# What is Probability

"..the key to the relation between statistics and truth may be found in a reasonable definition of probability"–
R. von Mises (1928/1951) Probability, Statistics, and Truth

"Probability does not exist"– De Finetti (1970) Theory of Probability

James L. Wayman, Ph.D., FIET, FIEEE
Office of Research
San José State University

# Outline

1. Equivocation
2. Probability
3. Conditional Probability
4. Inverse Probability
5. Likelihood Ratio
6. Bayes Factor
7. Options for Quantifying Evidence

# Equivocation

SJSU SAN JOSÉ STATE UNIVERSITY

Equivocal:  having two or more possible meanings

Merriam –Webster Online Dictionary

Equivocation leads to non sequitur:

Everyone must believe in something:

I believe I'll go canoeing  –  H.D. Thoreau

I believe I'll have a beer – Groucho Marx

Tradition of posthumous probability publication perhaps owing in part to recognition by authors of equivocation and non sequiturs

# Probability

SJSU SAN JOSÉ STATE UNIVERSITY

Often heard (disjunctive?) taxonomies:

- Objective vs Subjective (aka Epistemic)
- Post -1950: Frequentist vs. Bayesian (aka Personal)

    Gillies (1987) ”Was Bayes a Bayesian?

    Glymour(1981) “Why I am Not a Bayesian”

    Berger (2006) “The case for objective Bayesian analysis."

Gillies (2000) <u>Philosophical Theories of Probability</u>

1. Classical
2. Logical
3. Subjective
4. Frequency
5. Propensity
6. Various approaches by Gillies

I will take historical approach following

- Stigler (1986) <u>The History of Statistics</u>
- Hacking (1990) <u>Taming of Chance</u> ; (2006) <u>Emergence of Probability;</u>
- Feinberg (1992) “A Brief History of Statistics in Three and One-Half Chapters”
- Howie (2002) <u>Interpreting Probability: Controversies and Developments in Early 20th Century</u>

# Circa 18th Century Probability

SJSU SAN JOSÉ STATE UNIVERSITY

- Bernoulli/Leipniz/de Moivre/Bayes/Laplace
  - Games of chance and legal testimony/juries
  - Probability as a measure of ignorance− Locke(1690), Essay CHU Book IV
  - “Classical” definition: Probability as proportion of equally possible cases
- Principle of: Insufficient Reason (Leipniz) or Indifference (Keynes)
  - 19th century objections by Venn and Boole (*Ex nihilo nihil)*
  - 20th century objections
    - Is the book black/white/red?− Keynes(1921), <u>Treatise on Probability</u>
      - “necessary condition.. **indivisible alternatives”**
    - Loaded die do not have equiprobable states – von Mises(1928/1951) <u>Prob, Stats &Truth</u>
  - Additional modern objection:
  - Uniformity of the distribution of a parameter is not invariant under very reasonable transformations of the parameter

# 19th Century Probability

SJSU SAN JOSÉ STATE UNIVERSITY

- Poisson (1837), Research on the Probability of Judgments in Criminal and Civil Matters

  Probability: the reason we have to believe that an event will or will not occur

  Chance: events in themselves and independent of the knowledge we have of them

- A.A. Cournot (1843) as translated in Daston (1994), "How probabilities came to be objective and subjective"

  "the double sense of probability, which at once refers to a certain measure of our knowledge, and also to a measure of the possibility of things independently of the knowledge we have of them…. *subjective* and *objective*…"

- J.S Mill (1843), System of Logic, "Of the calculation of chances"

  "We must remember that the probability of an event is not a quality of the event itself, but is a mere name for the degree of ground which we, or some one else, have for expecting it"

- Venn (1866), The Logic of Chance

  Limiting ratio of an infinite series of events

- Apologies for omitting Galton, Peirce…..

# 20th Century Probability

SJSU SAN JOSÉ STATE UNIVERSITY

Von Mises(1928/1951) Probability, Statistics and Truth

"… probability….applies only to problems in which either the same event repeats itself again and again or a great number of uniform elements are involved at the same time". Non-repeatable events do not have a probability.

P(x based on a finite Collective) = $\lim_{N \to \infty} \left( \frac{\#x}{N} \right)$

Von Mise's response: "theory...not logically definable but sufficiently exact in practice"

Cambridge School: Russell=> Keynes => Jeffreys (1939) Theory of Probability,

Branch of logic: formal logical relationships

"degree of belief"

"I believe I'll have a beer" as a statement of probability

Fisher (1935), "The Logic of Inductive Inference"

"I mean by mathematical probability only that objective quality of the individual which corresponds to frequency in the population, of which the individual is spoken of as a typical member"

See also Zabell (1992) "RA Fisher and the Fiducial Argument"

Popper (1959), "The Propensity Interpretation of Probability,"

Objective, but not frequentist

P(x based on repeatable conditions with a tendency to produce sequences with frequencies equal to probabilities) Which conditions are important?

