Overview

Logic as a theory of:

- truth-preserving inference
- consistency
- definability
- proof / deduction
- rationality

Probability as a theory of:

- uncertain inference
- learning
- information
- induction
- rationality
Some questions and points of contact:

- In what ways might probability be said to extend logic?

- How do probability and various logical systems differ on what they say about rational inference?

- Can logic be used to gain a better understanding of probability? Say, through standard techniques of formalization, definability, questions of completeness, complexity, etc.?

- Can the respective advantages of logical representation and probabilistic learning and inference be combined to create more powerful reasoning systems?
Course Outline

- Day 1: Probability as (Extended) Logic
- Day 2: Probability, Nonmonotonicity, and Graphical Models
- Day 3: Beyond Boolean Logic
- Day 4: Qualitative Probability
- Day 5: Logical Dynamics
Interpretations of Probability

- **Frequentist**: Probabilities are about ‘limiting frequencies’ of events.

- **Propensity**: Probabilities are about physical dispositions, or propensities, of events.

- **Logical**: Probabilities are determined objectively via a logical language and some additional principles, e.g., of ‘symmetry’.

- **Bayesian**: Probabilities are subjective and reflect an agent’s degree of confidence concerning some event.

We will mostly remain neutral, focusing on mathematical questions, but will note when the interpretation is critical.
A measurable space is a pair \((W, \mathcal{E})\) with
- \(W\) is an arbitrary set
- \(\mathcal{E}\) is a \(\sigma\)-algebra over \(W\), i.e., a subset of \(\wp(W)\) closed under complement and infinite union.

A probability space is a triple \((W, \mathcal{E}, \mu)\), with \((W, \mathcal{E})\) a measurable space and \(\mu : \mathcal{E} \to [0, 1]\) a measure function satisfying:
- \(\mu(W) = 1\);
- \(\mu(E \cup F) = \mu(E) + \mu(F)\), whenever \(E \cap F = \emptyset\).

A random variable from \((W, \mathcal{E}, \mu)\) to some other space \((V, \mathcal{F})\) is a function \(X : W \to V\) such that:

\[
\text{whenever } X(E) = F \text{ and } F \in \mathcal{F}, \text{ we have } E \in \mathcal{E}.
\]

Then \(\mu\) induces a natural distribution for \(X\), typically written
\[
P(X = \nu) = \mu(\{w \in W : X(w) = \nu\}).
\]
Suppose we have a propositional logical language $\mathcal{L}$ generated as follows

$$\phi ::= A \mid B \mid \ldots \mid \phi \land \phi \mid \neg \phi$$

We can define a probability $P : \mathcal{L} \rightarrow [0, 1]$ directly on $\mathcal{L}$, requiring

- $P(\phi) = 1$, for any tautology $\phi$;
- $P(\phi \lor \psi) = P(\phi) + P(\psi)$, whenever $\models \neg(\phi \land \psi)$.

Equivalent set of requirements:

- $P(\phi) = 1$ for any tautology;
- $P(\phi) \leq P(\psi)$ whenever $\models \phi \rightarrow \psi$;
- $P(\phi) = P(\phi \land \psi) + P(\phi \land \neg \psi)$. 
It is then easy to show:

- $P(\varphi) = 0$, for any contradiction $\varphi$;
- $P(\neg \varphi) = 1 - P(\varphi)$;
- $P(\varphi \lor \psi) = P(\varphi) + P(\psi) - P(\varphi \land \psi)$;
- A propositional valuation sending atoms to 1 or 0 is a special case of a probability function;
- A probability on $\mathcal{L}$ gives rise to a standard probability measure over ‘world-states’, i.e., maximally consistent sets of formulas from $\mathcal{L}$.
With this setup we can easily define, e.g., conditional probability:

\[
P(\varphi|\psi) = \frac{P(\varphi \land \psi)}{P(\psi)},
\]

provided \(P(\psi) > 0\). With this it is easy to prove Bayes’ Theorem:

**Theorem (Bayes)**

\[
P(\varphi|\psi) = \frac{P(\psi|\varphi)P(\varphi)}{P(\psi)}.\]
Today’s Theme: Probability as Logic

Three senses in which probability has been claimed as part of logic:

1. Logical Deduction and Probability Preservation
2. Consistency and Fair Odds
3. Patterns of Plausible Reasoning
PART I: Logical Deduction and Probability Preservation
Suppose we can show $\Gamma \vdash \varphi$ in propositional logic, but the assumptions $\gamma \in \Gamma$ themselves are uncertain.

