Propositional Logic Reminder

- **language:** determines the well-formed formulas (wffs)
  - *language defined inductively in terms of set of prop. symbols and connectives:* $\neg$, $\vee$, $\wedge$, $\supset$, $\equiv$, *. . . . E.g.,* $(p \vee \neg q) \supset (r \wedge s)$ *is a* **wff**
- **semantics:** defines when A **entails** B, written $A \models B$
  - *Related notions: when a world* **satisfies** *a formula* A*, written* $\models_w A$*, when* A *is a* **tautology***, etc. E.g.,* $\models \neg(p \wedge \neg p)$, $(p \supset q) \models (\neg q \supset \neg p)$, $\ldots$
- **proof-theory:** when A is **derivable** B, written $A \vdash B$
  - *Derivations defined in terms of* **axioms** *and* **rules of inference**
  - *E.g.,* **resolution** *yields* $C \vee C'$ *from* **clauses** $p \vee C$ *and* $\neg p \vee C'$
  - *A proof-theory is* **sound** *if* $A \vdash B$ *implies* $A \models B$*, and* **complete** *if* $A \models B$ *implies* $A \vdash B$

From Propositional Logic to Probabilities: Semantics

- Logic is an **all-or-nothing** affair; e.g., A is a tautology if A is true in all worlds
- How to say that A is true in, say, **most** of the worlds, or to take into account the **likelihood** of the different worlds?
- **Probabilities** allows us to do just that; weighting each world w with a number, $P(w)$, the probability of w being the 'actual' world.
- The probability of a formula A being true, is then simply:

$$P(A) = \sum_{w: \quad \models_w A} P(w)$$

- The probability weights $P(w)$ are **normalized** so that

$$0 \leq P(w) \leq 1 \qquad \text{and} \qquad \textstyle\sum_{w \in W} P(w) \ = \ 1$$

Some Properties

- $P(\text{tautology}) = 1$; $P(\text{contradiction}) = 0$
- $P(\neg A) = 1 - P(A)$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- $P(A) = P(A \wedge B) + P(A \wedge \neg B)$
- $P(A) = P(B)$ if $A \equiv B$
- $P(A \wedge B) = P(A)$ if A implies B
- . . .

Probability Calculus: Axioms

Like in logic, all valid properties of probability functions can be **derived** from suitable **axioms**. We need **three** axioms only:

[A1] $0 \leq P(A) \leq 1$ for any formula A

[A2] $P(A) = 1$ ($P(A) = 0$) if A is necessarily **true** (**false**)

[A3] $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

E.g., let's prove $P(\neg A) = 1 - P(A)$ from these axioms:

1. $P(A \vee \neg A) = 1$, from A2
2. $P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$, from A3
3. $P(A \wedge \neg A) = 0$, from A2
4. $1 = P(A) + P(\neg A)$, from 1--3
5. $P(\neg A) = 1 - P(A)$, from 4 $\square$

Language revised: Multivalued and Random Variables

- In propositional logic, the language is built from a set of **boolean variables** p, q, . . . , that can take **true** or **false** as values.
- In probability theory, the language is built from a set of **multivalued variables** X, Y, . . . that can take an arbitrary range of values
- Here we consider multivalued variables over **finite integer domains**; the atoms in the resulting language have the form $X = i$, $X > i$, etc.
- In probability theory, such basic formulas are called **events**, and the variables are called **random variables**
- The value of an unobserved random variable cannot be predicted with **certainty** but can be characterized by a **probability distribution** P_X; namely values $P(X = x)$ for each possible value x of X

Uniform Probabilities and Counting

If probabilities $P(w)$ are **uniform**; i.e., $P(w) = k$ for all w, then the probability $P(A)$ of a formula A can be obtained by **counting** as the **fraction of worlds satisfying** A

Example: rolling two unloaded dice (dados), with D_1 and D_2 standing for the outcomes

- $P(D_1 = 4)$ = ?
- $P(D_1 = 4 \land D_2 = 4)$
- $P(D_1 = 4 \lor D_2 = 4)$
- $P(D_1 > 4)$
- $P(D_1 = D_2)$
- $P(D_1 > D_2)$

What would be the probabilities if 4 twice as likely as other outcomes?

Conditional Probabilities: Definition

- The **conditional probability** $P(A|B)$, read ''probability of A **given** B'', **defined** as

$$P(A|B) \overset{\text{def}}{=} P(A \wedge B)/P(B) \quad \text{provided } P(B) \neq 0$$

E.g., $P(D_1 = 1 \mid 4 \leq D_1 \leq 6) = ?$; $P(D_1 = 6 \mid 4 \leq D_1 \leq 6) = ?$

- If the 'joint probability distribution' $P(w)$ is **uniform**, then $P(A|B)$ given by the fraction of worlds satisfying A **among the worlds satisfying** B

- $P(A|B)$ is the 'revised' probability on A once B has been **observed**

Marginalization and Conditionalization

- **Marginalization:** $P(X = x) = \sum_y P(X = x \wedge Y = y)$

  recall that in logic we would have $X = x \equiv \bigvee_y (X = x \wedge Y = y)$*, from where this principle easily follows through axioms A2 and A3*

- **Conditionalization:** $P(X = x) = \sum_y P(X = x|Y = y)P(Y = y)$

  this follows from above & definition of conditional probabilities

- Same ideas in presence of ''evidence'' E:

$$P(X = x|E) = \sum_y P(X = x \wedge Y = y|E) = \sum_y P(X = x|Y = y, E)P(Y = y|E)$$

Bayes Rule

- Bayes rule refers to principle that allows us to invert conditional probabilities:

$$P(H|E) = P(E|H)P(H) \ / \ P(E)$$

- It easily follows from the definition of conditional probabilities; yet it's very meaningful and useful, and has a long history

- E.g., in a diagnostic system, H is a possible disease, and E an observed symptom: $P(E|H)$ easier to assess than $P(H|E)$ and easier to combine (more about this when we study Bayesian Networks)

Example

Given an urn with 6 red balls and 4 white balls, and a color sensor with 90% accuracy, determine the probability of having extracted a red ball at random, given that the sensor reports ‘red’

Independence and Conditional Independence

- Two variables X and Y **independent** (a priori) iff

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y), \quad \text{for all } x \text{ and } y$$

- X and Y are **conditionally independent** given E iff

$$P(X = x \wedge Y = y|E) = P(X = x|E) \cdot P(Y = y|E), \quad \text{for all } x \text{ and } y$$

e.g., Vars D_1 and D_2 are **independent** a priori, but **not** conditionally independent given $E = D_1 + D_2 > 10$

- **Corollary:** if X and Y (conditionally) independent given E then

$$P(X = x|Y = y, E) = P(X = x|E)$$

Independence judgments are often simple to come by, and they are qualitative and psychological meaningful, unlike probability numbers

Example: Urns, Balls, Sensors, . . .

Consider two urns now: U_1 with 6 red balls and 4 white, and U_2 with 3 red and 7 white.

- Draw at random a ball B_1 from U_1
- Place extracted ball in U_2, and draw at random a ball B_2 from U_2
- What's is the probability of B_2 being Red given that the sensor reports Red, provided that the sensor has 90% color accuracy?
- What's the probability of B_1 being Red in that case?

More Examples

- You throw two (unloaded) dice
- If sum of the outcomes is greater than 10, then you enter a lottery
- In the lottery, with probability 0.1 you win a Madonna CD (!)

-- What's the probability of winning the prize?

-- What's the probability of having obtained at least one 6, given that you won the prize?