SOME REMARKS ON THE PHILOSOPHY OF STATISTICS

BRANDEN FITELSON

Department of Philosophy

San José State University

branden@fitelson.org

http://fitelson.org/

Overview of Presentation

- What are the ends of statistical experiment, analysis, and inference?
- What are the most effective means for achieving these ends?
- Several paradigms for statistics have been developed — each of these presupposes answers to these key "philosophical" questions about statistics.
- Existing paradigms for statistics include the following:
  - Likelihoodist (*relative* evidential support *via* likelihood ratios)
  - Naïve Bayesian (posterior probability maximization)
  - Fisherian (significance testing/rejection trials)
  - Neyman–Pearsonian (Type I/Type II error minimization)
  - Sophisticated Bayesian (expected cognitive utility maximization)
  - Predictivist (predictive accuracy maximization/divergence minimization)
- I will try to classify each of these paradigms wrt our questions above…

An Elementary Example & Some Initial Distinctions I

- John Doe is about to be tested for some disease D. The experimental design (or model) $\mathcal{M}$ of the diagnostic test has the following error characteristics:

		Test Result	
		Positive	Negative
Disease D	Present	0.95	0.02
	Absent	0.05	0.98

- Let $H_0 = \neg H$ = John Doe does not have D, H = he has D, $+$ = test is positive, and $-$ = test is negative. Then, our experimental model $\mathcal{M}$ is such that:

$$\Pr_{\mathcal{M}}(+\mid H) = 0.95 \quad \Pr_{\mathcal{M}}(+\mid H_0) = 0.05$$
$$\Pr_{\mathcal{M}}(-\mid H) = 0.02 \quad \Pr_{\mathcal{M}}(-\mid H_0) = 0.98$$

- $\mathcal{M}$ does *not* tell us the prior probability (or "base rate") $\Pr(H)$ of H. If $\Pr(H)$ is very low, then $\Pr(H\mid +)$ will be low (physicians often get this wrong [16], [22]).

An Elementary Example & Some Initial Distinctions II

- Bayes' Theorem allows us to calculate $\Pr(H\mid +)$, as follows:

$$\Pr(H\mid +) = \frac{\Pr(+\mid H)\cdot\Pr(H)}{\Pr(+\mid H)\cdot\Pr(H) + \Pr(+\mid H_0)\cdot\Pr(H_0)}$$
$$\stackrel{?}{=} \frac{0.95\cdot\Pr(H)}{0.95\cdot\Pr(H) + 0.02\cdot(1-\Pr(H))}$$

- So, if $\Pr(H)$ is very small, then $\Pr(H\mid +)$ will also be small, even though the diagnostic test is (intuitively) "well designed". Now, some initial distinctions:
  - Naïve Bayesian: a $+$ alone is insufficient to determine the posterior of H. We can't properly interpret a $+$ without information about the prior of H.
  - Likelihoodist: a $+$ alone yields a large likelihood-ratio $\left[\frac{\Pr(+\mid H)}{\Pr(+\mid H_0)} = 47.5\right]$ in favor of H. So, a $+$ means strong evidence in favor of H (*versus* H_0).
  - The reactions of the other Paradigms will be more subtle and complex. I

will treat them next. I'll return to Bayesianism and Likelihoodism later...

Possible Reactions of "Fisherians" to Our Toy Example I

- A Fisherian would tend to interpret a +-result in our toy example in one of the following two ways (see [29, chapter 3] for detailed critical discussion):

1. Significance Test: If we take $H_0 = \neg H$ to be the null hypothesis, then a Fisherian might respond to a + by saying that we have observed a result which is significant at the 2% level, or with a p-value of 0.02. Fisher [10, p. 39] says:

   (*) "*Either* a rare event [$\Pr_{\mathcal{M}}(+ \mid H_0) = 0.02$] has occurred, *or* H_0 is false."

- Many statisticians (including Fisher himself) have interpreted p-values as measures of evidential strength. According to Fisherians, the lower the p-value, the stronger the evidence against the null hypothesis [9, p. 80].
- Let $\mathcal{M}$ = the tosses of a coin c are $Bin(1, \theta)$ (*viz.*, i.i.d., Bernoulli), E = a sequence of n tosses of c, and H_0: $\theta = \frac{1}{2}$. Then, $\Pr_{\mathcal{M}}(E \mid H_0) = (\frac{1}{2})^n$, for *any* E.
- $\therefore$ For large n, *any* outcome E is "strong evidence against H_0!" This takes the sting out of (*), and the evidential interpretation of p-values (see [19, p. 82]).