Knowing about the probabilities of formulas in $\Gamma$, what can we conclude about the probability of $\varphi$?

**Example (Adams & Levine)**

A telephone survey is made of 1,000 people as to their political party affiliation. Supposing 629 people report being Democrat, with some possibility of error in each case. What can we say about the probability of the following statement?

At least 600 people are Democrats.
Proposition (Suppes 1966)
If $\Gamma \vdash \varphi$, then for all $P$ we have $P(\neg \varphi) \leq \sum_{\gamma \in \Gamma} P(\neg \gamma)$. 
Example (Kyburg’s 1965 ‘Lottery Paradox’)
Imagine a fair lottery with \( n \) tickets. Let \( \gamma_1, \ldots, \gamma_n \) be sentences:

\[
\gamma_i : \text{ticket } i \text{ is the winner.}
\]

Each sentence \( \neg \gamma_i \) might seem reasonable, since it has probability \( \frac{n-1}{n} \).
However, putting these all together rapidly decreases probability. Indeed,

\[
P(\bigwedge_{i \leq n} \neg \gamma_i) = 0.
\]

This demonstrates that the (logical) conclusion of all these individually reasonable premises—that none of the tickets will win the lottery—has probability 0.
Definition (Suppes Entailment)
\[ \Gamma \vdash_s \varphi, \text{ iff for all } P: \]
\[ P(\neg \varphi) \leq \sum_{\gamma \in \Gamma} P(\neg \gamma). \]

Definition (Adams Entailment)
\[ \Gamma \vdash_a \varphi, \text{ iff for all } \epsilon > 0, \text{ there is a } \delta > 0, \text{ such that for all } P: \]
\[ \text{if } P(\neg \gamma) < \delta \text{ for all } \gamma \in \Gamma, \text{ then } P(\neg \varphi) < \epsilon. \]

Lemma
\[ \Gamma \vdash_s \varphi, \text{ iff } \Gamma \vdash_a \varphi. \]
Theorem (Adams 1966)

Classical logic is sound and complete for Suppes/Adams entailment.
The following are equivalent:

(a) $\Gamma \vdash \varphi$  \hspace{2cm} (b) $\Gamma \models \varphi$  \hspace{2cm} (c) $\Gamma \models_s \varphi$  \hspace{2cm} (d) $\Gamma \models_a \varphi$.

Proof.

(a) $\iff$ (b) is just standard completeness.
The following are equivalent:

\begin{align*}
\text{(a)} & \quad \Gamma \vdash \varphi & \text{(b)} & \quad \Gamma \models \varphi & \text{(c)} & \quad \Gamma \models_s \varphi & \text{(d)} & \quad \Gamma \models_a \varphi.
\end{align*}

Proof.

For (b) \implies (c), suppose \( \Gamma \not\models_s \varphi \), so there is a \( P \), such that

\[ 1 - P(\varphi) > \sum_{\gamma \in \Gamma} P(\neg \gamma). \]

From this it follows that

\[ 1 > \sum_{\gamma \in \Gamma} P(\neg \gamma) + P(\varphi). \]
The following are equivalent:

\[(a) \; \Gamma \vdash \varphi \quad (b) \; \Gamma \models \varphi \quad (c) \; \Gamma \models_s \varphi \quad (d) \; \Gamma \models_a \varphi.\]

**Proof.**

For \((b) \Rightarrow (c)\), suppose \(\Gamma \not\models_s \varphi\), so there is a \(P\), such that

\[
1 - P(\varphi) > \sum_{\gamma \in \Gamma} P(\neg \gamma).
\]

