Using Mixture Models to Examine Group Differences - An Illustration Involving the Perceived Strength of Forensic Science Evidence

Naomi Kaplan-Damary

Joint work with William Thompson, Rebecca Hofstein Grady and Hal Stern

Department of Statistics,
University of California, Irvine



December 2021

Motivation

- Suppose you are a juror and see this evidence

- A fingerprint expert compared the crime scene print with the fingerprint of a suspect and found them indistinguishable
- The strength of this evidence depends on the quality and number of distinguishing features in the prints
- The expert needs to explain the strength of the evidence

Motivation

- Examples of conclusion statements
  - "Suspect was the source"
  - "The observed evidence is 10 million times more likely if the suspect rather than a random person is the source"
  - "It is highly probable that the suspect was the source"
- Which statement is the strongest?
- Which statement is best understood by jurors?
- What are the strengths and weaknesses of these different kind of ways of expressing opinions?
- This has become an important issue in forensics

Thompson et. al. study

- Thompson et. al. (2018)[1] examined how lay people perceive the strength of examiner conclusion statements
- Their goal: compare different statements and determine their ranking
- This study will be reviewed more formally later
- The underlying assumption of Thompson et. al analysis: all people are governed by the same preferences (single homogeneous population)

[1] Perceived strength of forensic scientists' reporting statements about source conclusions. Law, Probability & Risk, 17:2

Background

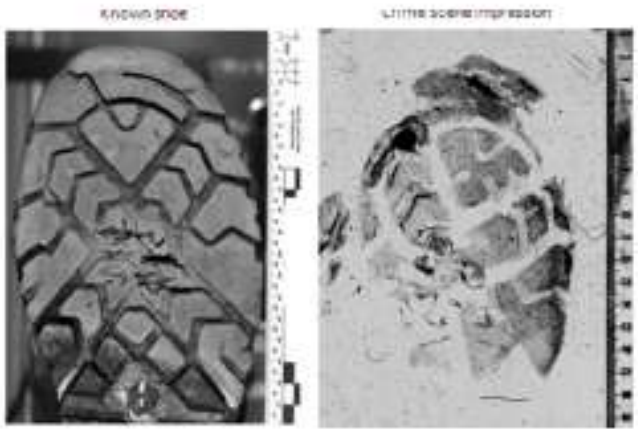
- National Academy of Sciences report (2009)[2] and President's Council of Advisors on Science and Technology (PCAST) report (2016)[3]
  - Question the scientific foundations of many forensic disciplines
  - Have led to new research on the scientifically rigorous analysis of forensic evidence
- Suppose:
  - We build up the scientific foundations
  - We have reliable quantitative conclusions about the strength of the evidence
- How would the court perceive the evidence?
  - Do jurors understand statistical testimony?
  - What value do jurors give qualitative versus quantitative statements about the weight of the evidence?

[2] National Research Council, Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington, DC, 2009

[3] Executive Office of the President President's Council of Advisors on Science and Technology, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, PCAST, Washington, DC, 2016

The setting

- Forensic scientists often compare items of evidence to assess whether they originate from a common source



The setting

- E: The evidence
- H_s: "Same source" proposition (two samples have the same source)
  H_d: "Different source" proposition (two samples have different sources)

The setting

- Logic of forensic examinations
  - Examine two samples to identify similarities and differences
  - Assess similarities and differences to see if they are expected (or likely) under the same source hypothesis
  - Assess similarities and differences to see if they are expected (or likely) under the different source hypothesis

The setting

- Approaches to the presentation of statistical evidence
  - Forensic conclusions as expert opinion - Expert assessment based on experience, training, use of accepted methods
  - Two-stage procedure
    - Similarity (binary decision of whether two samples can be distinguished based on distance or score)
    - Identification (if indistinguishable, how unusual is that?)
  - Likelihood ratio $\mathrm{LR} = \frac{\mathrm{Pr(E|H_s)}}{\mathrm{Pr(E|H_d)}}$
    - Related: Posterior odds (people sometimes report posterior probabilities)

$$\frac{Pr(H_s|E)}{Pr(H_d|E)} = \frac{Pr(E|H_s)}{Pr(E|H_d)} \cdot \frac{Pr(H_s)}{Pr(H_d)}$$

$$\underset{Odds}{Posterior} = \underset{Ratio}{Likelihood} \cdot \underset{Odds}{Prior}$$

Forensic Conclusions as Expert Opinion

- Status quo in pattern disciplines
  (fingerprints, shoe prints, toolmarks, questioned documents, etc.)
- Expert analyzes evidence based on
  - Experience
  - Training
  - Use of accepted methods in the field
- Assessment of the evidence reflects examiner's expert opinion
- Opinions typically reported as categorical conclusions
  - Identification, inconclusive, exclusion
  - Some support, strong support, very strong support, etc.

