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 on Friday (by midnight, *via* Blackboard)**
    - This HW consists of two sets of exercises from *Skyrms's Chapter 2*.
  - 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*
  - Tying up some loose ends from last week (thanks Cosmo!).
  - Objective Interpretations of Probability (see Hájek's SEP entry)
  - Inverse Probability and Bayes's Theorem
  - Our Two Factors and Two Infamous "Reasoning Fallacies"

Zero Probabilities for Logical Contingencies: An Example

- Recall this example I gave (sampling a card at random from a deck):
  - $E \stackrel{\text{def}}{=}$ card is black, $P \stackrel{\text{def}}{=}$ card is an ace, and $Q \stackrel{\text{def}}{=}$ card is a spade.
- Here is the full probability distribution over $\{E, P, Q\}$.

State (s_i)	E	P	Q	$\Pr(s_i)$
s_1	$\top$	$\top$	$\top$	$\Pr(s_1) = a_1 = 1/52$
s_2	$\top$	$\top$	$\bot$	$\Pr(s_2) = a_2 = 1/52$
s_3	$\top$	$\bot$	$\top$	$\Pr(s_3) = a_3 = 12/52$
s_4	$\top$	$\bot$	$\bot$	$\Pr(s_4) = a_4 = 12/52$
s_5	$\bot$	$\top$	$\top$	$\Pr(s_5) = a_5 = 0$
s_6	$\bot$	$\top$	$\bot$	$\Pr(s_6) = a_6 = 2/52$
s_7	$\bot$	$\bot$	$\top$	$\Pr(s_7) = a_7 = 0$
s_8	$\bot$	$\bot$	$\bot$	$\Pr(s_8) = a_8 = 24/52$

- Note that two of these "states" have *zero* probability (s_5 and s_7). These two "states" would correspond to situations in which the card was a *non-black spade. But, there are no non-black spades in standard decks.*
- Strictly speaking, these states are *not **logical** impossibilities.* It is *not a **logical** contradiction* for a card to be a non-black spade.
- Analogy back to Part I of the course: *it is not a **logical** contradiction* (*i.e., not a contradiction **in terms***) for a person to be a married bachelor. But, we know (by conventional meaning) that *there are no married bachelors.*

☞ Because the rules of probability calculus allow logical contingencies to receive probability zero, we may encode implications of (known) conventional meanings *via* probability assignments.

- $\Pr(P \,\&\, Q \mid E) \stackrel{\text{def}}{=} \frac{\Pr(P \& Q \& E)}{\Pr(E)} = \frac{a_1}{a_1+a_2+a_3+a_4} = \frac{1/52}{1/52+1/52+12/52+12/52} = 1/26$
- $\Pr(P \,\&\, Q) \stackrel{\text{def}}{=} a_1 + a_5 = 1/52 + 0 = 1/52$
- $\Pr(P \mid E) \stackrel{\text{def}}{=} \frac{\Pr(P \& E)}{\Pr(E)} = \frac{a_1+a_2}{a_1+a_2+a_3+a_4} = \frac{1/52+1/52}{1/52+1/52+12/52+12/52} = 2/26 = 1/13$
- $\Pr(P) \stackrel{\text{def}}{=} a_1 + a_2 + a_5 + a_6 = 1/52 + 1/52 + 0 + 2/52 = 4/52 = 1/13$

Theoretical Comparison of Our "Two Factors" III

- Here is another property satisfied by Factor 1, but not Factor 2.

  The Sure Thing Principle. If X constitutes a strong argument for Z *given* Y and X constitutes a strong argument for Z *given* ${\sim}Y$, then X constitutes a strong argument for Z (*unconditionally*).

- The reason Factor 1 satisfies The Sure Thing Principle is that, in general

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

- Let's prove this claim using our algebraic method.
- Factor 2 can *violate* The Sure Thing Principle. In other words,

$$[\Pr(Z \mid X \,\&\, Y) > \Pr(Z \mid Y) \text{ and } \Pr(Z \mid X \,\&\, {\sim}Y) > \Pr(Z \mid {\sim}Y)] \not\Rightarrow \Pr(Z \mid X) > \Pr(Z).$$

- See the next slide for an "urn-style" counterexample.

