Philosophy 148 — Announcements & Such

• Administrative Stuff
  - Branden’s office hours today will be 2:30–3:30.
  - 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.
  - We have a permanent location for the Tuesday section: 206 Wheeler. Stay tuned for the permanent location for the Wednesday section.

• Last Time: More Overview Stuff & Algebraic Probability (Intro.)

• Today’s Agenda
  - An Algebraic Approach to Probability Calculus, Continued
    ∗ “The Algebraic Method” and a Decision Procedure for PC (PrSAT)
    ∗ Systematic vs Extra-Systematic Logical Relations in Algebraic PC
  - Next: An Axiomatic Approach to Probability Calculus
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:

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

- 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 “$Y$”:

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
<th>States</th>
<th>$\Pr(s_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>$s_1$</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>$s_2$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>$s_3$</td>
<td>$\frac{1}{8}$</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>$s_4$</td>
<td>$\frac{11}{24}$</td>
</tr>
</tbody>
</table>

The area of the box is 1, since $\Pr(T) = 1$.

- 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 $B$ — 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 $'\Pr(p)'$ into a sum of some of the $a_i$, and thus we can reduce probabilistic claims about the $p$'s in $B/\mathcal{L}$ 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.** \( \Pr(X \lor Y) = \Pr(X) + \Pr(Y) - \Pr(X \land Y) \).

**Proof.** \( \Pr(X \lor Y) = a_1 + a_2 + a_3 = (a_1 + a_2) + (a_1 + a_3) - a_1 \).

**Theorem.** \( \Pr(X) = \Pr(X \land Y) + \Pr(X \land \lnot Y) \).

**Proof.** \( a_1 + a_2 = a_1 + a_2 \).

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

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

**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 \( \Pr(p) \) 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.** $\Pr(p \mid q) \overset{\text{def}}{=} \frac{\Pr(p \& q)}{\Pr(q)}$, 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 \( \Pr(\cdot \mid q) \) a probability function, for all \( q \) [if \( \Pr(q) > 0 \)]. We can (algebraically) “check” this, as follows:

<table>
<thead>
<tr>
<th>( p )</th>
<th>( q )</th>
<th>( \Pr(s_i) )</th>
<th>( \Pr(s_i \mid q) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>( a_1 )</td>
<td>( \frac{\Pr(s_1 &amp; q)}{\Pr(q)} = \frac{a_1}{a_1 + a_3} )</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>( a_2 )</td>
<td>( \frac{\Pr(s_2 &amp; q)}{\Pr(q)} = 0 )</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>( a_3 )</td>
<td>( \frac{\Pr(s_3 &amp; q)}{\Pr(q)} = \frac{a_3}{a_1 + a_3} )</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>( a_4 )</td>
<td>( \frac{\Pr(s_4 &amp; q)}{\Pr(q)} = 0 )</td>
</tr>
</tbody>
</table>

• 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 VII

- We can also use the algebraic method to verify that theorems which hold for \( \text{Pr}(\cdot) \) also hold for \( \text{Pr}(\cdot \mid q) \), for any \( q \) [provided \( \text{Pr}(q) > 0 \)].

- Recall the following theorem (trivial from an algebraic perspective).

\[
\text{Pr}(X \lor Y) = \text{Pr}(X) + \text{Pr}(Y) - \text{Pr}(X \land Y).
\]

- If \( \text{Pr}(\cdot \mid q) \) is to be a probability function for all \( q \) [where \( \text{Pr}(q) > 0 \)], then we must also have the following theorem, for all \( Z \) [where \( \text{Pr}(Z) > 0 \)]:

\[
\text{Pr}(X \lor Y \mid Z) = \text{Pr}(X \mid Z) + \text{Pr}(Y \mid Z) - \text{Pr}(X \land Y \mid Z).
\]

- Indeed, any theorem that holds for unconditional probabilities \( \text{Pr}(\cdot) \) must also hold for conditional probabilities, that is, when \( \text{Pr}(\cdot) \) is replaced by \( \text{Pr}(\cdot \mid q) \), so long as \( \text{Pr}(q) > 0 \). This will always be the case.

