Announcements and Such

- Administrative Stuff
  - **HW #4 is due today (by midnight, *via* Blackboard)**
    - ∗ This one consists of six (6) validity testing problems. The last 3 *require* the "short method" (either method is OK for the first 3).
    - ∗ **Consult my "Short Method" handout for detailed examples and their presentation. Please follow my guidelines for answers.**
  - **HW #5 has been posted. It's due on April 8**
    - ∗ This HW consists of two sets of exercises from Skyrms's Chapter 2 (which you should have read by now).
  - The times & locations are now known for our Final Exam
    - ∗ Morning Section: **8-10am, April 28 @ Dodge Hall 150**
    - ∗ Afternoon Section: **8-10am, April 29 @ Dodge Hall 119**
- Unit #4 — *Probability & Inductive Logic, Continued*

The Four Rules/Assumptions of Probability Calculus

1. The algebraic definition of *unconditional* probability, in terms of the (basic) probabilities assigned to the states [$\Pr(s_i) = a_i$].

$$\Pr(p) \stackrel{\text{def}}{=} \sum_{s_i \vDash p} \Pr(s_i) = \sum_{s_i \vDash p} a_i$$

2. The def. of *conditional* probability (in terms of unconditional Pr).

$$\Pr(p \mid q) \stackrel{\text{def}}{=} \frac{\Pr(p \,\&\, q)}{\Pr(q)}, \text{ provided that } \Pr(q) > 0.$$

3. Each of the a_i must lie on the unit interval $[0, 1]$.

$$a_1, \ldots, a_{2^n} \in [0, 1]$$

4. The a_i must sum to 1.

$$\sum_{i=1}^{2^n} a_i = 1$$

The Algebraic Method for Probabilisitic Reasoning

- There are two kinds of probabilistic reasoning we'll be doing: *numerical* and *abstract* (or general). Both involve *two steps*.
  - **First Step**: **Algebraic Translation**. Translate probabilistic expressions or claims into their algebraic equivalents (using our definitions of unconditional and conditional probability).
  - **Second Step**: **Algebraic Calculation**. Here, we **either** "*plug & chug*" a given *numerical* assignment to the state probabilities a_i **or** we *reason generally/asbtractly* about the a_i, using our two constraints.
- In the numerical case, we're just *calculating* numerical probabilities.
- In the abstract/general case, we're **either** *proving a general theorem* of probability calculus, **or** we're *providing a counterexample* to a general probabilistic claim (which shows it is *not* a general Pr-theorem).
- We'll be doing various examples of both kinds.

Skyrms's Two Accounts of Inductive Argument Strength

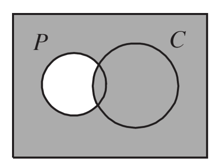
- Skyrms considers the following proposal for "inductive strength":

  Proposal #1. An argument $P \therefore C$ is inductively strong just in case the claim $P \to C$ is *probable*.

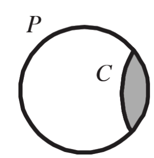
- This first proposal is inadequate, since an argument will be judged as strong if P is improbable (or C is probable). He moves to the following:

  Proposal #2. An argument $P \therefore C$ is inductively strong just in case C is probable, *given that* (*i.e.*, *on the supposition that*) P is true.

- It helps to visualize examples in which these two proposals *diverge*.



vs

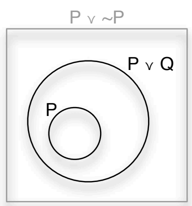


Proposal #1 (strong) Proposal #2 (weak)

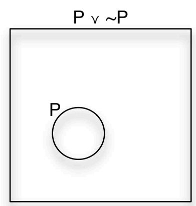
Both Proposals Get Our Foundational Example Right

(1) $P \vee Q.$ $\therefore P$ (2) $P \vee {\sim}P.$ $\therefore P$

- We can picture these two arguments (probabilistically), as follows.



(1)



(2)

- If we apply *either proposal #1 or proposal #2* to (1) and (2), we get the intuitively correct verdict that (1) *is stronger than* (2).
- We can *prove both*: (I) $\Pr[(P \vee Q) \to P] > \Pr[(P \vee {\sim}P) \to P]$, *and* (II) $\Pr(P \mid P \vee Q) > \Pr(P \mid P \vee {\sim}P)$. Let's prove these now...

Generalizing Skyrms's Objection to Proposal #1

- The probability of C ***given that*** P is a much better guide to the inductive strength of "$P \therefore C$" than the probability of $P \to C$.
- But, there is still something lacking in Skyrms's second proposal.
- This defect can be illustrated *via* the following inductive argument.