The Two-Stage Approach

- One common statistical approach solves the forensic problem in two stages
- Stage 1 (Similarity)
  - Determine if the crime scene and suspect objects agree on one or more characteristics of interest
  - Two samples "are indistinguishable", "can't be distinguished", "match"
- Stage 2 (Identification)
  - Assess the significance of this agreement by finding the likelihood of such agreement occurring by chance
    - Assume we have found a "match" : Assessment of the probability that two samples from different sources would "match" or be found "indistinguishable" by chance / coincidence

Likelihood Ratio Approach

- Goal for trier of fact in courtroom setting is decision about the relative likelihood of two hypotheses (e.g., same or different source) given data
- Formally using E (evidence) and S (same source) (with $\bar{S}$ denoting different source)
- $LR = \frac{\Pr(E|S)}{\Pr(E|\bar{S})}$, the relative likelihood of the evidence under the two hypotheses

Likelihood Ratio Approach

- $LR = \frac{\Pr(E|S)}{\Pr(E|\bar{S})}$
- Numerator assumes common source and asks about the likelihood of the evidence in that case
- Denominator assumes no common origin and asks about the likelihood of the evidence in that case
  - Analogous to finding coincidence probability
- LR statement: "The evidence (e.g., level of agreement) is LR times more likely if the objects have the same source than if the objects have different sources"

Likelihood Ratio Approach

- The $LR = \frac{\Pr(E|S)}{\Pr(E|\bar{S})}$ **does not** make any direct statement about the probability of the hypotheses
- If we wish to talk about the probability of the hypothesis, then we are interested in the posterior probabilities of the hypotheses
- Bayes' rule is a way of reversing direction of conditional probabilities
  - We can go from statements about the **likelihood of the evidence given the hypotheses** to statements about **the likelihood of the hypotheses given the evidence**
- Bayes' Theorem written in terms of the odds in favor of the same source hypothesis

$$\frac{\Pr(S|E)}{\Pr(\bar{S}|E)} = \frac{\Pr(E|S)}{\Pr(E|\bar{S})} \times \frac{\Pr(S)}{\Pr(\bar{S})}$$

- In words: Posterior odds = Likelihood ratio $\times$ Prior odds

Likelihood Ratio Approach

- Bayes' Theorem written in terms of the odds in favor of the same source hypothesis

$$\frac{\Pr(S|E)}{\Pr(\bar{S}|E)} = \frac{\Pr(E|S)}{\Pr(E|\bar{S})} \times \frac{\Pr(S)}{\Pr(\bar{S})}$$

- In words: Posterior odds = Likelihood ratio $\times$ Prior odds
- But to talk about posterior probabilities,
  we must have had prior probabilities to start with
  and we should be willing to say what they are
- Does not seem that this is the role of the forensic examiner

The setting

- The different approaches lead to different types of statements

  Conclusion list

  - Expert assessment based on experience, training, use of accepted methods - typically summarized by a categorical conclusion
  - Two-stage procedure - summarized by random match probability (RMP) or similar statements
    - Similarity (binary decision of whether two samples can be distinguished based on distance or score)
    - Identification (if indistinguishable, how unusual is that?)
  - Likelihood ratio $\mathrm{LR} = \frac{\Pr(\mathrm{E}|\mathrm{H_s})}{\Pr(\mathrm{E}|\mathrm{H_d})}$ - summarized by likelihood ratios (LR) or strength of support (SOS)
    - Related: Posterior odds (people sometimes report posterior probabilities) - summarized by source probability (SP) statements

$$\frac{Pr(H_s|E)}{Pr(H_d|E)} = \frac{Pr(E|H_s)}{Pr(E|H_d)} \cdot \frac{Pr(H_s)}{Pr(H_d)}$$

$$\underset{Odds}{Posterior} = \underset{Ratio}{Likelihood} \cdot \underset{Odds}{Prior}$$

The Statements 16 21 26 27 36 37 38 39

Study	Statement	
		Random match probability (RMP)
2,3	RMP4:	"one person in 10 million" expected to provide this degree of similarity
1, 2, 3	RMP3:	"one person in 100,000"
1, 2	RMP2:	"one person in 1,000"
1	RMP1:	"one person in 10"
		Likelihood ratios (LR)
3	LR4:	"10,000,000 times more likely" if suspect rather than random person is source
3	LR3:	"100,000 times more likely"
		Categorical conclusions (CC)
3	CC5:	"suspect was the source"
2	CC4:	"individualized . . . as coming from the finger of the suspect"
2	CC3:	"identified . . . to the finger of the suspect"
2	CC2:	"matches the fingerprint of the suspect"
3	CC1:	"suspect could have been the source"
		Likelihood of observed similarity (LOS)
2	LOS1:	"likelihood of observing this amount of corresponding ridge detail when two fingerprints are made by different people is considered extremely low"
		Source probability statements (SP)
1	SP3:	"a practical certainty" that suspect was the source
1, 2, 3	SP2:	"highly probable"
1	SP1:	"moderately probable"
		Strength of support (SOS)
1, 2, 3	SOS4:	"extremely strong support" for same source theory
3	SOS3:	"very strong support"
1	SOS2:	"moderate support"
1	SOS1:	"weak support"

