

Philosophy 148 — Announcements & Such

- Administrative Staff
 - Branden’s Thursday office hours will be 2:30–3:30 this week.
 - Raul’s office hours will be 10–12 Wed., and by appointment.
 - Section times have been determined. Sections will meet Tuesday, 10–11 and Wednesday, 9–10. You should have received an email assigning you to a section. Otherwise, please see Raul about this.
 - Section locations will be announced soon. Meanwhile, 301 Moses.
- Last Time: Finite Boolean Algebras & Some Overview Stuff
- Today’s Agenda
 - Review of Key Facts About Finite Propositional Boolean Algebras
 - Some Additional “Big Picture” Stuff (on Logic & Epistemology)
 - An Algebraic Approach to Probability Calculus
 - Next: An Axiomatic Approach to Probability Calculus

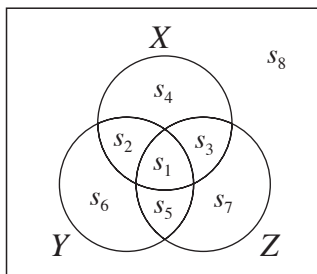
Overview of Finite Propositional Boolean Algebras I

- Consider a logical language \mathcal{L} containing n atomic sentences. These may be sentence letters ($X, Y, Z, \text{etc.}$), or they may be atomic sentences of monadic or relational predicate calculus ($Fa, Gb, Rab, Hcd, \text{etc.}$).
- 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 \emptyset of state descriptions corresponds to “the empty disjunction”, which corresponds to *the logical falsehood*: \perp .
 - Singleton 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



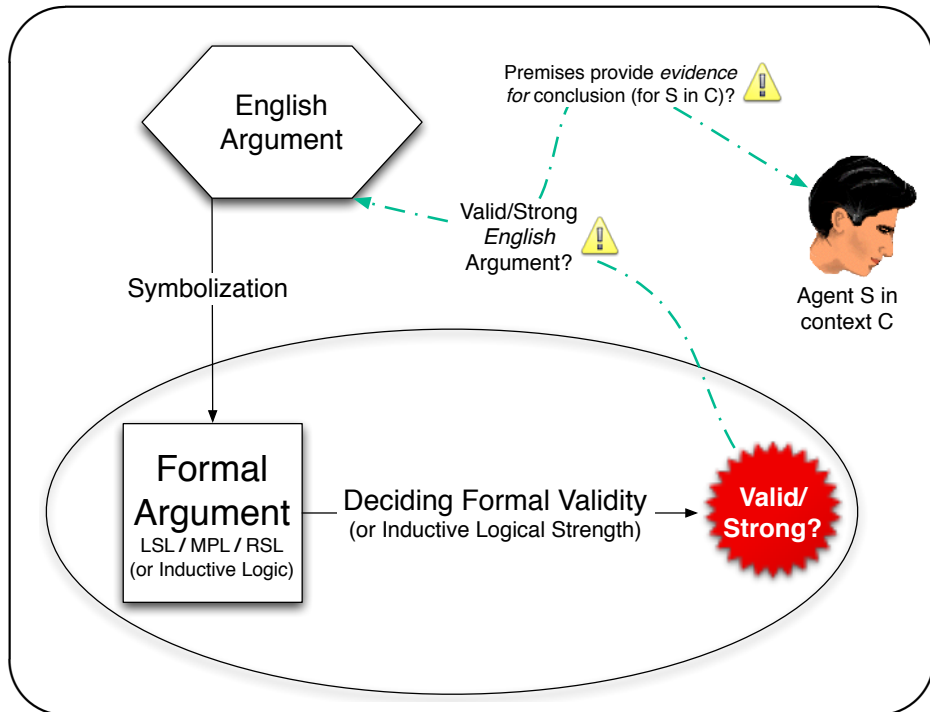
- Examples of reduction to disjunctions of state descriptions of \mathcal{L} :
 - ‘ $X \ \& \ \sim X$ ’ is equivalent to the *empty* disjunction: \perp .
 - ‘ $X \ \& \ (\sim Y \ \& \ Z)$ ’ is equivalent to the *singleton* disjunction: s_3 .
 - ‘ $X \ \leftrightarrow \ (Y \ \vee \ Z)$ ’ is equivalent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