State (s_i)	X	Y	Z	$\Pr(s_i)$
s_1	T	T	T	$\Pr(s_1) = a_1 = \frac{31}{192}$
s_2	T	T	⊥	$\Pr(s_2) = a_2 = \frac{59}{192}$
s_3	T	⊥	T	$\Pr(s_3) = a_3 = \frac{40}{192}$
s_4	T	⊥	⊥	$\Pr(s_4) = a_4 = \frac{14}{192}$
s_5	⊥	T	T	$\Pr(s_5) = a_5 = \frac{1}{192}$
s_6	⊥	T	⊥	$\Pr(s_6) = a_6 = \frac{5}{192}$
s_7	⊥	⊥	T	$\Pr(s_7) = a_7 = \frac{24}{192}$
s_8	⊥	⊥	⊥	$\Pr(s_8) = a_8 = \frac{18}{192}$

Theoretical Comparison of Our “Two Factors” IV

- The fact that Factor 2 can violate The Sure Thing Principle is known as “Simpson’s Paradox”. Here is a real-life example from a medical study comparing the success rates of two treatments for kidney stones.
- We can interpret the STT above (with X, Y, Z), as follows. Let X be the claim that a patient is given a treatment t for disease d. Let Z be the claim that a patient recovers from d. And, let Y be the claim that a patient is male. If we calculate the salient probabilities, we get:

(1) $\Pr(Z \mid X \,\&\, Y) > \Pr(Z \mid Y)$. [$^{31}/_{90} > {}^{1}/_{3}$]

(2) $\Pr(Z \mid X \,\&\, {\sim}Y) > \Pr(Z \mid {\sim}Y)$. [$^{20}/_{27} > {}^{2}/_{3}$]

(3) $\Pr(Z \mid X) < \Pr(Z)$. [$^{71}/_{144} < {}^{1}/_{2}$]

☞ (1) implies that the treatment is (somewhat) effective *for males*, and (2) implies that the treatment is (somewhat) effective *for females*. But, (3) implies that the treatment is *counter-productive for humans*!

Theoretical Comparison of Our “Two Factors” V

- Although Simpson’s Paradox implies that Factor #2 can violate The Sure Thing Principle, there is a related principle that *both* Factors *do* satisfy.

  The Unconditional Sure Thing Principle. If $X \,\&\, Y$ constitutes a strong argument for Z (unconditionally) and $X \,\&\, {\sim}Y$ constitutes a strong argument for Z (unconditionally), then X *alone* constitutes a strong argument for Z (unconditionally).

- In terms of Factor 1, The Unconditional Sure Thing Principle *is equiavlent to* The Sure Thing Principle (thus it satisfies both).
- From the point of view of Factor 2, these principles are *not* equivalent. Indeed, The Unconditional Sure Thing Principle *holds* for Factor 2, since

$$[\Pr(Z \mid X \,\&\, Y) > \Pr(Z) \text{ and } \Pr(Z \mid X \,\&\, {\sim}Y) > \Pr(Z)] \Longrightarrow \Pr(Z \mid X) > \Pr(Z).$$

- So, this disagreement trades *essentially* on the “*given*”s in the STP.

	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

A Peculiar Probability Distribution

- All of the ("urn-style") numerical probability distributions we've seen so far have involved *rational numbers*. Not all examples are like this.
- Consider the following three constraints: (1) $\Pr(Y \mid X) = \Pr(X \vee Y)$, (2) $\Pr(Y) = \Pr(\sim Y)$, (3) $\Pr(X \,\&\, Y) = \Pr(\sim X \,\&\, Y)$.

  Fact. (1)–(3) are satisfied by a *unique* numerical probability distribution, and this distribution assigns some *irrational* numbers to some states.
- In order to show this, one just needs to solve the following system of three equations in three unknowns (a_1, a_2, a_3 in the STT over $\{X, Y\}$): (1) $\dfrac{a_1}{a_1 + a_2} = a_1 + a_2 + a_3$, (2) $a_1 + a_3 = 1 - (a_1 + a_3)$, (3) $a_1 = a_3$.

X	Y	$\Pr(s_i)$
$\top$	$\top$	$a_1 = 1/4$
$\top$	$\bot$	$a_2 = \frac{1}{8}\left(\sqrt{17} - 3\right)$
$\bot$	$\top$	$a_3 = 1/4$
$\bot$	$\bot$	$a_4 = \frac{1}{8}\left(7 - \sqrt{17}\right)$

Objective Interpretations of Probability I

- The simplest objective theory is the *actual (finite) frequency* theory.
- First, we must verify that actual frequencies in finite populations satisfy the probability calculus (otherwise, they aren't *probabilities* at all!).
- Let $\mathbf{P}$ be an actual (non-empty, finite) population, let χ be a property, and let $\boldsymbol{\chi}$ denote the set of (all) objects that actually have property χ.
- Let $\#(S) \stackrel{\text{def}}{=}$ the number of objects in a set S. Using $\#(\cdot)$, we can define the actual frequency of χ in such a population $\mathbf{P}$ in the following way:
  - $f_{\mathbf{P}}(\chi) \stackrel{\text{def}}{=} \dfrac{\#(\boldsymbol{\chi} \cap \mathbf{P})}{\#(\mathbf{P})}$
- Next, let X be the proposition that an (arbitrary) object $a \in \mathbf{P}$ has property χ. Using $f_{\mathbf{P}}(\chi)$, we can define $\Pr_{\mathbf{P}}(X)$, as follows:
  - $\Pr_{\mathbf{P}}(X) \stackrel{\text{def}}{=} f_{\mathbf{P}}(\chi)$.
- We need to show that $\Pr_{\mathbf{P}}(X)$ is in fact a *probability* function. There are various ways to do this. Let's think in terms of *state descriptions*, *etc.*

Objective Interpretations of Probability II

(1) **Contradictions have probability zero**.
  - If χ is a *contradictory* property, then *nothing* in any population $\mathbf{P}$ will instantiate χ. So, by definition, we will have $\Pr_{\mathbf{P}}(X) = 0$.