Thompson et. al. study design

- Thompson et. al. (2018)[4] examined how lay people perceive the strength of these statements
- Their goal: compare different statements and determine their ranking
- They conducted three separate online studies with jury-eligible U.S. adults from mTurk
- Studies used different evidence types:
  - Fingerprint comparisons (Study 1, 2)
  - A DNA comparison (Study 3)
  - Study 1: N = 120; Study 2: N = 121; Study 3: N = 138

[4] Perceived strength of forensic scientists' reporting statements about source conclusions. Law, Probability & Risk, 17:2

Thompson et. al. study design

- For example, participants were:
  - Asked to imagine that a fingerprint expert compared a crime scene print with the fingerprint of a suspect and had found them indistinguishable
  - Told that the strength of this evidence depends on the quality and number of distinguishing features in the prints
  - Told that experts need to explain how strong the evidence is in a particular case
  - Asked to evaluate a series of statements that an expert might make about the strength of fingerprint evidence

Thompson et. al. study design

- Original attempt: have people rank many statements
- Hard for people to reliably rank multiple statements
- People provide meaningful responses when asked to evaluate the relative strength of two statements
- Shifted to paired comparison approach

Which of the following two conclusions would seem STRONGER if you heard it, meaning more convincing to you that the suspect is the source of the print?

- ○ I individualized the crime scene fingerprint as coming from the finger of the suspect.
- ○ It is highly probable that the suspect is the person who made the crime scene fingerprint.

Thompson et. al. study design

- Nine statements were to be compared, creating a pool of 36 possible paired comparisons
- Each participant evaluated 16 pairs; selected randomly
- Some statements were used in all studies; some in just one or two studies Conclusion list

Paired comparison models

- $Y_{kij} = 1$: if statement i if preferred to j by subject k and 0 otherwise
- K: the number of subjects
- m: the number of statements
- $\lambda_i \in \mathbb{R}$: the notional strength of the statements
  - typical perception
  - actual perception in a comparison includes randomness

