Scientific Explanation — More Background Notes

- Administrative:
  - I've added some readings on *Inductive Logic* to our background.
- Probability and its Interpretations
  - Probability Calculus
    - Algebraic treatment (and my decision procedure `PrSAT`)
    - Axiomatic treatment
  - Some (Salient) Interpretations
    - Frequency and Propensity Interpretations
    - Subjective Interpretations (?)
- More historical background leading up to Woodward's book
  - Hempel's I–S Model
  - Salmon's S–R Model

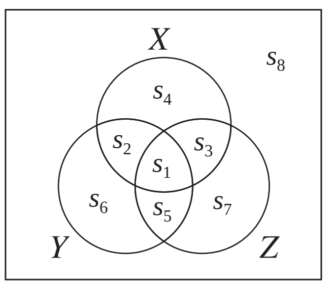
Overview of Finite Propositional Boolean Algebras I

- Consider a logical language $\mathcal{L}$ containing n atomic sentences. These may be sentence letters (X, Y, Z, *etc.*), or they may be atomic sentences of monadic or relational predicate calculus (Fa, Gb, Rab, Hcd, *etc.*).
- The Boolean Algebra $\mathcal{B}_{\mathcal{L}}$ set-up by such a language will be such that:
  - $\mathcal{B}_{\mathcal{L}}$ will have 2^n *states* (corresponding to the *state descriptions* of $\mathcal{L}$)
  - $\mathcal{B}_{\mathcal{L}}$ will contain 2^{2^n} *propositions*, in total.
    - This is because each proposition p in $\mathcal{B}_{\mathcal{L}}$ is equivalent to a disjunction of state descriptions. Thus, each subset of the set of state descriptions of $\mathcal{L}$ corresponds to a proposition of $\mathcal{B}_{\mathcal{L}}$.
    - Note: there are 2^{2^n} subsets of a set of size 2^n.
      - The empty set $\varnothing$ of state descriptions corresponds to "the empty disjunction", which corresponds to *the logical falsehood*: $\bot$.
      - Singelton sets of state descriptions correspond to "disjunctions with one member". [All other subsets are "normal" disjunctions.]

Overview of Finite Propositional Boolean Algebras II

- Example. Let $\mathcal{L}$ have three atomic sentences: X, Y, and Z. Then, $\mathcal{B}_{\mathcal{L}}$ is:

X	Y	Z	States
T	T	T	s_1
T	T	F	s_2
T	F	T	s_3
T	F	F	s_4
F	T	T	s_5
F	T	F	s_6
F	F	T	s_7
F	F	F	s_8



- Examples of reduction to disjunctions of state descriptions of $\mathcal{L}$:
  - '$X \,\&\, {\sim}X$' is equivalent to the *empty* disjunction: $\bot$.
  - '$X \,\&\, ({\sim}Y \,\&\, Z)$' is equivalent to the *singleton* disjunction: s_3.
  - '$X \leftrightarrow (Y \vee Z)$' is equivalent to: $s_1 \vee s_2 \vee s_3 \vee s_8$.
- In general: $p \dashv\vDash \bigvee\{s_i \mid s_i \vDash p\}$. And, if $\{s_i \mid s_i \vDash p\} = \varnothing$, then $p \dashv\vDash \bot$.

The Probability Calculus: An Algebraic Approach I

- Once we grasp the concept of a finite Boolean algebra of propositions, understanding the probability calculus *algebraically* is very easy.
- The central concept is a *finite probability model*. A finite probability model $\mathcal{M}$ is a finite Boolean algebra of propositions $\mathcal{B}$, together with a function $\Pr(\cdot)$ which maps elements of $\mathcal{B}$ to the unit interval $[0,1] \in \mathbb{R}$.
- This function $\Pr(\cdot)$ must be a *probability function*. It turns out that a probability function $\Pr(\cdot)$ on $\mathcal{B}$ is just a function that assigns a real number on $[0,1]$ to each state s_i of $\mathcal{B}$, such that $\sum_i \Pr(s_i) = 1$.
- Once we have $\Pr(\cdot)$'s *basic assignments* to the states of $\mathcal{B}$ (s.d.'s of $\mathcal{L}$), we define $\Pr(p)$ for *any* statement $\mathcal{L}$ of the language of $\mathcal{B}$, as follows:

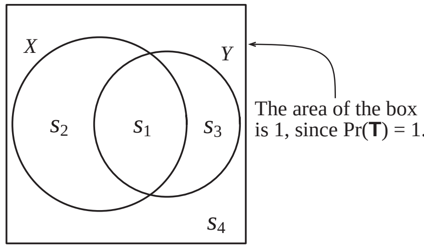
$$\Pr(p) = \sum_{s_i \vDash p} \Pr(s_i) \qquad \text{[note: if } p \dashv\vDash \bot\text{, then } \Pr(p) = 0]$$

- In other words, $\Pr(p)$ is the sum of the probabilities of the state descriptions in p's (equivalent) disjunction of state descriptions.

The Probability Calculus: An Algebraic Approach II

- Here's an example of a finite probability model $\mathcal{M}$, whose algebra $\mathcal{B}$ is characterized by a language $\mathcal{L}$ with two atomic letters "X" and "Y":

X	Y	States	$\Pr(s_i)$
T	T	s_1	$\frac{1}{6}$
T	F	s_2	$\frac{1}{4}$
F	T	s_3	$\frac{1}{8}$
F	F	s_4	$\frac{11}{24}$



- On the left, a *stochastic truth-table* (STT) representation of $\mathcal{M}$; on the right, a *stochastic Venn Diagram* (SVD) representation, in which *area is proportional to probability*. This is a *regular* model: $\Pr(s_i) > 0$, for all i.
- $\mathcal{M}$ determines a *numerical* probability for *each* p in $\mathcal{L}$. Examples?
- We can also use STTs to furnish an algebraic method for *proving* ***general*** *facts* about *all* probability models — *the algebraic method*.

