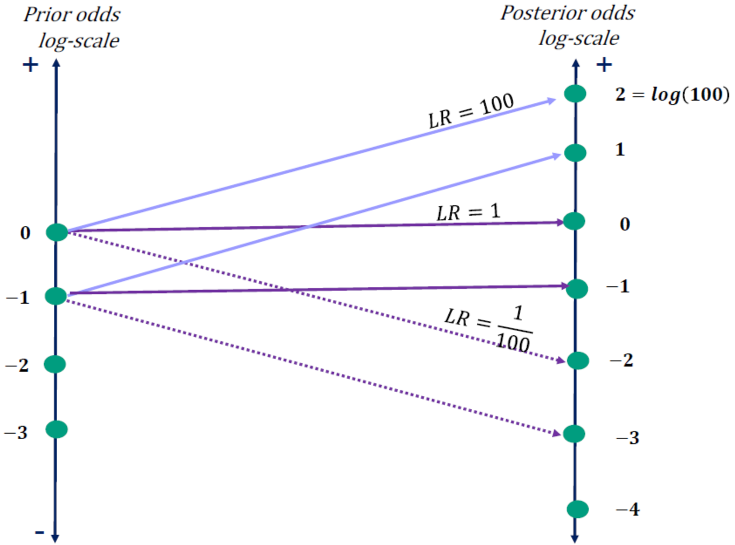
- Assume prior odds > 0 and $LR > 0$. Then

$$\log(\text{prior odds}) + \log(\text{LR}) = \log(\text{posterior odds})$$

- $\log(\text{LR}) = \log_{10}(\text{LR})$ (unit called "ban" - Alan Turing)

*Good IJ (1985)

Adding evidence II



References

- Egeland, Kling, Mostad. Academic Press, 2015.
- IJ Good. Bayesian Statistics, 1985.
- Making Sense of Forensic genetics
- Tillmar, Mostad. FSI: Genetics, 2014.