Paired comparison models

- π_{ij}: the probability that statement i is preferred to statement j

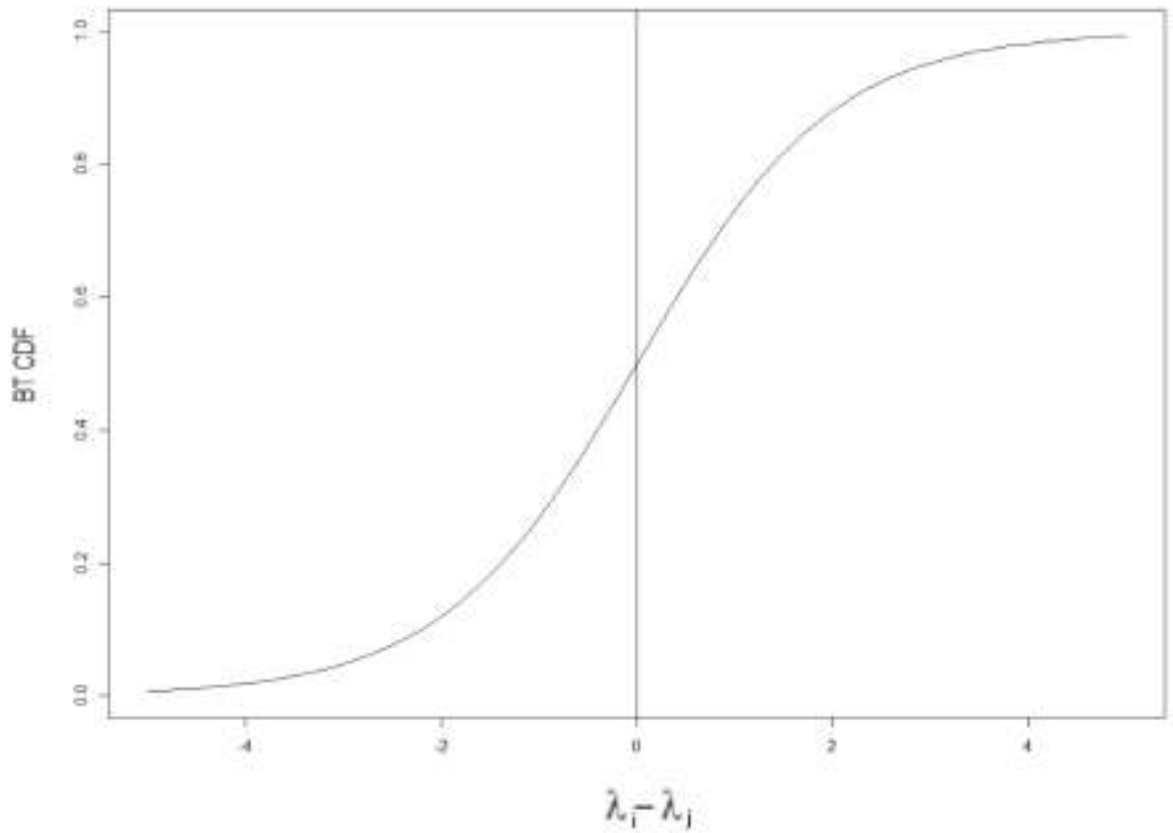
$$\pi_{ij} = F(\lambda_i - \lambda_j)$$

  known as a paired comparison model
- F: the cumulative distribution function of a zero-symmetric RV
  - Thurstone model (1927): F is the normal CDF
  - Bradley–Terry model (1952): F is the logistic CDF

Paired comparison models

- We are using the Bradley–Terry (BT) model:

$$\pi_{ij} = \frac{e^{\lambda_i - \lambda_j}}{1 + e^{\lambda_i - \lambda_j}}$$



Inference

- The aim of the analysis: make inference on $\lambda = (\lambda_1, \ldots, \lambda_m)$ to understand the relative positions of the statements
- The likelihood (BT model):

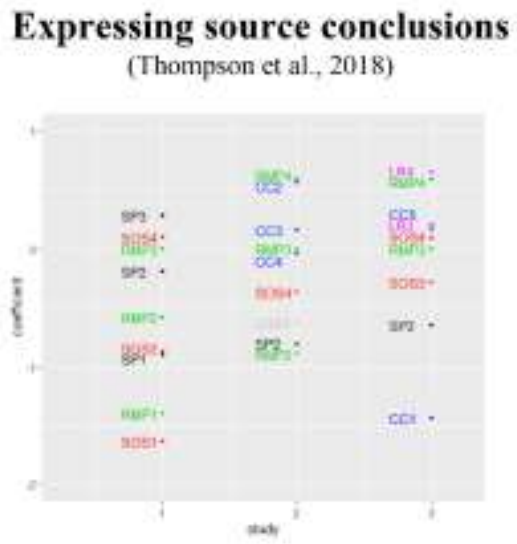
$$\prod_{k \in K} \prod_{ij \in I_k} \left(\frac{e^{\lambda_i - \lambda_j}}{1 + e^{\lambda_i - \lambda_j}} \right)^{y_{kij}} \cdot \left(\frac{1}{1 + e^{\lambda_i - \lambda_j}} \right)^{1 - y_{kij}}$$

- An identification constraint, $\lambda_i = 0$ for one statement $i \in 1, ..., m$ (reference statement constraint) is used

Thompson et. al. Results

Conclusion list

- The ranking of the statements:



- Statements the authors viewed as stronger were ranked significantly higher - e.g. $S0S4 > SOS2 > SOS1$ extremely strong $>$ moderate $>$ weak
- In no category were all statements perceived as stronger or weaker than the statements in any other category

Thompson et. al. Results

Conclusion list

- Surprisingly: CC3 (identified) and CC4 (individualized) , believed to be the strongest statements, were not perceived that way
- "Match" is dangerous: CC2 (match) was given similar weight as RMP4 ($1 : 10$ *million*)
- Likelihood ratios are powerful:
  - Strong as or stronger than comparable RMPs (one person in X)
  - $LR4 > CC5$ (10 *million times more likely* $>$ *suspect was source*)
  - $LR3 > SP2$ ($100,000$ *times more likely* $>$ *highly probable*)
  - $LR3 \approx SOS4$ ($100,000$ *times more likely* $\approx$ *extremely strong support*)

Limitations of the Thompson et. al. study

- The underlying assumption of Thompson et. al analysis: all people are governed by the same preferences
- Thompson et. al. have some information about each participant (was not used in their study)
  - Five characteristics of each participant were measured:
    1. Age
    2. SNS: self-rated numeracy on a scale from 1-6 (average of responses to 8 survey items)
    3. Gender
    4. Education level, on a scale from 1-9
    5. Forensic knowledge: "How knowledgeable are you about forensic science?" 1: (Not Knowledgeable) to 7: (Extremely Knowledgeable)
- Perhaps there are subsets of the population which interpret statements differently (e.g young, educated)

