Score-based Likelihood Ratios For Handwriting Evidence

Christopher P. Saunders
Assistant Professor of Statistics
South Dakota State University

Disclaimer

This work was supported in part under a Contract Award from the Counterterrorism and Forensic Science Research Unit of the Federal Bureau of Investigation's Laboratory Division.

Names of commercial manufacturers are provided for information only and inclusion does not imply endorsement by the FBI.

Points of view in this presentation are those of the authors and do not necessarily represent the official position of the FBI or the US Government.

Disclaimer continued

The research detailed in this presentation was supported in part by Award No. 2009-DN-BX-K234 awarded by the National Institute of Justice, of Justice Programs, US Department of Justice.

Background

- In 2008, we began exploring likelihood ratios as a tool for presenting handwriting evidence in a forensic context.
- Focus has been on Score-Based Likelihood Ratios (SLRs)
  - Three general classes of SLRs
  - Currently working on characterizing properties of SLRs

Related Research

This presentation is a summary of results from the fourth in a series of papers related to handwriting evidence.

Saunders, C.P., Davis, L.J., Lamas, A.C., Miller, J.J., Gantz, D.T. (2011) Construction and Evaluation of Classifiers for Forensic Document Analysis. *Annals of Applied Statistics*. 5, 1.

Saunders, C.P., Davis, L.J., Buscaglia, J. (2011) A Comparison between Biometric and Forensic Handwriting Individuality. *Journal of Forensic Sciences* 56, 3.

Davis, L.J., Saunders, C.P., Hepler, A.B., Buscaglia, J. (2012) Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios. *Forensic Science International*. 216 (1-3):146-157.

Hepler, A.B., Saunders, C.P., Davis, L.J., Buscaglia, J. (2012) Score-based likelihood ratios for handwriting evidence. *Forensic Science International*. 219 (1-3):129-140.

Overview

We will discuss:

- Bayesian paradigm for interpreting evidence
- Score-based likelihood ratios as approximations to the value of evidence
- Statistical interpretations of forensic hypotheses and the corresponding classes of SLRs

The Bayesian Paradigm

What do we believe the likelihood of observing the questioned document is if the suspect wrote the questioned document, given what we know about the suspect?

vs.

What do we believe the likelihood of observing the questioned document is if it was written by a writer in the alternative source population, given what we know about the alternative source population?

- **In a formal Bayesian paradigm, this is the comparison we have to address as forensic statisticians with respect to handwriting identification.**

Evaluation of Handwriting Evidence

$$\underbrace{\frac{P(H_p|E,I)}{P(H_d|E,I)}}_{\text{Posterior Odds}} = \underbrace{\frac{P(E|H_p,I)}{P(E|H_d,I)}}_{\text{Likelihood Ratio and/or Bayes Factor}} \times \underbrace{\frac{P(H_p|I)}{P(H_d|I)}}_{\text{Prior Odds}}$$

E : Evidence

H_p : Suspect wrote the questioned document (QD)

H_d : Suspect did not write the QD

I : Background information

"The evidence is LR (BF) = 100 times more probable if the suspect wrote the QD than if some unknown person wrote it."

Evidence

We partition the evidence: $E = \{E_S, E_U, E_A\}$

where:

$E_S =$ Sample(s) obtained from the **S**pecific source

$E_U =$ Sample(s) of **U**nknown source obtained at the crime scene.

$E_A =$ Sample(s) taken from **A**lternative sources, comprising a relevant database of other potential sources.

