- Technical Colloquium: Quantifying the Weight of Forensic Evidence
- May 5, 2016, NIST, Gaithersburg, MD

THE WEIGHT OF EVIDENCE IN LAW, STATISTICS, AND FORENSIC SCIENCE

David H Kaye
Penn State Law

WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic scientists speak about WoE

Weight versus Admissibility

Whether the spill or the initially mislabeled autoradiograph affected the reliability of the test is a question of fact. Alleged infirmities in the performance of a test usually go to the weight of the evidence, not to its admissibility. … [T]he irregularities which occurred here do not warrant … exclusion.

- State v. Kalakosky, 852 P.2d 1064, 1073-74 (Wash. 1993)

Weight vs Relevance

Evidence is relevant if:

(a) it has any tendency to make a fact more or less probable than it would be without the evidence; and

(b) the fact is of consequence in determining the action.

P(H|E) <> P(H)

- Fed R Evid 401

Weight = Probative Value

The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.

- Fed R Evid 403

Weight vs Sufficiency

But your honor, the verdict is against the weight of the evidence!

E.g., Pa. R. Crim. P. 607. ("A claim that the verdict was against the weight of the evidence shall be raised with the trial judge in a motion for a new trial")

Probative value = weight

Relevance

- Change in probability

Probative value

- Strength-weight

Sufficiency

- Burden of persuasion

Measures of Probative Value Proposed in Legal Literature

$P(H|E) - P(H)$

- Cullison 1969; Gerjuoy 1977; Friedman 1986

$P(E|H) / P(E|{\sim}H)$

- Lempert 1977?; Kaye 1986; Kaye & Koehler 2003

$P(E|H) - P(E|{\sim}H)$

- Davis & Follette 2002

WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic scientists speak about WoE

Theories of Inference

A statistical problem

- Data (x) are observed
- Probability model $P_\theta(x) = f(x|\theta)$ for generating x has unknown values for its parameters θ
- Having observed x, what can we say about θ?

Three Approaches to Inference

Frequentist

Fisher, Neymann, Pearson

Likelihoodist

Barnard 1949; Edwards 1972; Royall 1997

Bayesian

Bayes 1764

Frequentist

N-P hypothesis tests yield a decision

- A binary measure of the weight or strength of the evidence?

P-values indicate how surprising the evidence is under H_o: the smaller the p-value, the stronger the evidence against H_o.

- Characterizations of WoE: significant, highly significant, etc.

Example

- A fair and a biased coin: H_o ($\theta = ½$) and H_1 ($\theta = 1$)
- Data: 5 heads on 5 tosses
- $P = (1/2)^5 = 1/32 = 0.03$

Likelihood Defined

Probability distribution $P_\theta(x) = f(x|\theta)$

- x varies, θ is fixed

Likelihood function $L(\theta) \propto f(x|\theta)$

- θ varies, x is fixed

Example

- A fair and a biased coin: H_0 ($\theta = ½$) and H_1 ($\theta = 1$)
- Data: 5 heads on 5 tosses
- Likelihood ratio $LR = \frac{L(H_1)}{L(H_0)} = \frac{k\,P(x|H_1)}{k\,P(x|H_0)} = \frac{1}{\left(\frac{1}{2}\right)^5} = 32$

Three Approaches to Inference

- **Frequentist**
  - Fisher, Neymann, Pearson
- **Likelihoodist**
  - Barnard 1949; Edwards 1972; Royall 1997
- **Bayesian**
  - Bayes 1764

Likelihoodism (Edwards 1972)

Likelihood Principle

- All the information which the data provide concerning the relative merits of the hypotheses is contained in the likelihood ratio of those hypotheses on the data.

Law of Likelihood

- [W]ithin the framework of a statistical model, a particular set of data *supports* one statistical hypothesis better than another if the likelihood of the first hypothesis, on the data, exceeds the likelihood of the second hypothesis.

Support function S(θ)

- ln LR

Three Approaches to Inference

Frequentist

Fisher, Neymann, Pearson

Likelihoodist

Barnard 1949; Edwards 1972; Royall 1997

Bayesian

Bayes 1764

Bayesian inference

Likelihood function

$$f(\theta|x) = a\, \mathbf{L(\theta)}\, f(\theta)$$

Posterior distribution

Prior distribution

$$\text{where } 1/a = \int L(\theta) f(\theta)\, d\theta$$

Bayes Rule for Binary θ

$$\text{Odds}(H_1|E) = \frac{P(E|H_1)}{P(E|H_0)}\,\text{Odds}(H_1)$$

posterior odds on H_1

Bayes factor for H_1

prior odds on H

$$\text{Log odds}(H_1|E) = \text{Log}\,\frac{P(E|H_1)}{P(E|H_0)} + \text{Log odds}(H_1)$$

posterior log-odds

WoE (Good 1991)

prior log-odds

Verbal Tags for BFs

Log-BF	BF	Verbal tag
0 to ½	1 to 3.16	*barely worth mentioning*
½ to 1	3.16 to 10	*substantial*
1 to 1½	10 to 31.6	*strong*
1½ to 2	31.6 to 100	*very strong*
> 2	> 100	*decisive*

