Announcements and Such

- Administrative Stuff
  - **HW #5 will be graded soon (and I will post solutions soon)**
  - **HW #6 is due next Friday (April 22)**
    * Consists of two (sets of) probability problems: one involving general algebraic reasoning, one involving numerical calculation.
  - **I will distribute a Practice Final Exam on Friday (4/15). We will go over it in class on the last day of the semester (4/19).**
- Unit #4 — *Probability & Inductive Logic, Continued*
  - Review of two "reasoning fallacies" and how they involve Factor #1 *vs* Factor #2 assessements of strength.
  - Measuring Factor #2 — relevance measures
  - Measuring "Overall Argument Strength"?
  - Probabilism and the Accuracy of Credences

Two Infamous "Reasoning Fallacies" and our Two Factors I

- The *Base Rate Fallacy* occurs when one doesn't give proper weight to the base rate/prior/unconditional probability of an improbable hypothesis.
- For instance, Let $H \overset{\text{def}}{=}$ a woman (of age 40 who participates in routine screening) has breast cancer, and $E \overset{\text{def}}{=}$ such a woman has had a positive mammogram in routine screening. And, let us suppose that:

(1) The likelihood of H is: $\Pr(E \mid H) = 0.8$.

(2) The likelihood of $\sim H$ is: $\Pr(E \mid \sim H) = 0.1$,

(3) The base rate/prior probability of H is: $\Pr(H) = 0.01$.

- It follows from *Bayes's Theorem* (or a direct algebraic calculation) that (1)–(3) determine the following value for the posterior probability of H:

(4) The posterior probability of H is: $\Pr(H \mid E) = 0.075$.

- Many people make the (false) judgment that the (1)–(3) imply that posterior of H is *high*. (around 0.8) *This* is the *Base Rate Fallacy*.

- Note: (a) it's a non-trivial calculation to determine that (1)-(3) imply (4); and, (b) Claims (1) & (2) *immediately imply* that E is *strongly positively relevant to H*. So, although the argument from E to H is weak — in Factor #1 terms — *it is actually doing very well, from a Factor #2 perspective*.

- To my mind, it's not surprising that in cases such as these, people tend to latch onto the "Factor #2 perspective." Not only for reasons (a) and (b).

- It is also significant that the likelihoods $\Pr(E \mid H)$ and $\Pr(E \mid {\sim}H)$, which determine the reliability of the test (and the Factor #2 strength of the argument from E to H), are *more robust and invariant* than the base rate.

- After all, the reliability of the test is something that depends *only on the causal structure of the test apparatus*, which is *invariant* across samples drawn from different populations, *etc.*

- On the other hand, the posterior probability of H, $\Pr(H \mid E)$ depends (sensitively) on the base rate/prior probability of H, which will *vary wildly* from one population to another. This is why the likelihoods — *and not the posterior!* — are reported by the manufacturers of diagnostic tests.

Two Infamous "Reasoning Fallacies" and our Two Factors II

- Another infamous case in which our Two Factors pull in opposite directions (causing errors to be made) is *The Conjunction Fallacy*.
- Consider the following evidence E regarding a woman named Linda.

(E) Linda is 31, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and she also participated in antinuclear demonstrations.

- **Question**. Consider the following two hypotheses:

(B) Linda is a bank teller.

($F \,\&\, B$) Linda is a feminist bank teller.

which of these two hypotheses is *more probable, given E*?

- Formally, the question reduces to a comparison of the following to *conditional probabilities* (Factor #1): $\Pr(B \mid E)$ *vs* $\Pr(F \,\&\, B \mid E)$.
- It is easy to show that: $\Pr(B \mid E) \geq \Pr(B \,\&\, F \mid E)$.

- This just follows from *logic*. Because $F \,\&\, B \vDash B$, $F \,\&\, B$ cannot be true in *a larger set of* possible worlds than B is. Thus, generally, we can *never* have $\Pr(B \mid E) < \Pr(B \,\&\, F \mid E)$. But, many people give just this answer!
- We think it has to do with the distinction between conditional probability (Factor #1) and probabilistic relevance (Factor #2).
- Intuitively, (i) E is *positively relevant* to F (*even given* B), but (ii) E is *not positively relevant* to B. (i) & (ii) jointly entail that *E is* ***more*** *relevant to* $F \,\&\, B$ than it is to B — *on any (reasonable) relevance measure*.
- *E.g.*, consider the relevance measure $d(X, E) \overset{\text{def}}{=} \Pr(X \mid E) - \Pr(X)$.
- $d(X, E)$ is *one* possible measure of *how relevant* E is to X. If E is positively relevant to X, then $d(X, E) > 0$. If E is negatively relevant to X, then $d(X, E) < 0$. And, if E is irrelevant to X, then $d(X, E) = 0$.
- So, again, Factor #1 and Factor #2 *cut in opposite directions*:
  - **Factor #1**. $\Pr(B \mid E) > \Pr(B \,\&\, F \mid E)$.
  - **Factor #2**. $d(B, E) < d(B \,\&\, F, E)$.

