Announcements

• Midterm exam March 1, one week from tomorrow
• Topics:
  – Models
  – Search & Optimization
  – Logic, Planning, Knowledge Representation
  – Reinforcement Learning
  – Statistics
• Practice homework HW4 is out
• Solutions for HW3 will be posted soon
The Two Envelope Problem

• Some found this confusing
• x $ in one envelope, 2x $ in another
• You choose one (no information, Pr=0.5 for each); call your amount A (A is either x or 2x but we don’t know which)
• Do you want to keep it or trade?
  – Compute expected utility of each action
  – Keep: expect A
  – Trade:
    • 0.5 expect 2A
    • 0.5 expect A/2
    • In all expect 1.25 A
  – Trading is best
• STRANGE!
• Now would you like to trade again?
• Yes (by the same reasoning)
• Even Stranger!!
Weak Law of Large Numbers

Let $\bar{c}_n$ be the average of $n$ samples from “reasonable” random variable $C$ (well defined mean, bounded variance)

$$\lim_{n \to \infty} Pr(|\bar{c}_n - E[C]| > \varepsilon) = 0$$

• Compare a sample average to the true mean
• It is increasingly unlikely to find differences larger than any positive number $\varepsilon$ as $n$ grows without bound
• Later we will see similar expressions (probability concentration bounds) for finite samples.
Strong Law of Large Numbers

Let $\overline{c}_n$ be the average of $n$ samples from “reasonable” random variable $C$ (well defined mean, bounded variance)

$$Pr\left(\lim_{n \to \infty} \overline{c}_n = E[C]\right) = 1$$

- In the limit as the number of samples grows without bound, each sample average almost surely converges to the true mean
- Both laws hold
- The strong law implies the weak but not vice versa (Why?)
Statistical Independence

• Random variables A and B are Independent iff  \( \Pr(A, B) = \Pr(A) \Pr(B) \) or  \( \Pr(A \mid B) = \Pr(A) \)

• B provides no information about A

• A and B are independent
¡iClickers!

A and B are independent; which best captures this?

A) $\text{U} \quad \text{A} \quad \text{B}$

B) $\text{U} \quad \text{A} \quad \text{B}$

C) $\text{U} \quad \text{A} \quad \text{B}$

D) $\text{U} \quad \text{B} \quad \text{A}$
Statistical Independence

• Random variables $A$ and $B$ are Independent iff $\Pr(A,B) = \Pr(A) \Pr(B)$ or $\Pr(A \mid B) = \Pr(A)$

• $B$ provides no information about $A$

• $A$ and $B$ are independent

A

U

B
Events A and B are independent

- Random variables A and B are Independent iff $\Pr(A \mid B) = \Pr(A)$
- B provides no information about A $\Pr(A \mid B) = \frac{\Pr(A \land B)}{\Pr(B)}$
Statistical Independence

- Random variables A and B are Independent iff $\Pr(A \mid B) = \Pr(A)$
- B provides no information about A
- This is a very special relationship
- Independence can be known for certain only analytically (the probabilities are known for certain)
- Thus independence cannot be known for certain empirically from a sample of any size
- ...but we can guess that the influence is small empirically
Independences are Redundancies in the Joint

• Random variables A and B are Independent
  \( \Pr(A \mid B) = \Pr(A) \)

Suppose

Weather distinctions: \( m \)

Fun levels: \( n \)

Times of the day: \( k \)

How many degrees of freedom

If random variables interact? If they are independent?

\[ (m \cdot n \cdot k) - 1 \]

\[ (m-1) + (n-1) + (k-1) \]
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...
Statistical Inference

• Diagnosing diseases from symptoms
• Are they independent?
• We hope not!
  (no diagnoses; no statistical inference)
• What could be independent?
• Where do the symptoms come from?
Conditional Independence

• Perhaps model the disease influencing the symptoms
• One option: symptoms do not interact *given* the disease
• The disease probabilistically results in a set of symptoms
• But the symptoms do not interact with each other
• Conditioned on the disease, the symptoms are independent
Conditional Independence

• A somewhat more useful property

• A is conditionally independent of B given C

\[ \Pr(A|B,C) = \Pr(A|C) \]

• Knowledge of B yields no additional information beyond C wrt A

• Wrts A, If we know C, B does not help
iClickers!

If A is conditionally independent of B given C:

$$\Pr(A|B,C) = \Pr(A|C)$$

Is B conditionally independent of A given C?

A. Yes

B. No

C. Impossible to say
Suppose $A$ is conditionally independent of $B$ given $C$. Is $B$ necessarily conditionally independent of $A$ given $C$?

$$\Pr(A|B,C) = \Pr(A|C) \implies ?$$

$$\Pr(B|A,C) = \Pr(B|C)$$

$$\Pr(A,B,C) = \Pr(A|C) \Pr(B,C) = \Pr(A,C) \Pr(B|C)$$

$$\Pr(B|A,C) = \Pr(B|C)$$

YES, they are equivalent.
More Conditional Independence

A is conditionally independent of B given C
C can still depend on both A and B
  \( \Pr(C|A) \neq \Pr(C) \) and \( \Pr(C|B) \neq \Pr(C) \)

A and B are *not* necessarily independent:
  \( \Pr(B|A) \neq \Pr(B) \) and \( \Pr(A|B) \neq \Pr(A) \)

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

Knowledge of C separates the A and B influences

Suppose we do NOT know C,
  then A and B MAY influence each other
  they are not independent

(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 all more likely but not necessary
  \[
  \Pr(\text{congestion} \mid \text{cold}) \gg \Pr(\text{congestion}) \]
  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:
  \[
  \Pr(\text{congestion} \mid \text{sore throat, cold}) \sim \Pr(\text{congestion} \mid \text{cold})
  \]
- It is the cold and not so much the sore throat that is responsible for the congestion
- We may choose not to model this and other weak interactions
Some Distributions

- Normal (aka Gaussian)
- Binomial
- Beta
- Multinomial / Dirichlet
- Moments (for well-behaved distributions)
  - first moment = mean $\mu$
  - second central moment = variance $\sigma^2$ ("dispersion about the mean")
    sometimes more convenient: standard deviation $\sigma$ (units)
  - skew, kurtosis
- Discrete probability mass function
- Continuous probability density function
  (integrate over a region to get a probability)
- Support (domain where nonzero)
Normal Distribution

$X \sim N(\mu, \sigma^2)$

Continuous
Location $\mu$
Shape $\sigma^2$
Support ($-\infty, +\infty$)

$P(x; \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$
Binomial Distribution

\[ X \sim \text{Bin}(p, n) \]

\[
Pr(x; p, n) = \binom{n}{x} (p)^x (1-p)^{(n-x)} \quad \text{for } x = 0, 1, 2, \cdots, n
\]

recall

\[
\binom{n}{x} = \frac{n!}{x!(n-x)!}
\]

Discrete
Mean \( np \)
Variance \( np(1-p) \)
Support \([0, n]\)
Beta Distribution

- N flips of a coin with weight p
- Of the N, h are heads and t are tails
- So N = h + t
- Given p and N what are the distributions of h and t?
- Then \( h \sim \text{Binomial}(p, N) \)
  \( t \sim \text{Binomial}(1-p, N) \)
- h, t, and p are related
- Given h and t, how is p distributed? (recall N = h + t)
- Then \( p \sim \text{Beta}(h+1, t+1) \)
- \( p \sim \text{Beta}(a, b) \) (where \( a=h+1 \) and \( b=t+1 \))
- \( a, b > 0 \)