The Role of Databases in Forensic Science

Karen Kafadar

Department of Statistics

University of Virginia

kkafadar@virginia.edu

http://www.stat.virginia.edu

OUTLINE

1. Purposes of Databases
2. Method Development (Sufficiency: Realistic examples)
3. Method Validation (Representativeness)
4. Method Implementation (Completeness)
5. Illustrations
6. Summary

Statistics & Data

Science of analyzing data, characterizing uncertainties

- **Biology**: extinction/abundance of species; characterizing genetic expression (millions of SNPs) in response to stimuli; associating genotypes with phenotypes
- **Physics**: data analysis of high-energy physics (HEP) experiments to discover new particles; estimating 'big G' with uncertainty; existence of global warming
- **Engineering**: product design & development; nuclear safety programs; production efficiency
- **Medicine**: clinical trials of new drugs; evaluation of treatment and screening programs; estimating disease prevalence, incidence, spread

1. Purposes of Databases

- Develop methods
- Validate methods
- Implement methods: Reference Database (Exemplars)
- Information sharing
- Identify shortcomings
- Improve methods

2. Method development

DNA (NRC-2, 1996):

- Identification of 13 markers (presumed independent)
- Assure ability to separate "signal" peaks (allele identification) from noise
- Identify challenges: resolving mixtures; lab errors

Latent Print Analysis (NIST SD-27a; Neumann, JRSS-B 2012):

- Assumes pre-selected minutiae: **distinctive, specific**
- Calculate metrics among features (minutiae)
- Calculate "likelihood ratio"

"Proof of concept": Does not require representativeness

2. Method Validation

- *Sensitivity*: Given two specimens from *same source*, how likely does the method claim "same source"?
- *Specificity*: Given two specimens from *different sources*, how likely does the method claim "different source"?

Note: Not the questions of practical interest:

- PPV: If analysis on two specimens concludes *"same source"*, were the two specimens really from same source?
- NPV: If analysis on two specimens concludes *different sources*, were the two specimens from "different sources"?

In real life, *we will never know for sure.*
(Even DNA analysis has uncertainty – but very tiny.)

For validation: Need representative data base

- Estimate distribution of genotypes (DNA), features (latents)
- Address unsolved challenges:
  resolving mixtures; allelic drop-out (DNA)
  overlapping prints (latents)
- Improve analysis process:
  Minimize lab process and measurement

3. Implement methods: Reference Database

- **Completeness**: Does database have full set of all DNA signatures, latent prints?
- If so: Need good search algorithms
- If not: May end up selecting "nearest match" (but wrong)

A miss is as good as a mile.

4. Example: CBLA

- Crime → evidence → bullets
- Gun recovered: match striations on bullet and gun barrell (separate NRC committee)
- *No gun*: **Comparative Bullet Lead Analysis** (CBLA)
- "Working hypothesis": chemical concentration of lead used to make "batch" of bullets provides "unique signature" ⇒ "equal" concentrations of elements in Crime Scene (CS) bullets and Potential Suspect (PS) bullets may indicate "guilt"
- FBI measures (in triplicate) concentrations of 7 elements (As, Sb, Sn, Bi, Cu, Ag, Cd); "analytically indistinguishable concentrations" in CS & PS bullets if "mean ± 2·SD intervals overlap for *all* 7 elements"

What went wrong?

- Statistical procedure
- Validation on "1837-bullet database": "*one specimen from each combination of bullet caliber, style, and nominal alloy class was* **selected**" for database; found 693 "matches" out of (1837·1836/2) = 1,686,366 pairs of bullets
- FBI **selected** 1837 bullets to be as **different** as possible
- 1837-bullet set = FBI's attempt at different "melts"
- Only 854 of 1837 had all 7 elements (1997 or later)

FBI "Notes on 1837-bullet data set"