Measuring Factor 2: Degrees of Confirmation I

- In the contemporary literature, our "Factor 2" is called *confirmation*:

$$E \textit{ confirms } H \text{ if and only if } \Pr(H \mid E) > \Pr(H).$$

- If $\Pr(H \mid E) < \Pr(H)$, then E *dis*confirms H, and if $\Pr(H \mid E) = \Pr(H)$, then E is *irrelevant to* H.

- There are *many* logically equivalent (but syntactically different) ways of saying that E confirms H. Here are three of these ways:
  - E confirms H iff $\Pr(H \mid E) > \Pr(H)$.
  - E confirms H iff $\Pr(E \mid H) > \Pr(E \mid {\sim}H)$.
  - E confirms H iff $\Pr(H \mid E) > \Pr(H \mid {\sim}E)$.

- By taking differences, ratios, *etc.*, of the left/right sides of such inequalities, *many quantitative* Bayesian *relevance measures* $\mathfrak{c}(H, E)$ of the *degree* to which E confirms H can be constructed.

Measuring Factor 2: Degrees of Confirmation II

- *Dozens* of $\mathfrak{c}$'s have been proposed in the literature. Here are the four most popular measures (each based on one of the three inequalities above, and each defined on a $[-1, +1]$ scale, for ease of comparison).
  - The *Difference*: $d(H,E) = \Pr(H \mid E) - \Pr(H)$
  - The *Ratio*: $r(H,E) = \dfrac{\Pr(H \mid E) - \Pr(H)}{\Pr(H \mid E) + \Pr(H)}$
  - The *Likelihood-Ratio*: $l(H,E) = \dfrac{\Pr(E \mid H) - \Pr(E \mid {\sim}H)}{\Pr(E \mid H) + \Pr(E \mid {\sim}H)}$
  - The *Normalized-Difference*:

$$s(H,E) = \Pr(H \mid E) - \Pr(H \mid {\sim}E) = \frac{1}{\Pr({\sim}E)} \cdot d(H,E)$$

- *A fortiori*, *all* Bayesian confirmation measures agree on *qualitative* judgments like "E confirms/disconfirms/is irrelevant to H". But, these measures *disagree* with each other in various ways — *comparatively*.

Measuring Factor 2: Degrees of Confirmation III

- Consider the following two propositions concerning a card c, drawn at random from a standard deck of playing cards:

E: c is the ace of spades. H: c is *some* spade.

- I take it as intuitively clear and uncontroversial that ($K = \top$ is omitted):

(S_1) The degree to which E supports $H \neq$ the degree to which H supports E, since $E \vDash H$, but $H \nvDash E$. Intuitively, we have $\mathfrak{c}(H, E) \gg \mathfrak{c}(E, H)$.

(S_2) The degree to which E confirms $H \neq$ the degree to which $\sim E$ *dis*confirms H, since $E \vDash H$, but $\sim E \nvDash \sim H$. Intuitively, $\mathfrak{c}(H, E) \gg -\mathfrak{c}(H, \sim E)$.

- Therefore, *no adequate relevance measure of support* $\mathfrak{c}$ *should be such that either* $\mathfrak{c}(H, E) = -\mathfrak{c}(H, \sim E)$ *or* $\mathfrak{c}(H, E) = \mathfrak{c}(E, H)$ (for all E and H and all Pr-functions). I'll call these two desiderata S_1 and S_2, respectively.

- Note: $r(H, E) = r(E, H)$ and $s(H, E) = -s(H, \sim E)$. So, r violates S_1 and s violates S_2. d and l satisfy these desiderata. [This is interesting, *if* such symmetry desiderata hold for measures of *evidential support*.]

