Philosophy 148 — Announcements & Such

- Administrative Stuff
  - Raul's office is 5323 Tolman (not 301 Moses).
  - We have a permanent location for the Tues. section: 206 Wheeler.
  - We have a permanent location for the Wed. section: 2304 Tolman.

☞ **I have moved the quiz from Tuesday (2/12) to Thursday (2/14)**

- Last Time: Algebraic Probability (Part III) & Intro. to Axiomatic Approach
- Today's Agenda
  - The (Orthodox) Axiomatic Approach to Probability Calculus
    - The axioms, and axiomatic proofs of some basic theorems (again)
    - Perils of Axiomatizing Probability (Unconditional and Conditional)
    - Probabilistic independence and correlation
    - Inverse Probability and Bayes's Theorem
    - That will exhaust the quiz material (more Tuesday about quiz).

Axiomatic Treatment of Probability Calculus I

- Our *algebraic* characterization of $\Pr(\cdot)$ was in terms of *basic probability* assignments $\Pr(s_i) = a_i$ to the *state descriptions* s_i of a language $\mathcal{L}$.
- The only two *systematic* constraints on the a_i are that $a_i \in [0, 1]$, and $\sum a_i = 1$. Assuming the a_i satisfy these two systematic constraints, we then define $\Pr(p)$ for an *arbitrary* sentence $p \in \mathcal{L}$, as follows:

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

- The *axiomatic* characterization of $\Pr(\cdot)$ is as a function from $\mathcal{L}$ to $\mathbb{R}$, which satisfies the following three axioms, for all sentences $p, q \in \mathcal{L}$:

(1) $\Pr(p) \geq 0$. [non-negativity]

(2) If $p \dashv\vDash \top$, then $\Pr(p) = 1$. [normality]

(3) If $p \,\&\, q \dashv\vDash \bot$, then $\Pr(p \vee q) = \Pr(p) + \Pr(q)$. [additivity]

- Conditional probability is *defined* as $\Pr(p \mid q) \overset{\text{def}}{=} \dfrac{\Pr(p \,\&\, q)}{\Pr(q)}$ [if $\Pr(q) > 0$].

Axiomatic Treatment of Probability Calculus II

- The axiomatic method is the standard way people prove Pr-theorems.
- I want to go over the proofs of Theorems ①–④ again now (a bit slower).

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

Proof.

- Since $p \vee \sim p$ is a tautology, (2) implies $\Pr(p \vee \sim p) = 1$.
- Since $p \mathbin{\&} \sim p \mathrel{=\!\mid\models} \bot$, (3) implies $\Pr(p \vee \sim p) = \Pr(p) + \Pr(\sim p)$.

$\therefore$ $1 = \Pr(p) + \Pr(\sim p)$, which implies $\Pr(\sim p) = 1 - \Pr(p)$ by algebra. □

Theorem ②. If $p \mathrel{=\!\mid\models} q$, then $\Pr(p) = \Pr(q)$.

Proof.

- Assume $p \mathrel{=\!\mid\models} q$. Then, $p \mathbin{\&} \sim q \mathrel{=\!\mid\models} \bot$, and $p \vee \sim q \mathrel{=\!\mid\models} \top$.
- So by axioms (2) and (3), and Theorem ①, we have the following:

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

$\therefore$ $1 = \Pr(p) + 1 - \Pr(q)$, which implies $\Pr(p) = \Pr(q)$ by algebra. □

Axiomatic Treatment of Probability Calculus III

Theorem ③. If $p \dashv\vDash \bot$, then $\Pr(p) = 0$.

Proof.

- Assume $p \dashv\vDash \bot$. Then, $\sim p \dashv\vDash \top$, and, by (2), $\Pr(\sim p) = 1$.
- Then, by Theorem ①, $\Pr(\sim\sim p) = 1 - \Pr(\sim p) = 1 - 1 = 0$.

∴ By Theorem ③, $\Pr(\sim\sim p) = \Pr(p) = 0$. □

Theorem ④. If $p \vDash q$, then $\Pr(p) \leq \Pr(q)$.

Proof.