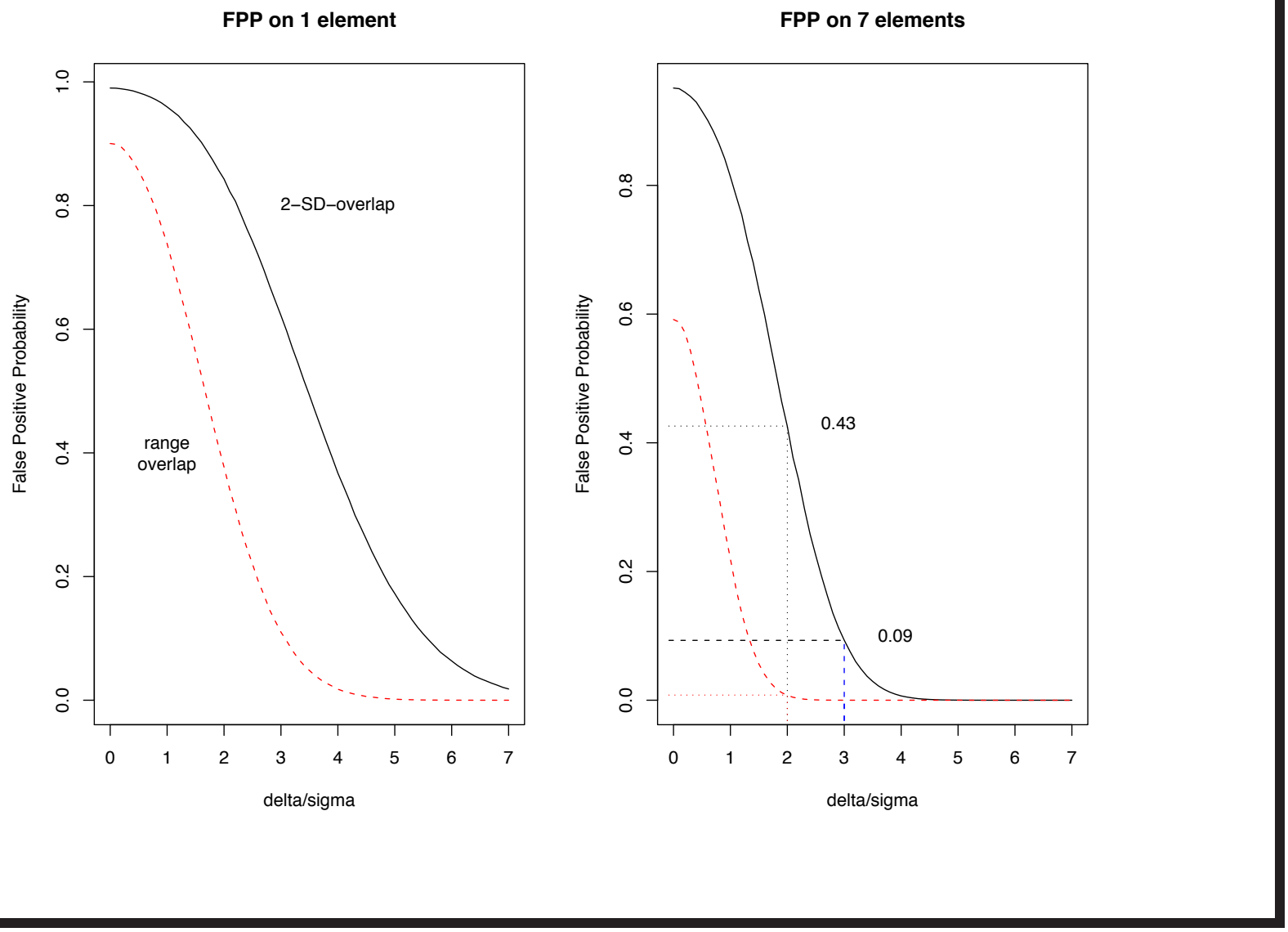
"To assure independence of samples, the number of samples in the full database was reduced by removing multiple bullets from a given known source in each case. To do this, evidentiary submissions were considered one case at a time. For each case, one specimen from each combination of bullet caliber, style, and nominal alloy class was **selected** *and that data was placed into the test sample set. In instances where two or more bullets in a case had the same nominal alloy class, one sample was randomly selected from those containing the maximum number of elements measured. ... The test set in this study, therefore, should represent an unbiased sample in the sense that each known production source of lead is represented by only one randomly selected specimen."*