  (P) Fred Fox (who is a man) is on birth control pills.

  Therefore, (C) Fred Fox (who is a man) will not get pregnant.
- The probability of C *given that* P is very high (as is the probability that $P \to C$). So, proposal #2 (and proposal #1) says "$P \therefore C$" is *strong*.
- But, intuitively, P is *irrelevant* to C, and so (intuitively) P *does not provide evidence in favor of* C. This suggests a third proposal.

  Proposal #3. "$P \therefore C$" is strong just in case (1) the probability of C *given that* P is *high*, *and* (2) P is *positively relevant to* C — *i.e.*, the probability of C *given that* P is high***er*** than the probability of C.

Adding a *Relevance Requirement* to Skyrms's Proposal #2

- My third proposal adds a *positive relevance* requirement.
- It is helpful to think about examples involving *games of chance*. Suppose card c is going to be sampled from a standard deck of cards.
- The probability that c is a spade (C), *given that* c is black (P) is $\frac{1}{2}$. I will abbreviate this *conditional probability* claim as: $\Pr(C \mid P) = \frac{1}{2}$.
- This is *not high* (*i.e.*, it is *not greater than* $\frac{1}{2}$). But, it *is* high***er*** than the probability that c is a spade (*i.e.*, the probability of C), which is $\frac{1}{4}$.
- So, in this case, P *is positively relevant to* C (note: the probability of C *given that* $P = \frac{1}{2}$, which is greater than the probability of $C = \frac{1}{4}$).
- So, "$P \therefore C$" does *not* come out *strong* on proposal #3, since $\Pr(C \mid P)$ is *not* high. But it *does* satisfy the *positive relevance* requirement. That is:
  - $\Pr(C \mid P) = \frac{1}{2}$, which is *not* high.
  - But, $\Pr(C \mid P) > \Pr(C) = \frac{1}{4}$, so P *is positively relevant to* C.

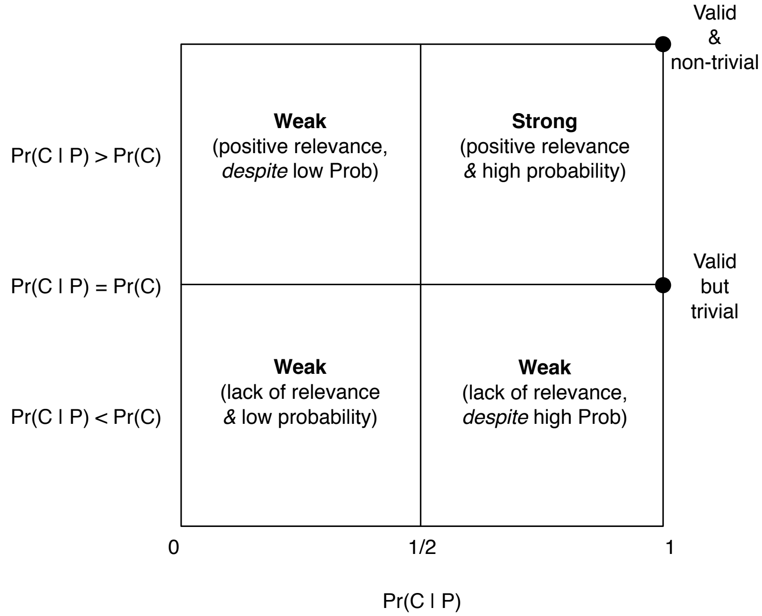
Our "Two Factor" Approach For Inductive Strength

- With our $\Pr(\cdot)$ notations in hand, we can formally state proposal #3.

  Proposal #3. An argument $P \therefore C$ is *inductively strong* iff

  (1) C is probable, *given* P, *i.e.*, $\Pr(C \mid P) > \frac{1}{2}$, *and*

  (2) P is *positively relevant* to C, *i.e.*, $\Pr(C \mid P) > \Pr(C)$.
- This proposal is superior to Skyrms's, as it requires *both* that $\Pr(C \mid P)$ be *high* ($> \frac{1}{2}$), *and* that $\Pr(C \mid P)$ be high*er* than $\Pr(C)$. This means P has to *raise the probability of* C *to a number that is greater than* $\frac{1}{2}$.
- I won't offer a measure of *degree* of inductive strength [$\mathfrak{c}(C, P)$], but, presumably, $\mathfrak{c}(C, P)$ would be some function of $\Pr(C \mid P)$ and $\Pr(C)$.
- The important thing is that one must think about *two factors* when assessing whether an argument $\ulcorner P \therefore C \urcorner$ is strong.
  - **Factor #1**. $\Pr(C \mid P)$ must be *high* ($> \frac{1}{2}$).
  - **Factor #2**. $\Pr(C \mid P)$ must be high***er*** than $\Pr(C)$.