Measuring Factor 2: Degrees of Confirmation IV

- There is a relatively simple way of narrowing the field of competing measures of degree of confirmation, which is based on *thinking of inductive logic as a generalization of deductive logic*.
- The likelihood-ratio measure l stands out from the other relevance measures in the literature, since l is the only relevance measure that gets the (non-trivial) deductive cases right (as cases of *extreme relevance*).
- That is, l is the only measure (defined on the scale $[-1, +1]$) that satisfies:

$$\mathfrak{c}(H, E) \text{ should be } \begin{cases} +1 & \Leftarrow E \text{ entails } H \text{ (non-trivially).} \\ > 0 \text{ (confirmation)} & \Rightarrow \Pr(H \mid E) > \Pr(H). \\ = 0 \text{ (irrelevance)} & \Rightarrow \Pr(H \mid E) = \Pr(H). \\ < 0 \text{ (disconfirmation)} & \Rightarrow \Pr(H \mid E) < \Pr(H). \\ -1 & \Leftarrow E \text{ entails } {\sim}H \text{ (non-trivially).} \end{cases}$$

- Here, we assume that $\mathfrak{c}$ is *defined*, which constrains the unconditional Pr's.

Measuring Factor 2: Degrees of Confirmation V

- Here's how our 4 relevance measures handle non-trivial deductive cases.

- $$l(H,E) = \begin{cases} +1 & \text{if } E \vDash H,\ \Pr(E) > 0,\ \Pr(H) \in (0,1) \\ -1 & \text{if } E \vDash \sim H,\ \Pr(E) > 0,\ \Pr(H) \in (0,1) \end{cases}$$

- $$d(H,E) = \begin{cases} \Pr(\sim H) & \text{if } E \vDash H,\ \Pr(E) > 0 \\ -\Pr(H) & \text{if } E \vDash \sim H,\ \Pr(E) > 0 \end{cases}$$

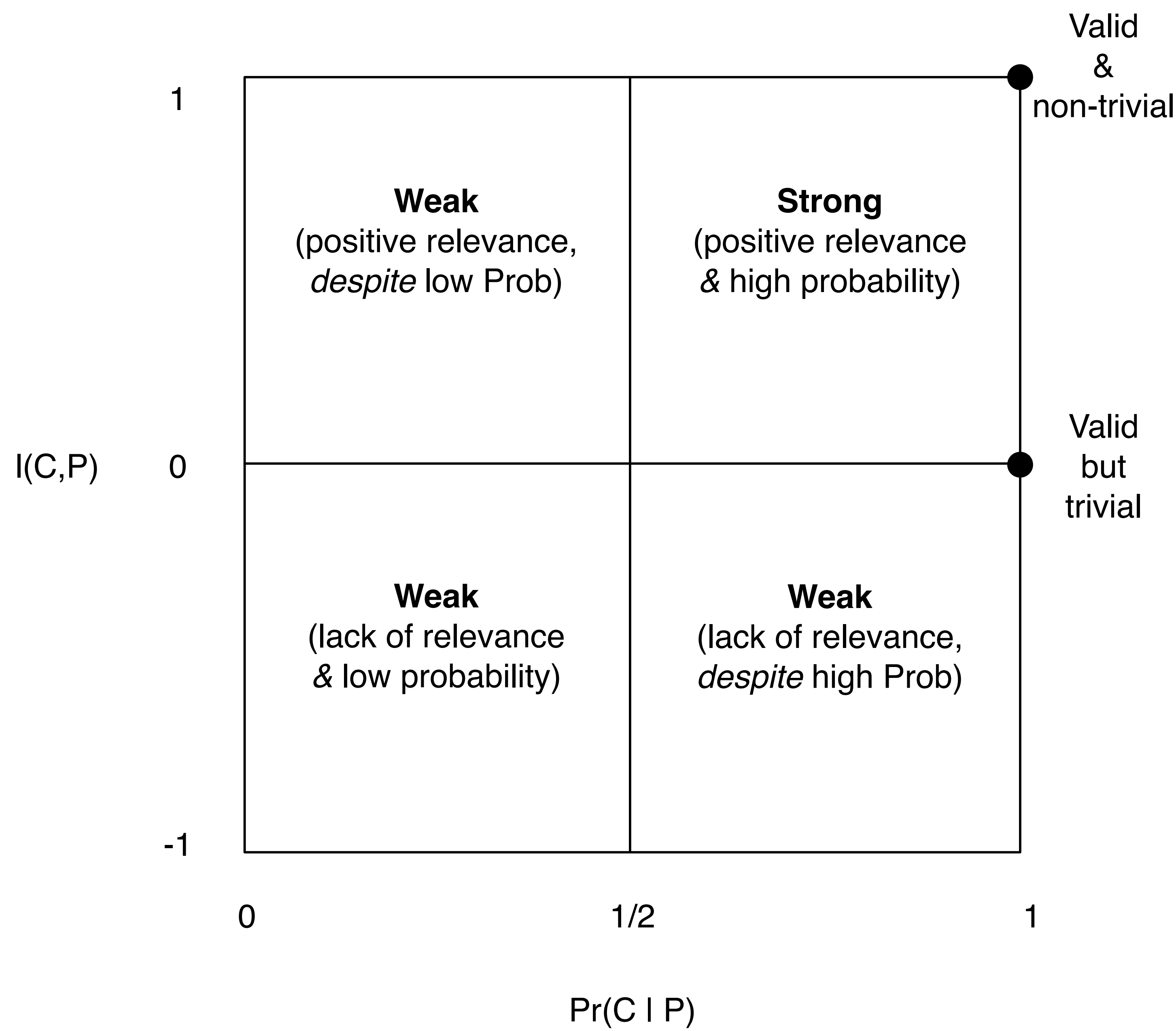
- $$r(H,E) = \begin{cases} \dfrac{1 - \Pr(H)}{1 + \Pr(H)} & \text{if } E \vDash H,\ \Pr(E) > 0,\ \Pr(H) > 0 \\ -1 & \text{if } E \vDash \sim H,\ \Pr(E) > 0,\ \Pr(H) > 0 \end{cases}$$