Inductive Logic — Basic Motivation and Ideas

- Intuitively, not all “logically good” arguments are deductively valid. Some invalid arguments seem (intuitively) logically *better than* others:

(6) p . Someone is wise. (7) r . Someone is either wise or unwise.
 $\therefore q$. Socrates is wise. $\therefore q$. Socrates is wise.
- *Inductive* logic should *theoretically ground* our intuition that (6) is a *logically stronger* argument than (7) is. Neither argument is *valid*.
- More ambitiously, an inductive logician might aim for a theory of “the *degree* to which the premises of an argument *confirm* its conclusion”.
- This ambitious project would aim to characterize a *function* $c(\mathcal{C}, \mathcal{P})$. And, an intuitive requirement would be that this function be such that:

$$c(q, p) > c(q, r)$$
- This course is (mainly) about *inductive logic*. We will examine how *probabilities* might be used to *quantitatively generalize* deductive logic.



Logic and Epistemology — A Prelude I

- As I mentioned, some have worried about the adequacy of classical logic as a formal explication of our informal “following-from” relation.
- Here’s a fact about classical deductive logic that may seem “odd”:
 (†) If p and q are (classically) logically inconsistent, then the argument from p and q to r is (classically) valid — for *any* r .
- There’s *something* “odd” about the fact that *everything follows-from inconsistent premises*, according to the classical formal explication of following-from. But, what, exactly, is supposed to be “odd” about it?
- Here’s an *epistemological* principle that is downright *crazy*:
 (‡) If one’s beliefs are inconsistent (and one knows that they are), then one should believe everything (*i.e.*, every proposition).
- It is clear that (‡) is false. There are things I *know* to be false, and I shouldn’t believe those things — no matter what else is true of me.

Logic and Epistemology — A Prelude II

- OK, (‡) is clearly false. So? What does that have to do with (†)?
 - After all, (†) is about *logic*, and (‡) is about *epistemology*.
 - Perhaps those worried about (†) are assuming that logic and epistemology are connected, or bridged by something like:
 (*) If an agent S ’s belief set B is such that $B \models p$ (and S knows that $B \models p$), then it would be reasonable for S to infer/believe p .
 - If (*) were true, then (†) would imply (‡), and — as a result — classical logicians who accepted (*) would seem to be stuck with (‡) too.
 - More precisely, classical logicians who believe (*) should find it reasonable to believe (‡). But, they don’t (at least, they shouldn’t!).
- ☞ But, *this* doesn’t *force* classical logicians to give up (†). They could give up (*) instead. In such contexts, logic (alone) doesn’t seem to tell us whether to infer something new, or reject something we already believe.

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 $\text{Pr}(\cdot)$ which maps elements of \mathcal{B} to the unit interval $[0, 1] \in \mathbb{R}$.
- This function $\text{Pr}(\cdot)$ must be a *probability function*. It turns out that a probability function $\text{Pr}(\cdot)$ 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 $\sum_i \text{Pr}(s_i) = 1$.
- Once we have $\text{Pr}(\cdot)$ ’s *basic assignments* to the states of \mathcal{B} (s.d.’s of \mathcal{L}), we define $\text{Pr}(p)$ for *any* statement \mathcal{L} of the language of \mathcal{B} , as follows:

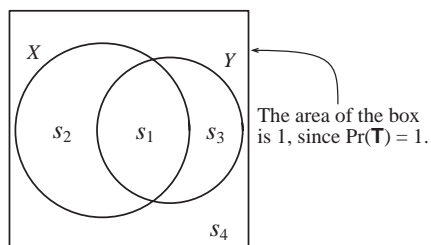
$$\text{Pr}(p) = \sum_{s_i \models p} \text{Pr}(s_i) \quad [\text{note: if } p \models \perp, \text{ then } \text{Pr}(p) = 0]$$

- In other words, $\text{Pr}(p)$ 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*: $\text{Pr}(s_i) > 0$, 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 $a_i = \text{Pr}(s_i)$ be the probability [under the probability assignment $\text{Pr}(\cdot)$] 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 ' $\text{Pr}(p)$ ' 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. $\text{Pr}(X \vee Y) = \text{Pr}(X) + \text{Pr}(Y) - \text{Pr}(X \& Y)$.