# 20th Century Probability

SJSU SAN JOSÉ STATE UNIVERSITY

Kolmogorov (1933), Foundations of the Theory of Probability

"Probability" as undefined "primitive", like the concept of "point" in geometry

All probability as mathematics: field theory (an algebraic concept with operations +, -, X, ÷ )

5 Axioms

I & II: If $\mathscr{F}$ is a set of subsets of E (collection of elementary events), then $\mathscr{F}$ is a field and contains E.

III. For each set A in $\mathscr{F}$, $P(A) \in \mathscr{R}$, $P(A) \geq 0$

IV. $P(E) = 1$

V. If A and B have no element in common (are mutually exclusive), then $P(A \cup B) = P(A) + P(B)$

Added definition: $P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)}$

L. "Jimmy" Savage, Foundations of Statistics (1972), Frank Ramsey (1926?), De Finetti(1931)

"Personal" probability

How much would you bet to win $1?

Apologies for omitting Pearson², Neyman, Carnap, Jaynes…..

# Conditional Probability

- P(x|y) probability of event x given that event y has occurred

  Notation owed to Jeffreys; T. Bayes' concept of rolling balls
- P(E|H) probability of E given Hypothesis

  Duhem-Quine Theory: every hypothesis entails background assumptions, conditions….

  For subjectivists, estimation of probability considers background information
- P(E|H, I) _Does H rely on an *a priori* model?

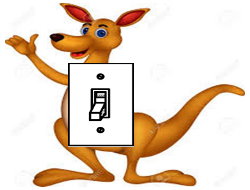
"In a problem of estimation we start with knowledge of the mathematical form of the population sampled, but without knowledge of one or more parameters which enter into this form" Fisher (1935), "The Logic of Inductive Inference".

"All models are wrong"− Box (1976), "Science & Statistics"

"The assumption of a normal distribution for Ɵ is unrealistic"− Lindley (1977) "A Problem in Forensic Science"

"This paper demonstrates that, for large-scale tests, the match and non-match similarity scores have no specific underlying distribution function", Wu&Wilson (2005) NISTIR 7226

# Inverse Probability



P(H|E, I) calculated from P(E|H, I)  via “Bayes Theorem”

Because H is either 100% true or 100% false, this P must certainly be subjective even if P(E|H, I) calculated from objective measures

“…the theory of inverse probability is founded upon error and must be wholly rejected”– Fisher (1925), “Statistical Methods for Research Workers”

“I know of only one case in mathematics of a doctrine which has been accepted and developed by the most eminent men of their time, and is now perhaps accepted by men now living, which is at the same time has appeared to a succession of sound writers to be fundamentally false and devoid of foundation.  Yet that is exactly quite the position with respect of inverse probability…reduces all probability to subjective judgement…The underlying cause is…that we learn by experience that science has its inductive processes, so it is naturally thought that such inductions, being uncertain, must be expressible in terms of probability ” – Fisher (1930) “Inverse Probability”

See Zabell (1989) "RA Fisher on the History of Inverse Probability.“

# Likelihood Ratio

SJSU SAN JOSÉ STATE UNIVERSITY

$$\frac{P(E|H,I)}{P\left(E\middle|\widetilde{H},I\right)}$$

Fisher's attempt to avoid inverse probability

H and $\widetilde{H}$ are taken as exclusive and exhaustive

"When we speak of the *probability* of a certain object fulfilling a certain condition, we imagine all such objects to be divided into two classes, according as they do or do not fulfil the condition. This is the only characteristic in them of which we take cognizance". – "On the Mathematical Foundations of Theoretical Statistics" (1921)

If evidence metric is continuous, P(E|H) is evaluated at the point on the CDF given H at which the value E' is observed. P(E≤E'|H)

# Bayes Factor

SJSU SAN JOSÉ STATE UNIVERSITY

$$\frac{P(H|E,I)}{P(\widetilde{H}|E,I)} = \frac{\mathrm{P(E|H,I)}}{\mathrm{P}\left(\mathrm{E}\middle|\widetilde{\mathrm{H}}, I\right)} \times \frac{P(H,I)}{P(\widetilde{H}),I)}$$

Good (1958), "Significance tests in parallel and in series." *J. Am. Stat. Assoc.*

"The Bayes factor is equal to the likelihood ratio when H and $\widetilde{\mathrm{H}}$ are both simple statistical hypotheses, but not otherwise in general…We shall think of $\widetilde{\mathrm{H}}$ as the logical disjunction of all the alternatives to H that are worth considering. Each alternative is a non-null hypothesis whereas $\widetilde{\mathrm{H}}$ itself is *the* non-null hypothesis. As a special case $\widetilde{\mathrm{H}}$ may be a simple statistical hypothesis. Otherwise it is composite and is the logical disjunction of a number of simple hypotheses, where this number may be finite, enumerable, of the power of the continuum, or even more." (notation slightly adjusted)

So how do we empirically determine distribution of E over the disjunction of **all alternatives** to H worth considering?

