Uncertainty

Chapter 13

Read Ch 13!

- This is core material for the rest of this course
- Understanding Bayes Nets and DBNs requires intuitive understanding of some basic probability concepts

Uncertainty

- Until now....propositions are T, F, or unknown
- Real environments are not so certain
  - partially observable
  - noisy sensors
  - unexpected events in dynamic environments

Probability Interpretations

Probability as frequency: *k* out of *n* possibilities

- Suppose we're drawing cards from a standard deck:
  - P(card is the Jack ♥ | standard deck) = 1/52
  - P(card is a ♣ | standard deck) = 13/52 = 1/4
- What's the probability of a drawing a pair in 5-card poker?
  - P(hand contains pair | standard deck) =

$$\frac{\text{\# of hands with pairs}}{\text{total \# of hands}}$$

  - Counting can be tricky (take a course in combinatorics)
  - Other ways to solve the problem?
- General probability of *event* given some conditions:

  P(event | conditions)

Probabilities

Probability of uncountable events

- How do we calculate probability that it will rain tomorrow?
  - Look at historical trends?
  - Assume it generalizes?

- What's the probability that there was life on Mars?
- What was the probability the sea level will rise 1 meter within the century?
- What's the probability that candidate X will win the election?

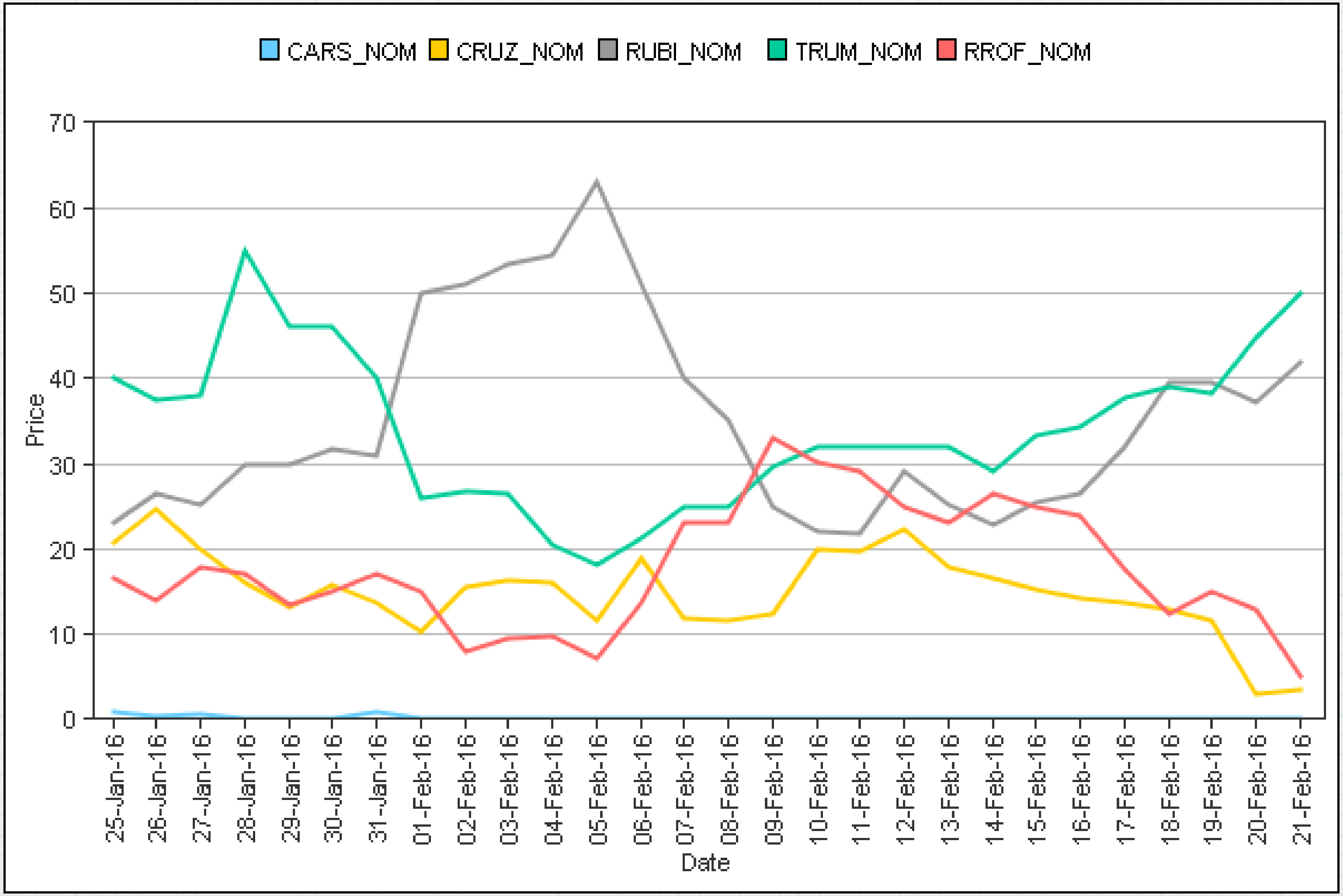
The Iowa Electronic Markets: placing probabilities on single events

