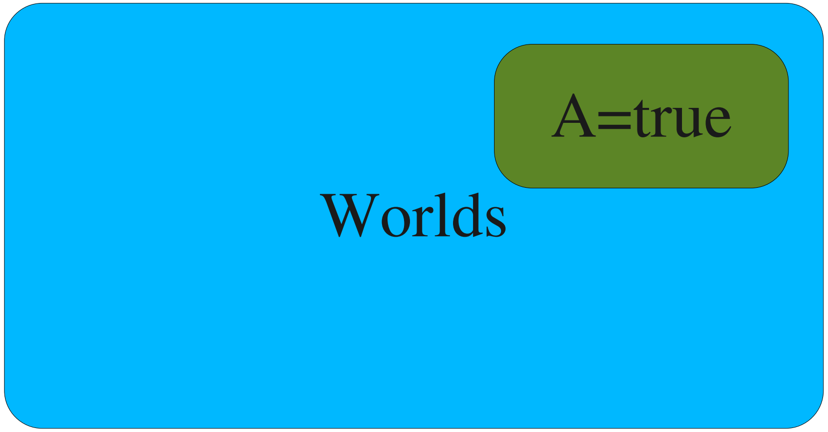
Probability

Discrete Random Variables

- We denote discrete random variables with capital letters.
- A boolean random variable may be either true or false
  - A = true or A=false.
- P(a), or P(A=true) denotes the probability that A is true.
- Also unconditional probability or prior probability.
- P(¬a) or P(A=false) denotes the probability that A is not true.

Discrete Random Variables

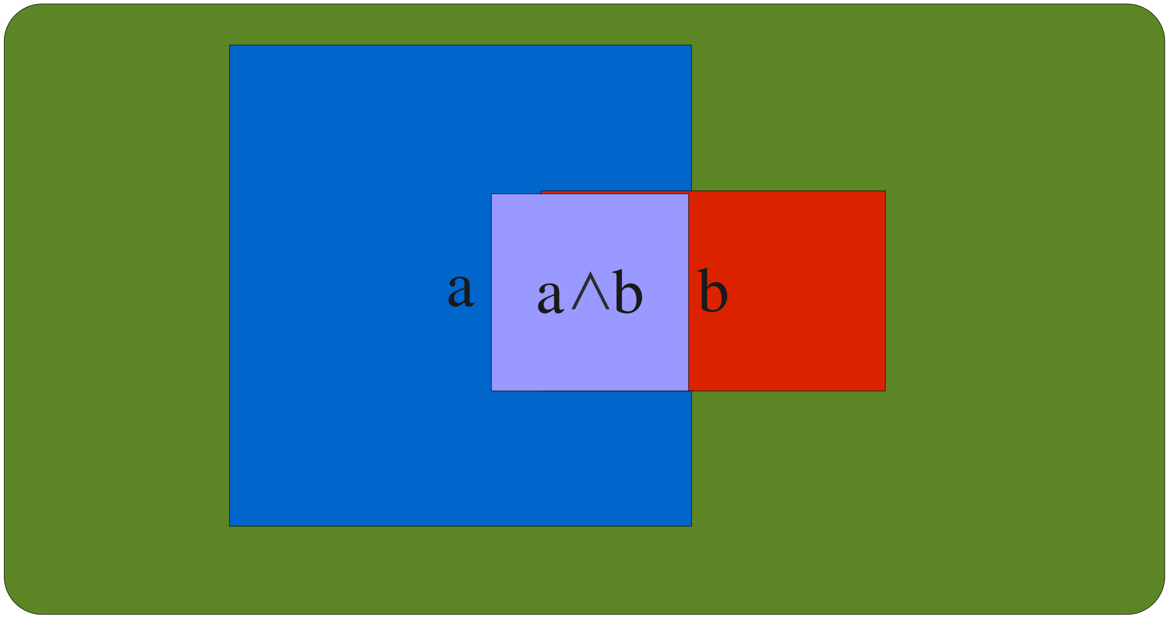
- P(a): the fraction of worlds in which A is true.