- First, note the following two Boolean equivalences:

$$p \dashv\vDash (p \,\&\, q) \vee (p \,\&\, \sim q)$$

$$q \dashv\vDash (p \,\&\, q) \vee (\sim p \,\&\, q)$$

– Thus, by Theorem ②, 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:

(i) $$\Pr(p) = \Pr(p \,\&\, q) + \Pr(p \,\&\, {\sim}q)$$

(ii) $$\Pr(q) = \Pr(p \,\&\, q) + \Pr({\sim}p \,\&\, q)$$

– Now, assume $p \vDash q$. Then, $p \,\&\, {\sim}q \;\exists\!\vDash\; \bot$. Hence, by Theorem ③, $\Pr(p \,\&\, {\sim}q) = 0$. And, substituting this into identity (i) above yields:

(i′) $$\Pr(p) = \Pr(p \,\&\, q)$$

– Now, combining (i′) and (ii) yields the following identity:

(iii) $$\Pr(q) = \Pr(p) + \Pr({\sim}p \,\&\, q)$$

∴ Since axiom (1) implies $\Pr({\sim}p \,\&\, q) \geq 0$, we have $\Pr(q) \geq \Pr(p)$. □

Axiomatic Treatment of Probability Calculus IV

- Skyrms uses a different-looking set of axioms from ours. One of our first exercises will be to show that Skyrms's axioms follow from ours (but, see below). Here are the six axioms Skyrms adopts (he calls them "rules").
  1. If a statement is a tautology, then its probability is equal to 1.
  2. If a statement is a self-contradiction, then its probability is equal to 0.
  3. If two statements are logically equivalent, they have the same probability.
  4. If p and q are mutually exclusive, then $\Pr(p \vee q) = \Pr(p) + \Pr(q)$.
  5. $\Pr({\sim}p) = 1 - \Pr(p)$
  6. $\Pr(p \vee q) = \Pr(p) + \Pr(q) - \Pr(p \,\&\, q)$
- Skyrms uses the same definition of conditional probability as we do, namely, that $\Pr(p \mid q) = \frac{\Pr(p \& q)}{\Pr(q)}$, if $\Pr(q) > 0$ (otherwise, it's undefined).
- Note: Skyrms's "rules" are *not complete*! That is, there are some theorems of probability calculus that do *not* follow from Skyrms's "rules". Example: our axiom (1)! I have posted a handout on our website that shows this.

Axiomatic Treatment of Probability Calculus V

- When it comes to unconditional probability, there is widespread agreement about the axioms we have adopted. Not so, when it comes to *conditional* probability. Some people think, for instance, that $\Pr(p \mid p)$ should be equal to 1, *even if* $\Pr(p) = 0$. Our theory disagrees.
- For this reason, some people have opted to take *conditional* probability as primitive and then to define *unconditional* probability in terms of *it*. See Hájek's paper "What Conditional Probability Could Not Be" (on website).
- There are various ways of doing this. These alternative axiomatizations tend to differ on how they treat $\Pr(p \mid q)$ when q is a *contradiction*. Most theories of conditional probability agree that: $\Pr(p \mid \bot) = 1$, for all p.
- This is a carry-over from classical deductive logic, which says that $\bot \vDash p$, for all p. Indeed, classical "logical" interpretations of (conditional) probability go hand-in-hand with this kind of assumption. We'll come back to this in the sections on "logical" probability and inductive logic.

Axiomatic Treatment of Probability Calculus VI

- There are various axiomatizations of conditional probability $\Pr(p \mid q)$. Unfortunately, many of them are *inconsistent*, because they are not careful about the $q = \bot$ case. Salmon & Earman (which we'll read later) use:

(A1) $\Pr(p \mid q) \in [0,1]$.

(A2) If $q \vDash p$, then $\Pr(p \mid q) = 1$.

(A3) If p and q are mutually exclusive, then $\Pr(p \vee q \mid r) = \Pr(p \mid r) + \Pr(q \mid r)$.

(A4) $\Pr(p \,\&\, q \mid r) = \Pr(p \mid r) \cdot \Pr(q \mid r \,\&\, p)$.