The Probability Calculus: An Algebraic Approach III

- Let $a_i = \Pr(s_i)$ be the probability [under the probability assignment $\Pr(\cdot)$] of state s_i in $\mathcal{B}$ — *i.e.*, the area of region s_i in our SVD.
- Once we have real variables (a_i) for each of the basic probabilities, we can not only calculate probabilities relative to *specific* numerical models — *we can say* ***general*** *things, using only simple high-school algebra.*
- That is, we can *translate* any expression $\ulcorner\Pr(p)\urcorner$ into a *sum* of some of the a_i, and thus we can *reduce probabilistic* claims about the p's in $\mathcal{B}/\mathcal{L}$ into simple, high-school-*algebraic* claims about the real variables a_i.
- This allows us to be able to prove general claims about *probability functions*, by proving their corresponding *algebraic theorems*.
- Method: translate the probability claim into a claim involving sums of the a_i, and determine whether the corresponding claim is a theorem of algebra (assuming only that the a_i are on $[0,1]$ and that they sum to 1).

The Probability Calculus: An Algebraic Approach IV

- Here are two simple/obvious examples involving two atomic sentences:

  Theorem. $\Pr(X \vee Y) = \Pr(X) + \Pr(Y) - \Pr(X \,\&\, Y)$.

  Proof. $\Pr(X \vee Y) = a_1 + a_2 + a_3 = (a_1 + a_2) + (a_1 + a_3) - a_1$.

  Theorem. $\Pr(X) = \Pr(X \,\&\, Y) + \Pr(X \,\&\, {\sim}Y)$.

  Proof. $a_1 + a_2 = a_1 + a_2$.

- Here are two general facts that are also obvious from the set-up:

  Theorem. If $p \mathrel{=\!\!|\!\!\models} q$, then $\Pr(p) = \Pr(q)$.

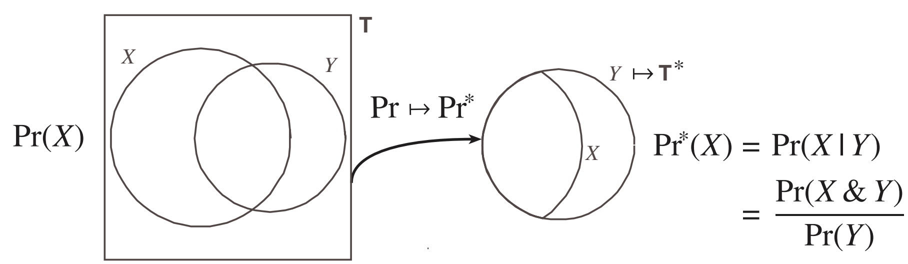
  Proof. Obvious, since the same regions always have the same areas, and the algebraic translation is *the same* for logically equivalent p/q.

  Theorem. If $p \vDash q$, then $\Pr(p) \leq \Pr(q)$.

  Proof. Since $p \vDash q$, the set of state descriptions entailing p is a subset of the set of state descriptions entailing q. Thus, the set of a_i in the summation for $\Pr(p)$ will be a subset of the a_i in the summation for $\Pr(q)$. Thus, since all the $a_i \geq 0$, $\Pr(p) \leq \Pr(q)$.

The Probability Calculus: An Algebraic Approach V

- **Conditional Probability**. $\Pr(p \mid q) \overset{\text{def}}{=} \dfrac{\Pr(p \,\&\, q)}{\Pr(q)}$, *provided that* $\Pr(q) > 0$.
- Intuitively, $\Pr(p \mid q)$ is supposed to be the probability of p ***given that*** q *is true*. So, *conditionalizing* on q is like "supposing q to be true".
- Using Venn diagrams, we can explain: "Supposing Y to be true" is like "treating the Y-circle as if it is the bounding box of the Venn Diagram".
- This is like "moving to a new $\Pr^*(\cdot)$ such that $\Pr^*(Y) = 1$." Picture:



The Probability Calculus: An Algebraic Approach VI

- There may be other ways of defining conditional probability, which may also seem to capture the "supposing q to be true" intuition.
- But, any such definition must make $\Pr(\cdot \mid q)$ itself a *probability function*, *for all* q. We will look at this important constraint again (and in more generality), when we discuss the axiomatic approach to probability.
- But, algebraically, we can see that this is a strong constraint. Recall:

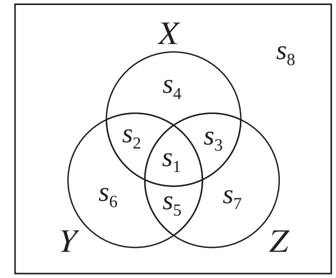
$$\Pr(X \vee Y) = \Pr(X) + \Pr(Y) - \Pr(X \,\&\, Y).$$

- Therefore, if $\Pr(\cdot \mid q)$ is to be a *probability* function *for all* q, then we must also have the following equality (in general), *for all* Z:

$$\Pr(X \vee Y \mid Z) = \Pr(X \mid Z) + \Pr(Y \mid Z) - \Pr(X \,\&\, Y \mid Z).$$