- http://www.biz.uiowa.edu/iem/
- "The Iowa Electronic Markets are real-money futures markets in which contract payoffs depend on economic and political events such as elections."
- Typical bet: predict vote share of candidate X - "a vote share market"

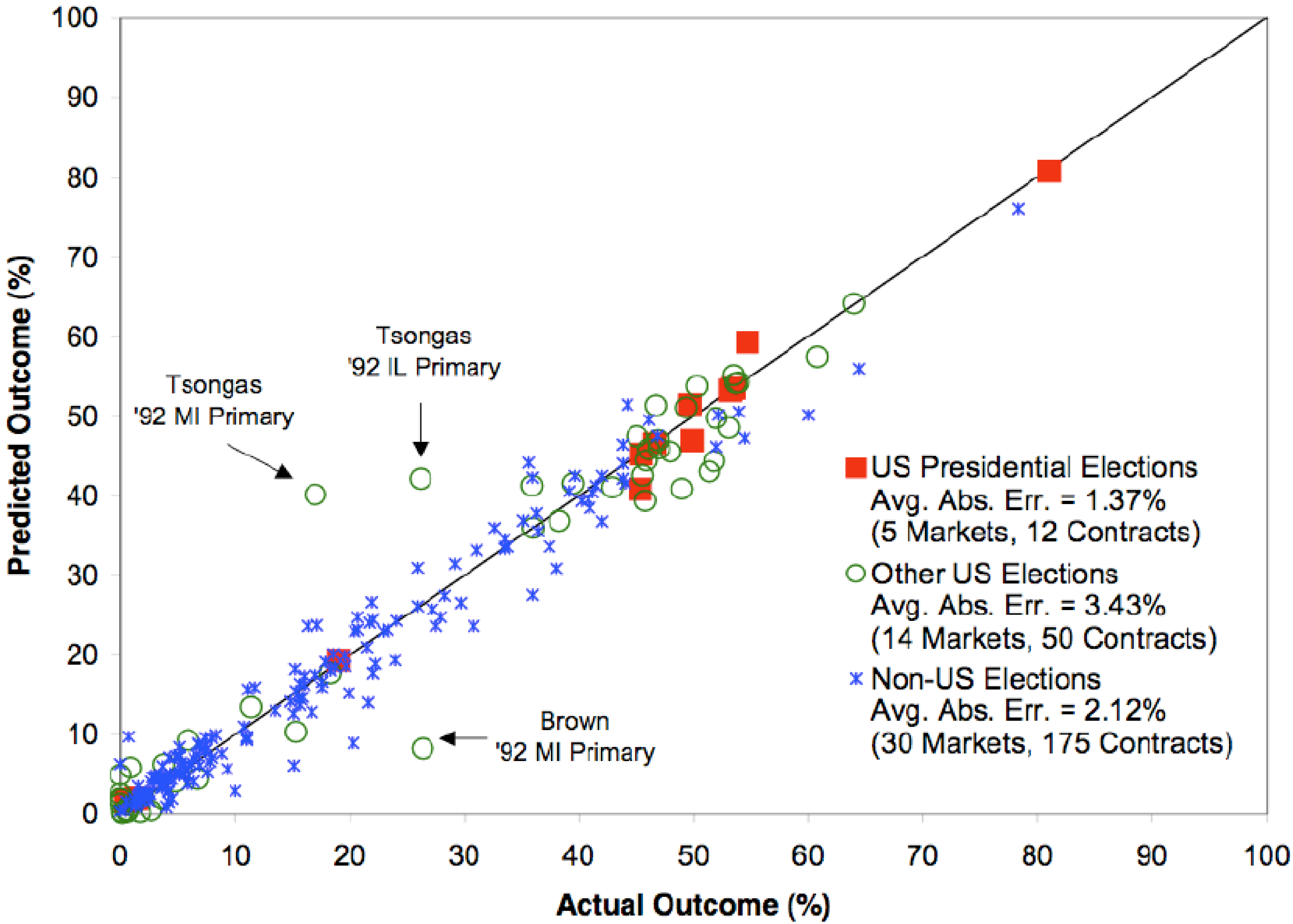
De Finetti's definition of probability

- Was there life on Mars?
- You promise to pay \$1 if there is, and \$0 if there is not.
- Suppose NASA will give us the answer tomorrow.
- Suppose you have an oppenent
  - You set the odds (or the "subjective probability") of the outcome
  - But your oppenent decides which side of the bet will be yours
- de Finetti showed that the price you set has to obey the axioms of probability or you face certain loss, i.e. you'll lose every time.

Probability and Wagers



Political futures market predicted vs actual outcomes



Probability

* Instead of absolute statements, use probability to summarize uncertainty
* Probabilities relate to the degree that an agent believes a statement to be true $P(A_{25}|\text{no reported accidents}) = 0.06$
* The probability changes with new information (evidence)

$$P(A_{25}|\text{no reported accidents, 5 a.m.}) = 0.15$$

Axioms of probability

- Axioms (Kolmogorov):

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

  $P(\text{true}) = 1$

  $P(\text{false}) = 0$

  $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

- Corollaries:
  - A single random variable must sum to 1:

  $$\sum_{i=1}^{n} P(D = d_i) = 1$$

  - The joint probability of a set of variables must also sum to 1.
  - If A and B are mutually exclusive:

  $$P(A \text{ or } B) = P(A) + P(B)$$

Rules of probability