- Jeffreys 1961

Example: The 2 Coins

BF (LR)	WoE (log-LR)	Jeffreys tag
32	1.51	very strong

- $\mathrm{BF} = \mathrm{LR} = \frac{P(E|H_1)}{P(E|H_0)} = {}^{1}/_{(\frac{1}{2})^5} = 32$
- $\mathrm{WoE} = \log(\mathrm{BF}) = \log\frac{P(E|H_1)}{P(E|H_0)} = 1.51$

WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic scientists speak about WoE

How FSWs Describe PV in Court

They do not	They go directly to hypotheses	They state values
• Features only • Match	• Categorical conclusion • Posterior probability	• Freq or P quant • Freq or P qual • LR (quant & qual)

How FSWs Describe PV in Court

They do not

- Features only
- Match

Features only testimony

- [A hair examiner] displayed an enlarged photograph of one of the defendant's hairs and one of the hairs recovered from the victim's clothing as they appeared side-by-side under the comparison microscope. [He] explained to the jurors how the hairs were similar and what particular features of the hairs were visible. He also drew a diagram of a hair on a courtroom blackboard for the jurors. The jurors were free to make their own determinations as to the weight they would accord the expert's testimony in the light of the photograph and their own powers of observation and comparison.

State v. Reid, 757 A.2d 482, 487 (Conn. 2000)

Match testimony

testified that some of those hairs were consistent, meaning had the same characteristics, with known hair samples provided by [the defendant] and some of those hairs were consistent with hair samples from the victim . . . ."

- Brown v. State, 999 So.2d 853 (Miss. Ct. App. 2008)

testified that . . . a piece of cord taken from the scene of the crime [and] a piece of cord taken from the hood of a jacket 'matched each other in component structure, . . . were similar and could have . . . originated from the same jacket.'

- State v. Gomes, 881 A.2d 97 (R.I. 2005)

not error to admit "testimony that [defendant] could not be excluded as the source of the DNA obtained from the sneakers [even without] testimony explaining the statistical relevance of the nonexclusion result, such as the percentage of the population that could be excluded."

- Rodriguez v. State, 273 P.3d 845, 850–51 (Nev. 2012)

How FSWs Describe PV in Court

They go directly to hypotheses

- Categorical conclusion
- Posterior probability

Categorical Conclusions

[A] ballistics expert, testified that he examined and compared the single shell casing found at the scene with the shell casings from the test firing of the gun found in the backpack. Walsh gave an opinion that to a 'reasonable degree of certainty in the ballistics community' the spent shell casing from the scene and the shell casings from the test firing were fired from the same weapon.

- Commonwealth v. Carnes, 967 N.E.2d 148, 154 (Mass. App. Ct. 2012).

For each elemental ratio, compare the average ratio for the questioned specimen to the average ratio for the known specimens ±3s. This range corresponds to 99.7 % of a normally distributed population.

- ASTM 2926-13 (μ-XRF spectrometry)

Posterior Probabilities

cf. Hummel et al. 1981

The blood genetic marker tests … registered a composite 99.99% probability that he is the biological father of Baby C, … the chance of someone else … is one in ten thousand.

- Comm'r of Social Serv. v. Hector S., 628 N.Y.S.2d 270 (App. Div. 1995)

Christina Buettner from the Wyoming State Crime Lab first testified "the probability of paternity" is "99.9999998638" that Mr. Snyder is the father of JL's baby.

- Snyder v. State, 2015 WY 91, 353 P.3d 693, 694 (2015)

How FSWs Describe PV in Court

They state values

- Freq or P quant
- Freq or P qual
- LR (quant & qual)

Frequency or P (Qualitative)

"an uncommon type of glass"

- People v. Smith, 968 N.E.2d 1271, 1277 (Ill. App. Ct. 2012)

"no doubt that" the impressions of "two very experienced forensic scientists" that 20 DNA alleles consistent with the defendant's full genotype were "rare" or at least "somewhat unusual" was "of assistance to a jury"

- R. v. Dlugosz, [2013] EWCA Crim 2, at ¶25

Frequency or P (Quant)

[O]nly 3.8 out of 100 samples could have the same physical properties, based upon the refractive index test alone, which was performed.