(2) **The probability of any state description s_i will lie on the unit interval**.
  - Suppose we have n (logically) independent properties: $\chi_1, \ldots, \chi_n$. Then, we can form 2^n *state descriptions* $s_1, \ldots, s_{2^n}$ using the n atomic sentences $X_1, \ldots, X_n$. Each of these state descriptions will have a probability, given by our frequency definition above. For instance:
  $$\Pr_{\mathbf{P}}(s_1) = \Pr_{\mathbf{P}}(X_1 \,\&\, \cdots \,\&\, X_n) \stackrel{\text{def}}{=} \frac{\#(\boldsymbol{\chi_1} \cap \cdots \cap \boldsymbol{\chi_n} \cap \mathbf{P})}{\#(\mathbf{P})} \in [0, 1]$$

(3) **The sum of the probabilities of (all) the state descriptions equals one**.
  - By definition, the sum $\sum_{i=1}^{2^n} \Pr_{\mathbf{P}}(s_i)$ is just the proportion of objects in $\mathbf{P}$ which instantiate *some* state description. Because the state description properties form a *partition* of $\mathbf{P}$, this proportion *must equal one*.

Objective Interpretations of Probability III

- OK, so actual frequencies in populations determine *probabilities*. But, they are rather peculiar probabilities, in several respects.
- First, they are *population-relative*. If an object a is a member of multiple populations $\mathbf{P}_1, \ldots, \mathbf{P}_n$, then this may yield different values for $\Pr_{\mathbf{P}_1}(X)$, …, $\Pr_{\mathbf{P}_n}(X)$. This is related to the *reference class problem* from last time.
- Another peculiarity of finite actual frequencies is that they sometimes seem to be misleading about intuitive objective probabilities.
- For instance, imagine tossing a coin n times. This gives a population $\mathbf{P}$ of size n, and we can compute the $\mathbf{P}$-frequency-probability of heads $\Pr_{\mathbf{P}}(H)$.
- As n gets larger, the value of this frequency tends to "settle down" to some small range of values (see *Mathematica* notebook). Intuitively, none of these finite actual frequencies is exactly equal to the bias of the coin.
- So, finite frequencies seem, at best, to provide "estimates" of probabilities in some deeper objective sense. What might such a "deeper sense" be?

Objective Interpretations of Probability IV

- The *law of large numbers* ensures that (given certain underlying assumptions about the coin) the "settling down" we observe in many actual frequency cases (coin-tossing) will converge *in the limit* ($n \to \infty$).
- If we do have convergence to some value (say $\frac{1}{2}$ for a fair coin), then this value seems a better candidate for the "intuitive" objective probability. This leads to the *hypothetical limiting frequency theory* of probability.
- According to the hypothetical limiting frequency theory, probabilities are frequencies we *would* observe in a population — *if* that population were extended indefinitely (*e.g.*, if we were to toss the coin ∞ times).
- There are various problems with this theory. First, convergence is not always guaranteed. In fact, there are *many* hypothetical infinite extensions of any **P** for which the frequencies do *not* converge as $n \to \infty$.
- Second, even among those extensions that *do* converge, there can be *many different* possible convergent values. Which is "the" probability?

Objective Interpretations of Probability V

- *Propensity* or *chance* theories of probability posit the existence of a deeper kind of physical probability, which manifests itself empirically in finite frequencies, and which constrains limiting frequencies.
- Having a theory that makes sense of quantum mechanical probabilities was one of the original inspirations of propensity theorists (Popper).
- In quantum mechanics, probability seems to be a fundamental physical property of certain systems. The theory entails exact *probabilities* of certain token events in certain experimental set-ups/contexts.
- These probabilities seem to transcend both finite and infinite frequencies. They seem to be basic *dispositional properties* of certain physical systems.
- In classical (deterministic) physics, all token events are *determined* by the physical laws + initial conditions of the universe. In quantum mechanics, only *probabilities* of token events are determined by the laws + i.c.'s.
- This leaves room for (non-extreme) *objective chances* of token events.

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:

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

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

E	H	$\Pr(s_i)$
⊤	⊤	$a_1 = 0.008$
⊤	⊥	$a_2 = 0.099$
⊥	⊤	$a_3 = 0.002$
⊥	⊥	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 $d(X, E) \overset{\text{def}}{=} \Pr(X \mid E) - \Pr(X)$.
- 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.