- conditional probability

$$Pr(A|B) = \frac{Pr(A \,\text{and}\, B)}{Pr(B)}, \qquad Pr(B) > 0$$

- corollary (Bayes' rule)

$$\begin{aligned} Pr(B|A)Pr(A) &= Pr(A \,\text{and}\, B) = Pr(A|B)Pr(B) \\ \Rightarrow Pr(B|A) &= \frac{Pr(A|B)Pr(B)}{Pr(A)} \end{aligned}$$

Probability Basics

Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

Probability Basics

Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

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

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

Probability Basics

Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

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

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

An event A is any subset of Ω

$$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$$

E.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Propositions

Think of a proposition as the event (set of sample points) where the proposition is true

Propositions

Think of a proposition as the event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(\omega) = true$

event $\neg a$ = set of sample points where $A(\omega) = false$

event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

Propositions

Think of a proposition as the event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(\omega) = true$
event $\neg a$ = set of sample points where $A(\omega) = false$
event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

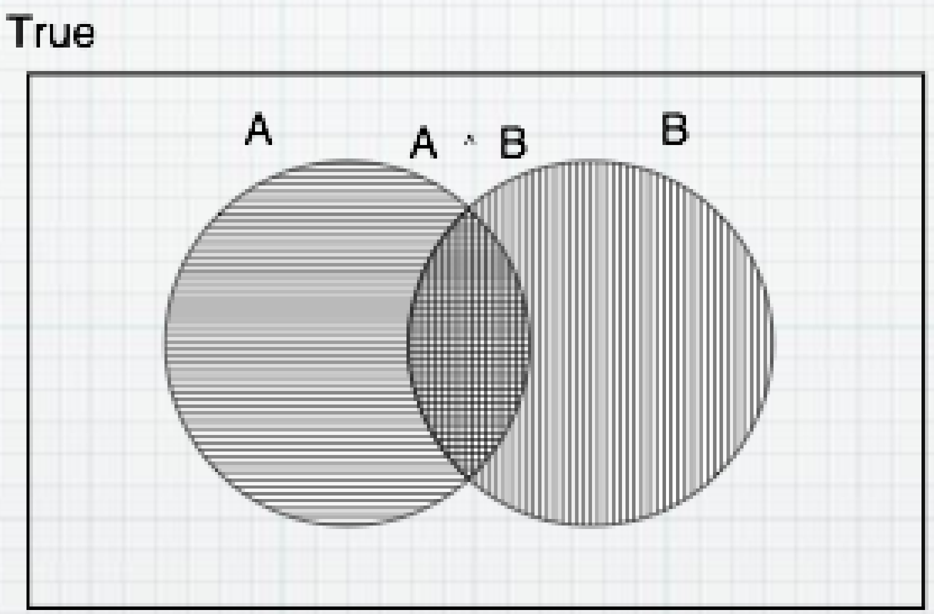
Often in AI applications, the sample points are **defined** by the values of a set of random variables, i.e., the sample space is the Cartesian product of the ranges of the variables