- Using our algebraic method, we can prove the above theorem. 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 …
By our definition of conditional probability, we have:

\[
\Pr(X \lor Y \mid Z) = \frac{\Pr((X \lor Y) \land Z)}{\Pr(Z)} = \frac{\Pr((X \land Z) \lor (Y \land Z))}{\Pr(Z)} = \frac{a_1 + a_3 + a_5}{a_1 + a_3 + a_5 + a_7}
\]

and

\[
\Pr(X \mid Z) + \Pr(Y \mid Z) - \Pr(X \land Y \mid Z) = \frac{\Pr(X \land Z)}{\Pr(Z)} + \frac{\Pr(Y \land Z)}{\Pr(Z)} - \frac{\Pr(X \land Y \land Z)}{\Pr(Z)}
\]

\[
= \frac{\Pr(X \land Z) + \Pr(Y \land Z) - \Pr(X \land Y \land Z)}{\Pr(Z)}
\]

\[
= \frac{(a_1 + a_3) + (a_1 + a_5) - a_1}{a_1 + a_3 + a_5 + a_7} = \frac{a_1 + a_3 + a_5}{a_1 + a_3 + a_5 + a_7}
\]
Here’s a neat theorem of the probability calculus, proved algebraically.

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

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

\[
\Pr(Y \mid X) = \frac{\Pr(Y \& X)}{\Pr(X)} = \frac{\Pr(s_1)}{\Pr(s_1 \lor s_2)} = \frac{a_1}{a_1 + a_2}.
\]

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):**
    \[
    \Pr(X \mid Y) = \Pr(X \mid Y \& Z) \cdot \Pr(Z \mid Y) + \Pr(X \mid Y \& \sim Z) \cdot \Pr(\sim Z \mid Y)
    \]
  - Note: \(\Pr(X \mid \top) = \Pr(X)\). Why? So, the LTP has a special case:
    \[
    \Pr(X \mid \top) = \Pr(X) = \Pr(X \mid \top \& Z) \cdot \Pr(Z \mid \top) + \Pr(X \mid \top \& \sim Z) \cdot \Pr(\sim Z \mid \top)
    \]
    \[
    = \Pr(X \mid Z) \cdot \Pr(Z) + \Pr(X \mid \sim Z) \cdot \Pr(\sim Z)
    \]
  - Two forms of **Bayes’s Theorem.** The second one follows, using (LTP):
    \[
    \Pr(X \mid Y) = \frac{\Pr(Y \mid X) \cdot \Pr(X)}{\Pr(Y)}
    \]
    \[
    = \frac{\Pr(Y \mid X) \cdot \Pr(X)}{\Pr(Y \mid Z) \cdot \Pr(Z) + \Pr(Y \mid \sim Z) \cdot \Pr(\sim Z)}
    \]
One more interesting theorem (due to Popper & Miller), algebraically.

Let \( d(X, Y) \overset{\text{def}}{=} \Pr(X \mid Y) - \Pr(X) \). Then, we have the following theorem:

**Theorem** (PM). \( d(X, Y) = d(X \lor Y, Y) + d(X \lor \neg Y, Y) \).

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

\[
d(X, Y) \overset{\text{def}}{=} \Pr(X \mid Y) - \Pr(X) = \frac{a_1}{a_1 + a_3} - (a_1 + a_2)
\]

\[
d(X \lor Y, Y) \overset{\text{def}}{=} \Pr(X \lor Y \mid Y) - \Pr(X \lor Y) = 1 - a_1 - a_2 - a_3
\]

\[
d(X \lor \neg Y, Y) \overset{\text{def}}{=} \Pr(X \lor \neg Y \mid Y) - \Pr(X \lor \neg Y) = \frac{a_1}{a_1 + a_3} - (a_1 + a_2 + a_4)
\]

\[
\therefore d(X \lor Y, Y) + d(X \lor \neg Y, Y) = 1 - a_1 - a_2 - a_3 + \frac{a_1}{a_1 + a_3} - a_1 - a_2 - a_4
\]

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

\[
= \frac{a_1}{a_1 + a_3} - (a_1 + a_2). \quad \square
\]
The Probability Calculus: An Algebraic Approach XI

- 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 \( \Pr(X | Y \& Z) = \Pr(X | Y \lor Z) \) can be *false*. Here’s a model \( \mathcal{M} \):

<table>
<thead>
<tr>
<th>( X )</th>
<th>( Y )</th>
<th>( Z )</th>
<th>States</th>
<th>( \Pr(s_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>( s_1 )</td>
<td>( a_1 = 1/6 )</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>( s_2 )</td>
<td>( a_2 = 1/6 )</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>( s_3 )</td>
<td>( a_3 = 1/4 )</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>( s_4 )</td>
<td>( a_4 = 1/16 )</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>( s_5 )</td>
<td>( a_5 = 1/6 )</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>( s_6 )</td>
<td>( a_6 = 1/12 )</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>( s_7 )</td>
<td>( a_7 = 1/24 )</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>( s_8 )</td>
<td>( a_8 = 1/16 )</td>
</tr>
</tbody>
</table>

(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 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.
Systematic vs Extra-Systematic Logical Relations I

- The entailment relation \( \vdash \) that we’ve been talking about is just the Boolean entailment relation that is in force within the algebra over which \( \Pr(\cdot) \) is defined. I will call this relation **systematic entailment**.