- Using our algebraic method, we can *prove* this. We just need to remind ourselves of what the 3-atomic sentence algebra looks like, and how the algebraic translation of this equation would go. Let's do that …

X	Y	Z	States	$\Pr(s_i)$
T	T	T	s_1	a_1
T	T	F	s_2	a_2
T	F	T	s_3	a_3
T	F	F	s_4	a_4
F	T	T	s_5	a_5
F	T	F	s_6	a_6
F	F	T	s_7	a_7
F	F	F	s_8	a_8



- By our definition of conditional probability, we have:

$$\Pr(X \vee Y \mid Z) = \frac{\Pr((X \vee Y) \,\&\, Z)}{\Pr(Z)} = \frac{\Pr((X \,\&\, Z) \vee (Y \,\&\, Z))}{\Pr(Z)} = \frac{a_1 + a_3 + a_5}{a_1 + a_3 + a_5 + a_7}$$

and

$$\Pr(X \mid Z) + \Pr(Y \mid Z) - \Pr(X\&Y \mid Z) = \frac{\Pr(X \,\&\, Z)}{\Pr(Z)} + \frac{\Pr(Y \,\&\, Z)}{\Pr(Z)} - \frac{\Pr(X \,\&\, Y \,\&\, Z)}{\Pr(Z)}$$

$$= \frac{\Pr(X \,\&\, Z) + \Pr(Y \,\&\, Z) - \Pr(X \,\&\, Y \,\&\, Z)}{\Pr(Z)}$$

$$= \frac{(a_1 + a_3) + (a_1 + a_5) - a_1}{a_1 + a_3 + a_5 + a_7} = \frac{a_1 + a_3 + a_5}{a_1 + a_3 + a_5 + a_7}$$

The Probability Calculus: An Algebraic Approach VII

- We can use our algebraic method to demonstrate that our definition of $\Pr(\cdot \mid q)$ yields a probability function, for all q, in the following way.
- Intuitively, think about what an "unconditional" and a "conditional" stochastic truth-table must look like, for any pair of sentences p and q.

p	q	$\Pr(s_i)$
T	T	a_1
T	F	a_2
F	T	a_3
F	F	a_4

$\xrightarrow{\cdot \mid q}$

p	q	$\Pr(s_i \mid q)$
T	T	$\Pr(s_1 \mid q) \overset{\text{def}}{=} \frac{\Pr(s_1 \& q)}{\Pr(q)} = \frac{a_1}{a_1+a_3}$
T	F	$\Pr(s_2 \mid q) \overset{\text{def}}{=} \frac{\Pr(s_2 \& q)}{\Pr(q)} = 0$
F	T	$\Pr(s_3 \mid q) \overset{\text{def}}{=} \frac{\Pr(s_3 \& q)}{\Pr(q)} = \frac{a_3}{a_1+a_3}$
F	F	$\Pr(s_4 \mid q) \overset{\text{def}}{=} \frac{\Pr(s_4 \& q)}{\Pr(q)} = 0$

- Note: the new basic probabilities assigned to the state descriptions, under our "conditionalized" $\Pr(\cdot \mid q)$ satisfy the requirements for being a *probability* function, since $\frac{a_1}{a_1+a_3} + \frac{a_3}{a_1+a_3} = 1$, and $\frac{a_1}{a_1+a_3}, \frac{a_3}{a_1+a_3} \in [0, 1]$.

The Probability Calculus: An Algebraic Approach VIII

- Here's a neat theorem of the probability calculus, proved algebraically.

Theorem. $\Pr(X \to Y) \geq \Pr(Y \mid X)$. [Provided that $\Pr(X) > 0$, of course.]

Proof. $\Pr(X \to Y) = \Pr({\sim}X \vee Y) = \Pr(s_1 \vee s_3 \vee s_4) = a_1 + a_3 + a_4$.

$$\Pr(Y \mid X) = \frac{\Pr(Y \,\&\, X)}{\Pr(X)} = \frac{\Pr(s_1)}{\Pr(s_1 \vee s_2)} = \frac{a_1}{a_1 + a_2}.$$

So, we need to prove that $\boxed{a_1 + a_3 + a_4 \geq \frac{a_1}{a_1 + a_2}}$.

- First, note that $a_4 = 1 - (a_1 + a_2 + a_3)$, since the a_i's must sum to 1.
- Thus, we need to show that $a_1 + a_3 + 1 - a_1 - a_2 - a_3 \geq \frac{a_1}{a_1 + a_2}$.
- By simple algebra, this reduces to showing that $\boxed{1 - a_2 \geq \frac{a_1}{a_1 + a_2}}$.
- If $a_1 + a_2 > 0$ and $a_i \in [0, 1]$, this must hold, since then we must have: $\boxed{a_2 \geq a_2 \cdot (a_1 + a_2)}$, and then the boxed formulas are equivalent. □

The Probability Calculus: An Algebraic Approach IX

- Here are some further fundamental theorems of probability calculus, involving 2 or 3 atomic sentences and CP. Easy, given defn. of CP.
  - **The Law of Total Probability** (LTP):

$$\Pr(X \mid Y) = \Pr(X \mid Y \,\&\, Z) \cdot \Pr(Z \mid Y) + \Pr(X \mid Y \,\&\, {\sim}Z) \cdot \Pr({\sim}Z \mid Y)$$