With Boolean variables, sample point = propositional logic model
e.g., $A = true$, $B = false$, or $a \wedge \neg b$.
Proposition = disjunction of atomic events in which it is true
e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Why use Probability?

The definitions imply that certain logically related events must have related probabilities

E.g., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Syntax for Propositions

Propositional or Boolean random variables

e.g., $Cavity$ (do I have a cavity?)

$Cavity = true$ is a proposition, also written $cavity$

Discrete random variables (finite or infinite)

e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow\rangle$

$Weather = rain$ is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions

Probability Model

* Tuple of $\{\Omega, \mathcal{F}, P\}$
* Used to define random variable X
* And probability density function p(X)

Prior Probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Prior Probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Prior Probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) =$ 0.2 and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

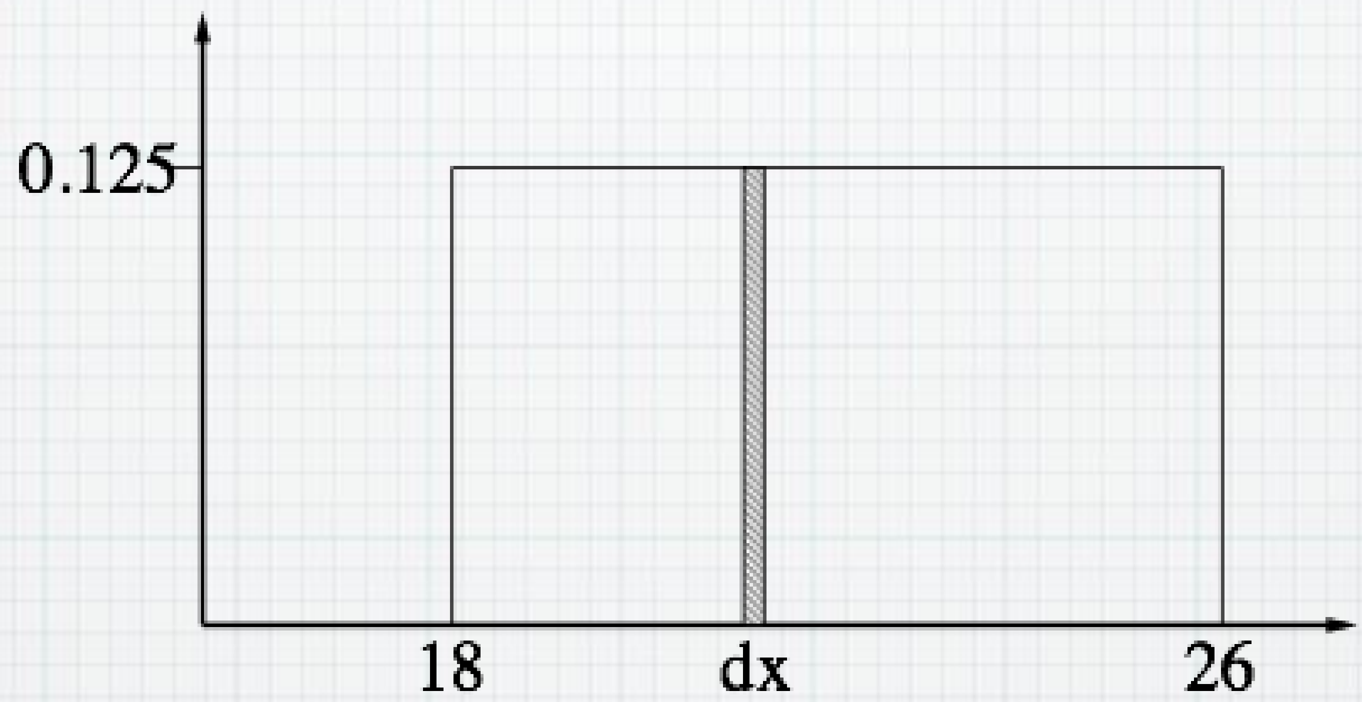
$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Continuous Variables