Because **probability zero is not the same thing as systematic logical falsehood**, there is room to emulate extra-systematic logical relations using probability models. This is an important “trick” we’ll use often.

- Here’s an example. Consider a propositional language with three atomic letters: \( X, Y, Z \). This sets-up the standard 3-atomic-sentence Boolean algebra \( \mathcal{B} \) that we’ve seen several times already. Now, we’ll add a twist.

- Let’s extra-systematically interpret ‘\( X \)’ as \( (\forall x)(Rx \to Bx) \), ‘\( Y \)’ as \( Ra \), and ‘\( Z \)’ as \( Ba \). This extra-systematic interpretation of the atomic sentences has no effect on the systematic logical relations in \( \mathcal{B} \).

- But, we can use a suitable \( \Pr(\cdot) \) over \( \mathcal{B} \) to emulate the extra-systematic (MPL) entailment relations \( (\vdash) \) between \( Ra, Ba, \) and \( (\forall x)(Rx \to Bx) \).
Systematic vs Extra-Systematic Logical Relations II

- **Example.** Extra-systematically, we have: \((\forall x)(Rx \rightarrow Bx) \land Ra \vdash Ba\).

- We do *not* have the corresponding *systematic* entailment: \(X \land Y \not\equiv Z\)!

- But, we can *emulate* this \(\vdash\) relation, by assigning \(\Pr(X \land Y \land \neg Z) = 0\).

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>States</th>
<th>(\Pr(s_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>(s_1)</td>
<td>(a_1)</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>(s_2)</td>
<td>(a_2 = 0)</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>(s_3)</td>
<td>(a_3)</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>(s_4)</td>
<td>(a_4)</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>(s_5)</td>
<td>(a_5)</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>(s_6)</td>
<td>(a_6)</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>(s_7)</td>
<td>(a_7)</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>(s_8)</td>
<td>(a_8)</td>
</tr>
</tbody>
</table>

- By enforcing the *extra-systematic constraint* \(\Pr(X \land Y \land \neg Z) = 0\), we can investigate features of our extra-systematic (*monadic-predicate-logical*) interpretation of \(X\), \(Y\), and \(Z\), using only *sentential* probability calculus.

- This very useful “trick” will be used throughout the course.
Axiomatic Treatment of Probability Calculus I

- A probability model $\mathcal{M}$ is a Boolean algebra of propositions $\mathcal{B}$, together with a function $\Pr(\cdot) : \mathcal{B} \to \mathbb{R}$ 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 $\Pr(p \lor q) = \Pr(p) + \Pr(q)$. [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 $p \vDash \top$, then $\Pr(p) = 1$.
  3. For all $p, q \in \mathcal{L}$, if $p \& q \vDash \bot$, then $\Pr(p \lor q) = \Pr(p) + \Pr(q)$.
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.** \( \Pr(\neg p) = 1 - \Pr(p) \).

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

**Theorem.** If \( p \models q \), then \( \Pr(p) = \Pr(q) \).

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

\[
1 = \Pr(p \lor \neg q) = \Pr(p) + \Pr(\neg q) = \Pr(p) + 1 - \Pr(q)
\]

So, \( 1 = \Pr(p) + 1 - \Pr(q) \), and \( 0 = \Pr(p) - \Pr(q) \). \( \therefore \Pr(p) = \Pr(q) \). \( \square \)
Axiomatic Treatment of Probability Calculus III

- Here are two more axiomatic proofs:

**Theorem.** If \( p \models \bot \), then \( \Pr(p) = 0 \).

**Proof.** Assume \( p \models \bot \). 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:

\[
\begin{align*}
p & \models (p \land q) \lor (p \land \sim q) \\
q & \models (p \land q) \lor (\sim p \land q)
\end{align*}
\]

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

\[
\begin{align*}
\Pr(p) & = \Pr[(p \land q) \lor (p \land \sim q)] \\
\Pr(q) & = \Pr[(p \land q) \lor (\sim p \land q)]
\end{align*}
\]
By axiom (3), this yields the following two identities:

\[ \Pr(p) = \Pr(p \& q) + \Pr(p \& \neg q) \]
\[ \Pr(q) = \Pr(p \& q) + \Pr(\neg p \& q) \]

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

\[ \Pr(p) = \Pr(p \& q) \]
\[ \Pr(q) = \Pr(p \& q) + \Pr(\neg p \& q) \]

That is to say, we must have the following:

\[ \Pr(q) = \Pr(p) + \Pr(\neg p \& q) \]

But, by axiom (1), \( \Pr(\neg p \& q) \geq 0 \). So, by algebra, \( \Pr(q) \geq \Pr(p) \). \( \Box \)

- 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 \( p \).