  - Note: $\Pr(X \mid \top) = \Pr(X)$. Why? So, the LTP has a *special case*:

$$\begin{aligned}\Pr(X \mid \top) = \Pr(X) &= \Pr(X \mid \top \,\&\, Z) \cdot \Pr(Z \mid \top) + \Pr(X \mid \top \,\&\, {\sim}Z) \cdot \Pr({\sim}Z \mid \top) \\ &= \Pr(X \mid Z) \cdot \Pr(Z) + \Pr(X \mid {\sim}Z) \cdot \Pr({\sim}Z)\end{aligned}$$

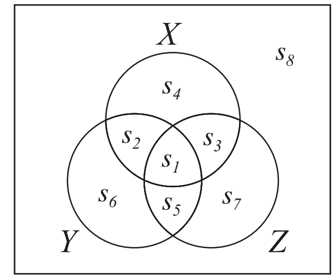
  - Two forms of **Bayes's Theorem**. The second one *follows*, using (LTP):

$$\begin{aligned}\Pr(X \mid Y) &= \frac{\Pr(Y \mid X) \cdot \Pr(X)}{\Pr(Y)} \\ &= \frac{\Pr(Y \mid X) \cdot \Pr(X)}{\Pr(Y \mid Z) \cdot \Pr(Z) + \Pr(Y \mid {\sim}Z) \cdot \Pr({\sim}Z)}\end{aligned}$$

The Probability Calculus: An Algebraic Approach X

- The algebraic approach for *refuting* general claims involves two steps:
  1. Translate the claim from probability notation into algebraic terms.
  2. Find a (numerical) probability model on which the translation is *false*.
- Show that $\Pr(X \mid Y \,\&\, Z) = \Pr(X \mid Y \vee Z)$ can be *false*. Here's a model $\mathcal{M}$:

X	Y	Z	States	$\Pr(s_i)$
T	T	T	s_1	$a_1 = 1/6$
T	T	F	s_2	$a_2 = 1/6$
T	F	T	s_3	$a_3 = 1/4$
T	F	F	s_4	$a_4 = 1/16$
F	T	T	s_5	$a_5 = 1/6$
F	T	F	s_6	$a_6 = 1/12$
F	F	T	s_7	$a_7 = 1/24$
F	F	F	s_8	$a_8 = 1/16$



(1) Algebraic Translation: $\dfrac{a_1}{a_1 + a_5} = \dfrac{a_1 + a_2 + a_3}{a_1 + a_2 + a_3 + a_5 + a_6 + a_7}$.

(2) This claim is *false* on $\mathcal{M}$, since $1/2 \neq 2/3$. I used `PrSAT` to find $\mathcal{M}$.

The Probability Calculus: An Algebraic Approach XI

- There are *decision procedures* for Boolean propositional logic, based on truth-tables. These methods are *exponential* in the number of atomic sentences (n), because truth-tables grow exponentially in n (2^n).
- It would be nice if there were a decision procedure for probability calculus, too. In algebraic terms, this would require a decision procedure for the salient fragment of high-school (real) algebra.
- As it turns out, high-school (real) algebra (HSA) *is* a decidable theory. This was shown by Tarski in the 1920's. But, it's only been very recently that computationally feasible procedures have been developed.
- In my "A Decision Procedure for Probability Calculus with Applications", I describe a user-friendly decision procedure (called `PrSAT`) for probability calculus, based on recent HSA procedures.
- My implementation is written in *Mathematica* (a general-purpose mathematics computer programming framework). It is freely downloadable from my website, at: `http://fitelson.org/PrSAT/`.

The Probability Calculus: An Algebraic Approach XII

- I encourage the use of `PrSAT` as a tool for finding counter-models and for establishing theorems of probability calculus. It is not a requirement of the course, but it is a useful tool that is worth learning.
- `PrSAT` doesn't give readable proofs of theorems. But, it will find concrete numerical counter-models for claims that are not theorems.
- `PrSAT` will also allow you to calculate probabilities that are determined by a *given* probability assignment. And, it will allow you to do algebraic and numerical "scratch work" without making errors.
- I have posted a *Mathematica* notebook which contains the examples from algebraic probability calculus that we have seen in this lecture. I will be posting further notebooks as the course goes along.
- Let's have a look at this first notebook (`examples_1.nb`). I will now go through the examples in this notebook, and demonstrate some of the features of `PrSAT`. I encourage you to play around with it.

Axiomatic Treatment of Probability Calculus

- A probability model $\mathcal{M}$ is a Boolean algebra of propositions $\mathcal{B}$, together with a function $\Pr(\cdot) : \mathcal{B} \mapsto \mathbb{R}$ satisfying the following three *axioms*.
  1. For all $p \in \mathcal{B}$, $\Pr(p) \geq 0$. [non-negativity]
  2. $\Pr(\top) = 1$, where $\top$ is the tautological proposition. [normality]
  3. For all $p, q \in \mathcal{B}$, if p and q are mutually exclusive (inconsistent), then $\Pr(p \vee q) = \Pr(p) + \Pr(q)$. [additivity]
- Conditional probability is *defined* in terms of unconditional probability in the usual way: $\Pr(p \mid q) \overset{\text{def}}{=} \frac{\Pr(p \& q)}{\Pr(q)}$, provided that $\Pr(q) > 0$.
- We could also state everything in terms of a (propositional) *language* $\mathcal{L}$ with a finite number of atomic *sentences*. Then, we would talk about *sentences* rather than *propositions*, and the axioms would read:
  1. For all $p \in \mathcal{L}$, $\Pr(p) \geq 0$.
  2. For all $p \in \mathcal{L}$, if $p \dashv\vDash \top$, then $\Pr(p) = 1$.
  3. For all $p, q \in \mathcal{L}$, if $p \,\&\, q \dashv\vDash \bot$, then $\Pr(p \vee q) = \Pr(p) + \Pr(q)$.

