Announcements and Such

- Administrative Stuff
  - **HW #4 grades and solutions have been posted**
    - People (generally) did pretty well on this HW.
  - **HW #5 is due tonight (by midnight, *via* Blackboard)**
    - This HW consists of two sets of exercises from *Skyrms's Chapter 2*.
    - These are *informal* exercises — you're not meant to apply our theoretical/probabilistic analyses of argument strength here.
  - **HW #6 has been posted (it's due in 2 weeks – on 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 next 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*

Theoretical Comparison of Our "Two Factors": Summary

	Does Factor satisfy property?	
Property	Factor 1?	Factor 2?
The Conjunction Condition	YES	NO
The Disjunction Condition	YES	NO
The Sure Thing Principle	YES	NO
$\dfrac{P}{\therefore Q \vee {\sim}Q}$ is *weak*.	NO	YES
$\dfrac{P \,\&\, {\sim}P}{\therefore Q}$ is *weak*.	YES	YES
$\dfrac{{\sim}X}{\therefore X}$ is *weak*.	YES	YES
$\dfrac{P \vee Q}{\therefore P}$ is (generally) strong*er* than $\dfrac{P \vee {\sim}P}{\therefore P}$	YES	YES
The Unconditional Sure Thing Principle	YES	YES

Objective (Physical) Interpretations of Probability

- The simplest physical interpretation of probability interprets probabilities as finite relative (actual) frequencies of events.
- All finite relative frequencies are probabilities, but the converse does not hold. There can be *irrational-valued* (objective/physical) probabilities.
- Irrational values *can* be achieved as *limiting* relative frequencies, in *hypothetical infitite extensions* of (actual, finite) experiments.
- But, nothing guarantees that such limiting frequencies always exist (or that they always converge to the objectively correct values).
- So, some deeper physical property of systems is required to ensure (a) the existence of these limiting relative frequencies, and (b) their correct convergence. These properties are called *propensities* (or *chances*).
- Propensities are analogous to other physical properties (like mass). They are *reflected* in (finite, actual, observed) relative frequencies, but they are *not identical to* these (finite, actual, observed) relative frequencies.

"Subjective" Interpretations of Probability

- We often make judgments regarding the likelihood of events. These judgments involve *degrees of confidence in propositions.*
- Degrees of confidence can be *reported directly* (as we'll see below), or they can be *inferred from behavior* (*e.g.*, from betting behavior).
- There are various arguments that can be given in support of the claim that these "degrees of confidence" (a.k.a., *credences*) *ought to* obey the laws of the probability calculus (*i.e.*, the formal principles we've learned).
- We won't discuss these general arguments for "probabilism" here. But, we will consider some simple examples of probabilistic constraints that seem correct (*i.e.*, legislative) for degrees of confidence.
- However, we will see that even very simple constraints such as these are often *violated* — even by expert judges. Such violations of simple probabilistic laws are often called "reasoning fallacies". We'll discuss two.

Inverse Probability and Bayes's Theorem

- $\Pr(H \mid E)$ is called the *posterior* H (on E). $\Pr(H)$ is called the *prior* of H. $\Pr(E \mid H)$ is called the *likelihood* of H (on E).
- By the definition of $\Pr(\bullet \mid \bullet)$, we can write the posterior and likelihood as:

$$\Pr(H \mid E) = \frac{\Pr(H \,\&\, E)}{\Pr(E)} \qquad \text{and} \qquad \Pr(E \mid H) = \frac{\Pr(H \,\&\, E)}{\Pr(H)}$$

- So, the posterior and the likelihood are related by *Bayes's Theorem*:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E)}$$

- **Law of Total Probability**. If $\Pr(H)$ is non-extreme, then:

$$\begin{aligned}\Pr(E) &= \Pr((E \,\&\, H) \vee (E \,\&\, {\sim}H)) \\ &= \Pr(E \,\&\, H) + \Pr(E \,\&\, {\sim}H) \\ &= \Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid {\sim}H) \cdot \Pr({\sim}H)\end{aligned}$$

- This allows us to write a more perspicuous form of *Bayes's Theorem*:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid {\sim}H) \cdot \Pr({\sim}H)}$$

Our Two Factors and The Base Rate Fallacy

- Here's a famous example, illustrating the subtlety of Bayes's Theorem:

  The (unconditional) probability of breast cancer is 1% for a woman at age forty who participates in routine screening. The probability of such a woman having a positive mammogram, given that she has breast cancer, is 80%. The probability of such a woman having a positive mammogram, given that she does not have breast cancer, is 10%. What is the probability that such a woman has breast cancer, given that she has had a positive mammogram in routine screening?