- $$s(H,E) = \begin{cases} \Pr(\sim H \mid \sim E) & \text{if } E \vDash H,\ \Pr(E) \in (0,1) \\ -\Pr(H \mid \sim E) & \text{if } E \vDash \sim H,\ \Pr(E) \in (0,1) \end{cases}$$

- From an inductive-logical point of view, this favors l over the other measures. Other considerations can also be used to narrow the field.

Can We Measure *Argument Strength* (Numerically)? I

- We know how to measure Factor #1 — this is just the conditional probability of the conclusion, given the premise: $\Pr(C \mid P)$.
- We have some idea of how we might go about measuring Factor #2 — a measure like $l(C, P)$ seems a plausible candidate. Let's run with that.
- This allows us to give a *numerical* version of our "Two-Factor" Chart for graphing the two components of argument strength (next slide).
- Every argument will have associated with it an *ordered pair/vector*: $\langle \Pr(C \mid P), l(C, P) \rangle$, which records values for both Factors.
- However, it is not at all clear how we might *combine* these two measures to yield a *single measure* of *overall* argument strength.
- Presumably, such a measure would be *some function f of* $\Pr(C \mid P)$ and $l(C, P)$. The challenge is to say *which function f is*. Let's think about this a bit, by thinking about shapes of the function in the 4 quadrants.


Valid
&
non-trivial
1
Weak
(positive relevance,
despite low Prob)
Strong
(positive relevance
& high probability)
Valid
but
trivial
I(C,P)
0
Weak
(lack of relevance
& low probability)
Weak
(lack of relevance,
despite high Prob)
-1
0
1/2
1
Pr(C | P)

Probabilism and The Accuracy of Credences I

- Many philosophers have argued for **Probabilism**, which is the claim that one's degrees of confidence (*i.e.*, one's credences) *should obey the probability calculus*. I will discuss one argument for probabilism.
- In epistemology (the theory of knowledge and rational belief), it is typical to suppose that *accuracy* in one's judgments is a virtue.
- For instance, when it comes to (qualitative) *belief*, it is better to have true beliefs than false beliefs. If a belief is false, then it *misrepresents* the world, and this is generally agreed to be (epistemically) *bad*.
- Something similar can be said for credences. Here is a principle.

  The Principle of Gradational Accuracy (qualitaitve rendition). One ought to be more confident in truths than in falsehoods.

- Ideally, one would assign maximal confidence to all the truths and minimal confidence to all the falsehoods (think: omniscient agents).

Probabilism and The Accuracy of Credences II

- Of course, it would be far too strong to require all rational agents to live up to this ideal. But, we can use this ideal notion to generate an interesting argument for probabilism.
- Let's call the ideal credence function (in a possible world) the vindicated credence function. I will use $v_w(\cdot)$ to denote this ideal function.

$$v_w(p) = \begin{cases} 1 & \text{if } p \text{ is true in } w, \\ 0 & \text{if } p \text{ is false in } w. \end{cases}$$

- We can use $v_w(\cdot)$ to state a quantitative form of the PGA.