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

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

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

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

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

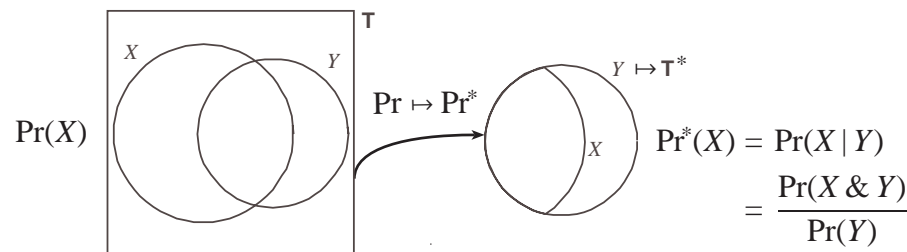
Proof. Obvious, since the same regions always have the same areas, and the algebraic translation is *the same* for logically equivalent p/q .

Theorem. If $p \models q$, then $\text{Pr}(p) \leq \text{Pr}(q)$.

Proof. Since $p \models q$, 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 $\text{Pr}(p)$ will be a subset of the a_i in the summation for $\text{Pr}(q)$. Thus, since all the $a_i \geq 0$, $\text{Pr}(p) \leq \text{Pr}(q)$.

The Probability Calculus: An Algebraic Approach V

- Conditional Probability.** $\text{Pr}(p | q) \stackrel{\text{def}}{=} \frac{\text{Pr}(p \& q)}{\text{Pr}(q)}$, provided that $\text{Pr}(q) > 0$.
- Intuitively, $\text{Pr}(p | q)$ 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 $\text{Pr}^*(\cdot)$ such that $\text{Pr}^*(Y) = 1$." 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 $\Pr(\cdot | q)$ 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:

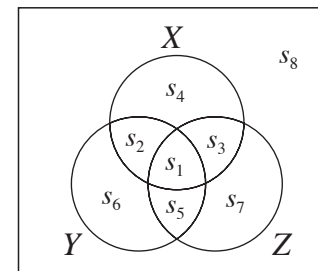
$$\Pr(X \vee Y) = \Pr(X) + \Pr(Y) - \Pr(X \& Y).$$

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

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

- 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:

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

and

$$\begin{aligned} \Pr(X | Z) + \Pr(Y | Z) - \Pr(X \& Y | Z) &= \frac{\Pr(X \& Z)}{\Pr(Z)} + \frac{\Pr(Y \& Z)}{\Pr(Z)} - \frac{\Pr(X \& Y \& Z)}{\Pr(Z)} \\ &= \frac{\Pr(X \& Z) + \Pr(Y \& Z) - \Pr(X \& Y \& Z)}{\Pr(Z)} \\ &= \frac{(a_1 + a_3) + (a_1 + a_5) - a_1}{a_1 + a_3 + a_5 + a_7} = \frac{a_1 + a_3 + a_5}{a_1 + a_3 + a_5 + a_7} \end{aligned}$$

The Probability Calculus: An Algebraic Approach VII

- We can use our algebraic method to demonstrate that our definition of $\Pr(\cdot | q)$ 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)	$\cdot q$	p	q	Pr($s_i q$)
T	T	a_1		T	T	$\Pr(s_1 q) \stackrel{\text{def}}{=} \frac{\Pr(s_1 \& q)}{\Pr(q)} = \frac{a_1}{a_1 + a_3}$
T	F	a_2		T	F	$\Pr(s_2 q) \stackrel{\text{def}}{=} \frac{\Pr(s_2 \& q)}{\Pr(q)} = 0$
F	T	a_3		F	T	$\Pr(s_3 q) \stackrel{\text{def}}{=} \frac{\Pr(s_3 \& q)}{\Pr(q)} = \frac{a_3}{a_1 + a_3}$
F	F	a_4		F	F	$\Pr(s_4 q) \stackrel{\text{def}}{=} \frac{\Pr(s_4 \& q)}{\Pr(q)} = 0$