From this it follows that

\[
1 > \sum_{\gamma \in \Gamma} P(\neg \gamma) + P(\varphi) \geq P(\gamma_1 \land \cdots \land \gamma_n \rightarrow \varphi),
\]

for any subset \(\{\gamma_1, \ldots, \gamma_n\} \subseteq \Gamma\). Since all tautologies must have probability 1, no such implication \(\gamma_1 \land \cdots \land \gamma_n \rightarrow \varphi\) is a tautology.
The following are equivalent:

\[(a) \quad \Gamma \vdash \varphi \quad (b) \quad \Gamma \models \varphi \quad (c) \quad \Gamma \models_s \varphi \quad (d) \quad \Gamma \models_a \varphi.\]

**Proof.**

For \((c) \Rightarrow (d)\), suppose \(\Gamma \models_s \varphi\), and \(\epsilon > 0\). If \(\Gamma\) is finite, choose

\[\delta = \frac{\epsilon}{|\Gamma|}.\]

Then we have

\[P(\neg \varphi) \leq \sum_{\gamma} P(\neg \gamma) < \sum_{\gamma} \delta = \epsilon.\]

Hence \(P(\neg \varphi) < \epsilon\), and \(\Gamma \models_a \varphi\). The case of \(\Gamma\) infinite is left as exercise.
The following are equivalent:

\[(a) \quad \Gamma \vdash \varphi \quad (b) \quad \Gamma \models \varphi \quad (c) \quad \Gamma \models_s \varphi \quad (d) \quad \Gamma \models_a \varphi.\]

**Proof.**

Finally, for (d) \(\Rightarrow\) (b), suppose \(\Gamma \not\models \varphi\). For any \(\epsilon > 0\), we can simply take a valuation witnessing \(\Gamma \not\models \varphi\), which is itself a probability function \(P\) making \(P(\varphi) = 0\) and \(P(\gamma) = 1\), for all \(\gamma \in \Gamma\). Hence \(\Gamma \not\models_a \varphi\).

(Note that essentially the same proof works to show (c) \(\Rightarrow\) (b).)

Thus we have shown (a) \(\Leftrightarrow\) (b) \(\Rightarrow\) (c) \(\Rightarrow\) (d) \(\Rightarrow\) (b). \(\dashv\)
Example

In the telephone survey example, suppose each person’s statement has probability 0.999. Then the best upper bound we can get on probability for the negation of the conclusion is

\[ 629 \times 0.001 = 0.629. \]

That is, our best lower bound on the probability that there are at least 600 Democrats is 0.371. This seems weak.
Definition
Where $\Gamma \models \varphi$, we say $\Gamma' \subseteq \Gamma$ is essential if $\Gamma - \Gamma' \not\models \varphi$.

Definition
Given $\Gamma \models \varphi$, the degree of essentialness of $\gamma \in \Gamma$ is given by

$$E(\gamma) = \frac{1}{|S_\gamma|},$$

where $|S_\gamma|$ is the size of the smallest essential set containing $\gamma$.

Theorem (Adams & Levine 1975)
If $\Gamma \models \varphi$, then

$$P(\neg \varphi) \leq \sum_{\gamma \in \Gamma} E(\gamma) P(\neg \gamma).$$
Example

Back to the telephone survey example, we can now obtain a much better lower bound on what seems to be a strong conclusion. The upper bound on the negation of the conclusion is now

$$629 \times 0.001 \times \frac{1}{30} \approx 0.02,$$

making the lower bound on probability just below 0.98.
Part II: Consistency and Fair Odds
We strive to make judgments as dispassionate, reflective, and wise as possible by a doctrine that shows where and how they intervene and lays bare possible inconsistencies between judgments. There is an instructive analogy between [deductive] logic, which convinces one that acceptance of some opinions as ‘certain’ entails the certainty of others, and the theory of subjective probabilities, which similarly connects uncertain opinions.

–Bruno de Finetti, 1974
Consistency and Fair Odds

de Finetti’s central idea:

Probabilities as chance-based odds

Example (Shakespeare, via Girotto & Gonzalez; Howson)

*Speed*: Sir Proteus, save you! Saw you my master?