Inductive Strength and Trivial *vs* Non-Trivial Validities

- Usually, valid arguments will come out strong on our proposal. This is because most valid arguments are non-trivial, and so they satisfy both of our requirements of high probability and positive relevance.
- But, there are some (fringe) validities that are *trivial*, and *not* strong.
- Recall that the following two LSL arguments are *valid*:

  (i) $P \,\&\, {\sim}P \therefore Q$.

  (ii) $P \therefore Q \vee {\sim}Q$.
- I will call these *trivial* validities, because they do *not* count as strong on our *two-factor* approach. Argument (ii) is easier to think about.
- What is $\Pr(Q \vee {\sim}Q \mid P)$? Because $Q \vee {\sim}Q$ is a *tautology*, we will have $\Pr(Q \vee {\sim}Q \mid P) = 1$. So, on Proposal #2, argument (ii) comes out strong. But, this is incorrect, since $\Pr(Q \vee {\sim}Q \mid P) = \Pr(Q \vee {\sim}Q) = 1$.
- That is, P is *irrelevant to* $Q \vee {\sim}Q$. So, (ii) is *not* a strong argument.

Non-Monotonicity of Probability/Inductive Strength I

- A crucial difference between validity *vs* strength is that deductive validity is *monotonic*, but inductive strength is *non*-monotonic.
- If an argument is valid, then it cannot be rendered invalid merely by adding additional premises to it (remind yourself why this is true).
- On the other hand, an inductively strong argument *can* be rendered inductively weak merely by adding premises to it. Example, ($\mathcal{A}$):

  (P) It is January.

  $\therefore$ (C) If I walk out onto Lake Mendota, I will not get wet.
- Usually, Lake Mendota is frozen through in January. So, ($\mathcal{A}$) is *strong*.

  (P) It is January.

  (Q) Lake Mendota is not frozen through.

  $\therefore$ (C) If I walk out onto Lake Mendota, I will not get wet.
- This argument is *weak*. So, adding premise (Q) to ($\mathcal{A}$) *made it weak*.

Non-Monotonicity of Probability/Inductive Strength II

- In fact, inductive arguments can oscillate in strength as new premises are added. Here is an example that illustrates the phenomenon.

  (P) John has lung cancer.

  $\therefore$ (C) John is (or was) a smoker.
- By learning more things about John (*i.e.*, by adding premises about John), the strength of this argument can *oscillate.* Consider:

  (Q) John has a certain gene that would make him susceptible to lung cancer even if he were not a smoker.

  (R) John has yellow stains on his fingers and smells like tobacco.

  (S) John rolls cigars for a living.
- And so on... This non-monotonicity of inductive strength is a key feature. Diagramming can help explain why this can happen...

A Real-Life Example of Non-Monotonicity

- After the O.J. trial, Alan Dershowitz remarked on T.V. that "fewer than 1 in 1,000 women who are abused by their mates go on to be killed by them".
- He suggested that "the *probability*" that Nicole Brown Simpson (N.B.S.) was killed by mate (O.J.) — *given that he abused her* — was less than 1 in 1,000.
- Presumably, this was supposed to have some consequences for people's *degrees of confidence* (*degrees of belief*) in the hypothesis of O.J.'s guilt.
- The debate that ensued provides a nice example of non-monotonicity.
- Let A be the proposition that N.B.S. is abused by her mate (O.J.), let K be the proposition that N.B.S. is killed by her mate (O.J.). Dershowitz claims:

  (1) The probability that K is true, *given* A, is very small (less than 1/1000).
- Thus, the argument from A to K is *inductively (very) weak.*
- I.J. Good wrote a brief response in *Nature.* Good pointed out that, while Dershowitz's (1) may be true, *it leaves out some crucial known premises.*

- Good argues that what's relevant here is not the probability that N.B.S. was killed by O.J., given that she was abused by O.J., but the probability that she was killed by O.J., given that she was abused by O.J. *and that she was killed.*
- After all, we do know that Nicole was killed, and (plausibly) this information should be taken into account in our probabilistic musings about the case.
- So, let K' be the proposition that N.B.S was killed (by *someone*). Using Dershowitz's (1) as a starting point, Good does some "back-of-the-envelope calculations," and he comes up with the following "guesstimate":

  (2) The probability that K is true, *given both* A *and* K', is approximately 1/2, which is much greater than 1/1000.
- This would seem to make it far more probable that O.J. is the killer than Dershowitz's claim would have us believe. Independently, and using statistical data about murders committed in 1992, Merz & Caulkins show:

  (3) The probability that K is true, *given both* A *and* K', is approximately 4/5.
- This would seem to provide us with an *even greater* "estimate" of "the probability" that N.B.S. was killed by O.J.