Possible Reactions of "Fisherians" to Our Toy Example II

2. Rejection Trial: Sometimes, Fisherians take observations with small p-values as reasons to *reject* null hypotheses. Is there "probabilistic *modus tollens*?"

$$\text{(MT)}\ \frac{H \Rightarrow E}{\therefore \neg E \Rightarrow \neg H} \qquad \text{(PMT)}\ \frac{\Pr(E \mid H) \approx 1}{\therefore \Pr(\neg H \mid \neg E) \approx 1}$$

- While (MT) is valid, its inductive analogue (PMT) is not. One must assume $\Pr(E \mid H) = 1$ to ensure $\Pr(\neg H \mid \neg E) \approx 1$ (*pace* [5, §4.3] & [20, §1.7]).
- This illegitimate form of inference has been used several times in the history of statistics ([5, §4.3] & [20, §1.7]). More recently, it has been used by creationists to argue against evolution theory ([6], [12], [26], [11], [33]).
- $\mathcal{M}$: $X \sim Bin(n, \theta)$, H_0: $\theta = \frac{1}{2}$, and H_0': $\theta \le \frac{1}{2}$. E: $X = x$ sanctions rejection of H_0 at level α if $\Pr_{\mathcal{M}}(X \ge x \mid H_0) \le \frac{\alpha}{2}$ and of H_0' if $\Pr_{\mathcal{M}}(X \ge x \mid H_0') \le \alpha$. So, an x such that $\frac{\alpha}{2} < \Pr_{\mathcal{M}}(X \ge x \mid H_0) \le \Pr_{\mathcal{M}}(X \ge x \mid H_0') \le \alpha$ sanctions rejection of H_0' but *not* H_0 [29, p. 77]. We may reject "A or B", but we may *not* reject A!

Reaction of "Neyman–Pearsonians" to Our Toy Example

- N–P Reaction: The experiment is designed (*viz.*, $\mathcal{M}$) for the purpose of recommending rejection of H_0 if + is observed and acceptance of H_0 if − is observed. There are two types of errors we could make (in so using $\mathcal{M}$):
  - Type I error: rejecting H_0 (& accepting H) on the basis of + when H_0 is true.
  - Type II error: accepting H_0 (& rejecting H) on the basis of − when H_0 is false.
- For our $\mathcal{M}$, the probability of a Type I error (size) is $\alpha = \Pr_{\mathcal{M}}(+ \mid H_0) = 0.02$. And, the probability of a Type II error (power) is $\beta = \Pr_{\mathcal{M}}(- \mid H) = 0.05$.
- In rejecting H_0 on the basis of + (using $\mathcal{M}$), we are not saying that we should (strongly) *believe* H_0; nor are we saying that + constitutes *strong evidence against* H_0 (*vs* H). Statistics is not in the businesses of grounding such claims.
- Statistics is in the business of providing "performance characteristics of rules of inductive behavior based on random experiments" [25, p. 11]. In this case, $\mathcal{M}$ has size (or significance level) $\alpha = 0.02$ and power $\beta = 0.05$. End of story.

More on the "Neyman–Pearsonian" and "Fisherian" Approaches to Statistics

- The key difference between the N–P and Fisherian approaches is that N–P is *comparative*. N–P looks at *both* H_0 *and* its alternatives (*e.g.*, H), and ∴ seeks tests $\mathcal{M}$ with (simultaneously, sort of) low values of *both* α *and* β.
- The Fisherian focuses *only* on the null H_0, and ∴ worries *only* about α.
- The advantage of ignoring β is that β is often difficult to calculate. If H_0 is a simple hypothesis ($\theta = \frac{1}{2}$), its negation will be a messy, composite hypothesis ($\theta \neq \frac{1}{2}$). Calculating the likelihood (β) $\Pr(E \mid \neg H_0)$ in such cases is difficult.
- This problem of computing likelihoods of composite hypotheses plagues all of the Paradigms (see [29, ch. 7] on this problem for Likelihoodism, and [24, p. 194–5] on this problem for more traditional statistical testing Paradigms).
- One Paradigm faces this problem *head-on*, by endowing the model $\mathcal{M}$ with enough structure to compute *all* probabilities and *all* likelihoods in *all* cases. This is the Naïve Bayesian approach to statistics, to which I now turn.

