Reasoning under Uncertainty

Extrapolate Shift from Classical Planning
to Reinforcement Learning

- Classical Planning
  - Logic as reasoning engine
  - Model is purely analytic
  - Deep (first-order expressiveness)
  - Brittle (uncertainty is unavoidable and catastrophic)
- Reinforcement Learning
  - Decision Theory (statistics) as reasoning engine
  - Model is largely empirical
  - Shallow (propositional expressiveness)
  - Robust (model is fit / adapted to the real world)
- Apply this paradigm shift to reasoning more generally

Reasoning under Uncertainty

- Qualification Problem
- Laziness
- Theoretical Ignorance
- Practical Ignorance
- Computational Issues

It is sometimes better *not* to model the world accurately (consider a coffee cup…)

Our models are always "approximate"

"All models are wrong, but some are useful"

- George Box, famous statistician

Ontology / Semantics of Uncertainty

- Fuzzy logic – the world is imprecise
  - "John is tall"
  - Linguistic variables
  - Characteristic functions / fuzzy sets
- Frequentist statistics – the distribution is real (world)
  - Classical statistics; probability = long run average
  - Data helps us approximate it
- Bayesian statistics – the data is real
  - Subjective probabilities
  - Evidence for different distributions

Religious wars among Statisticians
(are you a Bayesian?)

- Frequentist / objectivist / classical statistics / Fisherian
  - World is some true distribution
  - Data are a random sampling
  - We can come to know the World approximately via data
  - Hypothesize; Observe; Evaluate
  - Changing the hypothesis taints the data (baseball, lottery)
  - Stock scam, wrong hypothesis
- Bayesian
  - Evidence can be objective (data) or subjective
  - Evidence can testify for / against different distributions
  - Will my plane crash? Chance of rain?
  - Two meter problem
  - Two envelope problem

Probability / Statistics

- Probability Space
  - Sample space (Atomic Events)
  - Event space
  - Probability measure

- Random Variables
- Distributions
- Statistical Inference

Probabilities

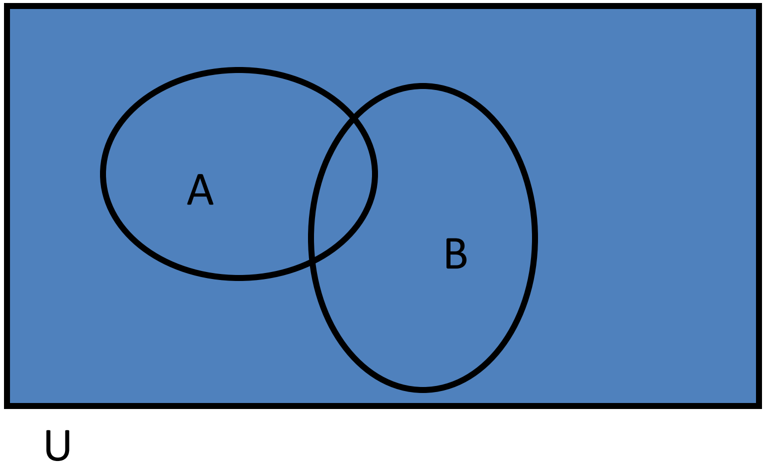
- Random variables – think features w/ prob. values
  - Discrete (Boolean, multi-valued, countably infinite)
  - Continuous
  - (Technically neither a variable nor random…)
- Events
  - Atomic events
    - "Sample space"
    - Exclusive & Exhaustive
    - Somewhat analogous to possible worlds of logic
  - Complex events are sets of atomic events
- Underlying population (all possible sequences of coin flips)
- Sample (a few observed sequences of coin flips)
- Joint probability distribution
- Inference – conditioning, marginalizing, parameter estimation…

