Propositional Logic and Probability Theory: Review

Vibhav Gogate

THE UNIVERSITY OF TEXAS AT DALLAS
Erik Jonsson School of Engineering and Computer Science

Logic

- Logics are formal languages for representing information such that conclusions can be drawn
- Syntax defines the sentences in the language
- Semantics define the "meaning" of sentences; i.e., define truth of a sentence in a world

Propositional logic: Syntax

Propositional logic is the simplest logic—illustrates basic ideas
The proposition symbols P_1, P_2 etc are sentences
If S is a sentence, $\neg S$ is a sentence
If S_1 and S_2 is a sentence, $S_1 \wedge S_2$ is a sentence
If S_1 and S_2 is a sentence, $S_1 \vee S_2$ is a sentence
If S_1 and S_2 is a sentence, $S_1 \Rightarrow S_2$ is a sentence
If S_1 and S_2 is a sentence, $S_1 \Leftrightarrow S_2$ is a sentence

Propositional Logic Example

Consider an alarm used for detecting burglaries. It may also be triggered by an earthquake.

- Burglary $\vee$ Earthquake
  is a propositional sentence where Burglary and Earthquake are called propositional variables and $\vee$ represents logical disjunction (or).
- Burglary $\vee$ Earthquake $\Rightarrow$ Alarm
  $\Rightarrow$ represents logical implication.
- $\neg$ Burglary $\wedge$ $\neg$ Earthquake $\Rightarrow$ $\neg$ Alarm

World

- A world (Truth assignment, a variable assignment, or a variable instantiation) is a particular state of affairs in which the value of each propositional variable is known.

world	Earthquake	Burglary	Alarm
w_1	true	true	true
w_2	true	true	false
w_3	true	false	true
w_4	true	false	false
w_5	false	true	true
w_6	false	true	false
w_7	false	false	true
w_8	false	false	false

Entailment and Models

$$KB \models \alpha$$

- A knowledge base KB is a conjunction of logical sentences.
- Knowledge base KB entails sentence α if and only if α is true in all worlds where KB is true
  E.g., the KB containing "the Giants won" and "the Reds won" entails "Either the Giants won or the Reds won"
- We say m is a model of a sentence α if α is true in m
  $M(\alpha)$ is the set of all models of α
  Then $KB \models \alpha$ if and only if $M(KB) \subseteq M(\alpha)$

Events

- Every sentence α can be viewed as representing a set of worlds $M(\alpha)$, which is called the event denoted by α.
- We will use sentence and event interchangeably.

Models and Set Theory

Using the definition of satisfaction $M(\alpha)$

- $M(\alpha \wedge \beta) = M(\alpha) \cap M(\beta)$
- $M(\alpha \vee \beta) = M(\alpha) \cup M(\beta)$
- $M(\neg\alpha) =??$

Models: Example

world	E	B	A
w_1	true	true	true
w_2	true	true	false
w_3	true	false	true
w_4	true	false	false
w_5	false	true	true
w_6	false	true	false
w_7	false	false	true
w_8	false	false	false

E: Earthquake,
B: Burglary
A: Alarm

M (Burglary) =???
M (Earthquake) =???
M (Burglary $\vee$ Earthquake) =???
M ($\neg$ Burglary $\wedge$ $\neg$ Earthquake) =???
M (Burglary $\vee$ Earthquake $\Rightarrow$ Alarm) =???

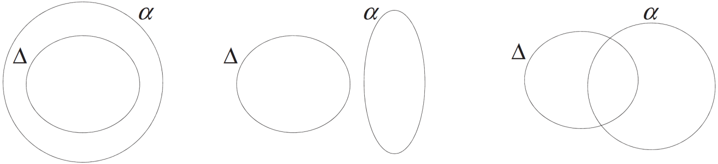
Validity and Satisfiability

- A sentence is valid if it is true in all models
  e.g., $A \vee \neg A$, $A \Rightarrow A$, $(A \wedge (A \Rightarrow B)) \Rightarrow B$
- A sentence is satisfiable if it is true in some model
  e.g., $A \vee B$, C
- A sentence is unsatisfiable if it is true in no models
  e.g., $A \wedge \neg A$

Fun facts:

- α is valid iff $\neg\alpha$ is unsatisfiable
- α is satisfiable iff $\neg\alpha$ is not valid.