In real world, "dynamic" probability allows update of P(H, I) based on reconsideration of old I −Diaconis and Zabell(1982) "Updating Subjective Probability", *J.Am Stat. Assoc.* "Evolving" probability−Good(1968), "Corroboration Explanation…")

# Bayes Factor with $H_2$ as Proxy for $\widetilde{H}_1$

But only if $H_2 \Rightarrow \widetilde{H}_1$ and $H_1 \Rightarrow \widetilde{H}_2$ (exclusive)

Such that $P(H_1) + P(H_2) = P(\widetilde{H}_1) + P(\widetilde{H}_2) = 1$ (exhaustive)

Otherwise, by Ramsey–De Finetti Theorem, system is "incoherent/inconsistent" (vulnerable to "Dutch book")

<u>The "Reference Class Problem"</u>

- Either $H_1$ = "from this source" (subject specific)

or "from someone who is the same source" (general population)
(within-class variation homogeneous across population?)

- $H_2$ = "from relevant population", where either "relevant" refers to

subject or to questioned (People v Pizarro, CA 5th Dist. Ct. Appeals, 1992)

Can $H_1$ and $H_2$ be chosen as to maintain exclusive/exhaustive requirements?

# Bayes Factor with $H_2$ as Proxy for $\widetilde{H}_1$

But only if $H_2 \Rightarrow \widetilde{H}_1$ and $H_1 \Rightarrow \widetilde{H}_2$ (exclusive)

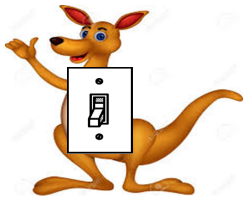
Such that $P(H_1) + P(H_2) = P(\widetilde{H}_1) + P(\widetilde{H}_2) = 1$ (exhaustive)

Otherwise, by Ramsey–De Finetti Theorem, system is "incoherent/inconsistent" (vulnerable to "Dutch book")

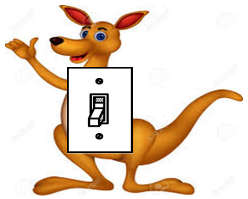
<u>The "Reference Class Problem"</u>

- Either $H_1$ = "from this source" (subject specific)

or "from someone who is the same source" (general population)
(within-class variation homogeneous across population?)



- $H_2$ = "from relevant population", where either "relevant" refers to subject or to questioned (People v Pizarro, CA 5th Dist. Ct. Appeals, 1992)



Can $H_1$ and $H_2$ be chosen as to maintain exclusive/exhaustive requirements?

# Weight of Evidence: Some probabilistic alternatives

SJSU SAN JOSÉ STATE UNIVERSITY

1. P(H|E,I) Subjective, NIST tests with human subjects, J.S. Mill might approve

2. P(E|H,I) or $P(E|\widetilde{H},I)$ Hypothesis test, *a priori* info used for specification, Does evidence lie in rejection region of the model?

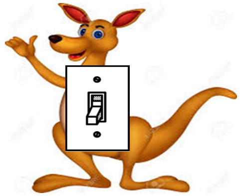
3. $\frac{P(E|H,I)}{P(E|\widetilde{H},I)}$ Fisher's Likelihood Ratio, avoids inversion and priors, not a probability.

4. $\frac{P(E|H_1,I)}{P(E|H_2,I)}$ Bayes Factor if $H_2 \equiv \widetilde{H_1}$ $H_2$ is logical disjunction of all alternatives worth considering.

5. $(P(E|H,I), P(E|\widetilde{H},I))$. Traditional decision theoretic ROC, probability of no error given "from someone who is the same" against probability of error given "from someone who is not the same".

6. $P(E|H_n, I)$ n=1,2…N "Objective", but needs subjective specification and understanding of associated context. See French and Harrison (2007) "Position Statement concerning use of impressionistic likelihood..."

# One More Alternative



Probabilities have no place in court

Tribe (1971), "Trial by mathematics: Precision and ritual in the legal process." *Harvard Law Rev.*

P. Tillers (2011), "Trial by mathematics—reconsidered." *Law, probability and risk*

"It could be…that judgements about the confirmation of a hypothesis by evidence are inherently qualitative rather than quantitative in nature" – Gilles (2000) <u>Philosophical Theories of Probability</u>

"The (incorrect) argument runs somewhat as follows: a number of uncertain but useful judgements can be expressed with exactitude in terms of probability; our judgements concerning causes or hypotheses are uncertain, therefore our rational attitude towards them is expressible in terms of probability" – Fisher (1930) "Inverse Probability"

R. v T [2010] EWCA Crim 2439; [2011] 1 Cr. App. R. 9

# Quiz

Based on the evidence just presented, state which of the following non-exclusive, non-exhaustive hypotheses is more probable:

$H_1$: We can all agree on a single definition of probability

$H_2$: I can agree on a single definition of probability

$H_3$: I’ll have a beer

$H_4$: Sorry, I’m a strict frequentist