- That Salmon's (A1)–(A3) are inconsistent can be shown, as follows:
  - By (A2), $\Pr(p \mid \bot) = 1$, and $\Pr({\sim}p \mid \bot) = 1$.
  - By (A3), $\Pr(p \vee {\sim}p \mid \bot) = \Pr(p \mid \bot) + \Pr({\sim}p \mid \bot)$.
  - Therefore, $\Pr(p \vee {\sim}p \mid \bot) = 1 + 1 = 2$, which contradicts (A1). □
- Consistency can be restored, by amending (A2) and/or (A3) in various ways. Different ways of doing this lead to different axiomatizations.
- See Hájek and Goosens (on website) for discussions of axiomatizing CP.

Axiomatic Treatment of Probability Calculus VII

- Exercise: Prove *axiomatically* that $\Pr(\cdot \mid q)$ is a probability function, assuming our axioms for $\Pr(\cdot)$ and our definition of $\Pr(\cdot \mid \cdot)$, *i.e.*, prove that $\Pr(\cdot \mid q)$ satisfies our axioms for $\Pr(\cdot)$, for all q such that $\Pr(q) > 0$.
  - Axiom (1) requires $\Pr(p \mid q) \geq 0$. By definition, $\Pr(p \mid q) \overset{\text{def}}{=} \dfrac{\Pr(p \,\&\, q)}{\Pr(q)}$. Since both $\Pr(p \,\&\, q) \geq 0$ and $\Pr(q) > 0$, it follows that $\Pr(p \mid q) \geq 0$. □
  - Axiom (2) requires $\Pr(\top \mid q) = 1$. By definition, $\Pr(\top \mid q) \overset{\text{def}}{=} \dfrac{\Pr(\top \,\&\, q)}{\Pr(q)}$. By theorem ②, $\Pr(\top \,\&\, q) = \Pr(q)$. Thus, $\Pr(\top \mid q) = \dfrac{\Pr(q)}{\Pr(q)} = 1$. □
  - Axiom (3) requires $\Pr(p \vee r \mid q) = \Pr(p \mid q) + \Pr(r \mid q)$, if p and r are mutually exclusive. Assume p and r are mutually exclusive. Then:

$$\Pr(p \vee r \mid q) = \frac{\Pr[(p \vee r) \,\&\, q]}{\Pr(q)} \text{ [Definition of } \Pr(\cdot \mid \cdot)]$$

$$= \frac{\Pr[(p \,\&\, q) \vee (r \,\&\, q)]}{\Pr(q)} \text{ [Theorem ②]}$$

$$= \frac{\Pr(p \,\&\, q) + \Pr(r \,\&\, q)}{\Pr(q)} \text{ [Axiom (3)]}$$

$$= \Pr(p \mid q) + \Pr(r \mid q) \text{ [Definition of } \Pr(\cdot \mid \cdot)] \quad \square$$

- So, our definition of $\Pr(\cdot \mid q)$ yields a *probability* function, for all q such that $\Pr(q) > 0$. This is also true for all non-orthdodox axiomatizations of $\Pr(\cdot \mid \cdot)$. But, those approaches can *fail* this requirement when $\Pr(q) = 0$.
- *E.g.*, $\Pr(\cdot \mid \bot)$ will *not* be a probability function on many non-orthodox accounts of $\Pr(\cdot \mid \cdot)$, since many of them assume $\Pr(p \mid \bot) = 1$, for *all* p.
- As in Salmon's theory, this leads to $\Pr(p \vee \sim p \mid \bot) = 2$, if we enforce axiom (3) on $\Pr(\cdot \mid \bot)$. Thus, non-orthodox $\Pr(\cdot \mid \bot)$'s are not probabilities.

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: $\dfrac{\Pr(p \,\&\, y)}{\Pr(y)} = \Pr(p)$]
  - $\Pr(q \mid p) = \Pr(q)$ [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 $\mathbf{P} = \{p_1, \dots p_n\}$ is *mutually independent* if all subsets $\{p_i, \dots, p_j\} \subseteq \mathbf{P}$ 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.