Independence, Correlation, and Anti-Correlation 1

Definition. p and q are probabilistically independent ($p \perp\!\!\!\perp q$) in a Pr-model $\mathcal{M}$ if $\mathcal{M} = \langle \mathcal{L}, \Pr \rangle$ is such that: $\Pr(p \,\&\, q) = \Pr(p) \cdot \Pr(q)$.

- If $\Pr(p) > 0$ and $\Pr(q) > 0$, we can express independence also as follows:
  - ∗ $\Pr(p \mid q) = \Pr(p)$ [Why? Because this is just: $\frac{\Pr(p \,\&\, y)}{\Pr(y)} = \Pr(p)$]
  - ∗ $\Pr(q \mid p) = \Pr(q)$ [ditto.]
  - ∗ $\Pr(p \mid q) = \Pr(p \mid \sim q)$ [Not as obvious. See next slide.]
  - ∗ $\Pr(q \mid p) = \Pr(q \mid \sim p)$ [ditto.]
- Exercise: prove this! Closely related fact about independence. If $p \perp\!\!\!\perp q$, then we also must have: $p \perp\!\!\!\perp \sim q$, $q \perp\!\!\!\perp \sim p$, and $\sim p \perp\!\!\!\perp \sim q$. Prove this too!
- A set of propositions $\mathbf{P} = \{p_1, \ldots p_n\}$ is *mutually independent* if all subsets $\{p_i, \ldots, p_j\} \subseteq \mathbf{P}$ are s.t. $\Pr(p_i \,\&\, \cdots \,\&\, p_j) = \Pr(p_i) \cdot \cdots \cdot \Pr(p_j)$. For sets with 2 propositions, pairwise independence is equivalent to mutual independence. But, not for 3 or more propositions. Example given below.

- **Theorem**. $\Pr(p \,\&\, q) = \Pr(p) \cdot \Pr(q) \Leftrightarrow \Pr(p \mid q) = \Pr(p \mid \sim q)$, provided that that $\Pr(q) \in (0, 1)$.
- A *purely algebraic* proof of this theorem can be obtained rather easily:

p	q	States	$\Pr(s_i)$
T	T	s_1	a_1
T	F	s_2	a_2
F	T	s_3	a_3
F	F	s_4	$a_4 = 1 - (a_1 + a_2 + a_3)$

$$\therefore \Pr(p \mid q) = \Pr(p \mid \sim q) \Leftrightarrow \frac{a_1}{a_1 + a_3} = \frac{a_2}{a_2 + a_4} = \frac{a_2}{1 - (a_1 + a_3)}$$
$$\Leftrightarrow a_1 \cdot (1 - (a_1 + a_3)) = a_2 \cdot (a_1 + a_3)$$
$$\Leftrightarrow a_1 = a_2 \cdot (a_1 + a_3) + a_1 \cdot (a_1 + a_3) = (a_2 + a_1) \cdot (a_1 + a_3)$$
$$\Leftrightarrow \Pr(p \,\&\, q) = \Pr(p) \cdot \Pr(q) \quad \square$$

- If p and q are independent, then so are p and $\sim q$. This is also pretty easy to prove algebraically. I'll leave this one as an exercise…

Independence, Correlation, and Anti-Correlation 2

- **Theorem**. Pairwise independence of a collection of three propositions $\{X, Y, Z\}$ does *not* entail mutual independence of the collection.
- That is: there exist probability models s.t. (1) $\Pr(X \,\&\, Y) = \Pr(X) \cdot \Pr(Y)$, (2) $\Pr(X \,\&\, Z) = \Pr(X) \cdot \Pr(Z)$, (3) $\Pr(Y \,\&\, Z) = \Pr(Y) \cdot \Pr(Z)$, but (4) $\Pr(X \,\&\, Y \,\&\, Z) \neq \Pr(X) \cdot \Pr(Y) \cdot \Pr(Z)$. *Proof*. Here's a counterexample.
- Suppose a box contains 4 tickets labelled with the following numbers:

  112, 121, 211, 222

  Let us choose one ticket at random (*i.e.*, each ticket has an *equal* probability of being chosen), and consider the following propositions:

  X = "1" occurs at the first place of the chosen ticket.
  Y = "1" occurs at the second place of the chosen ticket.
  Z = "1" occurs at the third place of the chosen ticket.

  Since the ticket #'s are 112, 121, 211, 222, we have these probabilities:

$$\Pr(X) = \tfrac{1}{2}, \Pr(Y) = \tfrac{1}{2}, \Pr(Z) = \tfrac{1}{2}$$

Moreover, each of the three conjunctions determines a unique ticket #:

$X \,\&\, Y$ = the ticket is labeled #112
$X \,\&\, Z$ = the ticket is labeled #121
$Y \,\&\, Z$ = the ticket is labeled #211

Therefore, since each ticket is equally probable to be chosen, we have:

$$\Pr(X \,\&\, Y) = \Pr(X \,\&\, Z) = \Pr(Y \,\&\, Z) = \tfrac{1}{4}$$

So, the three events X, Y, Z are pairwise independent (*why?*). But,

$X \,\&\, Y \,\&\, Z \dashv\vDash \bot$, since X, Y, and Z are jointly inconsistent.