- Note: the new basic probabilities assigned to the state descriptions, under our “conditionalized” $\Pr(\cdot | q)$ 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. $\Pr(X \rightarrow Y) \geq \Pr(Y | X)$. [Provided that $\Pr(X) > 0$, of course.]

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

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

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

- 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 $1 - a_2 \geq \frac{a_1}{a_1 + a_2}$.

- If $a_1 + a_2 > 0$ and $a_i \in [0, 1]$, this must hold, since then we must have: $a_2 \geq a_2 \cdot (a_1 + a_2)$, and then the boxed formulas are equivalent. \square

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):**

$$\Pr(X | Y) = \Pr(X | Y \ \& \ Z) \cdot \Pr(Z | Y) + \Pr(X | Y \ \& \ \sim Z) \cdot \Pr(\sim Z | Y)$$

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

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

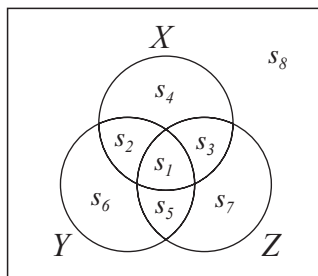
- Two forms of **Bayes's Theorem**. The second one *follows*, using (LTP):

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

The Probability Calculus: An Algebraic Approach XI

- The algebraic approach for *refuting* general claims involves two steps:
 - Translate the claim from probability notation into algebraic terms.
 - Find a (numerical) probability model on which the translation is *false*.
- Show that $\Pr(X | Y \ \& \ Z) = \Pr(X | Y \ \vee \ Z)$ 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: $\frac{a_1}{a_1 + a_5} = \frac{a_1 + a_2 + a_3}{a_1 + a_2 + a_3 + a_5 + a_6 + a_7}$.

(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 X

- One more interesting theorem (due to Popper & Miller), algebraically.
- Let $d(X, Y) \stackrel{\text{def}}{=} \Pr(X | Y) - \Pr(X)$. Then, we have the following theorem:

Theorem (PM). $d(X, Y) = d(X \vee Y, Y) + d(X \vee \sim Y, Y)$.

Proof (algebraic, using STT from X/Y language, above).

$$d(X, Y) \stackrel{\text{def}}{=} \Pr(X | Y) - \Pr(X) = \frac{a_1}{a_1 + a_3} - (a_1 + a_2)$$

$$d(X \vee Y, Y) \stackrel{\text{def}}{=} \Pr(X \vee Y | Y) - \Pr(X \vee Y) = 1 - a_1 - a_2 - a_3$$

$$d(X \vee \sim Y, Y) \stackrel{\text{def}}{=} \Pr(X \vee \sim Y | Y) - \Pr(X \vee \sim Y) = \frac{a_1}{a_1 + a_3} - (a_1 + a_2 + a_4)$$

$$\therefore d(X \vee Y, Y) + d(X \vee \sim Y, Y) = 1 - a_1 - a_2 - a_3 + \frac{a_1}{a_1 + a_3} - a_1 - a_2 - a_4$$

$$= \frac{a_1}{a_1 + a_3} + 1 - a_1 - a_2 - a_3 - a_1 - a_2 - (1 - (a_1 + a_2 + a_3))$$

$$= \frac{a_1}{a_1 + a_3} - (a_1 + a_2). \quad \square$$

The Probability Calculus: An Algebraic Approach XII

- 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 XIII

- 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 II

- Instead of using the algebraic approach for proving theorems, we can also give *axiomatic* proofs. This is the standard way of proving claims in probability calculus (PrSAT doesn't give proofs, so we need axioms).
- Here are two examples of theorems and their *axiomatic* proofs (see the Eells *Appendix*). Note: these are *trivial* from an *algebraic* point of view!