Entailment as a Picture



(a) $\Delta \models \alpha$ (b) $\Delta \models \neg\alpha$ (c) $\Delta \not\models \alpha$ and $\Delta \not\models \neg\alpha$

Equivalence, Mutually Exclusive and Exhaustive Events or Logical sentences

- Sentences α and β are equivalent iff they are true at the same set of worlds: $M(\alpha) = M(\beta)$.
- Sentences α and β are mutually exclusive iff they are never true at the same world: $M(\alpha) \cap M(\beta) = \emptyset$.
- Sentences α and β are exhaustive iff each world satisfies at least one of the sentences: $M(\alpha) \cup M(\beta) = \Omega$ where Ω is the set of all possible worlds.

Propositional Logic: What definitions you should know?

- Syntax, What is a world? What is an event?
- Model of a logical formula
- What is a KB?
- Entailment
- Validity and Satisfiability (Consistency). Relationship between them.
- equivalence, mutually exclusive events, exhaustive events

Logical Equivalences

Schema	*Equivalent Schema*	*Name*
¬true	false	
¬false	true	
false $\wedge \beta$	false	
$\alpha \wedge$ true	α	
false $\vee \beta$	β	
$\alpha \vee$ true	true	
$\neg\neg\alpha$	α	double negation
$\neg(\alpha \wedge \beta)$	$\neg\alpha \vee \neg\beta$	de Morgan
$\neg(\alpha \vee \beta)$	$\neg\alpha \wedge \neg\beta$	de Morgan
$\alpha \vee (\beta \wedge \gamma)$	$(\alpha \vee \beta) \wedge (\alpha \vee \gamma)$	distribution
$\alpha \wedge (\beta \vee \gamma)$	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	distribution
$\alpha \implies \beta$	$\neg\beta \implies \neg\alpha$	contraposition
$\alpha \implies \beta$	$\neg\alpha \vee \beta$	definition of $\implies$
$\alpha \iff \beta$	$(\alpha \implies \beta) \wedge (\beta \implies \alpha)$	definition of $\iff$

Logical Deduction or Inference

- Deriving Conclusions from a KB
- Question: Prove that KB entails α, namely $KB \Rightarrow \alpha$ is valid
- Proving $KB \Rightarrow \alpha$ is valid is the same as proving that $KB \wedge \neg\alpha$ is unsatisfiable. Why?
- Fun Question
  - Let us say we find that KB entails α and we also find that KB entails γ. Suppose we add α to the KB, will this new KB also entail γ?

Monotonicity of Propositional Logic

- Let us say we find that KB entails α and we also find that KB entails γ. Suppose we add α to the KB, will this new KB also entail γ?
- Answer: Yes, Learning new facts do not invalidate previous conclusions. Propositional logic is monotonic.

- $KB \Rightarrow \gamma$. This means that $M(KB) \subseteq M(\gamma)$
- Since $M(KB \wedge \alpha) \subseteq M(KB)$, we must also have, $M(KB \wedge \alpha) \subseteq M(\gamma)$. Namely, $KB \wedge \alpha \Rightarrow \gamma$

Why Study Probabilities?

- Logic is brittle and therefore not enough.
  - KB must be consistent! Not always possible.
  - Example:
    - Strong leaders are typically not pacifist.
    - Democrats are typically pacifist.
    - President "X" is a democrat and a strong leader.
- The world is full of uncertainty and nondeterminism
- Computers need to be able to handle it
- Probability: New foundation for AI/ML
- Many other proposals in logic: non-monotonic logic.

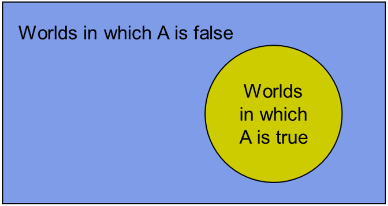
Elements of Probability Theory

- Events, Sample Space and Random Variables
- Axioms of Probability
- Conditional Probability
- Bayes Theorem
- Joint Probability Distribution
- Expectations and Variance
- Independence and Conditional Independence

Events, Sample Space and Random Variables

- A sample space is a set of possible outcomes in your domain.
  - All possible entries in a truth table.
  - Can be Infinite. Example: Set of Real numbers
- Random Variable is a function defined over the sample space S
  - A Boolean random variable X: $S \rightarrow \{True, False\}$
  - Stock price of Google G: $S \rightarrow$ Set of Reals
- An Event is a subset of S
  - A subset of S for which $X = True$.
  - Stock price of Google is between 575 and 580.