Hence,

$\Pr(X \,\&\, Y \,\&\, Z) = \Pr(\mathsf{F}) = 1 - \Pr(\mathsf{T}) = 0 \neq \Pr(X) \cdot \Pr(Y) \cdot \Pr(Z) = (\tfrac{1}{2})^3 = \tfrac{1}{8}$

- This information determines a *unique* probability function. Can you specify it? Algebra (7 equations, 7 unknowns — see STT below).

$$\Pr(X) = a_4 + a_2 + a_3 + a_1 = \tfrac{1}{2}, \Pr(Y) = a_2 + a_6 + a_1 + a_5 = \tfrac{1}{2}$$

$$\Pr(Z) = a_3 + a_1 + a_5 + a_7 = \tfrac{1}{2}, \Pr(X \,\&\, Y \,\&\, Z) = a_1 = 0$$

$$\Pr(X \,\&\, Y) = a_2 + a_1 = \tfrac{1}{4}, \Pr(X \,\&\, Z) = a_3 + a_1 = \tfrac{1}{4}, \Pr(Y \,\&\, Z) = a_1 + a_5 = \tfrac{1}{4}$$

- Here's the STT. [This (and other models) can be found with `PrSAT`.]

X	Y	Z	States	$\Pr(s_i)$
T	T	T	s_1	$\Pr(s_1) = a_1 = 0$
T	T	F	s_2	$\Pr(s_2) = a_2 = 1/4$
T	F	T	s_3	$\Pr(s_3) = a_3 = 1/4$
T	F	F	s_4	$\Pr(s_4) = a_4 = 0$
F	T	T	s_5	$\Pr(s_5) = a_5 = 1/4$
F	T	F	s_6	$\Pr(s_6) = a_6 = 0$
F	F	T	s_7	$\Pr(s_7) = a_7 = 0$
F	F	F	s_8	$\Pr(s_8) = a_8 = 1/4$

- **Theorem**. $\perp\!\!\!\perp$ is *not* transitive. Example in which $\Pr(X \,\&\, Y) = \Pr(X) \cdot \Pr(Y)$, $\Pr(Y \,\&\, Z) = \Pr(Y) \cdot \Pr(Z)$, but $\Pr(X \,\&\, Z) \neq \Pr(X) \cdot \Pr(Z)$ [$X \neq Y \neq Z$]:

X	Y	Z	States	$\Pr(s_i)$
T	T	T	s_1	$\Pr(s_1) = a_1 = 3/32$
T	T	F	s_2	$\Pr(s_2) = a_2 = 9/32$
T	F	T	s_3	$\Pr(s_3) = a_3 = 3/32$
T	F	F	s_4	$\Pr(s_4) = a_4 = 9/32$
F	T	T	s_5	$\Pr(s_5) = a_5 = 2/32$
F	T	F	s_6	$\Pr(s_6) = a_6 = 2/32$
F	F	T	s_7	$\Pr(s_7) = a_7 = 2/32$
F	F	F	s_8	$\Pr(s_8) = a_8 = 2/32$

$$\begin{aligned}\Pr(X \,\&\, Y) &= a_2 + a_1 = \frac{3}{8} = \frac{3}{4} \cdot \frac{1}{2} \\ &= (a_4 + a_2 + a_3 + a_1) \cdot (a_2 + a_1 + a_6 + a_5) = \Pr(X) \cdot \Pr(Y)\end{aligned}$$

$$\begin{aligned}\Pr(Y \,\&\, Z) &= a_1 + a_5 = \frac{5}{32} = \frac{1}{2} \cdot \frac{5}{16} \\ &= (a_2 + a_1 + a_6 + a_5) \cdot (a_3 + a_1 + a_5 + a_7) = \Pr(Y) \cdot \Pr(Z)\end{aligned}$$

$$\begin{aligned}\Pr(X \,\&\, Z) &= a_3 + a_1 = \frac{3}{16} \neq \frac{3}{4} \cdot \frac{5}{16} \\ &= (a_4 + a_2 + a_3 + a_1) \cdot (a_3 + a_1 + a_5 + a_7) = \Pr(X) \cdot \Pr(Z)\end{aligned}$$

Independence, Correlation, and Anti-Correlation 3

- So far, we've been talking about *unconditional* independence, correlation, and anti-correlation. There are also *conditional* notions.
- **Definition**. p and q are *conditionally* independent, *given* r [$p \perp\!\!\!\perp q \mid r$] iff:

$$\Pr(p \,\&\, q \mid r) = \Pr(p \mid r) \cdot \Pr(q \mid r)$$

- Similarly, we have conditional correlation and anti-correlation as well (just change the equal sign "=" above to a ">" or a "<", respectively).
- Conditional and unconditional independence are not related in any obvious way. In fact, they can come apart in rather strange ways!
- **Example**. It is possible to have all three of the following simultaneously:
  - $p \perp\!\!\!\perp q \mid r$
  - $p \perp\!\!\!\perp q \mid {\sim}r$
  - $p \not\perp\!\!\!\perp q$
- *Simpson's Paradox* (more below)! Exercise: use `PrSAT` to find one.