Handwriting Assumptions

We assume:

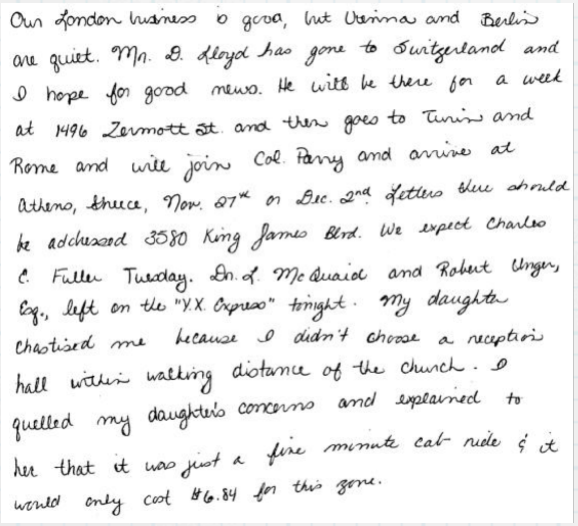
- Every individual has an unobservable ***writing profile***, that is constant over time.
- A document is a random sample from a writing profile.

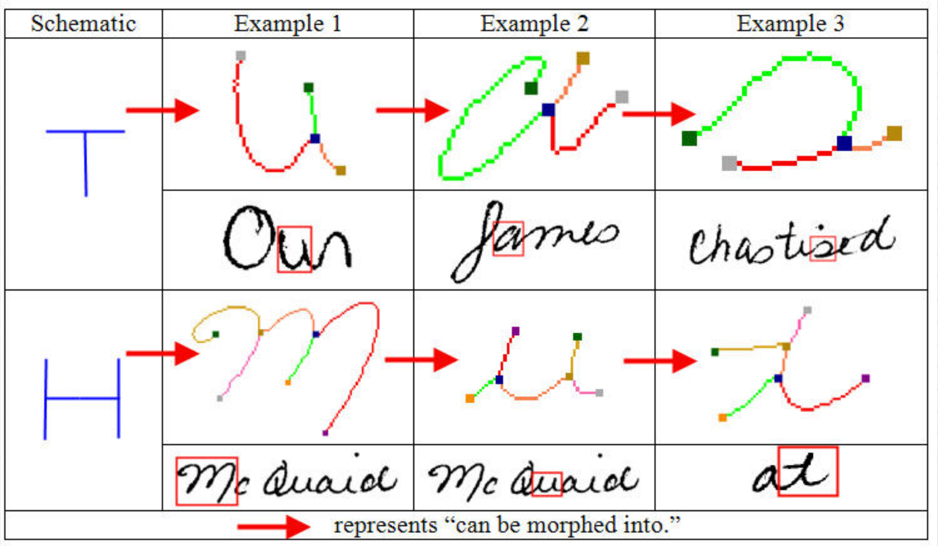
Estimate the profile by collecting many documents, forming a ***writing template*** (or simply ***template***).

Quantification

- A proprietary, automated process developed by Gannon Technologies Group processes handwritten documents.

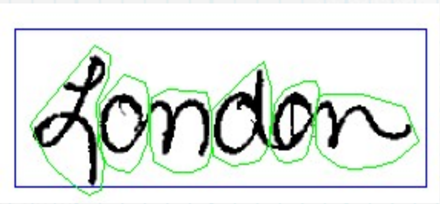
- This process uniquely associates each parsed character’s skeleton with a ***graphical isomorphism***.

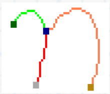

Our London business is good, but Vienna and Berlin
are quiet. Mr. D. Lloyd has gone to Switzerland and
I hope for good news. He will be there for a week
at 1496 Zermott St. and then goes to Turin and
Rome and will join Col. Parry and arrive at
Athens, Greece, Nov. 27th or Dec. 2nd. Letters there should
be addressed 3580 King James Blvd. We expect Charles
E. Fuller Tuesday. Dr. L. McQuaid and Robert Unger,
Esq., left on the "Y.X. Express" tonight. My daughter
chastised me because I didn't choose a reception
hall within walking distance of the church. I
quelled my daughter's concerns and explained to
her that it was just a five minute cab ride & it
would only cost $6.84 for this zone.



Quantification, cont.

Obtain ***measurements*** from E_s, E_u, and each alternate writer's template in E_a.