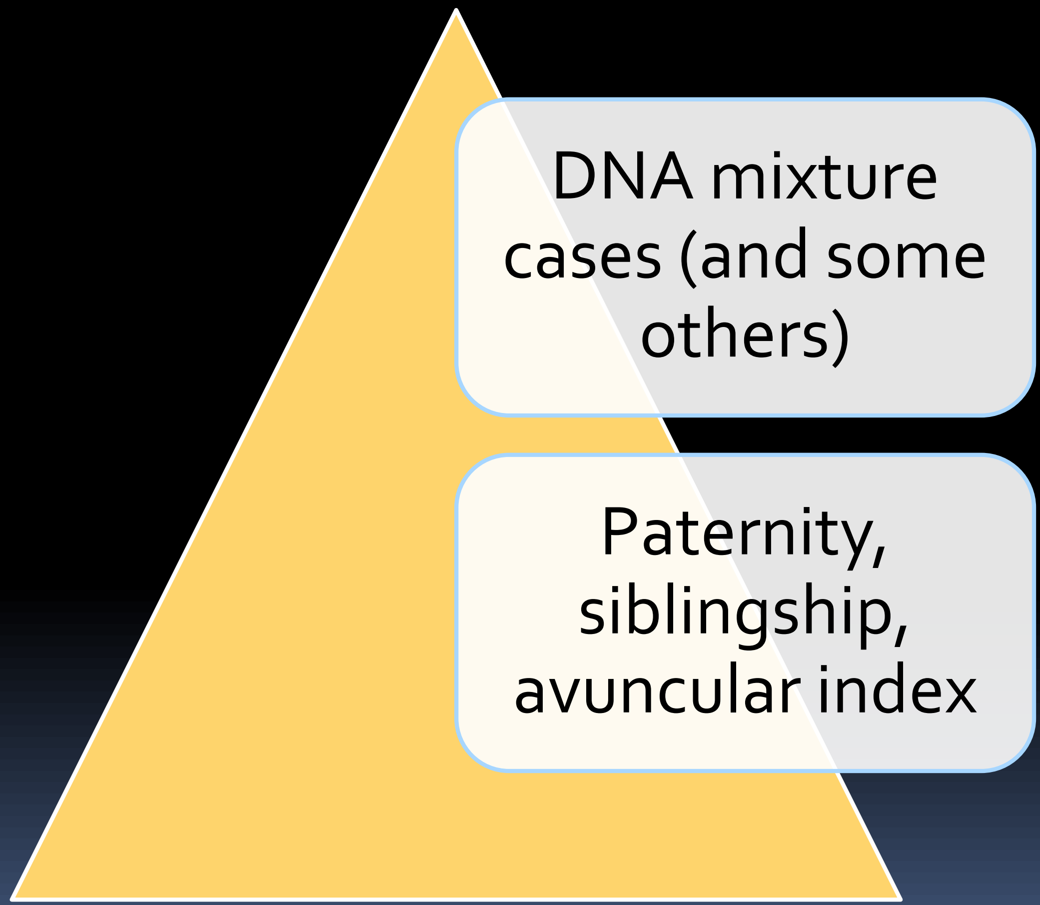
- Johnson v. State, 521 So.2d 1006, 1009 (Ala. Ct. Crim. App. 1986)

1 in many quadrillions, quintillions, sextillions, or septillions

Mixtures: 1/1 to 1/10, 1/80, 1/300, 1/500, 1/3,000, 1/8,000, 1/9,000, 1/15,000, 1/35,000, 1/120,000, and 1/180,000

- DNA cases

Likelihood Ratios



Qualitative LRs

Evertt 1991

Log LR	LR	Verbal Tag
0 to ½	1 to 33	weak
½ to 2	33 to 100	fair
2 to 2½	100 to 330	good
2½ to 3	330 to 1000	strong
>3	>1000	very strong

		Evett et al. 2000
0 to 1	1 to 10	limited
1 to 2	10 to 100	moderate
2 to 3	100 to 1000	moderately strong
3 to 4	1000 to 10000	strong
>4	>10000	very strong

ENFSI 2015

Log LR	Verbal tag
0	no support
0.3 to 1	weak
2 to 3	moderate
2 to 3	strong
4 to 6	very strong
>6	extremely strong

Approved of in NRC 2009

WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic s … about WoE

References

- Workshop bibliography +
- David H. Kaye, David E. Bernstein & Jennifer L. Mnookin, The New Wigmore: A Treatise on Evidence: Expert Evidence (2d ed. 2011) (updated annually)
- ---, Likelihoodism, Bayesianism, and a Pair of Shoes, 53 Jurimetrics J. 1-9 (2012)