Course Logisics

- We need more TAs
- Or less work for Dae Hoon

  Grading is the time-consuming portion…

- Machine Problem?

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

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)

Weather

Have Fun?		Sunny	Cloudy	Rainy	Snowy
	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 | ¬Sunny)

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

= 0.3 / 0.7

≈0.43

P(Fun | Sunny)

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

= 0.25 / 0.3

≈0.83

Do I prefer Sun or Snow?

P(Fun | Snowy)

= 0.13 / 0.15 ≈0.87 So I prefer snow

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 many numbers (parameters) to estimate from data?

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

$$nv^q - 1$$

We need a number for each distinction in each dimension (less 1 since it is a distribution and sums to 1.0)

How many dimensions for this problem?

How many distinctions in each dimension?

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

$$nv^q - 1$$

- How many dimensions for this problem?
  - One for each random variable
  - How many RV's?
  - One for the disease, one for each symptom
- How many distinctions in each dimension?
  - n for the disease
  - v for each of the q symptoms
- In total: n * v * v * v…(q times)… minus 1

How do we know the number or Random Variables?

- Identify the atomic events
- What is an atomic event for our problem?
- A person
- How many “random” attributes does a single person have?
- a disease and q symptoms

How do we know the number of parameters in the joint?

- One less than…
- The possible *combinations* of distinctions!
- Each RV makes some distinctions (takes on some number of different values)
- Combinations: $\prod_{rv \in RVs}$ distinctions(rv)
- One factor for each RV; factor is # distinctions
- So: $n * v_1 * v_2 * v_3 * \ldots * v_{1q} - 1$ Or: $nv^q - 1$
- The joint allows for arbitrary (all possible) interactions

Indepenences are Redundancies in the Joint

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

Suppose

Weather distinctions: m

Fun levels: n

Times of the day: k

How many numbers

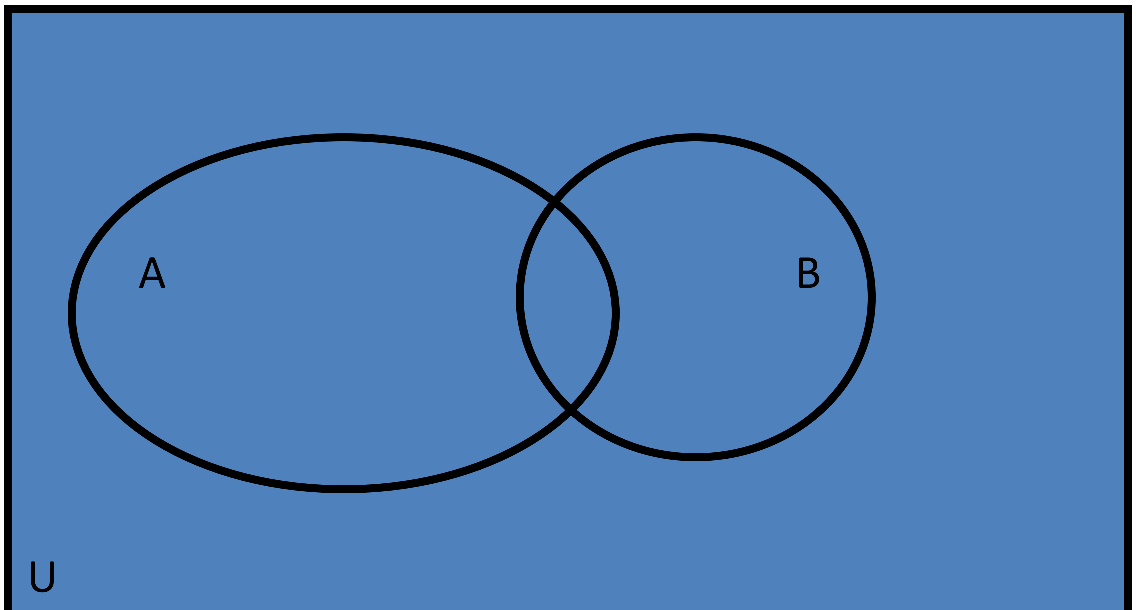
If random variables interact?