Shape = $\boldsymbol{n}$
Topology = ⊤

The Gannon Technologies Group

Accumulate across characters

Frequency Distribution of Letter/Isocode Usage in a Single Writing Sample

	Topology 1	Topology 2	• • •
1			
• • •			
9			
A			
• • •			
Z			
• • •			
a			
• • •			
z			

$\mathbf{X} =$ matrix of counts (MOC)

Likelihood Ratio or Bayes Factor

$$\frac{P(E|H_p, I)}{P(E|H_d, I)} = \frac{P(\mathbf{X}_S, \mathbf{X}_U | H_p, I)}{P(\mathbf{X}_S, \mathbf{X}_U | H_d, \mathbf{X}_A, I)}$$

$\mathbf{X}_S$ = MOC obtained from the **S**uspect's template

$\mathbf{X}_U$ = MOC of **U**nknown source (QD)

$\mathbf{X}_A$ = Collection of MOCs taken from **A**lternative sources, comprising a relevant database of other potential suspects.

Score Based Likelihood Ratios

- A score-based approach would reduce the dimensionality of the problem.
- The score-based likelihood ratio (*SLR*) is

$$SLR = \frac{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_p, I\right)}{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_d, X_A, I\right)}$$

where Δ is a dissimilarity score between two MOCs.

Generic SLR Algorithm

1. ***Select a statistic***, Δ, to assess dissimilarity between two items of evidence.

2. ***Develop a database*** of Δ's for items of known origin, under H_p and H_d.

3. ***Construct an empirical distribution*** for Δ, under H_p and H_d.

4. ***Evaluate*** a specific δ for two items of evidence.

5. ***Consider in context***: evaluate the numerator and denominator distributions from (3) at the point δ.

Adapted from Aitken & Taroni (2004) ***Statistics and the evaluation of evidence for forensic scientists***. Wiley and Sons.

Statistical vs. Forensic Propositions

Forensic propositions:

H_p: Suspect (S) wrote the QD.

H_d: S did not write the QD.

Statistical propositions:

H_p: Evidence score arises from the distribution of scores obtained by pairing a randomly selected QD and a randomly selected template ***both written by S***.

Statistical propositions, cont.

Crime scene anchored:

$\boldsymbol{H_{d1}}$: Evidence score arises from the distribution of scores obtained by pairing ***the QD*** with ***a template*** written by a random individual.

Statistical propositions, cont.

Crime scene anchored:

H_{d1}: Evidence score arises from the distribution of scores obtained by pairing ***the QD*** with ***a template*** written by a random individual.

Suspect anchored:

H_{d2}: Evidence score arises from the distribution of scores obtained by pairing ***a QD*** written by a random individual with ***the template*** written by S.

Statistical propositions, cont.

Crime scene anchored:

$\boldsymbol{H_{d1}}$: Evidence score arises from the distribution of scores obtained by pairing ***the QD*** with ***a template*** written by a random individual.

Suspect anchored:

$\boldsymbol{H_{d2}}$: Evidence score arises from the distribution of scores obtained by pairing ***a QD*** written by a random individual with ***the template*** written by *S*.

General match:

$\boldsymbol{H_{d3}}$: Evidence score is a realization from the distribution of scores obtained by pairing ***a QD*** written by a random individual with ***a template*** from a <u>different</u> random individual.

Specific SLR Algorithm

1. ***Kullback-Leibler*** (KL) divergence
   - Obtain KL, by letter, between a QD and a template.
   - The overall dissimilarity is then a weighted average, over all letters.
     - Weighting accounts for differing numbers of observed characters across letters. (Letter ‘a’ is used more than ‘z’.)
   - One advantage of the KL is that it is non-symmetric:
     - QD is typically much smaller than the template.

   - Many other choices exist – additional research needed.

Specific SLR Algorithm

2. ***Database of scores: Numerator***
   - Under H_p, QD and template written by suspect
   - Assume we have a collection of prior writings obtained from the suspect – ***suspect's template***.
   - Ideally, we would also have a collection of samples from the suspect 'similar' to the QD (e.g. same number of characters).
   - Then, to generate our database, we would compute the KL between each of the 'QD-type' documents and the suspect's template.