*Proteus*: But now he parted hence, to embark for Milan.

*Speed*: Twenty to one, then, he is shipp’d already.

(*Two Gentlemen of Verona*, 1592)

*Gloss*: A bet that *collects* $Y$ if he is gone, and *pays* $X$ if not, will be *fair*, provided $X : Y = 20 : 1$. These are *fair betting odds*.

The normalized betting odds, $\frac{20}{21}$ and $\frac{1}{21}$, are called *betting quotients*. 
Consistency and Fair Odds

- \( S \): stake
- \( \mathbb{1}_\varphi \): indicator function (= 1 if \( \varphi \) holds, 0 otherwise)
- \( p \): betting quotient for \( \varphi \)

A bet on \( \varphi \) at (positive) stake \( S \):
- Pay \( S \ast p \)
- Receive \( S \) if \( \varphi \) is true, nothing otherwise

A bet against \( \varphi \) at (negative) stake \( S \):
- Receive \( S \ast p \)
- Pay \( S \) if \( \varphi \) is true, nothing otherwise

In both cases the value of the gamble is given by:

\[
S(\mathbb{1}_\varphi - p)
\]
The probability calculus is about probabilities of logically complex propositions. What can we say about fair betting odds in this setting?

de Finetti’s critical assumption (cf. Skyrms 1975, Howson 2007, etc.):

(A) Fair gambles do not become collectively unfair on collection into a joint gamble.

Assumption (A)—a kind of agglomerativity assumption—will allow us to reason about probabilities, by reasoning about which gambles are fair on the basis of the assumed fairness of other gambles.
Proposition

1. If \( p \) is fair betting quotient for \( \varphi \), then \( 1-p \) is fair for \( \neg \varphi \).

2. If \( \models \neg(\varphi \land \psi) \), with fair betting quotients \( p \) and \( q \) on \( \varphi \) and \( \psi \), respectively, then \( p + q \) is fair betting quotient for \( \varphi \lor \psi \).

Proof.

For 1., note that

\[
S(1_\varphi - p) = -S[1_{\neg \varphi} - (1 - p)].
\]

For 2., note that

\[
S(1_\varphi - p) + S(1_\psi - q) = S[1_{\varphi \lor \psi} - (p + q)].
\]

We now invoke principle \((A)\).
Fundamental idea:

Take fair betting quotients $p$ for formulas $\varphi$ to be probabilities $P(\varphi)$.

The axioms of probability are thus axioms of consistency.
Dutch Book Theorem

If we conceive of principle \( \mathcal{A} \) as telling us what combinations of bets a person ought to accept, then we can prove:

**Theorem (de Finetti 1937)**

Someone with judgments \( P : \mathcal{L} \rightarrow [0,1] \) is subject to a sure loss, iff \( P \) is inconsistent with the probability axioms.
Consistency and Fair Odds

Dutch Book Theorem

Example
Suppose $\models \neg (\varphi \land \psi)$, and $P(\varphi \lor \psi) = r$, while $P(\varphi) = p$, $P(\psi) = q$, and $r > p + q$. Such a person would then sell bets:

$$-1(1_{\varphi} - p) \quad \text{and} \quad -1(1_{\psi} - q)$$

while buying

$$1(1_{\varphi \lor \psi} - r)$$

which results in a sure loss:

$$-r + p + q < 0.$$
Suppose $P : \mathcal{L}' \to [0, 1]$ is defined on some finite sublanguage $\mathcal{L}' \subseteq \mathcal{L}$.

**Theorem (de Finetti 1974)**

The following are equivalent:

1. $P$ does not conflict with the probability axioms (i.e., is consistent).
2. $P$ is the restriction of a probability function on all of $\mathcal{L}$.
3. No system of bets with odds as given by $P$ leads to certain loss.

In fact, de Finetti proved something stronger, for general (potentially uncountable) probability spaces.