- FBI used it to estimate FPP=False Positive Probability: 693 2-SD-overlap "matches" among 1,686,366 comparisons $\Rightarrow$ "about 1 in 2500"
- NRC Committee: This FPP (1 in 2500) is not valid
- 1837-bullet data set is not a random sample:
- Cochran, Mosteller, Tukey (1954), "Principles of Sampling"
- FBI study: 4 boxes (50 each) from 4 manufacturers; only 1 (Federal) had 6 of the 7 elements
- Simulation: Pooled standard deviations and estimated correlations among elements to calculate realistic error rates

Weighing the Evidence: Forensic Analysis of Bullet Lead, 2004

Sample correlation matrix: Federal bullets

	As	Sb	Sn	Bi	Cu	Ag	(Cd)
As	1.000	0.320	0.222	0.236	0.420	0.215	0.000
Sb	0.320	1.000	0.390	0.304	0.635	0.242	0.000
Sn	0.222	0.390	1.000	0.163	0.440	0.154	0.000
Bi	0.236	0.304	0.163	1.000	0.240	0.179	0.000
Cu	0.420	0.635	0.440	0.240	1.000	0.251	0.000
Ag	0.215	0.242	0.154	0.179	0.251	1.000	0.000
(Cd)	0.000	0.000	0.000	0.000	0.000	0.000	1.000


FPP on 1 element
False Positive Probability
0.0
0.2
0.4
0.6
0.8
1.0
2−SD−overlap
range
overlap
0
1
2
3
4
5
6
7
delta/sigma
FPP on 7 elements
False Positive Probability
0.0
0.2
0.4
0.6
0.8
0.43
0.09
0
1
2
3
4
5
6
7
delta/sigma

Using FBI "2-SD-match" criterion:
How often do bullets from different boxes "match"?