Our Goal

Goal:

- Examine whether there are subpopulations which interpret statements differently
- Examine whether these subpopulations can be characterized

- One approach: Look at different subsets of the population defined by specific covariates to see whether they seem to be described by the same preferences
- A more general approach: Allow the data to determine if there are subpopulations

Methods

- The first approach is an exploratory analysis, which considers subpopulations defined a priori by specific characteristics
- The original study sample was split into two groups according to each characteristic (Gender, Age, SNS, Level of Education, Forensic knowledge) based on the median values
- The primary finding: there appear to be differences based on SNS
- The BT model fitted separately to the two SNS groups was found to be a significant improvement over the BT model fitted to the entire population in all three studies

Methods

TABLE 3 *SNS Analysis. In each study, the estimated strength parameters of the low and high SNS groups are presented*

Study 1				Study 2				Study 3			
low		high		low		high		low		high	
SP3	0.41	SP3	0.79	CC2	0.99	RMP4	1.10	LR4	0.96	LR4	1.31
SOS4	0.05	SOS4	0.47	RMP4	0.93	CC2	1.02	RMP4	0.93	RMP4	1.07
RMP3	0.00	RMP3	0.00	CC3	0.30	CC3	0.25	SOS4	0.30	CC5	0.42
SP2	−0.27	SP2	−0.43	CC4	0.08	RMP3	0.00	CC5	0.27	LR3	0.38
RMP2	−0.76	RMP2	−1.57	RMP3	0.00	CC4	−0.27	LR3	0.21	RMP3	0.00
SP1	−1.09	SOS2	−2.29	SOS4	−0.40	SOS4	−1.07	RMP3	0.00	SOS4	−0.23
SOS2	−1.20	SP1	−2.46	LOS1	−0.70	LOS1	−1.60	SOS3	−0.20	SOS3	−1.07
RMP1	−1.81	RMP1	−3.77	SP2	−0.81	RMP2	−1.60	SP2	−1.02	SP2	−1.35
SOS1	−2.14	SOS1	−4.65	RMP2	−1.41	SP2	−2.29	CC1	−2.25	CC1	−3.46

- The ordering of the statements is similar in both groups but the high SNS group tends to distinguish more between stronger and weaker statements
  - High SNS group of Study 1: the estimated strength parameters range from –4.65 to 0.79
    $\rightarrow$ the estimated probability that the strongest statement is preferred to the weakest statement is 0.99
  - Low SNS group of Study 1: the range is between -2.14 to 0.41
    $\rightarrow$ the estimated probability that the strongest statement is preferred to the weakest statement is 0.93

Methods

- A limitation of this approach is that it focuses only on a single characteristic of the individual at a time
- A second approach that does not require us to a priori identify which characteristics might be relevant uses a mixture model to identify subpopulations
- Mixture model allows for various subpopulations (or clusters) with possibly different rankings of the statements
- Allow for possibility that covariates (SNS, Forens.Knowl.) explain subpopulation membership
- x_k: the covariate vector of subject k
- I_k: the statements subject k compared

Methods

- We are going to build a cluster model. Intuition:
  - Suppose there are two clusters each with a set of preferences, λ_1 and λ_2
  - We do not define clusters in advance, we explore and see if they exist
  - Let u_k define the cluster that person k belongs to, $u_k \in \{1, 2\}$
  - The model for cluster membership is a logistic regression
  - The probability of belonging to group 1: $Pr(U_k = 1) = \frac{e^{\beta_1 x_k}}{1+e^{\beta_1 x_k}}$
  - The probability of belonging to group 2: $Pr(U_k = 2) = 1 - \frac{e^{\beta_1 x_k}}{1+e^{\beta_1 x_k}}$

Methods

- It is possible to simultaneously estimate the cluster membership and the parameters that describe the cluster
- The EM algorithm is an approach for estimating parameters when there are missing data (here clusters are the missing data)
- The EM has two steps
  1. If you knew the λs you could estimate the cluster membership probabilities
  2. If you have the estimated cluster membership probabilities you can update the λs

Methods

- In general, u_k: the cluster membership, $u = 1, \ldots, U$ (u_k is not observed)
- Model for cluster membership:

$$Pr(U_k = u_k) = \frac{e^{\beta_{u_k} x_k}}{\sum_{u=1}^{U} e^{\beta_u x_k}}, \beta_U = 0$$