Many have argued that 3. is merely for dramatic effect, with consistency the fundamental notion. (See especially papers by Colin Howson.)
Part III:
Patterns of Plausible Reasoning
Our theme is simply: *Probability Theory as Extended Logic*. The “new” perception amounts to the recognition that the mathematical rules of probability theory are not merely rules for calculating frequencies of “random variables”; they are also the unique consistent rules for conducting inference (i.e., plausible reasoning) of any kind . . .

—E. T. Jaynes, 1995
Deductive inference

All wombats are wild animals
   No wild animals snore
       No wombats snore

José is either in Frankfurt or Paris
   José is not in Paris
       José is in Frankfurt

...
Patterns of Plausible Reasoning

Plausible Inference

Example (Jaynes 1995)
Suppose some dark night a policeman walks down a street, apparently deserted; but suddenly he hears a burglar alarm, looks across the street, and sees a jewelry store with a broken window. Then a gentleman wearing a mask comes crawling out through the broken window, carrying a bag which turns out to be full of expensive jewelry. The policeman doesn’t hesitate at all in deciding that this gentleman is dishonest.

\[
\varphi \text{ being true makes } \psi \text{ more plausible} \\
\psi \text{ is true} \\
\rightarrow \varphi \text{ becomes more plausible}
\]

(N.B. This is one of Pólya’s (1954) five patterns of plausible inference.)
Jaynes’ three assumptions guiding a theory of probability:

1. Degrees of plausibility should be real numbers.
2. We must respect logic in requiring consistency of assessment.
3. The aim is to capture intuitive reasoning patterns about plausibility.

\[
\phi | \psi
\]

How plausible is \( \phi \) given that \( \psi \) is true?
In order for $\varphi \land \psi$ to be true, $\psi$ should be true. Thus $\psi|\chi$ should somehow be involved.

If $\psi$ is true, we still need $\varphi$ to be true as well. So $\varphi|\psi \land \chi$ should somehow be involved.

Thus, we conclude that there should be some monotonic function $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, for which

$$\varphi \land \psi|\chi = F((\psi|\chi), (\varphi|\psi \land \chi)).$$

Consistency requires that $\varphi \land \psi|\chi$ and $\psi \land \varphi|\chi$ coincide, and thus

$$F((\psi|\chi), (\varphi|\psi \land \chi)) = F((\varphi|\chi), (\psi|\varphi \land \chi)).$$
As a matter of logic

\[ \models \left( (\phi \land \psi) \land \chi \right) \leftrightarrow (\phi \land (\psi \land \chi)) \, . \]

Hence consistency requires that

\[ (\phi \land \psi) \land \chi \mid \delta \ = \ \phi \land (\psi \land \chi) \mid \delta \, , \]

i.e., that

\[ F \left( F \left( F \left( (\chi \mid \delta), (\psi \mid \chi \land \delta) \right), (\phi \mid \psi \land \chi \land \delta) \right) \right) \]

\[ = \ F \left( (\chi \mid \delta), F \left( (\psi \mid \chi \land \delta), (\phi \mid \psi \land \chi \land \delta) \right) \right) \, . \]

Or more generally,

\[ F \left( F \left( x, y \right), z \right) \ = \ F \left( x, F \left( y, z \right) \right) \, . \]
Lemma (Cox 1946; Jaynes 1995)

Under certain further assumptions about $F$—e.g., that it is continuous—there is a unique set of solutions to these requirements. Namely there must be a function $w : \mathbb{R} \to \mathbb{R}$ such that

$$w(\varphi \land \psi | \chi) = w(\psi | \chi) w(\varphi | \psi \land \chi).$$
\[ w(\varphi \land \psi|\chi) = w(\psi|\chi)w(\varphi|\psi \land \chi) \quad (1) \]

Evidently, if \( \varphi \) is certain, then
\[ \varphi \land \psi|\chi = \psi|\chi \quad \text{and} \quad \varphi|\psi \land \chi = \varphi|\chi. \]

Hence, plugging into (1) we obtain \( w(\psi|\chi) = w(\varphi|\chi)w(\psi|\chi) \), i.e.,
\[ w(\varphi|\chi) = 1. \]