An Anecdotal Prelude to "Interpretations" of Probability

- After the O.J. trial, Alan Dershowitz remarked that "fewer than 1 in 1,000 women who are abused by their mates go on to be killed by them".
- He said "the *probability*" that Nicole Brown Simpson (N.B.S.) was killed by her 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 segué from our discussion of the formal theory of probability calculus to its "interpretation(s)".
- 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.), and let $\Pr(\cdot)$ be whatever probability function Dershowitz has in mind here, over the salient algebra of propositions. Dershowitz is saying the following:

(1) $$\Pr(K \mid A) < \frac{1}{1000}$$

- Shortly after Dershowitz's remark, the statistician I.J. Good wrote a brief response in *Nature*. Good pointed out that, while Dershowitz's claim may be true, it is not salient to the case at hand, since it *ignores evidence*.
- Good argues that what's relevant here is is 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.
- To wit: let K' be the proposition that N.B.S was killed (by *someone*). Using Dershowitz's (1) as a starting point, Good does some *ex cathedra* "back-of-the-envelope calculations," and he comes up with the following:

(2) $$\Pr(K \mid A \,\&\, K') \approx \frac{1}{2} \gg \frac{1}{1000}$$

- This would seem to make it far more probable that O.J. is the killer than

Dershowitz's claim would have us believe. Using statistical data about murders committed in 1992, Merz & Caulkins "estimated" that:

(3) $$\Pr(K \mid A \,\&\, K') \approx \frac{4}{5}$$

- 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 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) $\Pr(K \mid K') \approx \Pr(K \mid A \& K')$ [*i.e.*, K', not A, is doing the real work here]

- Not to be outdone, Merz & Caulkins give the following "estimate" of the

salient probabilities (again, this is based on statistics for 1992):

(5) $$\Pr(K \mid K') \approx 0.29 \ll \Pr(K \mid A \,\&\, K') \approx 0.8$$

- We could continue this dialectic *ad nauseam*. I'll stop here. This anecdote raises several key issues about "interpretations" and "applications" of Pr.
  - Our discussants want some kind of "objective" probabilities about *N.B.S.*'s murder (and murderer) *in particular*. But, the "estimates" trade on *statistics* involving *frequencies* of murders in some *population*.
  - Are there probabilities of token events, or only statistical frequencies in populations? If there are probabilities of token events how do they relate to frequencies? And, which population is "the right one" in which to include the event in question (*reference class problem*)?
  - Our discussants want these "objective" probabilities (whatever kind they are) to be relevant to *people's degrees of belief*. What is the connection (if any) between "objective" and "subjective" probabilities?
  - When it comes to probabilistic *explanation*, it is often assumed that subjective interpretations are *inappropriate*. More on this later . . .

Brief Overview of Theories of Truth

- According to objective theories of truth, p is true if it *corresponds* to "the way the world really is". In other words, there are *mind-independent truthmakers*, and these determine which statements are true.
- Subjective theories of truth tend to talk about *beliefs* being true if they are *justified*, *coherent* with one's beliefs, and/or *useful* for one to believe.
- Some have argued that subjective theories of truth face a regress problem. Moreover, it does seem that subjective theories seem to yield *incorrect* verdicts about various truth-related phenomena.
- One thing about truth that seems clear is that it is *redundant*. When I assert "p is true", this is just like asserting p itself. For instance, if I say "*it is true that* it is raining", this is equivalent to just saying "*it is raining*".
- Subjective theories seems to *violate* redundancy. Intuitively, when I say "p is justified" (useful, coherent, *etc.*), this is *not* equivalent to just saying p. [Intuitively, evidence can be *misleading*, and *wishful thinking* may be useful, *etc.*]

Probability-in-$\mathcal{M}$ as analogous to *truth-on-$\mathcal{I}$*

- Just as we can talk about p being *true-on-$\mathcal{I}_i$*, which is synonymous with $s_i \vDash p$, we can also talk about p having *probability-r-on-$\mathcal{M}$*.
- And, like *truth-on-$\mathcal{I}_i$*, *probability-on-$\mathcal{M}$* is a *logical/formal* concept.
- That is, once we have *specified* a probability model $\mathcal{M}$, this *logically determines* the *probability-on-$\mathcal{M}$* values of all sentences in $\mathcal{L}$.
- Moreover, just as the *truth-on-$\mathcal{I}_i$* of sentence p does not imply anything about p's *truth* (*simpliciter*), neither does the *probability-on-$\mathcal{M}$* of p imply anything about p's *probability* (*simpliciter*) — *if there be such a thing.*
- Just as we have different philosophical "theories" of truth, we will also have different (and analogous) philosophical "theories" of probability.
- And, as in the case of truth, there will be objective theories and subjective theories of probability. However, there are more compelling reasons for "going subjective" in the *probability* case than in the *truth* case.
- Let's begin by looking at some objective theories of probability.

Objective Theories 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 axioms (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 $\#(S) \overset{\text{def}}{=}$ 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:
  - $f_{\mathbf{P}}(\chi) \overset{\text{def}}{=} \dfrac{\#(\boldsymbol{\chi} \cap \mathbf{P})}{\#(\mathbf{P})}$
- 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:
  - $\Pr_{\mathbf{P}}(X) \overset{\text{def}}{=} f_{\mathbf{P}}(\chi)$.
- We need to show that $\Pr_{\mathbf{P}}(X)$ is in fact a *probability* function. There are various ways to do this. I will show that $\Pr_{\mathbf{P}}(X)$ satisfies our three axioms.