- P(a) = .2 P(¬a) = .8

More Notation

- We can apply boolean operators
  - Probability of a AND b: $P(a \wedge b)$ or $P(a,b)$
  - Probability of a OR b: $P(a \vee b)$
  - In general: $P(\alpha)$



The Axioms of Probability

- $0 \leq P(\alpha)$ for any proposition α .
- $P(\tau)=1$ if τ is a tautology.
- $P(\alpha \vee \beta)=P(\alpha)+P(\beta)$ if α and β are contradictory propositions

A Simple Proof

- $0 \le P(\alpha)$ for any proposition α .
- $P(\tau)=1$ if τ is a tautology.
- $P(\alpha \vee \beta)=P(\alpha)+P(\beta)$ if α and β are contradictory propositions

- Prove that $P(\neg\alpha)=1-P(\alpha)$

$$P(\neg\alpha \vee \alpha)=1 \quad \textit{(axiom 2)}$$

$$P(\neg\alpha)+P(\alpha)=1 \quad \textit{(axiom 3)}$$

$$P(\neg\alpha)=1-P(\alpha) \quad \textit{rearranging terms}$$

Multi-valued Random Variables

- We can define a random variable that can take on more than two possible values.
- E.g. C is one of $\{v_1, v_2, \ldots, v_N\}$
- Note, it must be the case that :

$$\sum_{1}^{N} P(v_i) = 1$$

- Example: W may have the domain {sunny, cloudy, rainy}.

Conditional Probability

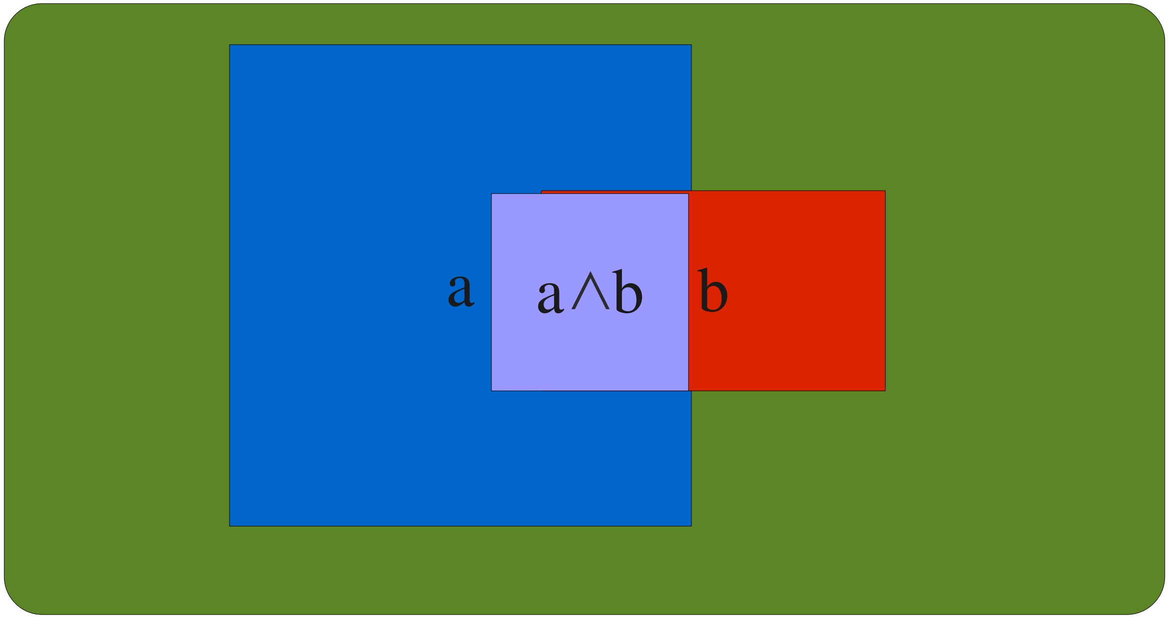
- P(a | b), the probability that A is true given that B is true.
  - P(sunny) = .1
  - P(sunny | warm) = .3
- Definition:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

- The fraction of worlds in which B is true, that also have A true.
- May also be written as the product rule:

$$P(a \wedge b) = P(a|b)\,P(b)$$

Conditional Probability



- P(a) = .3
- P(b) = .1
- P(a ∧ b) = .05
- P(a | b) = .5

Probability Distributions

- A probability distribution is a complete description of the probability of all possible assignments to a random variable.
- Examples:
- For a boolean variable
  - P(A=TRUE) = .1
  - P(A=FALSE) = .9
- Random variable W from the domain {sunny, cloudy, rainy}
  - **P**(W) = <.2 , .7, .1>

Joint Probability Distribution

- A complete description of the probability of all possible assignments to all variables (atomic event).