Specific SLR Algorithm

2. **Database of scores: Numerator**
   - Unfortunately, we typically do not have access to a sufficient number of 'QD-type' documents written by the suspect.

A subsampling routine:

A. ***Create:***

   Pseudo-QD: randomly select n characters from template
   (n = # of chars in QD)

   Pseudo-template: remaining characters

B. ***Compute KL*** between pseudo-QD and pseudo-template.

C. ***Repeat A, B*** many times to generate database of scores.

Specific SLR Algorithm

Database of scores: Denominator

- Assume we have access to a collection of templates from a large number of alternative sources.
- One approach would be to compute KL between the actual QD and each template in the collection.

Histogram Estimator *as empirical distribution*.

- Many other choices – additional research needed.

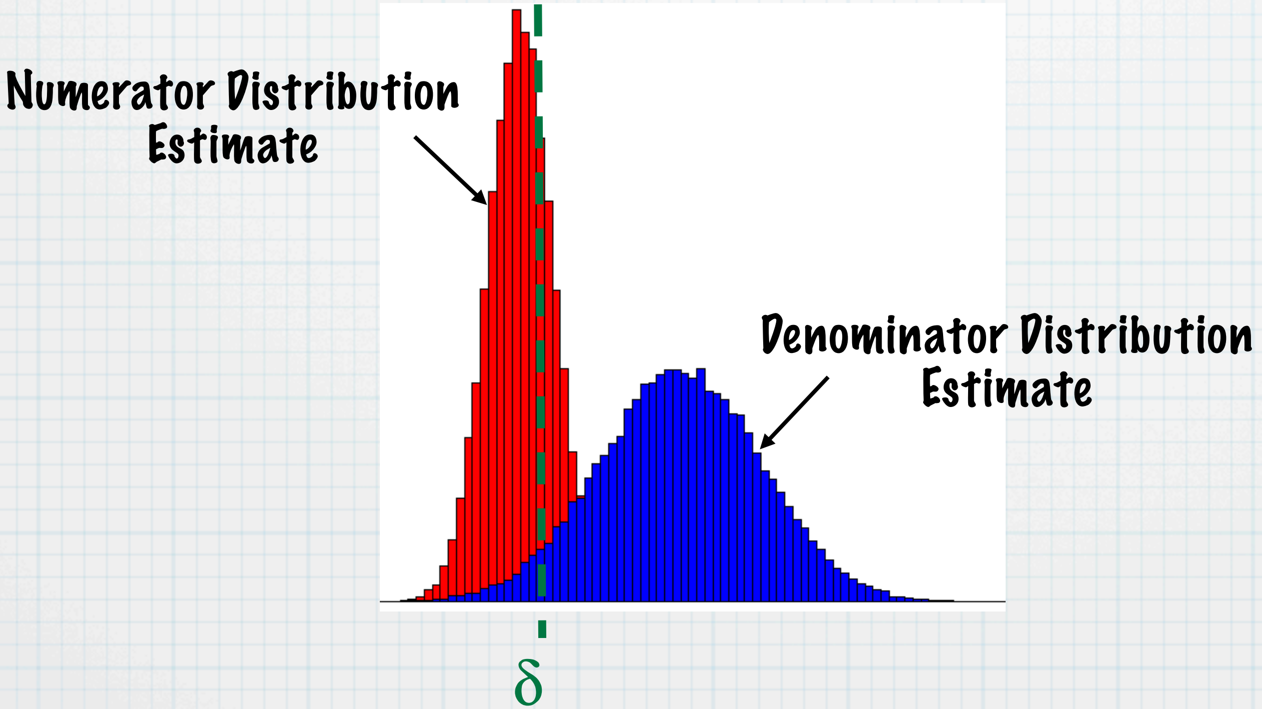
Evaluate

$$\Delta\left(\mathbf{X}_S, \mathbf{X}_U\right) = \delta.$$

Specific SLR Algorithm, cont.