Ex: CCI bullets – 4 boxes, 50 bullets per box

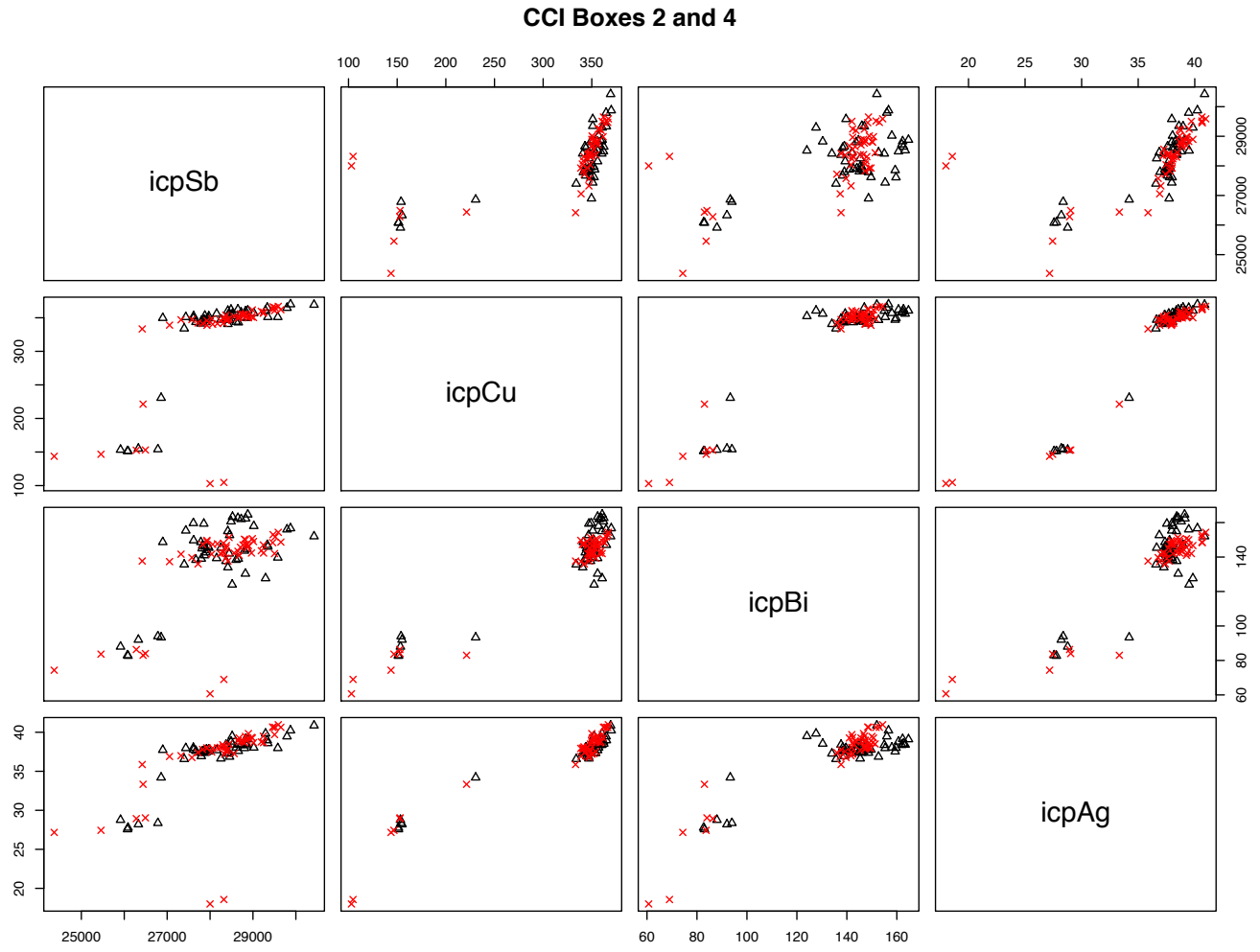
Sometimes FBI-"matches" are rare:

- Box 1 with Box 2: Bullet 45(1) "matches" Bullet 93(2)
- Box 1 with Box 3: None
- Box 1 with Box 4: Bullet 45(1) "matches" Bullet 194(4)

Sometimes frequent:

- Box 2 with Box 4: 1092 "matches"!
  $(50 \times 50 = 2500$ comparisons)

Consequences of Non-representative database:
Wrong error rates, Missed sources of variability


CCI Boxes 2 and 4
icpSb
icpCu
icpBi
icpAg
100 150 200 250 300 350
20 25 30 35 40
25000 27000 29000
100 200 300
60 80 100 140
20 25 30 35 40
25000 27000 29000
60 80 100 120 140 160

After report: "70,000 bullets" (56,260 records, 17,572 bullets)

"Resurrected" measurements:

Find Bullet #4 in 1837-bullet (ave, sd) data file in
"Full Database": Bullet Q67 (normalized to NIST S2416):

Case	year		As	Sb	Sn	Bi	Cu	Ag	Cd
2	1989	Ave	0.01260	2.37710	NA	0.0233	0.0596	0.00384	NA
		SD	0.00077	0.04110	NA	0.0006	0.0012	0.00014	NA

Case	year	bullet	As	Sb	Sn	Bi	Cu	Ag	Cd
2	1989	Q67A	NA	2.39388	NA	0.02392	0.06071	0.00400	NA
2	1989	Q67B	NA	2.33020	NA	0.02332	0.05968	0.00377	NA
2	1989	Q67C	NA	2.40718	NA	0.02262	0.05841	0.00375	NA

From where did the As measurement come?