- The likelihood $L(\lambda, \beta \mid x, y)$:

$$\prod_{k \in K} \sum_{u_k=1}^{U} \prod_{ij \in I_k} \left(\frac{e^{\lambda_i^{(u_k)} - \lambda_j^{(u_k)}}}{1 + e^{\lambda_i^{(u_k)} - \lambda_j^{(u_k)}}} \right)^{y_{kij}} \cdot \left(\frac{1}{1 + e^{\lambda_i^{(u_k)} - \lambda_j^{(u_k)}}} \right)^{1 - y_{kij}} \cdot \left(\frac{e^{\beta_{u_k} x_k}}{\sum_{u=1}^{U} e^{\beta_u x_k}} \right)$$

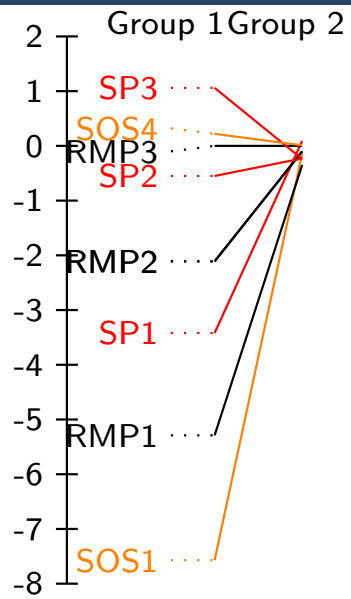
- Inference: We can use EM algorithm to find ML estimates
- We begin by assuming there are two clusters

Results - Study 1

Conclusion list

Study 1

Group 1 (≈75% of the population)	Estimate	SE	Group 2 (≈25% of the population)	Estimate	SE
SP3	1.06	0.27	SP1	0.09	0.27
SOS4	0.22	0.25	SOS4	0.01	0.28
RMP3	0	-	RMP3	0	-
SP2	-0.55	0.25	RMP2	-0.1	0.23
RMP2	-2.11	0.25	SOS1	-0.14	0.26
SOS2	-3.28	0.32	SP2	-0.23	0.26
SP1	-3.42	0.32	SP3	-0.24	0.28
RMP1	-5.29	0.4	SOS2	-0.29	0.26
SOS1	-7.57	0.62	RMP1	-0.35	0.23

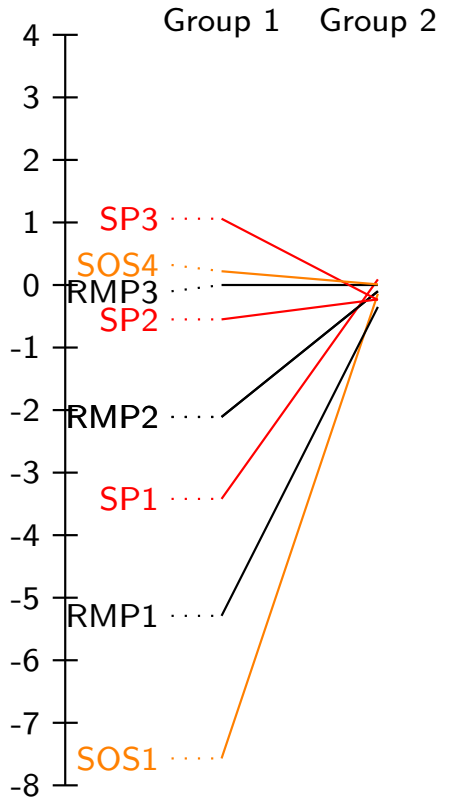


- A high estimate corresponds to a statement which was perceived as a particularly strong statement
- Group 2 ($\approx$ 25%): does not differentiate between the statements
- Group 1: statements are in the order found by Thompson et. al.
  - Possible explanation: Group 1 paid more attention to the details of what the expert was saying
  - Alternatively: Group 1 may have been more capable of processing the significance of the expert's statements

Results - Study 1

Conclusion list

	Estimate	Std. Error	Pr(>|z|)
(Intercept)	-6.22	1.73	0.00
SNS total	0.97	0.29	0.00
Age	0.06	-0.03	0.05
Gender	0.32	0.53	0.55
Education	0.33	0.16	0.04
Forensic knowledge	-0.28	0.18	0.12



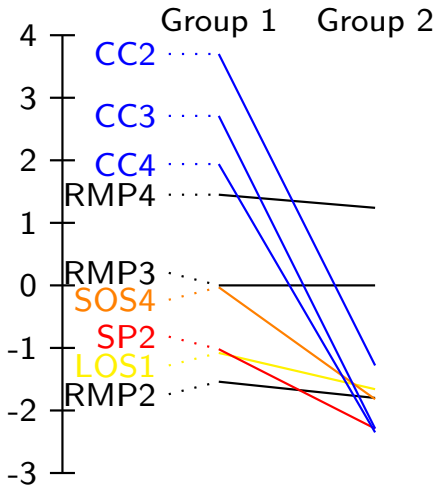
- Group 1 - people with higher SNS, age and education
- Every increase of 1 point in SNS increases the odds of belonging to group 1 by a factor of 2.6