Consider in context:

$$\hat{SLR} = \frac{\hat{g}\left(\Delta(\mathbf{X}_S,\mathbf{X}_U) = \delta \,\middle|\, H_p, I\right)}{\hat{g}\left(\Delta(\mathbf{X}_S,\mathbf{X}_U) = \delta \,\middle|\, H_d, \mathbf{X}_A, I\right)}$$



Calculating SLR for handwriting

SLR numerator:

- Obtain $(\mathbf{X}_{Si}, \mathbf{X}_{Ui})$ for $i = 1, \ldots, 500$ ***pseudo-QDs*** obtained via ***subsampling*** (AAFS09) from the suspect's template.

- Obtain $\Delta(\mathbf{X}_{Si}, \mathbf{X}_{Ui})$ for each pseudo-QD.

- Estimate the numerator distribution: $\hat{g}$.

- Evaluate $\hat{g}\left(\Delta\left(\mathbf{X}_S, \mathbf{X}_U\right) \middle| H_p, I\right)$.

Calculating SLR for handwriting

SLR denominator 1 - crime scene anchored:

- Obtain $\mathbf{X}_{Ai}$ from templates taken from each writer i in E_A.
- Obtain $\Delta(\mathbf{X}_{Ai}, \mathbf{X}_U)$ for each writer i in E_A, where $\mathbf{X}_U$ from the QD.
- Estimate the denominator distribution: $\hat{g}$.
- Evaluate

$$\hat{g}\left(\Delta\left(\mathbf{X}_S, \mathbf{X}_U\right) \middle| H_d, \mathbf{X}_A, I\right).$$

Calculating SLR for handwriting

SLR denominator 2 - suspect anchored:

- Obtain X_{Ui} from 500 pseudo-QD's sampled from E_A.
- Obtain $\Delta(X_S, X_{Ui})$ for each pseudo-QD, where X_S is obtained from suspect's template.
- Estimate the denominator distribution: $\hat{g}$.
- Evaluate

$$\hat{g}\left(\Delta\left(\mathbf{X}_S, \mathbf{X}_U\right) \middle| H_d, \mathbf{X}_A, I\right).$$

Calculating SLR for handwriting

SLR denominator 3 – general match:

- Obtain X_{Ai} from 500 templates sampled from E_A.
- For each i, obtain X_{Ui} by randomly selecting a pseudo-QD from E_A, ensuring that writer of $X_{Ui} \neq$ writer of X_{Ai}.
- Obtain $\Delta(X_{Ai}, X_{Ui})$.
- Estimate the denominator distribution: $\hat{g}$.
- Evaluate

$$\hat{g}\left(\Delta\left(\mathbf{X}_S, \mathbf{X}_U\right) \middle| H_d, \mathbf{X}_A, I\right).$$

Results: Specific Cases

QD = first 60 characters of document

H_p True

S ID:	0799	0772
SLR_1	1858	2370
SLR_2	1701	6
SLR_3	15	19

H_d True

S ID / True ID:	0273 / 0595	0797 / 0110
SLR_1	0.00006	0.9517
SLR_2	1.002	6.98
SLR_3	0.005	3.503

Implications

- SLR's are starting to be introduced in US legal system with respect to fingerprints.
- Likelihood ratios in general are very hard to defend against.

However-

To date we have empirically demonstrated that the three standard SLR's give different values of the Evidence.

Why are the SLRs different?

- SLR ≠ LR.
  - By replacing E_u and E_s with $\Delta\left(X_s, X_u\right)$ we are losing information.

$$SLR = \frac{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_p, I\right)}{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_d, I\right)}$$

$$LR = \frac{\Pr\left(X_s, X_u \middle| H_p, I\right)}{\Pr\left(X_s, X_u \middle| H_d, I\right)}$$

Why are the SLRs different?