Events, Sample Space and Random Variables: Picture



$P(A)$ is the area of the oval

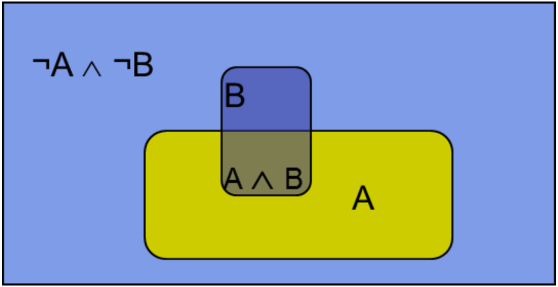
Sample Space: The Rectangle. Random variable: A. Event: A is *True*

Probability: A real function defined over the events in the sample space.

Axioms of Probability

Four Axioms of Probability:

- $0 \leq P(A) \leq 1$
- $P(True) = 1$ (i.e., an event in which all outcomes occur)
- $P(False) = 0$ (i.e., an event in no outcomes occur)
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Probability Distribution

- $A_1, \ldots, A_n$ is a set of mutually exclusive and exhaustive events

$$\sum_{i=1}^{n} P(A_i) = 1$$

- P is called a probability distribution.
- Example: P(Heads)=0.3, P(Tails)=0.7 (Univariate distribution)
- Example: Multi-variate distribution. Probabilities attached to each row in a truth table.

Sum Rule

- $0 \leq P(A) \leq 1$
- $P(True) = 1$ (i.e., an event in which all outcomes occur)
- $P(False) = 0$ (i.e., an event in no outcomes occur)
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

To prove that:

1. $P(A) = 1 - P(\neg A)$
2. $P(A) = P(A \wedge B) + P(A \wedge \neg B)$

SUM RULE:

$$P(A) = \sum_{i=1}^{n} P(A \wedge B_i)$$

where $\{B_1, \ldots, B_n\}$ is a set of of mutually exclusive and exhaustive events.

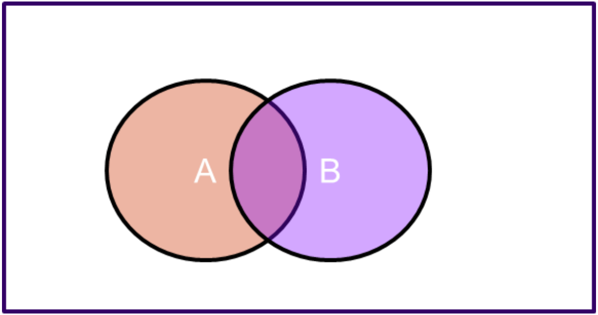
Using the SUM RULE

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

$$
\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= .2 \\
\Pr(\neg\text{Burglary}) &= .8 \\
\Pr(\text{Alarm}) &= .2442
\end{aligned}
$$

Conditional Probability

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Conditional Probability: Example

world	Earthquake	Burglary	Alarm	$\Pr(.)$	$\Pr(.\|\text{Alarm})$
ω_1	true	true	true	.0190	.0190/.2442
ω_2	true	true	false	.0010	0
ω_3	true	false	true	.0560	.0560/.2442
ω_4	true	false	false	.0240	0
ω_5	false	true	true	.1620	.1620/.2442
ω_6	false	true	false	.0180	0
ω_7	false	false	true	.0072	.0072/.2442
ω_8	false	false	false	.7128	0

Example

Our belief in Burglary increases:

$$\begin{aligned} \Pr(\text{Burglary}) &= .2 \\ \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \uparrow \end{aligned}$$

And so does our belief in Earthquake:

$$\begin{aligned} \Pr(\text{Earthquake}) &= .1 \\ \Pr(\text{Earthquake}|\text{Alarm}) &\approx .307 \uparrow \end{aligned}$$

Chain Rule

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

$$P(A \wedge B) = P(A|B)P(B)$$

$$P(A \wedge B \wedge C) = P(A|B \wedge C)P(B|C)P(C)$$

$$P(A_1 \wedge A_2 \wedge \ldots \wedge A_n) = \prod_{i=1}^{n} P(A_i|A_1 \wedge \ldots \wedge A_{i-1})$$

Independence and Conditional Independence

Independence:

- Two events are independent if $P(A \wedge B) = P(A)P(B)$
- Implies that: $P(A|B) = P(A)$ and $P(B|A) = P(B)$
- Knowing A tells me nothing about B and vice versa.
- A: Getting a 3 on the face of a die.
- B: New England Patriots win the Superbowl.