Results - Study 2

Conclusion list

Study 2

Group 1 (≈ 59% of the population)	Estimate	SE	Group 2 (≈ 41% of the population)	Estimate	SE
CC2	3.7	0.41	RMP4	1.24	0.28
CC3	2.71	0.33	RMP3	0	-
CC4	1.94	0.31	CC2	-1.28	0.29
RMP4	1.45	0.22	LOS1	-1.66	0.28
RMP3	0	-	RMP2	-1.8	0.25
SOS4	-0.03	0.24	SOS4	-1.82	0.28
SP2	-1.02	0.26	CC3	-2.29	0.32
LOS1	-1.08	0.27	SP2	-2.29	0.29
RMP2	-1.54	0.21	CC4	-2.35	0.3



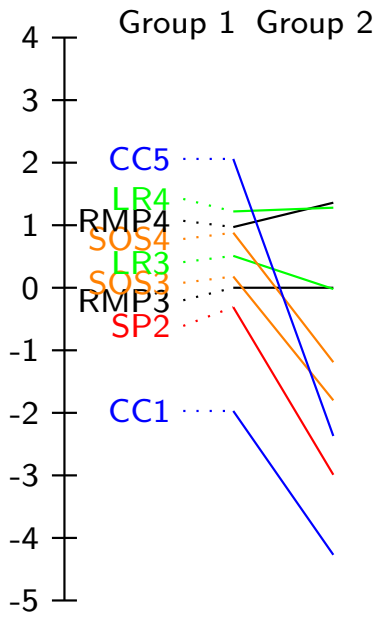
- Group 1 and 2 agree on the strength of RMP statements (similar estimates)
- Group 1 finds CC statements stronger than quantitative (RMP)
- Group 2 favors quantitative statements
  - Group 1: the probability CC4 (individualized) is preferred to RMP4 (1:10 million) is 0.62
  - Group 2: the probability CC4 is preferred to RMP4 is 0.03
- Interestingly, no covariates help distinguish between Group 1 and 2

Results - Study 3

Conclusion list

Study 3

Group 1 (≈ 56% of the population)	Estimate	SE	Group 2 (≈ 44% of the population)	Estimate	SE
CC5	2.06	0.29	RMP4	1.36	0.32
LR4	1.22	0.21	LR4	1.28	0.36
RMP4	0.97	0.19	RMP3	0	0
SOS4	0.88	0.21	LR3	-0.02	0.3
LR3	0.51	0.2	SOS4	-1.19	0.3
SOS3	0.18	0.21	SOS3	-1.8	0.32
RMP3	0	0	CC5	-2.37	0.48
SP2	-0.31	0.23	SP2	-2.99	0.35
CC1	-1.97	0.29	CC1	-4.27	0.42



- Study 3 resembles Study 2: the two groups tend to agree on the RMP statements and interpret the CC statements relatively different
- Group 1 finds CC5 ("suspect was the source") as the strongest statement as opposed to Group 2
- Group 2 does not differentiate between LR and RMP, Group 1 slightly prefers LR

Key findings

1. In essentially every subpopulation there are consistent rankings within a category ($RMP3 > RMP2 > RMP1$ etc.) but the spaces between the items vary
2. There are differences is the relative standing of categorical and quantitative opinions
3. These suggest there are different kinds of people

Goodness of fit test

- In a real data set how can we decide if a two-cluster model fits better than a one-cluster model?
- Likelihood ratio test (no covariates):