Two boolean variables A and B

A	B	Prob
T	T	.1
T	F	.2
F	T	.5
F	F	.2

Rooster Crows (C) and Weather (W)

C	W	Prob
T	sunny	.05
T	cloudy	.2
T	rainy	0
F	sunny	.05
F	cloudy	.4
F	rainy	.3

Inference

- Determining the probability of an event of interest, given everything that we know about the world.
- This is easy if we have the joint probability distribution.
- *The probability of a proposition is equal to the sum of the probabilities of worlds in which it holds.*

$$P(\alpha) = \sum_{\omega \in \{\omega : \omega \models \alpha\}} P(\omega)$$

Inference Example

A	B	Prob
T	T	.1
T	F	.2
F	T	.5
F	F	.2

- What is P(A = true)?
- P(a) = ??

Inference Example

A	B	Prob
T	T	.1
T	F	.2
F	T	.5
F	F	.2

- What is P(A = true)?
- P(a) = .1 + .2 = .3
- In general $P(Y)=\sum_{z} P(Y, z)$ marginalization.
- Here Y and Z may be sets of variables, and the sum is over all possible assignments to the variables Z.

Conditional Inference

C	W	Prob
T	sunny	.05
T	cloudy	.2
T	rainy	0
F	sunny	.05
F	cloudy	.4
F	rainy	.3

- P(C=true | W =sunny)?
- Remember that: $$P(a|b)=\frac{P(a \wedge b)}{P(b)}$$
- P(C=true | W = sunny) = ??

Conditional Inference

C	W	Prob
T	sunny	.05
T	cloudy	.2
T	rainy	0
F	sunny	.05
F	cloudy	.4
F	rainy	.3

- P(C=true | W =sunny)?
- Remember that: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$
- P(C=true | W = sunny) = .05 / (.05 + .05) = .5

"Learning" a Joint Probability Distribution

- Where does the joint probability distribution come from?
- Maybe we (or an expert) make it up.
- Or we can learn it:

$$\hat{P}(row) = \frac{\#\ instances\ that\ match\ row}{\#\ total\ instances}$$

C	W	#days	Prob
T	sunny	12	12/38 =.32
T	cloudy	3	3/38 =.08
T	rainy	0	0/38 =.0
F	sunny	8	8/38 =.21
F	cloudy	10	10/38 =.26
F	rainy	5	5/38 =.13

total: 38

ANY PROBLEMS?

Problems with Learning PD

- This will quickly break down if we have more than a few variables.

Independence

- Variables A and B are independent if *P(a | b) = P(a)*
- We can also write: $P(a \wedge b) = P(a)P(b)$
  - Remember the product rule: $P(a \wedge b) = P(a|b)P(b)$
- Independence is a big deal for probabilistic reasoning.
  - Specifying the full joint PD requires exponential storage.
  - Learning it requires an exponentially growing amount of data.
  - These both become linear if all variables are independent.
  - This is called factoring the joint distribution.

Bayes' Rule

- The most useful identity in AI:

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

- Think of *h* as hypothesis and *e* as evidence.
- $P(e|h)$ is called likelihood.

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

- Why would we know P(*e* | *h*) and not P(*h* | *e*)?

Diagnosis

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

- I have a cough, I want to know the probability that I have pneumonia.
- P(cough) = .1, P(pneumonia) = .001,
  P(cough | pneumonia) = .5
- P(pneumonia | cough) = (.5 * .001) / .1 = .005 = .5%

P(e)?

- That P(e) term seems a little odd. How do we know the prior probability of the evidence?
- We don't need to:
  - Maybe we don't care (if only want to know which hypothesis is most likely).
  - Otherwise:

$$P(e) = \sum_{h \in H} P(e|h)\,P(h)$$