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

If they are independent?

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

Events A and B are independent



Indepenences are Redundancies in the Joint

- Random variables A and B are Independent
  iff $P(A \mid B) = P(A)$
- 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?

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

Implications for AI Representations

- Joint contains just as many numbers
- Some are functions of others
- AI knowledge representations should make *all* and *only* the distinctions necessary for effective performance
- Fewer numbers
  - more efficient processing
  - faster learning
  - fewer inconsistencies
  - smaller data structure…

Diseases and Symptoms

- Are they independent?
- We hope not!
- What could be independent?
- Where do the symptoms come from?

Conditional Independence

- Model the disease influencing the symptoms
- But no symptom interactions *given* the disease
- Conditional independence:

$$P(A|B,C) = P(A|C)$$

"A is conditionally independent of B given C"

More Conditional Independence

Suppose A is conditionally independent of B given C

Is B necessarily conditionally independent of A given C?

$$P(A|B,C)=P(A|C) \Rightarrow ?$$
$$P(B|A,C)=P(B|C)$$

$$P(A|B,C) = P(A|C)$$

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

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

$$P(B|A,C) = P(B|C)$$

YES, they are equivalent

More Conditional Independence

A is conditionally independent of B given C

C can still depend on both A and B

$P(C|A) \neq P(C)$ and $P(C|B) \neq P(C)$

A and B are *not* necessarily independent:

$P(B|A) \neq P(B)$ and $P(A|B) \neq P(A)$

But, Given C,
A and B do not influence each other

Knowledge of C separates A and B

Suppose we do NOT know C,

then A and B MAY influence each other

(think about this one until it's intuitive
– what does "influence" mean here?)

Suppose that Symptoms are Conditionally Independent given the Disease

- We know John has a cold
- Congestion, a sore throat, headache, rash are more likely but not necessary

  P(congestion | cold) >> P(congestion)   likewise sore throat

- We discover he does in fact have a sore throat
- Is he now more or less likely to also have congestion?
- Not much:

  P(congestion | sorethroat, cold) ~ P(congestion | cold)

- We may choose not to model this and other weak interactions

Bayesian Networks
(belief networks,…)

- Graphical model
  - Probability + Graph theory
- Nodes – discrete random variables
- Directed Arcs – direct influence
  - Usually “causality” (class, text,…)
- Configuration of parents establishes contexts for a node
  - Different distribution for each context
  - Conditional Probability Tables (CPTs)
- No directed cycles (DAG)
- General graphical models
  - Markov Networks
  - Conditional random fields
  - Dynamic Bayesian nets
  - others

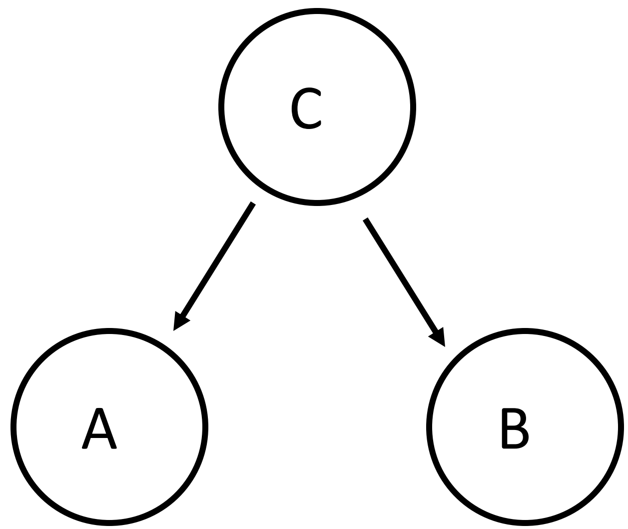
Dentist Example

3 Boolean Random Variables:

C – Patient has a cavity

A – Patient reports a toothache

B – Dentist's probe catches on tooth



C directly influences A and B

A and B are conditionally independent given C

CPT at each node specifies a probability distribution for each context (configuration of parents)

How many numbers do we need for the full joint?

How many for the Bayesian net?