- Here's an axiomatic proof that $\Pr(p \,\&\, q) = \Pr(p) \cdot \Pr(q) \Leftrightarrow \Pr(p \mid q) = \Pr(p \mid {\sim}q)$, provided that that $\Pr(q) \in (0, 1)$:

$$\Pr(p \mid q) = \Pr(p \mid {\sim}q) \Leftrightarrow \frac{\Pr(p \,\&\, q)}{\Pr(q)} = \frac{\Pr(p \,\&\, {\sim}q)}{\Pr({\sim}q)} \text{ [definition of CP]}$$

$$\Leftrightarrow \frac{\Pr(p \,\&\, q)}{\Pr(q)} - \frac{\Pr(p \,\&\, {\sim}q)}{\Pr({\sim}q)} = 0 \text{ [algebra]}$$

$$\Leftrightarrow \frac{\Pr(p \,\&\, q) \cdot \Pr({\sim}q) - \Pr(p \,\&\, {\sim}q) \cdot \Pr(q)}{\Pr(q)\Pr({\sim}q)} = 0 \text{ [algebra]}$$

$$\Leftrightarrow \Pr(p \,\&\, q) \cdot \Pr({\sim}q) - \Pr(p \,\&\, {\sim}q) \cdot \Pr(q) = 0 \text{ [algebra]}$$

$$\Leftrightarrow \Pr(p \,\&\, q) \cdot (1 - \Pr(q)) - \Pr(p \,\&\, {\sim}q) \cdot \Pr(q) = 0 \text{ [algebra]}$$

$$\Leftrightarrow \Pr(p \,\&\, q) - \Pr(q) \cdot [\Pr(p \,\&\, q) + \Pr(p \,\&\, {\sim}q)] = 0 \text{ [algebra]}$$

$$\Leftrightarrow \Pr(p \,\&\, q) - \Pr(q) \cdot \Pr((p \,\&\, q) \vee (p \,\&\, {\sim}q)) = 0 \text{ [additivity axiom]}$$

$$\Leftrightarrow \Pr(p \,\&\, q) - \Pr(q) \cdot \Pr(p) = 0 \text{ [Theorem ②]}$$

$$\Leftrightarrow \Pr(p \,\&\, q) = \Pr(p) \cdot \Pr(q) \text{ [algebra]} \quad \square$$

- 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 ${\sim}q$. Pretty easy proof. Assume that p and q are independent. Then, since $p \mathrel{=\!\!|\!\models} (p \,\&\, q) \vee (p \,\&\, {\sim}q)$

$$\Pr(p) = \Pr(p \,\&\, q) + \Pr(p \,\&\, {\sim}q) \text{ [Th. ② and additivity axiom]}$$

$$= \Pr(p) \cdot \Pr(q) + \Pr(p \,\&\, {\sim}q) \text{ [independence]}$$

$$\text{So, } \Pr(p \,\&\, {\sim}q) = \Pr(p) \cdot [1 - \Pr(q)] = \Pr(p) \cdot \Pr({\sim}q) \text{ [algebra \& Th. ①]} \quad \square$$

- More generally, if $\{p, q, r\}$ are mutually independent, then p is independent of *any* propositional function of q and r, *e.g.*, $p \perp\!\!\!\perp q \vee r$.

  Proof. $\Pr(p \,\&\, (q \vee r)) = \Pr((p \,\&\, q) \vee (p \,\&\, r))$ [Theorem ②]

  $$= \Pr(p \,\&\, q) + \Pr(p \,\&\, r) - \Pr(p \,\&\, q \,\&\, r) \text{ [general additivity]}$$

  $$= \Pr(p) \cdot \Pr(q) + \Pr(p) \cdot \Pr(r) - \Pr(p) \cdot \Pr(q \,\&\, r) \text{ [mutual } \perp\!\!\!\perp]$$

  $$= \Pr(p) \cdot [\Pr(q) + \Pr(r) - \Pr(q \,\&\, r)] \text{ [algebra]}$$

  $$= \Pr(p) \cdot \Pr(q \vee r) \text{ [general additivity]} \quad \Box$$

- This last proof makes heavy use of general additivity (Skyrms's rule 6): $\Pr(p \vee q) = \Pr(p) + \Pr(q) - \Pr(p \,\&\, q)$. This is one of our first exercises involving axiomatic proof (to prove this rule from our axioms).

- How might one prove the more general theorem above: that if $\{p, q, r\}$ are mutually independent, then p is independent of *any* propositional function of q and r? And, is there an even more general theorem here?

- To wit: is it the case that if $\mathbf{P} = \{p_1, \ldots, p_n\}$ is a mutually independent set, then *any p-functions of any two disjoint subsets of* $\mathbf{P}$ are independent?