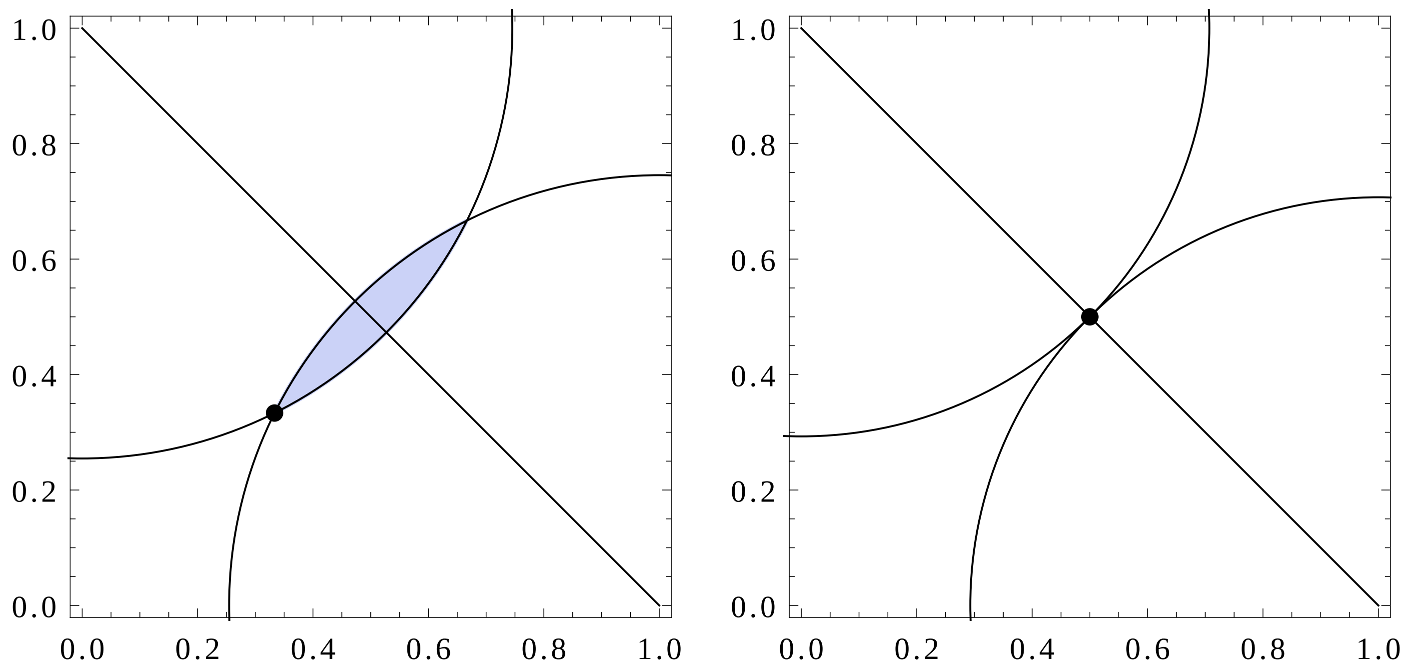
  The Principle of Gradational Accuracy (PGA, *quantitative* rendition). The closer a crecence function $b(\cdot)$ is to $v_w(\cdot)$, the better.

- To precisify PGA, we need a way to measure the *distance* between a credence function $b(\cdot)$ and the vindicated/ideal function $v_w(\cdot)$.

Probabilism and The Accuracy of Credences III

- Because we are only dealing with finite probability spaces, $b(\cdot)$ and $v_w(\cdot)$ will always be representable as *finite vectors of real numbers*.
- So, distance between $b(\cdot)$ and $v_w(\cdot)$ is just distance between finite vectors of real numbers. A very natural way to measure the distance between such vectors is *via* (squared) *Euclidean distance*.
- To make things easy, let's focus on the simplest possible example. Suppose we're assigning credences over a language with one atomic sentence: P. This means we'll have just *two states*: $\{P, \sim P\}$.
- So, any assignment of credence in this case will consist of vector containing two numbers: $\langle b(P), b(\sim P)\rangle$. This means we can visualize all such credences *via* a two-dimensional plot.
- On the next slide, I use such a plot to explain the simplest case of what I will call *the accuracy dominance argument for probabilism*.

Probabilism and The Accuracy of Credences IV



- The diagonal lines are the *probabilistic* b's (on $\langle P, {\sim}P\rangle$). The point $\langle 1, 0\rangle$ ($\langle 0, 1\rangle$) corresponds to the values assigned by $v_w(\cdot)$ in the P (${\sim}P$) world.

  Theorem (de Finetti). b is *non*-probabilistic $\Leftrightarrow$ there exists a $b'(\cdot)$ which is (Euclidean) *closer to* $v_w(\cdot)$ *in every possible world*.

- The plot on the left (right) explains the $\Rightarrow$ ($\Leftarrow$) direction.