- $SLR_1 \neq SLR_2 \neq SLR_3$.
  - The conditioning arguments are different for each SLR-

Crime scene anchored:

$$SLR_1 = \frac{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_p, I\right)}{\Pr\left(\Delta\left(X_a, x_u\right) = \delta \middle| H_d, I\right)}$$

Suspect anchored:

$$SLR_2 = \frac{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_p, I\right)}{\Pr\left(\Delta\left(x_s, X_a\right) = \delta \middle| H_d, I\right)}$$

General match:

$$SLR_3 = \frac{\Pr\left(\Delta\left(X_s, X_u\right) = \delta \middle| H_p, I\right)}{\Pr\left(\Delta\left(X_{a1}, X_{a2}\right) = \delta \middle| H_d, I\right)}$$

Implications

- SLR ≠ LR.
- It is not a straightforward task to statistically interpret forensic propositions.
- Different statistical interpretations can lead to very different conclusions.

Acknowledgements

- South Dakota State University
  - Douglas Armstrong
- The IC Post Doctoral Fellowship Program
  - Amanda Hepler
  - Eric Kalendra
- George Mason University
  - Linda Davis
  - John J. Miller
  - Don Gantz
- Gannon Technologies Group
- FBI, Counterterrorism and Forensic Science Research Unit
  - JoAnn Buscaglia
- National Institute of Justice

Additional Slides

- A Normal Digression
- Comparative Study- “FBI 500”
- Sub Sampling Data Bases
- Bayes Factors

Please see Helper et al. for details on these slides-

Hepler, A.B., Saunders, C.P., Davis, L.J., Buscaglia, J. (2012) Score-based likelihood ratios for handwriting evidence. *Forensic Science International*. 219 (1-3):129-140.

A Normal Digression

In general, it is technically difficult to solve for the exact distributions under the three stated statistical propositions.

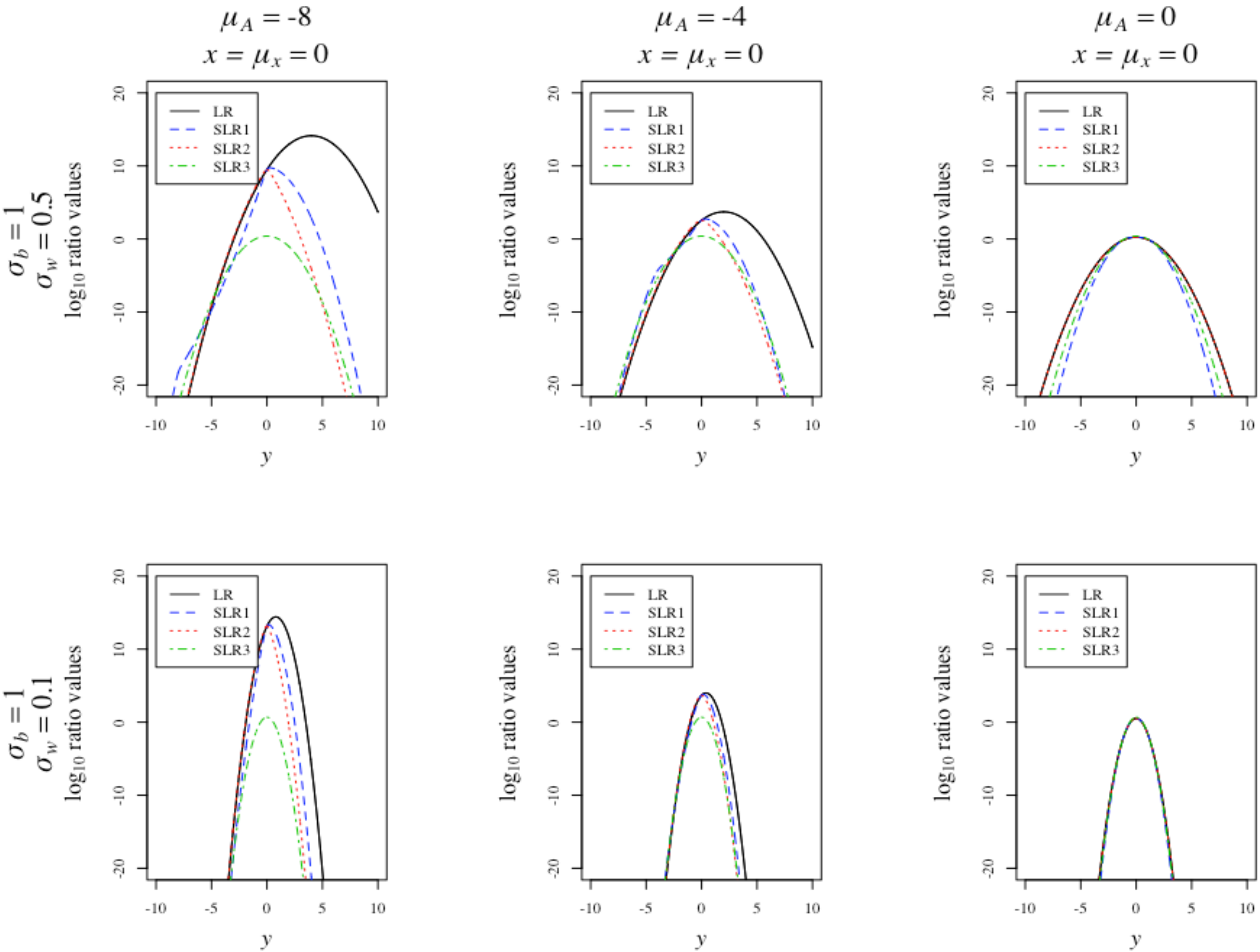
- If we use squared distance between the normal random variables *we can solve for the three SLR's in closed form.*