Prior or Unconditional Probability

- Long-run average
- For Bayesian also subjective likelihood

$$P(A) = \frac{\#A \text{ outcomes}}{\text{total } \# \text{ outcomes}}$$

in the underlying population



$0 \leq P(A) \leq 1$

$P(True) = 1; \quad P(False) = 0$

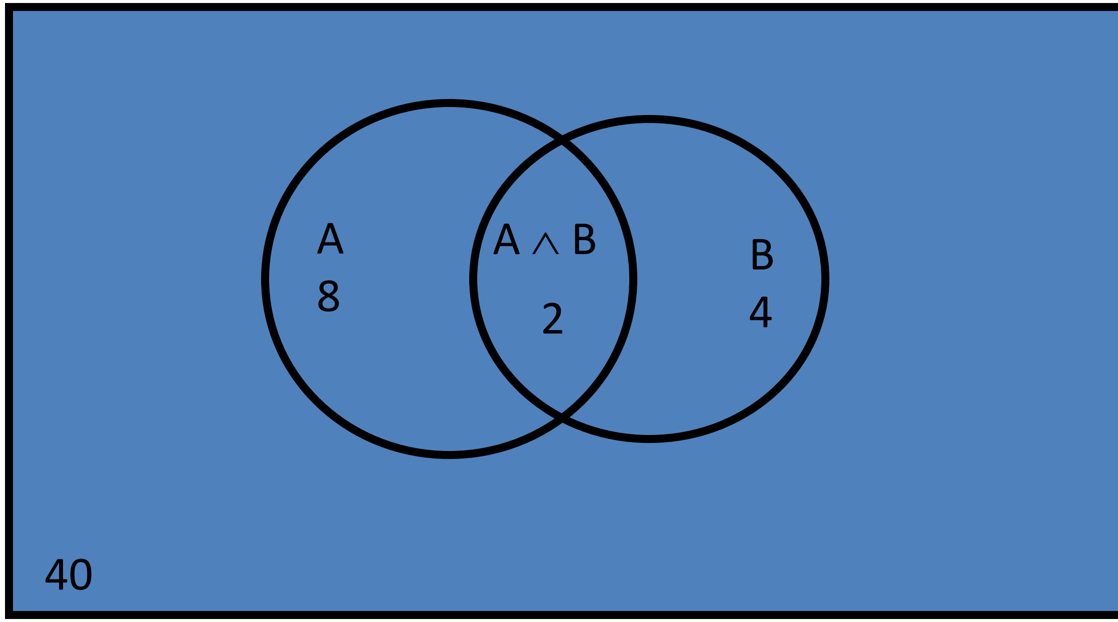
$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Posterior or Conditional Probabilities

- You don’t know Fred
- What’s the probability that Fred is taller than 6’?
- In fact Fred’s mother is 6’3” (> 6’)
- Now what’s the probability that Fred is taller than 6’?
- $P(A|B)$ – the universe becomes outcomes where B holds

A: Person > 6’

B: Parent > 6’



$P(A) = \#A / Total$

$= 8 / 40$ or 0.2

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

$= 2 / 4$ or 0.5

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

provided $P(B) > 0$

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

Note the difference between probabilities and sample–based estimates of probabilities

Joint Probability Distribution

		Weather			
		Sunny	Cloudy	Rainy	Snowy
Have Fun?	Yes	0.25	0.15	0.05	0.13
	No	0.05	0.1	0.25	0.02

Probabilities for each Atomic Event

Marginal Probabilities: P(Sunny), P(Fun), etc.

(Written in the margins)

Conditional Probabilities: P(Fun|Sunny)

How many degrees of freedom in the joint?

		Weather			
		Sunny	Cloudy	Rainy	Snowy
Have Fun?	Yes	0.25	0.15	0.05	0.13
	No	0.05	0.1	0.25	0.02

P(Rainy)

= 0.05 + 0.25

= 0.3

P(Rainy | $\neg$Sunny)

$$= \frac{P(\text{Rainy} \wedge \neg\text{Sunny})}{P(\neg\text{Sunny})}$$

= 0.3 / 0.7

$\approx$0.43

P(Fun | Sunny)

$$= \frac{P(\text{Fun} \wedge \text{Sunny})}{P(\text{Sunny})}$$

= 0.25 / 0.3

$\approx$0.83

Do I prefer Sun or Snow?

P(Fun | Snowy)

= 0.13 / 0.15 $\approx$0.87 So I prefer snow

Joint Probability Distribution

- Discrete random variables
- Encodes all information of interest
- Allows arbitrary dependencies
- Exhaustive and Exclusive Atomic Events
- Must sum to 1
  (one less degrees of freedom than entries)
- Diseases / Symptoms illustration
  (useful but overly specific; really evidence and conclusions)

Bayes Rule

(NB: Frequentists also believe in Bayes Rule)

- Symptoms: headache, stiff neck, fever, …
- Diseases: head cold, poor posture, meningitis, …
- $P(d_i \mid S)$: Find people with symptom combinations, determine if diseased
- But some (meningitis) are rare
- $P(S \mid d_i)$ is often easier to estimate
  go to hospital, ask sufferers of d_i about S

Bayes Theorem

- $P(A \mid B) = \dfrac{P(B \mid A)\, P(A)}{P(B)}$

- Easy to rederive (you should *never* get it wrong!)
- $P(A \wedge B) = P(A \mid B)\, P(B)$

  $= P(B \mid A)\, P(A)$
- Equate and solve for $P(A \mid B)$

How to Diagnose

- Estimate the Joint
- Observe patient (= set of symptoms)
- Compute P(d_i | observed symptoms)
  for each d_i
- Marginalize over the unobserved symptoms
- Condition on the observed symptoms
- Choose the most likely d_i

How many numbers (parameters) to estimate from data?

- Number of diseases: n
- Number of symptoms: q
- Each symptom takes on s values

$$ns^q - 1$$

Indepenences are Redundancies in the Joint

- Random variables A and B are Independent iff P(A | B) = P(A)
- B provides no information about A

Suppose

Weather distinctions: m

Fun levels: n

Times of the day: k

How many numbers

If random variables interact?

$(m \cdot n \cdot k) - 1$

If they are independent?

$(m-1) + (n-1) + (k-1)$