The Measurement of Statistical Evidence
Lecture 4 - part 2

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/sta4522/STA4522.html

2021

Statistical Reasoning

- the goal is a theory of statistical reasoning that addresses the issues raised and is based on a proper characterization of statistical evidence

- here is a sequence of steps to statistical reasoning concerning **E** and/or **H**

1. choose a model $\{f_\theta : \theta \in \Theta\}$
2. choose a prior π
3. measure bias and select the amount of data to collect to avoid bias
4. collect the data
5. check the model (modify if necessary)
6. check the prior (modify if necessary)
7. derive the inferences (based on principles of inference to be discussed)

- 7 and 3 are now discussed and based on the ingredients

$$(\{f_\theta : \theta \in \Theta\}, \pi, x)$$

- these ingredients lead to a probability model $(\theta, x) \sim \pi(\theta) f_\theta(x)$

- so all discussion of the principles of inference can take place within the context of a probability model $(\Omega, \mathcal{F}, P)$

- the first principle of inference

> *1.* **Principle of Conditional Probability***: initial belief that the unknown value of $\omega \in A \in \mathcal{F}$ is measured by $P(A)$ and after observing that $\omega \in C$ (via a known information generator), where $P(C) > 0$, then belief that $\omega \in A$ is measured by $P(A \,|\, C) = P(A \cap C)/P(C)$.*

- second principle of inference

> *2.* **Principle of Evidence***: if $P(A \,|\, C) > P(A)$, then the observation that C is true is evidence in favor of A being true, if $P(A \,|\, C) < P(A)$, then the observation that C is true is evidence against A being true, and $P(A \,|\, C) = P(A)$ is neither evidence in favor nor evidence against A being true.*

- note - $P(A \,|\, C) = P(A)$ iff A and C are statistically independent

- so the principle of evidence tells us when there is evidence in favor or evidence against only and sometimes more is needed as it will be necessary to order alternatives

- third principle of inference

> *3.* **Principle of the Relative Belief Ratio**: *when a numerical measure of evidence is required this is given by the relative belief ratio* $RB(A|C) = \frac{P(A\,|\,C)}{P(A)}$.

$$\text{- so } RB(A|C) \begin{array}{ll} > 1 & \text{evidence in favor} \\ < 1 & \text{evidence against} \\ = 1 & \text{no evidence either way} \end{array}$$

- principles 1 and 2 seem simple and sound whereas 3 is more controversial as there are other measures of evidence that are valid measures of evidence, namely, there is a clear cut-off that determines evidence in favor or against according to the principle of evidence

- is the principle of evidence sound?

Example *Card game.*

- two players in a card game, labeled I and II

- each is dealt m cards, where $2 \leq m \leq 26$, from a randomly shuffled deck of 52 playing cards

- player I, after seeing their hand, is concerned with the truth or falsity of

H_0 : *player II has exactly two aces*

- the hand of player I will contain evidence concerning this

- what is the evidence when $C_k =$ "the number of aces in the hand of player I is k" with $k = 0, 1$ or 2?

- two questions

(i) is there evidence in favor of or against H_0?
(ii) how strong is this evidence?

- we have $P(H_0)$ and $P(H_0 \mid C_k)$ available for this

	$P(H_0)$	$P(H_0 \mid C_k)$		$RB(H_0 \mid C_k)$
$m = 2$	0.0045	$k = 0$	0.0049	1.0824
		$k = 1$	0.0024	0.5412
		$k = 2$	0.0008	0.1804
$m = 5$	0.0399	$k = 0$	0.0483	1.2089
		$k = 1$	0.0259	0.6487
		$k = 2$	0.0093	0.2317
$m = 10$	0.1431	$k = 0$	0.1994	1.3934
		$k = 1$	0.1254	0.8765
		$k = 2$	0.0522	0.3652
$m = 20$	0.3481	$k = 0$	0.3487	1.0018
		$k = 1$	0.4597	1.3205
		$k = 2$	0.3831	1.1004
$m = 25$	0.3890	$k = 0$	0.0171	0.0439
		$k = 1$	0.2051	0.5274
		$k = 2$	0.8547	2.1974
$m = 26$	0.3902	$k = 0$	0.0000	0.0000
		$k = 1$	0.0000	0.0000
		$k = 2$	1.0000	2.5630

- other than $(m, k) = (25, 2), (26, 2)$, the conditional probability $P(H_0 \mid C_k)$ does not support H_0 being true and in many cases some would argue that the value of this probability indicates evidence against H_0

- also conditional probabilities do not satisfy the principle of evidence and so are not valid measures of evidence

- comparing $RB(H_0 \mid C_k)$ to 1 answers 1 and quoting $P(H_0 \mid C_k)$ answers (ii) as it measures how strongly we believe the evidence

Example 2. *Prosecutor's fallacy.*

- a uniform probability distribution on a population of size N of which some member has committed a crime

- DNA evidence has been left at the crime scene and suppose this trait is shared by $m \ll N$ of the population

- a particular member possesses the trait and the prosecutor concludes they are guilty because the trait is rare

- $P(\text{"has trait"} \mid \text{"guilty"}) = 1$ is misinterpreted as the probability of guilt rather than $P(\text{"guilty"} \mid \text{"has trait"}) = 1/m$ which is small if m is large

- **but** clearly there is evidence of guilt and probability does not indicate this (MAP suggests innocence) and

$$\begin{aligned} RB(\text{"guilty"} \mid \text{"has trait"}) &= N/m > 1 \\ P(\text{"guilty"} \mid \text{"has trait"}) &= 1/m \end{aligned}$$

- so there **is** evidence of guilt but the evidence is weak whenever m is large and a conviction does not then seem appropriate

- but suppose that "guilty" corresponds to being a carrier of a highly infectious deadly disease and "has trait" corresponds to some positive, but not definitive, test for this

- the same numbers should undoubtedly lead to a quarantine

- there is a difference between a decision and what the evidence says

- the Principle of Evidence has a long history but not in the statistical literature rather in the philosophy of science literature where it falls under discussions of Confirmation Theory

- Popper, K. (1968) The Logic of Scientific Discovery. Harper Torchbooks Appendix ix where, with x and y denoting events

> *"If we are asked to give a criterion of the fact that the evidence y supports or corroborates a statement x, the most obvious reply is: that y increases the probability of x."*

- Achinstein, P. (2001) The Book of Evidence. Oxford University Press.

> *"for a fact e to be evidence that a hypothesis h is true, it is both necessary and sufficient for e to increase h's probability over its prior probability"*

Example *Hempel's (the Raven) Paradox.*

- $\Omega =$ the universe of all objects

- $A =$ 'if an object is a crow, then it is black' or equivalently 'all crows are black'

- a black crow is observed so $C =$ 'a black crow is observed'

- naturally $RB(A \mid C) > 1$ and so this observation produces evidence in favor of A

- the contrapositive of A, namely, $B =$ 'if an object is not black, then it is not a crow' or equivalently 'all nonblack objects are not crows'

- the paradox supposedly arises due to the fact that, observing an nonblack object that isn't a crow, such as a white handkerchief, wouldn't necessarily be viewed as evidence in favor of A even though it is in favor of B

- resolution (?) bring bias calculations into the discussion and then it is seen that this is just a bad study if our purpose is to confirm A by viewing an object at random (see text)

- but not really "statistical" in nature