- Under H_p; Assume that: $X \sim N\left(\mu_x, \sigma_w^2\right)$

  $Y \sim N\left(\mu_x, \sigma_w^2\right)$

- Under H_d; Assume that: $X \sim N\left(\mu_x, \sigma_w^2\right)$

  $Y \sim N\left(\mu_A, \sigma_A^2\right)$



$\mu_A = -8$
$x = \mu_x = 0$
$\mu_A = -4$
$x = \mu_x = 0$
$\mu_A = 0$
$x = \mu_x = 0$
$\sigma_b = 1$
$\sigma_w = 0.5$
$\sigma_b = 1$
$\sigma_w = 0.1$
$\log_{10}$ ratio values
y
LR
SLR1
SLR2
SLR3
-10
-5
0
5
10
-20
20

Comparative study

- Alternative source writing samples: FBI provided 2120 script documents (a convenience sample)
  - 5 documents from 424 individuals
  - Each document has approximately 550 characters
  - Text selected to be representative of English language
- Randomly select a writer to serve as ‘suspect.’
- Obtain SLR1-3 values for both scenarios.
  - H_p True: S wrote QD
  - H_d True: S did not write QD
- Repeat 1000 times.

Results: H_p True

	Agreement			Disagreement
	Supports H_p	*Inconclusive*	*Supports H_d*	
	$SLR > 10$	$1/10 < SLR < 10$	$SLR < 1/10$	
1 vs 2 vs 3	**58%**	**4%**	**1%**	**37%**
1 vs 2	63%	5%	1%	31%
1 vs 3	67%	4%	1%	28%
2 vs 3	60%	24%	1%	17%

Correct

QD = first 60 characters of document

Results: H_d True

	Agreement			Disagreement
	Supports H_p	*Inconclusive*	*Supports H_d*	
	SLR > 10	1/10 < *SLR* < 10	*SLR* < 1/10	
1 vs 2 vs 3	**0%**	**16%**	**69%**	**14%**
1 vs 2	0%	18%	72%	10%
1 vs 3	0%	18%	72%	9%
2 vs 3	1%	19%	72%	8%
			Correct	