The Naïve Bayesian Approach I

- The Naïve Bayesian aims to "accept" hypotheses with *maximal posterior probability* among the available alternatives. [This aim will be explained in terms of a more sophisticated, decision-theoretic Bayesian framework, below.]
- In our example, $\mathcal{M}$ did not have enough structure to allow for calculation of the posterior $\Pr(H_0 \mid +)$ of H_0. Information about the prior $\Pr(H_0)$ is needed.
- The main problem for Naïve Bayesianism is the origin and status of the priors [32]. In diagnostic testing cases, "base rates" or frequencies from actual populations are often used as the "priors" in Bayes' Theorem. Problems:
  - How does one choose the appropriate reference class for such frequencies? The prior probability of my having D will depend on the me-containing class that we decide to use as a reference [27, §72], [17, pp. 119–125].
  - The likelihoods $\Pr_{\mathcal{M}}(+ \mid \pm H)$ are *resilient* [31], causal propensities, but the priors are mere actual frequencies. Should a Bayesian "mix" these [4]?

The Naïve Bayesian Approach II

- There have been many attempts to provide objective accounts of "invariant" or "informationless" priors [2, §5.6], [8]. Such an account (either logical or empirical) would place priors on an objective footing (like $\Pr_{\mathcal{M}}(+ \mid \pm H)$).
- Unfortunately, no satisfactory account has appeared, and the prospects for "Objective Bayesianism" do not look good (see [30] and [28] for discussion).
- This has lead most Bayesians to take a subjectivist line [2, pp. 99–102] in which $\Pr_{\mathcal{M}}(H)$ [$\Pr_{\mathcal{M}}(H \mid E)$] is taken to be a rational agent's degree of belief in H prior to [after] learning E (relative to background knowledge corpus $\mathcal{M}$).
- Sophisticated Bayesians move away from the unclear Fisherian or N–P notions of "acceptance", and even from the fundamental dogma that posterior probability distributions are the *sole* currency of statistical inquiry [23], [2].
- Such Bayesians think of statistical practice simply as a rational enterprise which may involve various (possibly competing) cognitive utilities [23].

Sophisticated Bayesianism, N–P, and Naïve Bayesianism

- The sophisticated Bayesian uses (personalistic) decision theory (*i.e.*, expected cognitive utility maximization) as their guide to inductive behavior.
- In this way, they are similar to Neyman, who viewed statistics as prescribing "well performing" inductive rules for practitioners to use and follow.
- The Bayesian has a more general (albeit *subjective*!) view than Neyman, since they allow *many* cognitive utilities (not just α/β min. [2, pp. 471–472]).
- For instance, say you assign cognitive utility 1 to "accepting" a true hypothesis and 0 to "accepting" a false hypothesis, and that truth and falsity (*simpliciter*) are *all* that you care about in the context of "acceptance".
- In order to maximize your expected cognitive utility, you should "accept" hypotheses with maximal posterior probability (among the alternatives).
- In this sense, naïve Bayesianism is a special case of sophisticated Bayesianism in which the agent has naïve, "truth-functional" cognitive utilities [2, §6.1.4].

Sophisticated Bayesianism and Predictivism

- What if your cognitive utilities are less naïve?
- What if you are interested in *quantitatively approximating* a true distribution (t), and you want to minimize the "distance" (really, *average* distance, as a rule) between your approximation ($\hat{t}$) and the true distribution?
- Then, you'll need a finer-grained cognitive utility function — one which is inversely proportional to some measure of the *divergence* between t and $\hat{t}$.
- This kind of utility function might be called predictive (by statisticians like [15]) or verisimilitudinous (by philosophers of science/statistics like [23]).
- There is a very lively debate currently raging on in the philosophy of statistical inference (and in philosophy of science generally) between (subjective) Bayesians and (objective or frequentist) non-Bayesians who both share predictive/verisimilitudinous leanings in this sense [13], [1].
- Additional (good) Bayes/non-Bayes discussions: [14], [18], [7], [3], [21].

Extra Slide: Explaining Fisher's False Dilemma

- Fisher claimed that in examples like our diagnostic testing example, the observation of a + allowed us to infer the following disjunction:

  (*) "*Either* a rare event [$\Pr_{\mathcal{M}}(+ \mid H_0) = 0.02$] has occurred, *or* H_0 is false."

- But, arguments of Hacking [19, p. 82] and Royall [29, p. 77] show this disjunction to have little force. Where does Fisher go wrong?

(a)
$\mathcal{M}$ is correct.
(*i*) If $\mathcal{M}$ is correct, then $H_0 \Rightarrow +$ is improbable [$\Pr(+) = 0.02$].
$\therefore$ Either H_0 is false or + is improbable.

(b)
$\mathcal{M}$ is correct.
(*ii*) If $\mathcal{M}$ is correct, then $\Pr(+ \mid H_0) = 0.02$.
$\therefore$ Either H_0 is false or + is improbable.