Independence, Correlation, and Anti-Correlation 2

- So far, we've seen a some *proofs* of *true* general claims about independence, correlation, *etc.* Now, for some *counterexamples*!
- As always, these are numerical probability models in which some claim *fails*. We have seen two false claims about $\perp\!\!\!\perp$ already. Let's prove them.
- **Theorem**. Pairwise independence of a collection of three propositions $\{X, Y, Z\}$ does not entail mutual independence of the collection. That is to say, there exist probability models in which (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:

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

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:

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

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

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

Hence,

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

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

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

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

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

- 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:
  - $p \perp\!\!\!\perp q \mid r$
  - $p \perp\!\!\!\perp q \mid {\sim} r$
  - $p \not\perp\!\!\!\perp q$
- This is *Simpson's Paradox*. See `PrSAT` notebook for an example.

Inverse Probability and Bayes's Theorem I

- $\Pr(h \mid e)$ is sometimes called the *posterior* h (on e). $\Pr(h)$ is sometimes 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, we can see that the posterior and the likelihood are related as follows:

$$\Pr(h \mid e) = \frac{\Pr(e \mid h) \cdot \Pr(h)}{\Pr(e)}$$

- This is a simple form of *Bayes's Theorem*, which says *posteriors are proportional to likelihoods*, with priors determining the proportionality.
- **Law of Total Probability 1**. 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:

$$\Pr(h \mid e) = \frac{\Pr(e \mid h) \cdot \Pr(h)}{\Pr(e \mid h) \cdot \Pr(h) + \Pr(e \mid {\sim}h) \cdot \Pr({\sim}h)}$$

- More generally, a *partition* (of logical space) is a set of propositions $\{p_1, \ldots, p_n\}$ such that the p_i are pairwise mutually exclusive and the disjunction of the p_i is exhaustive (*i.e.*, $p_1 \vee \cdots \vee p_n \mathrel{=\!\mid\!\models} \top$).

- **Law of Total Probability 2.** If $\{p_1, \ldots, p_n\}$ is a partition of logical space, and $\Pr(p_i)$ is non-extreme for each of the p_i in the partition, then:

$$\begin{aligned}\Pr(e) &= \Pr((e \mathbin{\&} p_1) \vee \cdots \vee (e \mathbin{\&} p_n)) \\ &= \Pr(e \mathbin{\&} p_1) + \cdots + \Pr(e \mathbin{\&} p_n) \\ &= \Pr(e \mid p_1) \cdot \Pr(p_1) + \cdots + \Pr(e \mid p_n) \cdot \Pr(p_n)\end{aligned}$$

- **Bayes's Theorem (general).** If $\{p_1, \ldots, p_n\}$ is a partition of logical space, and $\Pr(p_i)$ is non-extreme for each of the p_i in the partition, then:

$$\Pr(h \mid e) = \frac{\Pr(e \mid h) \cdot \Pr(h)}{\Pr(e \mid p_1) \cdot \Pr(p_1) + \cdots + \Pr(e \mid p_n) \cdot \Pr(p_n)}$$

Inverse Probability and Bayes's Theorem II

- 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 (like Hacking's O.Q. #5): What is $\Pr(H \mid E)$? What would you guess? Most experts guess a pretty high number (near 0.8, usually).

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

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

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

E	H	$\Pr(s_i)$
T	T	$a_1 = 0.008$
T	F	$a_2 = 0.099$
F	T	$a_3 = 0.002$
F	F	$a_4 = 0.891$

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

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

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

- 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*. I will return to this example later in the course, and ask whether it really is a mistake to report a large number in this example. Perhaps it is not a mistake.

Inverse Probability and Bayes's Theorem III

- Hacking's O.Q. #6: You are a physician. You think it is quite probable (say 90% probable) that one of your patients has strep throat (S). You take some swabs from the throat and send them to the lab for testing. The test is imperfect, with the following likelihoods (Y is + result, N is −):
  - $\Pr(Y \mid S) = 0.7$, $\Pr(Y \mid {\sim}S) = 0.1$
- You send five successive swabs to the lab, from the same patient. You get the following results, in order: Y, N, Y, N, Y. What is $\Pr(S \mid YNYNY)$?
- Hacking: Assume that the 5 test results are *conditionally independent*, *given* both S and ${\sim}S$, *i.e.*, that S *screens-off* the 5 tests results. So:
  - $\Pr(YNYNY \mid S) = 0.7 \cdot 0.3 \cdot 0.7 \cdot 0.3 \cdot 0.7 \approx 0.03087$
  - $\Pr(YNYNY \mid {\sim}S) = 0.1 \cdot 0.9 \cdot 0.1 \cdot 0.9 \cdot 0.1 \approx 0.00081$