- Dershowitz replied to analyses like those of Good and Merz & Caulkins with the following sort of rejoinder:

  > ...whenever a woman is murdered, it is highly likely that her husband or her boyfriend is the murderer without regard to whether battery preceded the murder. The key question is how salient a characteristic is the battery as compared with the relationship itself. Without that information, the 80 percent figure [as in Merz & Caulkins' estimation] is meaningless. I would expect that a couple of statisticians would have spotted this fallacy.
- Dershowitz's rejoinder seems to trade on something like the following:

  (4) The probability that K is true, *given both* A *and* K', is approximately the same as the probability that K is true, *given* K' *alone.*
- Not to be outdone, Merz & Caulkins give the following "estimate" of the salient probabilities (again, their "estimate" is based on statistics for 1992):

  (5) The probability that K is true, *given* K' *alone*, is approximately 0.29 — *far less than* the probability that K is true, *given both* A *and* K' (≈ 0.8).
- We could continue this *ad nauseum.* But, I will not.

Two Key Differences Between Our "Two Factors" I

- The two factors that go into determining whether an inductive argument is *strong* are different in some crucial ways.

  The Conjunction Condition. If a claim (X) constitutes a strong argument for a conjunction ($Y \,\&\, Z$), then X also constitutes a strong argument for each of its conjuncts (Y, Z).
- Factor #1 *satisfies* The Conjunction Condition, because, in general

$$\Pr(Y \,\&\, Z \mid X) > \tfrac{1}{2} \Longrightarrow \Pr(Y \mid X) > \tfrac{1}{2} \text{ and } \Pr(Z \mid X) > \tfrac{1}{2}.$$

  – Because: $\Pr(Y \mid E) \geq \Pr(Y \,\&\, Z \mid E)$ and $\Pr(Z \mid E) \geq \Pr(Y \,\&\, Z \mid E)$.
- Factor #2 can *violate* The Conjunction Condition. That is:

$$\Pr(P \,\&\, Q \mid E) > \Pr(P \,\&\, Q) \not\Rightarrow \Pr(P \mid E) > \Pr(P).$$

  – Let E = card is black, P = card is an ace, and Q = card is a spade.

Two Key Differences Between Our "Two Factors" II

- Another crucial difference between our Two Factors involves
- **The Disjunction Condition** (DC). If $P \therefore X$ is a strong argument, and $P \therefore Y$ is a strong argument, then $P \therefore X \vee Y$ is a strong argument.
- If we measure strength using *only* Factor 1, then (DC) is true. This is because of the following fact (which we'll prove using our PTT method).

$$\text{If } \Pr(X \mid P) > \tfrac{1}{2} \text{ and } \Pr(Y \mid P) > \tfrac{1}{2}\text{, then } \Pr(X \vee Y \mid P) > \tfrac{1}{2}.$$

- But, if we think about the Factor 2 component of strength, then (DC) can *fail*. That is to say, there are examples (see next slide) in which
  - $\Pr(X \mid P) > \Pr(X)$ [$13/22 > 1/2$]
  - $\Pr(Y \mid P) > \Pr(Y)$ [$6/11 > 1/2$]
  - $\Pr(X \vee Y \mid P) < \Pr(X \vee Y)$ [$9/11 < 7/8$]

State (s_i)	P	X	Y	$\Pr(s_i) = a_i$
s_1	$\top$	$\top$	$\top$	$\Pr(s_1) = a_1 = \frac{7}{64}$
s_2	$\top$	$\top$	$\bot$	$\Pr(s_2) = a_2 = \frac{6}{64}$
s_3	$\top$	$\bot$	$\top$	$\Pr(s_3) = a_3 = \frac{5}{64}$
s_4	$\top$	$\bot$	$\bot$	$\Pr(s_4) = a_4 = \frac{4}{64}$
s_5	$\bot$	$\top$	$\top$	$\Pr(s_5) = a_5 = \frac{1}{64}$
s_6	$\bot$	$\top$	$\bot$	$\Pr(s_6) = a_6 = \frac{18}{64}$
s_7	$\bot$	$\bot$	$\top$	$\Pr(s_7) = a_7 = \frac{19}{64}$
s_8	$\bot$	$\bot$	$\bot$	$\Pr(s_8) = a_8 = \frac{4}{64}$