- Argument (a) is valid, but (*i*) is false. In (b), (*ii*) is true, but the argument is invalid. Fallacy: (*ii*) $\not\models$ (*i*). *i.e.*, $\Pr(+ \mid H_0) = 0.02 \not\models H_0 \Rightarrow \Pr(+) = 0.02$ (why?).

References

[1] P.S. Bandyopadhyay and R.J. Boik, *The curve fitting problem: a Bayesian rejoinder*, Philosophy of Science **66** (1999), no. 3, suppl., S390–S402, PSA 1998, Part I (Kansas City, MO).

[2] J.M. Bernardo and A.F.M. Smith, *Bayesian theory*, 2 ed., John Wiley & Sons Ltd., 2000.

[3] G. Casella and R.L. Berger, *Statistical inference*, Wadsworth & Brooks/Cole, 1990.

[4] L.J. Cohen, *Can human irrationality be experimentally demonstrated?*, Behavioral and Brain Sciences **4** (1981), 317–370.

[5] A.I. Dale, *A history of inverse probability: From Thomas Bayes to Karl Pearson*, Springer, 1991.

[6] W.A. Dembski, *The design inference: Eliminating chance through small probabilities*, Cambridge University Press, 1998.

[7] B. Efron, *Controversies in the foundations of statistics*, American Mathematical Monthly **85** (1978), 231–246.

[8] R. Festa, *Optimum inductive methods*, Kluwer Academic Publishers, 1993.

[9] R.A. Fisher, *Statistical methods for research workers*, Hafner Publishing Co., 1958.

[10] ______, *Statistical methods and scientific inference*, Hafner Publishing Co., 1959.

[11] B. Fitelson and E. Sober, *Plantinga's probability arguments against evolutionary naturalism*,

Pacific Philosophical Quarterly **79** (1998), 115–130.

[12] B. Fitelson, C. Stevens, and E. Sober, *How not to detect design – critical notice: William A. Dembski, "The design of inference"*, Philosophy of Science **66** (1999), 472–499.

[13] M. Forster and E. Sober, *How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions*, British Journal for the Philosophy of Science **45** (1994), 1–35.

[14] D. Freedman, *Some issues in the foundation of statistics*, Topics in the foundation of statistics (B. Van Fraassen, ed.), Kluwer Academic Publishers, 1997, pp. 19–39, 41–67, 69–83.

[15] S. Geisser, *Predictive inference: An introduction*, Chapman & Hall, 1993.

[16] G. Gigerenzer, *Adaptive thinking*, Oxford University Press, 2000.

[17] D. Gillies, *Philosophical theories of probability*, Routledge, 2000.

[18] I.J. Good, *The Bayes/non-Bayes compromise: a brief review*, Journal of the American Statistical Association **87** (1992), 597–606.

[19] I. Hacking, *Logic of statistical inference*, Cambridge University Press, 1965.

[20] T. Hailperin, *Sentential probability logic: Origins, development, current status, and technical applications*, Lehigh University Press, 1996.

[21] C. Howson and P. Urbach, *Scientific reasoning: The Bayesian approach*, Open Court, 1993.

[22] J. J. Koehler, *The base rate fallacy reconsidered: Normative, descriptive and methodological*

challenges, Behavioral and Brain Sciences **19** (1996), 1–53.

[23] P. Maher, *Betting on theories*, Cambridge University Press, 1993.

[24] D. Mayo, *Error and the growth of experimental knowledge*, University of Chicago Press, 1996.

[25] J. Neyman, *First Course in Probability and Statistics*, Henry Holt & Co., 1950.

[26] A. Plantinga, *Warrant and proper function*, Oxford University Press, 1993.

[27] H. Reichenbach, *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*, University of California Press, 1949.

[28] R.D. Rosenkrantz, *Bayesian theory appraisal: a reply to Seidenfeld*, Theory and Decision **11** (1979), no. 4, 441–451.

[29] R. Royall, *Statistical evidence: A likelihood paradigm*, Chapman & Hall, 1997.

[30] T. Seidenfeld, *Why I am not an objective Bayesian; some reflections prompted by Rosenkrantz*, Theory and Decision **11** (1979), no. 4, 413–440.

[31] B. Skyrms, *Pragmatics and empiricism*, Yale University Press, 1984.

[32] E. Sober, *Bayesianism — its scope and limits*, Bayes' Theorem (R. Swinburne, ed.), Oxford University Press, 2002, Proceedings of the British Academy Press, vol. 113.

[33] ______, *Intelligent design and probability reasoning*, International Journal for the Philosophy of Religion (forthcoming), 2002.