Express distribution as a parameterized function of value:

$P(X = x) = U[18, 26](x)$ = uniform density between 18 and 26

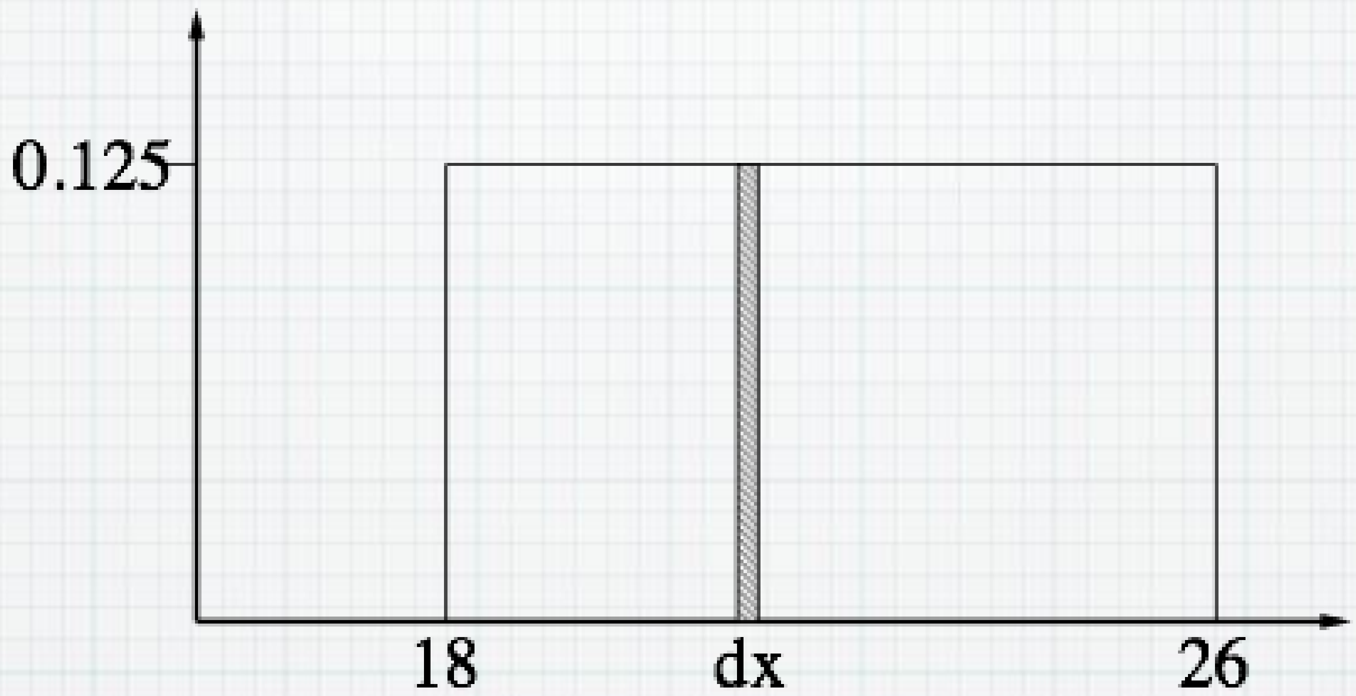


Here P is a density; integrates to 1.