Objective Theories of Probability II

- **Axiom 1**. We need to show that $\Pr_{\mathbf{P}}(X) \geq 0$, for any property χ. This is easy, since the ratio $\frac{\#(\boldsymbol{\chi} \cap \mathbf{P})}{\#(\mathbf{P})}$ must be non-negative, for any property χ. This is because $\mathbf{P}$ is non-empty [$\#(\mathbf{P}) > 0$], and $\#(\boldsymbol{\chi} \cap \mathbf{P})$ must be non-negative.
- **Axiom 2**. We need to show that, if $X \dashv\vDash \top$, then $\Pr_{\mathbf{P}}(X) = 1$. In this context, we're taking about properties χ that — by logic alone — must be satisfied by all objects in the universe (*e.g.*, $\chi x = Fx \vee \sim Fx$). In this case, we have $\boldsymbol{\chi} \cap \mathbf{P} = \mathbf{P}$, since *every* object is in $\boldsymbol{\chi}$. Therefore, $\Pr_{\mathbf{P}}(X) = \frac{\#(\mathbf{P})}{\#(\mathbf{P})} = 1$.
- **Axiom 3**. To be shown: If $X \,\&\, Y \dashv\vDash \bot$, then $\Pr_{\mathbf{P}}(X \vee Y) = \Pr_{\mathbf{P}}(X) + \Pr_{\mathbf{P}}(Y)$. In this context, $X \,\&\, Y \dashv\vDash \bot$ means we are talking about properties χ and ψ such that — by logic alone — no object can satisfy both properties at once (*e.g.*, $\chi a \,\&\, \psi a \vDash \bot$). In such a case, we will have the following:

$$\Pr_{\mathbf{P}}(X \vee Y) = \frac{\#[(\boldsymbol{\chi} \cup \boldsymbol{\psi}) \cap \mathbf{P}]}{\#(\mathbf{P})} = \frac{\#[(\boldsymbol{\chi} \cap \mathbf{P}) \cup (\boldsymbol{\psi} \cap \mathbf{P})]}{\#(\mathbf{P})} = \frac{\#(\boldsymbol{\chi} \cap \mathbf{P}) + \#(\boldsymbol{\psi} \cap \mathbf{P})}{\#(\mathbf{P})}$$
$$= \Pr_{\mathbf{P}}(X) + \Pr_{\mathbf{P}}(Y)$$

Objective Theories 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 $\Pr_{\mathbf{P}_1}(X)$, $\ldots, \Pr_{\mathbf{P}_n}(X)$. This is related to the *reference class problem* (see below!).
- 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 (assuming IID trials!). 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 Theories 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 $\mathbf{P}$ 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 Theories 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 QM, it seems only *probabilities* of (some) token events are determined by the laws + i.c.'s.
- This leaves room for (non-extreme) *objective chances* of token events.

Objective Theories of Probability VI

- We saw that finite frequencies satisfy the (classical) probability axioms.
- Infinite frequencies don't satisfy the (classical) axioms of (infinite) probability calculus, for two reasons (beyond the scope of our course).
  - The underlying (infinite) logical space is non-Boolean.
  - Infinite frequencies do not satisfy the (infinite) additivity axiom.
- Some have claimed that QM-probabilities are also non-classical, owing to the fact that the underlying "quantum *logic*" is non-Boolean. But, there are also interpretations of QM in terms of classical probabilities.
- It is often *assumed* that objective chances satisfy the probability axioms, but it is not quite clear *why* (especially, in light of the above remarks).
- We won't be dwelling *too* much on probabilistic explanation (*per se*). But, there will be some rather subtle and vexing questions about probabilistic explanation in Woodward's framework (especially, in QM systems!).
- Thus, Woodward will face an "interpretation of $\Pr(\cdot)$ problem".

Subjective Theories of Probability I

- I'll begin by motivating subjective probability with an example/context:

  I'm holding a coin behind my back. It is either 2-headed or 2-tailed. You do not know which kind of coin it is (and you have no reason to favor one of these possibilities over the other). I'm about to toss it. What probability (or odds) would you assign to the proposition that it will land heads?

- Many people have the intuition that $\frac{1}{2}$ (or even 50:50 odds) would be a reasonable answer to this question. However, it seems clear that $\frac{1}{2}$ *cannot* be the *objective* probability/chance of heads in this example.
- After all, we know that the coin is either 2-headed or 2-tailed. As such, the objective probability of heads is either 1 or 0 in this example.
- One might describe this as *epistemic* probability, because it seems *epistemically reasonable* to be *50% confident* that the coin will land heads.
- Also, taking a bet at even odds on heads seems *pragmatically* reasonable. This suggests a *pragmatic* theory of probability is also plausible here.

Subjective Theories of Probability II

- It seems clear that there is such a thing as "degree of belief". And, it also seems clear that there are *some* sorts of constraints on such degrees.
- But, should degrees of belief obey the *probability axioms*? If so, *Why*?
- There are arguments to the effect that epistemic (accuracy arguments) and pragmatic (dutch book arguments) dob's should be *probabilities*.
  - I will not be discussing such arguments in the seminar.
- Most commentators who write about scientific explanation seem to think that subjective probabilities (epistemic or pragmatic) have no place in any adequate account of explanation. We'll return to this below (Coffa).
- Moreover, so-called "logical" probabilities (if there be such) also seem out of place here, since explanatory relations do not seem to be "logical".
- However, *if* there is such a thing as *inductive logic* (*and it is grounded in "logical probabilities"*!) *then* a Hempelian might have use for them after all — in order to gauge "inductive strength" of *inductive arguments*.