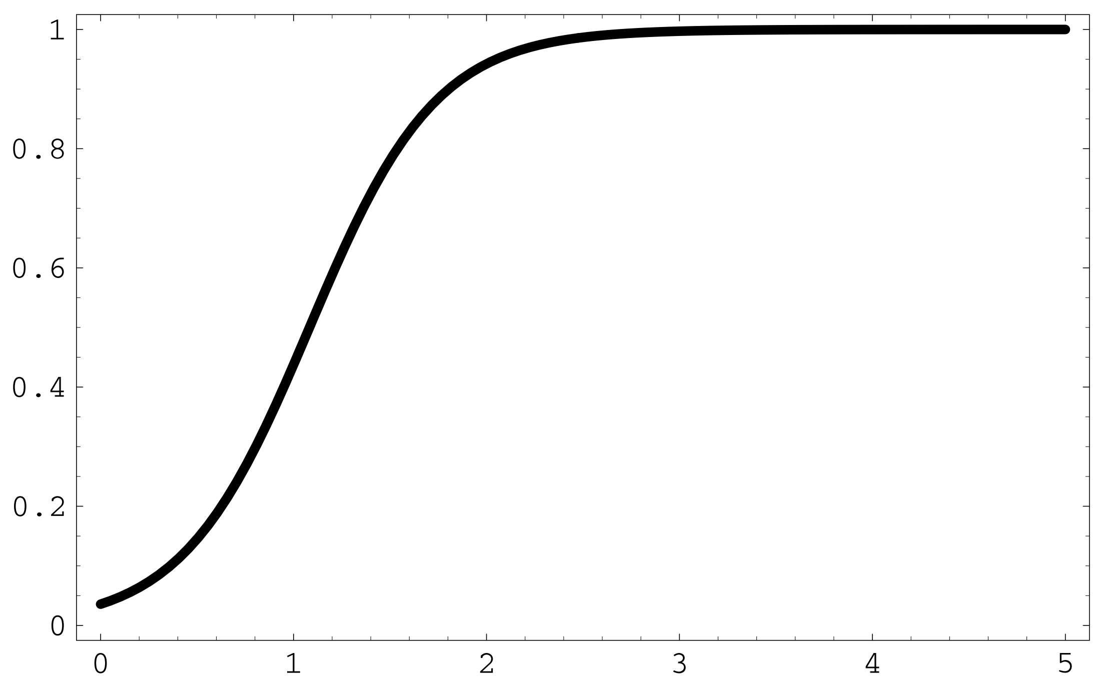
$$\Pr(S \mid YNYNY) = \frac{\Pr(YNYNY \mid S) \cdot \Pr(S)}{\Pr(YNYNY \mid S) \cdot \Pr(S) + \Pr(YNYNY \mid {\sim}S) \cdot \Pr({\sim}S)}$$

$$= \frac{0.03087 \cdot 0.9}{0.03087 \cdot 0.9 + 0.00081 \cdot 0.1} \approx 0.997$$

General Analysis of Hacking's "Odd Question #6"

- If n is the number of Y results, then $(5-n)$ is the number of N results (out of 5 results). Bayes's theorem allows us to calculate $\Pr(S \mid E_n)$, where E_n is evidence consisting of n Y results and $(5-n)$ N results (any order):

$$\frac{\Pr(E_n \mid S) \cdot \Pr(S)}{\Pr(E_n \mid S) \cdot \Pr(S) + \Pr(E_n \mid {\sim}S) \cdot \Pr({\sim}S)} = \frac{0.7^n \cdot 0.3^{5-n} \cdot 0.9}{0.7^n \cdot 0.3^{5-n} \cdot 0.9 + 0.1^n \cdot 0.9^{5-n} \cdot 0.1}$$



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:

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

- 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?
  - We'll be thinking more about some of these questions in the next unit.