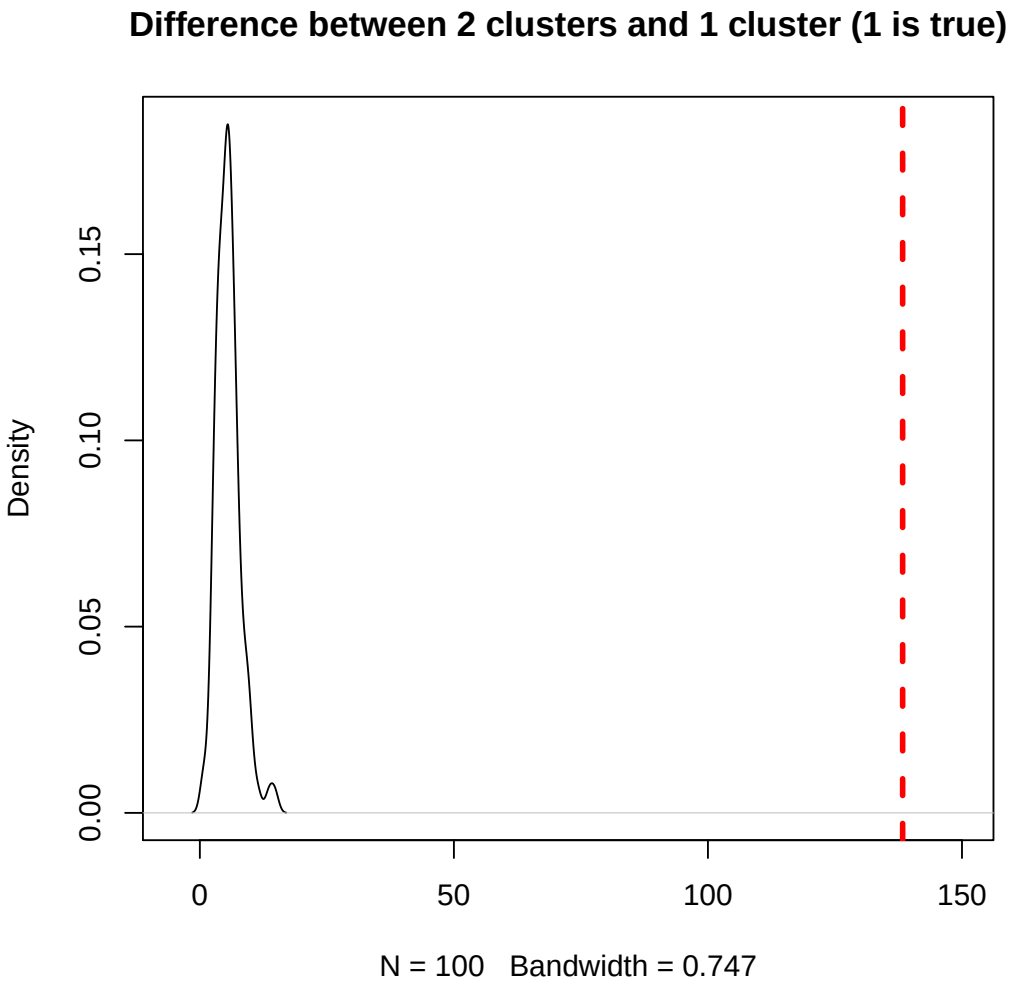
  Likelihood of one-cluster $= -951.5$,

  Likelihood of two-clusters$= -813.2$
- A statistical issue: the Likelihood ratio test does not have the usual null distribution so the P-value can not be calculated the usual way
- One approach to address this issue: conduct a Monte Carlo LR test to determine the number of clusters

Goodness of fit test

- The Monte Carlo LR test was calculated as follows:
  - Data was simulated when the one-cluster model is the true model
  - The two-cluster and the one-cluster models were fitted to the simulated data
  - The likelihoods based on these models were compared
  - This gives an idea of how big a change in the likelihood would be expected by chance
  - This was repeated, the distribution of the difference in likelihoods was calculated and compared to the difference based on the real data

Goodness of fit test



- There is evidence the data did not originate from a one-cluster model

Conclusions

- There is evidence that different subpopulations tend to perceive statements presented by an examiner differently
- Even when solid scientific evidence is presented, it may be interpreted at least by a certain percentage of the population in a way that is different from the way the examiner intended
- Studies 2 and 3 show that the groups seem to agree on the interpretation of RMP statements though one subpopulation prefers categorical statements
- This finding may reflect a general difference among individuals in their ability to make use of quantitative information

Conclusions

- If the population of potential jurors consists of subgroups that respond differently to expert testimony, then forensic scientists may need to explain their findings in multiple ways to assure that their testimony is more broadly understood

Future work - An alternative approach for dealing with heterogeneity

- Classical pair comparison model $\pi_{ij} = F(\lambda_i - \lambda_j)$
- Assumption: given the strength parameters, the results of all comparisons are independent (both within a single participant and across participants)
- It is conceivable that results of several paired comparisons performed by the same subject will not be independent

Future work - An alternative approach for dealing with heterogeneity

- Classic model: $\pi_{ij} = F(\lambda_i - \lambda_j)$
  - Everyone has the same order of λs
- $\pi_{ijk} = F\left[\lambda_i(u_k) - \lambda_j(u_k)\right]$
  - u_k: the cluster membership, $u = 1, \ldots, U$ (u_k is not observed)
  - limited data $\rightarrow$ only z groups
- $\pi_{ijk} = F\left[(\lambda_i - \lambda_j)\alpha_k\right]$
  - The α_k is a factor for an individual k
  - Everyone has the same order of λs but α_k allows for the strength of the opinions to vary
- λ and α can depend on the characteristics of an individual (x_k)

Thank you!