QD = first 60 characters of document

Results: Rates of Misleading Evidence

H_p True: Suspect wrote the QD

LR Range	QD = 20 Characters SLR_1	QD = 20 Characters SLR_2	QD = 20 Characters SLR_3	QD = 60 Characters SLR_1	QD = 60 Characters SLR_2	QD = 60 Characters SLR_3
(0, 0.001]	0.002	0.002	0.002	0.002	0.006	0.007
(0.001, 0.01]	0.007	0.007	0.008	0.006	0.011	0.007
(0.01, 0.1]	0.004	0.007	0.013	0.000	0.003	0.005
(0.1, 1]	<u>0.074</u>	<u>0.117</u>	<u>0.086</u>	<u>0.025</u>	<u>0.060</u>	<u>0.044</u>
RME	**0.087**	**0.133**	**0.109**	**0.033**	**0.08**	**0.063**
(1, 10]	0.235	0.446	0.403	0.040	0.279	0.255
(10, 100]	0.106	0.114	0.142	0.033	0.171	0.184
(100, 1000]	0.209	0.137	0.122	0.164	0.063	0.054
>1000	0.363	0.170	0.224	0.730	0.407	0.444

Misleading

Results: Rates of Misleading Evidence

H_d True: Suspect did not write the QD

LR Range	QD = 20 Characters SLR_1	SLR_2	SLR_3	QD = 60 Characters SLR_1	SLR_2	SLR_3
(0, 0.001]	0.195	0.139	0.148	0.493	0.354	0.401
(0.001, 0.01]	0.114	0.177	0.168	0.227	0.355	0.310
(0.01, 0.1]	0.093	0.050	0.060	0.042	0.047	0.039
(0.1, 1]	0.372	0.377	0.343	0.145	0.144	0.139
(1, 10]	0.201	0.239	0.240	0.079	0.084	0.089
(10, 100]	0.010	0.011	0.025	0.003	0.010	0.012
(100, 1000]	0.005	0.003	0.002	0.004	0.002	0.001
>1000	0.010	0.004	0.014	0.007	0.004	0.009
RME	**0.226**	**0.257**	**0.281**	**0.093**	**0.100**	**0.111**

Misleading

Subsampling Algorithm- Suspect Specific database

For a questioned document of size *n* characters and a suspect writer template of *N*, we then:

- Randomly divide the template into two sets of characters of sizes n and N-n.
  - These are the pseudo-document and pseudo-writing template, respectively.
- Compare the pseudo-document and pseudo-writing template with the score function.
- Record the similarity score

Repeat *k* times to obtain a dataset of size *k*

Parametric Approach

- Conditioning on the letter, we can assume the corresponding row follows a multinomial distribution.
- There are thousands of topologies that can be assigned to a given letter – each row could have thousands of cells.
  - This leads to some ambiguity when specifying the priors necessary for computing the Bayes factor.
- We have explored these issues with limited success.

A note on the Bayes Factor

$$\underbrace{\frac{P\left(E \middle| H_p, I\right)}{P\left(E \middle| H_d, I\right)}}_{\textit{Bayes Factor}} = \frac{P\left(E_S, E_U, E_A \middle| H_p, I\right)}{P\left(E_S, E_U, E_A \middle| H_d, I\right)}$$

$$= \frac{P\left(E_U, E_S \middle| H_p, I\right)}{P\left(E_S \middle| H_d, I\right)} \frac{P\left(E_A \middle| H_p, I\right)}{P\left(E_S, E_U, E_A \middle| H_d, I\right)}$$

$$= \frac{P\left(E_U \middle| E_S, H_p, I\right)}{P\left(E_U \middle| E_A, H_d, I\right)}$$

Theoretical Example

It is not a straightforward task to statistically interpret forensic propositions.

In the traditional setting we have-

$$LR = \frac{f\left(x, y \mid H_p, I\right)}{f\left(x, y \mid H_d, I\right)} = \frac{f\left(y \mid x, H_p, I\right)}{f\left(y \mid H_d, I\right)}$$

Compared with the SLR-

$$\mathrm{SLR} = \frac{g(\Delta(x,y) \mid H_p, I)}{g(\Delta(x,y) \mid H_d, I)} = \frac{\int g(\Delta(x,y) \mid x, H_p, I) f(x \mid H_p, I)\,dx}{\int g(\Delta(x,y) \mid x, H_d, I) f(x \mid H_d, I)\,dx},$$