4. Example: Anthrax

Sep-Oct 2001: Anthrax letters mailed to NYC (ABC, CBS, NBC*, NYPost*), FL (AMI), DC (Daschle*, Leahy*)

- 4 morphotypes of specific anthrax *Ames* strain found in Leahy* letter (A1, A3, D, E)
- 5 assays (present/absent); 2 for D (D_M, D_I)
- Feb'02: FBI subpoenas labs for samples of *B. anthracis-Ames*
- 1,070 samples in FBI Repository, *believed* complete
- "Smoking gun": Only 8 samples showed all 4 morphotypes; 7 from one lab at USAMRIID, 8th sent to BMI from that lab
- Inference: "Anthrax came from that lab"

"Statistics means never having to say you're certain"

- 1,070 samples came from 20 labs (17 U.S.)
- 11 samples not viable $\Rightarrow$ 1,059
- Lab-to-lab variation since "D" assayed by 2 labs
  $\Rightarrow$ Concordance: 975/1059 = 0.921 (0.903, 0.937) (not 1.000)
- Ignored D_I for vague reasons
- 947 samples had "conclusive" measurements A1,A3,D_M,E
- One suspect sample assayed 30 times $\Rightarrow$ measurement varibility: 16 of 30 reps showed all 4 morphotypes
- Dilution studies: sudden "appearance" of morphotype at higher dilution rates after disappearance at lower dilution rates

Distribution of #samples by Lab:

F	S	N	P	T	G	E	H	Q	A
598	74	62	50	49	31	24	18	15	6

J	K	I	M	O	R	B	C	D	L	F*
4	3	2	2	2	2	1	1	1	1	1

One Lab F submitted 598 samples (63%)

$\Rightarrow$ P{7 or 8 from Lab F} = 0.14 (hypergeometric distn)

Not an everyday occurrence, but certainly not rare.

Summary

Role of Databases

- For development: Realistic samples
- For validation: Representative of populations
- For implementation: Completeness

Involve Statisticians

- Recognize uncertainty
- Design experiments
- Validate methods
- Characterize "representativeness" of data

References

Box GEP; Hunter WF; Hunter JS (2005), *Statistics for Experimenters, 2nd ed.*, Wiley.

Cochran W; Mosteller F; Tukey JW (1954): Principles of Sampling, *JASA*

Haber L; Haber RN (2008): Scientific validation of fingerprint evidence under Daubert, *Law, Probability, and Risk* 7(2):87–109.

Kafadar K (2014): Statistical Issues in Assessing Forensic Evidence, *International Statistical Review*

Mearns GS: The NAS report: In pursuit of justice, *Fordham Urban Law Journal*, December 2010.

National Research Council (1996): *The Evaluation of DNA Evidence*, National Academies Press, ISBN-10: 0-309-12194-9.

National Research Council (2004): *Forensic Analysis: Weighing Bullet Lead Evidence*, National Academies Press, ISBN-10: 0-309-09079-2.

National Research Council (2008): *Ballistic Imaging*, National Academies Press, ISBN-10: 0-309-11724-0.

National Research Council (2009): *Strengthening Forensic Science in the United States: A Path Forward*, National Academies Press, ISBN-10: 0-309-13135-9.

Neumann C; Evett IW; Skerrett J (2011), Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, *J Royal Statistical Society A*175(2), 1–26.

Peskin, AP; Kafadar K: A new measurement of quality for minutiae in latent fingerprints, preprint.

Spiegelman CS; Kafadar K (2006), Data Integrity and the Scientific Method: The Case of Bullet Lead Data as Forensic Evidence, *Chance* 19(2):17–25.

Ulery BT; Hicklin RA; Buscaglia J; Roberts MA (2011), Accuracy and reliability of forensic latent fingerprint decisions, *Procedings of the National Academy of Sciences* 108(19), 7733–7738.

Ulery BT; Hicklin RA; Buscaglia J; Roberts MA (2012), Repeatability and reproducibility of decisions by latent fingerprint examiners, *PLoS ONE* 7(3).

Yoon S; Liu E; Jain AK (2012), "On latent fingerprint image quality," *Proceedings of the Fifth International Workshop on Computational Forensics*, Tsukuba, Japan, 11 Nov 2012.

Zabell SL (2006), Fingerprint Evidence, *Journal of Law & Policy*, 143–179