$P(X = 20.5) = 0.125$ really means

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)/dx = 0.125$$

Continuous Variables

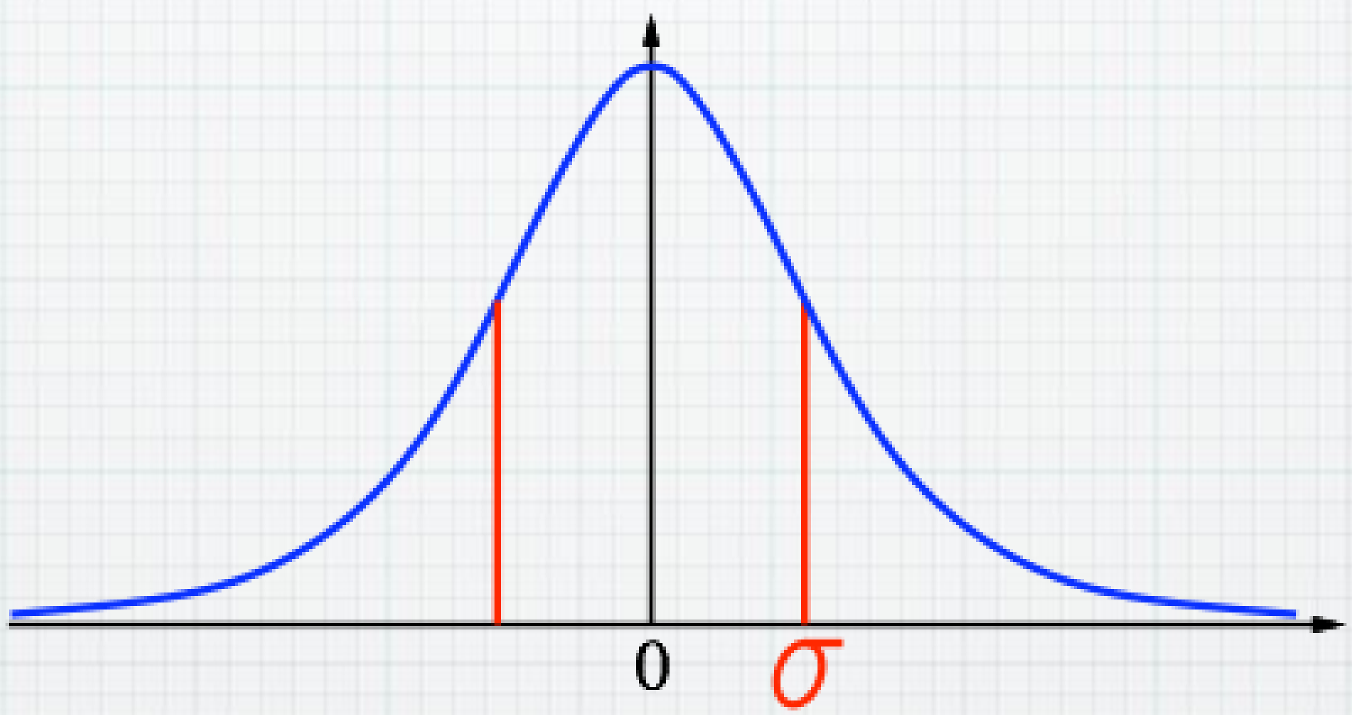


Density integrates to 1, in this case (26-18)*0.125 = 1

Now consider U[1.0,1.2]: We require (1.2-1.0)*x = 1
So x=5. Note that p(x)>1, different from discrete case

Continuous Variables

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



Mean μ controls where samples cluster on x axis
Variance σ^2 controls the spread, how tightly points are clustered. Smaller $\sigma \Rightarrow$ tighter clustering

Conditional Probability

Conditional or posterior probabilities

e.g., $P(cavity|toothache) = 0.8$

i.e., **given that** $toothache$ **is all I know**

NOT "if $toothache$ then 80% chance of $cavity$"

(Notation for conditional distributions:

$\mathbf{P}(Cavity|Toothache)$ = 2-element vector of 2-element vectors)

Conditional Probability

Conditional or posterior probabilities

e.g., $P(cavity|toothache) = 0.8$

i.e., **given that** $toothache$ **is all I know**

NOT "if $toothache$ then 80% chance of $cavity$"

(Notation for conditional distributions:

$\mathbf{P}(Cavity|Toothache)$ = 2-element vector of 2-element vectors)

If we know more, e.g., $cavity$ is also given, then we have

$P(cavity|toothache, cavity) = 1$

Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**

Conditional Probability

Conditional or posterior probabilities
  e.g., $P(cavity|toothache) = 0.8$
  i.e., **given that** *toothache* **is all I know**
  NOT "if *toothache* then 80% chance of *cavity*"

(Notation for conditional distributions:
  $\mathbf{P}(Cavity|Toothache) =$ 2-element vector of 2-element vectors)

If we know more, e.g., *cavity* is also given, then we have
  $P(cavity|toothache, cavity) = 1$
Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**

New evidence may be irrelevant, allowing simplification, e.g.,
  $P(cavity|toothache, 49ersWin) = P(cavity|toothache) = 0.8$
This kind of inference, sanctioned by domain knowledge, is crucial

Summary

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every atomic event

Queries can be answered by summing over atomic events

For nontrivial domains, we must find a way to reduce the joint size

Independence and conditional independence provide the tools