Likewise, if \( \varphi \) is impossible, then
\[ \varphi \land \psi|\chi = \varphi|\chi \quad \text{and} \quad \varphi|\psi \land \chi = \varphi|\chi. \]

Hence, again appealing to (1) we obtain \( w(\varphi|\psi) = w(\varphi|\psi)w(\psi|\chi) \), i.e.,
\[ w(\varphi|\chi) = 0. \]

That is, in any case, we always have \( 0 \leq w(\varphi|\chi) \leq 1. \)
Under basic assumptions, we have shown that there should always be weights $w$ for plausibilities $\varphi|\psi$ in between 0 and 1.

These weights should obey the following equation:

$$w(\varphi \land \psi|\chi) = w(\psi|\chi)w(\varphi|\psi \land \chi).$$

What about $w(\neg \varphi|\psi)$?
Jaynes argues that there should be some $S$ such that:

$$w(\neg \varphi | \psi) = S(w(\varphi | \psi)),$$

with $S$ continuous and monotone decreasing.

Furthermore, by consistency we must have

$$w(\varphi \land \psi | \chi) = w(\varphi | \chi)S(w(\neg \psi | \varphi \land \chi)) .$$

**Lemma (Cox 1946; Jaynes 1995)**

The only way of satisfying these requirements is if

$$w(\neg \varphi | \psi) = 1 - w(\varphi | \psi) .$$
Theorem (Cox 1946; Jaynes 1995)

The requirements discussed above ensure there is $P : \mathcal{L} \times \mathcal{L} \to [0, 1]$:

- $0 \leq P(\varphi | \psi) \leq 1$;
- $P(\neg \varphi | \psi) = 1 - P(\varphi | \psi)$;
- $P(\varphi \land \psi | \chi) = P(\psi | \chi)P(\varphi | \psi \land \chi)$.

As Jaynes says, we can now use logic and the rules above to assign plausibilities/probabilities to any complex sentence on the basis of those assigned to simpler sentences.
Example (Sum Rule)

\[
P(\varphi \lor \psi | \chi) = 1 - P(\neg \varphi \land \neg \psi | \chi) \\
= 1 - P(\neg \varphi | \chi)P(\neg \psi | \neg \varphi \land \chi) \\
= 1 - P(\neg \varphi | \chi)[1 - P(\psi | \neg \varphi \land \chi)] \\
= P(\varphi | \chi) + P(\neg \varphi \land \psi | \chi) \\
= P(\varphi | \chi) + P(\psi | \chi)P(\neg \varphi | \psi \land \chi) \\
= P(\varphi | \chi) + P(\psi | \chi)[1 - P(\varphi | \psi \land \chi)] \\
= P(\varphi | \chi) + P(\psi | \chi) - P(\psi | \chi)P(\varphi | \psi \land \chi) \\
= P(\varphi | \chi) + P(\psi | \chi) - P(\varphi \land \psi | \chi) .
\]
Example (Jewelry Thief)

Returning to the other example, recall the pattern we wanted to capture:

\[
\begin{align*}
\varphi \text{ being true makes } \psi \text{ more plausible} \\
\psi \text{ is true} \\
\hline
\varphi \text{ becomes more plausible}
\end{align*}
\]

This now comes out as a natural result of the calculus. For, suppose

\[
P(\psi|\varphi \land \chi) > P(\psi|\chi).
\]

Then, because it is now derivable that

\[
P(\varphi|\psi \land \chi) = P(\varphi|\chi) \frac{P(\psi|\varphi \land \chi)}{P(\psi|\chi)},
\]

it follows that

\[
P(\varphi|\psi \land \chi) > P(\varphi|\chi).
\]
Many have claimed that probability should be seen as a part of (extended) logic. We have seen three sources of such claims:

- Logical inference preserves not just truth, but also probability.
- Probability, like logic, is about consistency.
- Probability is the theory of reasonable inference, going beyond logic.

Next time we will look more closely at these ‘patterns of reasonable inference’, focusing on the relation between probabilistic notions and logical systems concerned with nonmonotonicity.