Conditional Independence:

- A and C are conditionally independent given B iff $P(A|B \wedge C) = P(A|B)$
- Knowing C tells us nothing about A given B.

Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Proof.

$P(A|B) = \frac{P(A \wedge B)}{P(B)}$ – (1)
$P(B|A) = \frac{P(A \wedge B)}{P(A)}$ – (2)
Therefore,
$P(A \wedge B) = P(B|A)P(A)$ – (3)
Substituting $P(A \wedge B)$ in Equation (1), we get Bayes Rule. □

Other Forms of Bayes Rule

Form 1:

$$P(A|B) = \frac{P(B|A)P(A)}{P(A \wedge B) + P(\neg A \wedge B)} \quad (1)$$

$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} \quad (2)$$

Form 2:

$$P(A|B \wedge C) = \frac{P(B|A \wedge C)P(A \wedge C)}{P(B \wedge C)}$$

Applying Bayes Rule: Example

- The probability that a person fails a lie detector test given that he/she is cheating on his/her partner is 0.98. The probability that a person fails the test given that he/she is not cheating on his/her partner is 0.05.
- You are a CS graduate student and the probability that a CS graduate student will cheat on his/her partner is 1 in 10000 (CS grads are boring!).
- A person will break up with his/her partner if the probability that the partner is cheating is greater than 0.005 (i.e., $> 0.5\%$).

Today, you come home and you find out that you have failed the lie detector test. Convince him/her that he/she should not break up with you.

Another Interpretation of the Bayes Rule

$$posterior = \frac{likelihood \times prior}{Probability\ \ of\ \ evidence}$$

$$P(Cheating = yes|Test = Fail) = \frac{P(Test = Fail|Cheating = yes) \times P(Cheating = yes)}{P(Test = Fail)}$$

- Prior probability of cheating
- Likelihood of failing the test given that a person is cheating
- Test=Fail is the evidence

Expectation and Variance

Expectation:

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

Conditional Expectation:

$$\mathbb{E}[f|y] = \sum_x p(x|y) f(x)$$

Variance:

$$\text{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

Joint Distribution

- Assign a probability value to joint assignments to random variables.
- If all variables are discrete, we consider Cartesian product of their sets of values For Boolean variables, we attach a value to each row of a truth table
- The sum of probabilities should sum to 1.

Outlook	Humidity	Tennis?	Value
Sunny	High	Yes	0.05
Sunny	High	No	0.2
Sunny	Normal	Yes	0.2
Sunny	Normal	No	0.1
Windy	High	Yes	0.2
Windy	High	No	0.05
Windy	Normal	Yes	0.05
Windy	Normal	No	0.15

The Joint Distribution

Represents complete knowledge about the domain
Can be used to answer any question that you might have about the domain

- $P(Event)$ = Sum of Probabilities where the Event is True
- $P(Outlook = Sunny) =$
- $P(Humidity = High \wedge Tennis? = No) =$
- $P(Humidity = High | Tennis? = No) =$

Outlook	Humidity	Tennis?	Value
Sunny	High	Yes	0.05
Sunny	High	No	0.2
Sunny	Normal	Yes	0.2
Sunny	Normal	No	0.1
Windy	High	Yes	0.2
Windy	High	No	0.05
Windy	Normal	Yes	0.05
Windy	Normal	No	0.15

Quiz

α and β are propositional variables

1. Given $\Pr(\alpha|\beta)$, do we have enough information to compute $\Pr(\alpha|\neg\beta)$?
2. Given $\Pr(\alpha|\beta)$ and $\Pr(\alpha|\neg\beta)$, do we have enough information to compute $\Pr(\alpha)$?
3. Given $\Pr(\alpha)$ and that α and β are independent, do we have enough information to compute $\Pr(\alpha|\beta)$?
4. Do we have enough information to compute the probability of $((\alpha \Rightarrow \beta) \wedge \alpha \wedge \neg\beta)$?