Theorem. $\Pr(\sim p) = 1 - \Pr(p)$.

Proof. Since $p \vee \sim p$ is a tautology, (2) implies $\Pr(p \vee \sim p) = 1$; and since p and $\sim p$ are m.e., (3) implies $\Pr(p \vee \sim p) = \Pr(p) + \Pr(\sim p)$. Therefore, $1 = \Pr(p) + \Pr(\sim p)$, and thus $\Pr(\sim p) = 1 - \Pr(p)$, by simple algebra. \square

Theorem. If $p \models q$, then $\Pr(p) = \Pr(q)$. *Proof.* Assume $p \models q$. Then, p and $\sim q$ are mutually exclusive (inconsistent), and $p \vee \sim q \models \top$. So by axioms (2) and (3), and the previous theorem [$\Pr(\sim p) = 1 - \Pr(p)$]:

$$1 = \Pr(p \vee \sim q) = \Pr(p) + \Pr(\sim q) = \Pr(p) + 1 - \Pr(q)$$

So, $1 = \Pr(p) + 1 - \Pr(q)$, and $0 = \Pr(p) - \Pr(q)$. $\therefore \Pr(p) = \Pr(q)$. \square

Axiomatic Treatment of Probability Calculus I

- A probability model \mathcal{M} is a Boolean algebra of propositions \mathcal{B} , together with a function $\Pr(\cdot) : \mathcal{B} \rightarrow \mathbb{R}$ 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: $\Pr(p | q) \stackrel{\text{def}}{=} \frac{\Pr(p \& q)}{\Pr(q)}$, 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 $p \models \top$, then $\Pr(p) = 1$.
 3. For all $p, q \in \mathcal{L}$, if $p \& q \models \perp$, then $\Pr(p \vee q) = \Pr(p) + \Pr(q)$.

Axiomatic Treatment of Probability Calculus III

- Here are two more axiomatic proofs:

Theorem. If $p \models \perp$, then $\Pr(p) = 0$.

Proof. Assume $p \models \perp$. Then, $\sim p \models \top$, and, by (2), $\Pr(\sim p) = 1$. Then, by the above theorem, $\Pr(\sim p) = 1 - \Pr(p) = 1$, and $\Pr(p) = 0$. \square

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

Proof. First, note the following two Boolean equivalences:

$$p \models (p \& q) \vee (p \& \sim q)$$

$$q \models (p \& q) \vee (\sim p \& q)$$

Thus, by our theorem above, we must have the following two identities:

$$\Pr(p) = \Pr[(p \& q) \vee (p \& \sim q)]$$

$$\Pr(q) = \Pr[(p \& q) \vee (\sim p \& q)]$$

By axiom (3), this yields the following two identities:

$$\Pr(p) = \Pr(p \& q) + \Pr(p \& \sim q)$$

$$\Pr(q) = \Pr(p \& q) + \Pr(\sim p \& q)$$

Now, assume $p \models q$. Then, $p \& \sim q \models \perp$. Hence, by our theorem above, $\Pr(p \& \sim q) = 0$. And, under these circumstances, we must have:

$$\Pr(p) = \Pr(p \& q)$$

$$\Pr(q) = \Pr(p \& q) + \Pr(\sim p \& q)$$

That is to say, we must have the following:

$$\Pr(q) = \Pr(p) + \Pr(\sim p \& q)$$

But, by axiom (1), $\Pr(\sim p \& q) \geq 0$. So, by algebra, $\Pr(q) \geq \Pr(p)$. \square

- This gives us an alternative way to prove $p \models q \implies \Pr(p) = \Pr(q)$. We just apply the previous theorem, in both directions (plus algebra).
- You should now be able to prove that $\Pr(p) \in [0, 1]$, for all p .