- We can formalize this, as follows. Let H = such a woman (age 40 who participates in routine screening) has breast cancer, and E = such a woman has had a positive mammogram in routine screening. Then:

$$\Pr(E \mid H) = 0.8,\ \Pr(E \mid {\sim}H) = 0.1, \text{ and } \Pr(H) = 0.01.$$

- **Question**: What is $\Pr(H \mid E)$? What would you guess? Most experts guess a pretty high number (near 0.8, usually).

- If we apply Bayes's Theorem, we get the following answer:

$$\begin{aligned}\Pr(H \mid E) &= \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E \mid H) \cdot \Pr(H) + \Pr(E \mid {\sim}H) \cdot \Pr({\sim}H)} \\ &= \frac{0.8 \cdot 0.01}{0.8 \cdot 0.01 + 0.1 \cdot 0.99} \approx 0.075\end{aligned}$$

- We can also use our algebraic technique to compute an answer.

E	H	$\Pr(s_i)$
$\top$	$\top$	$a_1 = 0.008$
$\top$	$\bot$	$a_2 = 0.099$
$\bot$	$\top$	$a_3 = 0.002$
$\bot$	$\bot$	0.891

$$\Pr(E \mid H) = \frac{\Pr(E \,\&\, H)}{\Pr(H)} = \frac{a_1}{a_1 + a_3} = 0.8$$

$$\Pr(E \mid {\sim}H) = \frac{\Pr(E \,\&\, {\sim}H)}{\Pr({\sim}H)} = \frac{a_2}{1 - (a_1 + a_3)} = 0.1$$

$$\Pr(H) = a_1 + a_3 = 0.01$$

- Note: The posterior is about eight times the prior in this case, but since the prior is *so* low to begin with, the posterior is still pretty low.
- This mistake is usually called the *base rate fallacy*. People tend to neglect base rates in their estimates of probability — *when E is strongly relevant to H*. Here, our Two Factors *pull in opposite directions*.

Our Two Factors and The Conjunction Fallacy

- Another infamous case in which our Two Factors pull in opposite directions is the so-called Conjunction Fallacy.
- Tversky & Kahneman discuss the following example, which was the first example of the "conjunction fallacy." Here is some evidence E:

  (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**. Is it more probable, given E, that Linda is (B) a bank teller, or ($B \,\&\, F$) a bank teller *and* an active feminist?
- Formally, the question reduces to a comparison of the following to conditional probabilities (Factor #1): $\Pr(B \mid E)$ *vs* $\Pr(B \,\&\, F \mid E)$.
- It is easy to show that: $\Pr(B \mid E) \geq \Pr(B \,\&\, F \mid E)$. But, many people answer the question by saying that $\Pr(B \mid E) < \Pr(B \,\&\, F \mid E)$.

- So, why do people commit this fallacy of probabilistic reasoning?
- We think it has to do with the distinction between conditional probability (Factor #1) and probabilistic relevance (Factor #2).
- Intuitively, E is *positively* (statistically) *relevant* to F, but E is *irrelevant* to B. As a result, it makes sense that E could be *more relevant to* $B \,\&\, F$ than it is to B. In fact, this is precisely what happens in such cases.
- To make this more precise, we can define $d(X,E) \stackrel{\text{def}}{=} \Pr(X \mid E) - \Pr(X)$.
- Then, we can use $d(X,E)$ to measure *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 $d(X,E) = 0$.
- Now, intuitively, we have the following two facts in the Linda case:
  - **Factor #1**. $\Pr(B \mid E) > \Pr(B \,\&\, F \mid E)$.
  - **Factor #2**. $d(B,E) < d(B \,\&\, F, E)$.
- Again, our Two Factors pull in opposite directions.

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

- 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 IV

- 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.

Measuring Factor 2: Degrees of Confirmation V

- 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*.]

Can We Measure *Argument Strength* (Numerically)?

- 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).
- However, it is not at all clear how we might *combine* these two measures to yield a single measure of *overall* argument strength.
- If we think of such a measure as a *function f of* $\Pr(C \mid P)$ and $l(C,P)$, then we can try to write down some desirable properties of f.
- We certainly want f to be *high* in the *upper-right* quadrant, and *low* in the *lower-left* quadrant. But, what else can we say about f?