Introduction to
Artificial Intelligence

Lecture 10 – Probability

CS/CNS/EE 154
Andreas Krause

Announcements

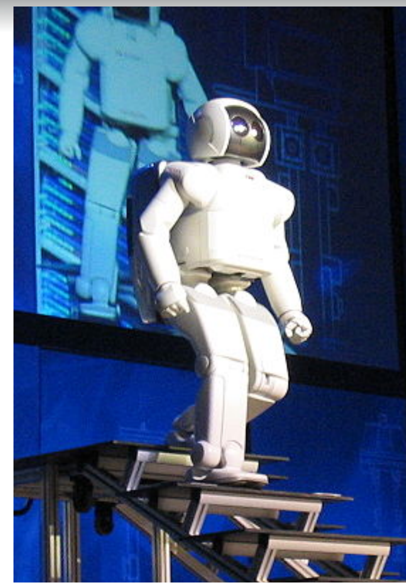
- Milestone due Nov 3. Please submit code to TAs
- Grading: PacMan
  - Compiles?
  - Correct? (Will clear maze within 5 minutes)
  - Reasonable performance?
    (Time to clear maze < 1.5*Time taken by baseline)
  - Beat baseline? (Eat more dots than baseline)
  - Top 4 teams
  - Winner

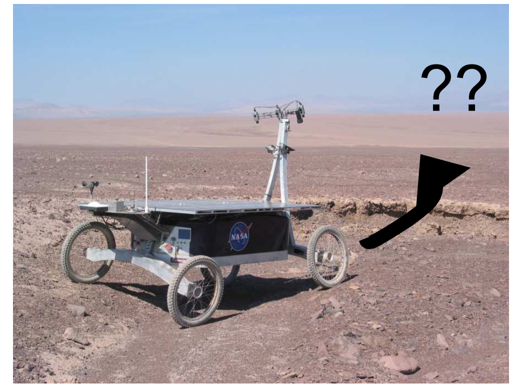
Announcements

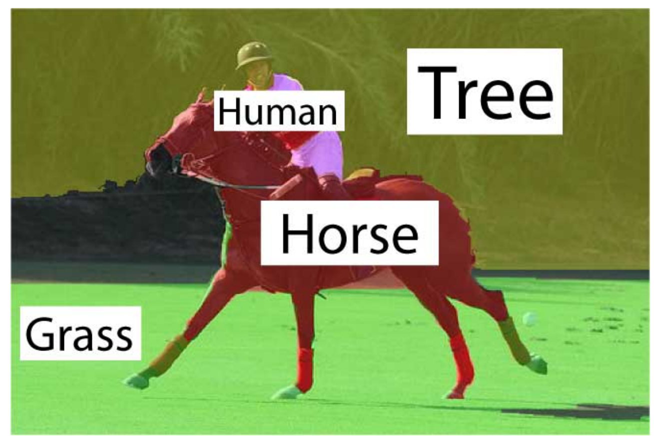
- Grading: AvianAsker. Goal: Minimize #questions/guesses
  - Compiles?
  - Correct? (Will correctly identify bird within 300 questions)
  - Reasonable performance? (#Questions/guesses < 1.5 Baseline)
  - Beat baseline? (<= Questions than baseline)
  - Top 4 teams
  - Winner

- Of course: No cheating (e.g., can't just submit baseline)

Probabilistic AI





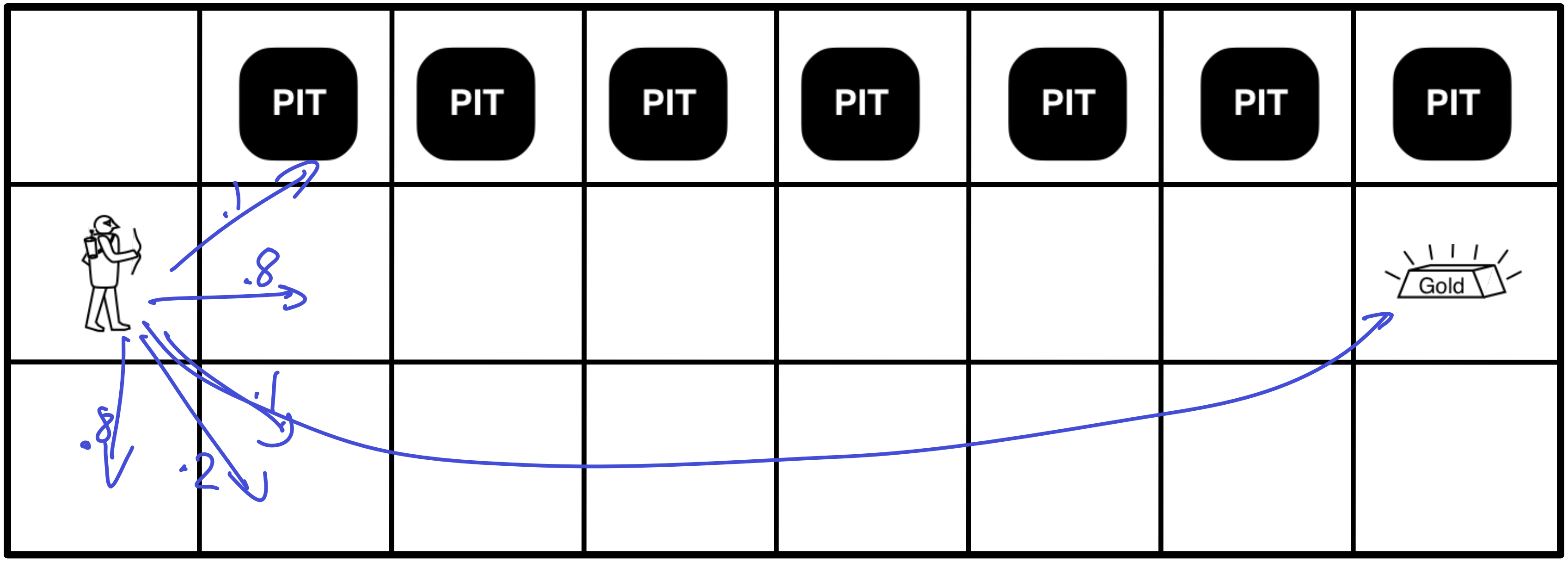


Quantifying Uncertainty

- So far, we have mainly focused on *deterministic* environments
- Often, actions can have *uncertain outcomes*
- Sensor observations are *noisy*

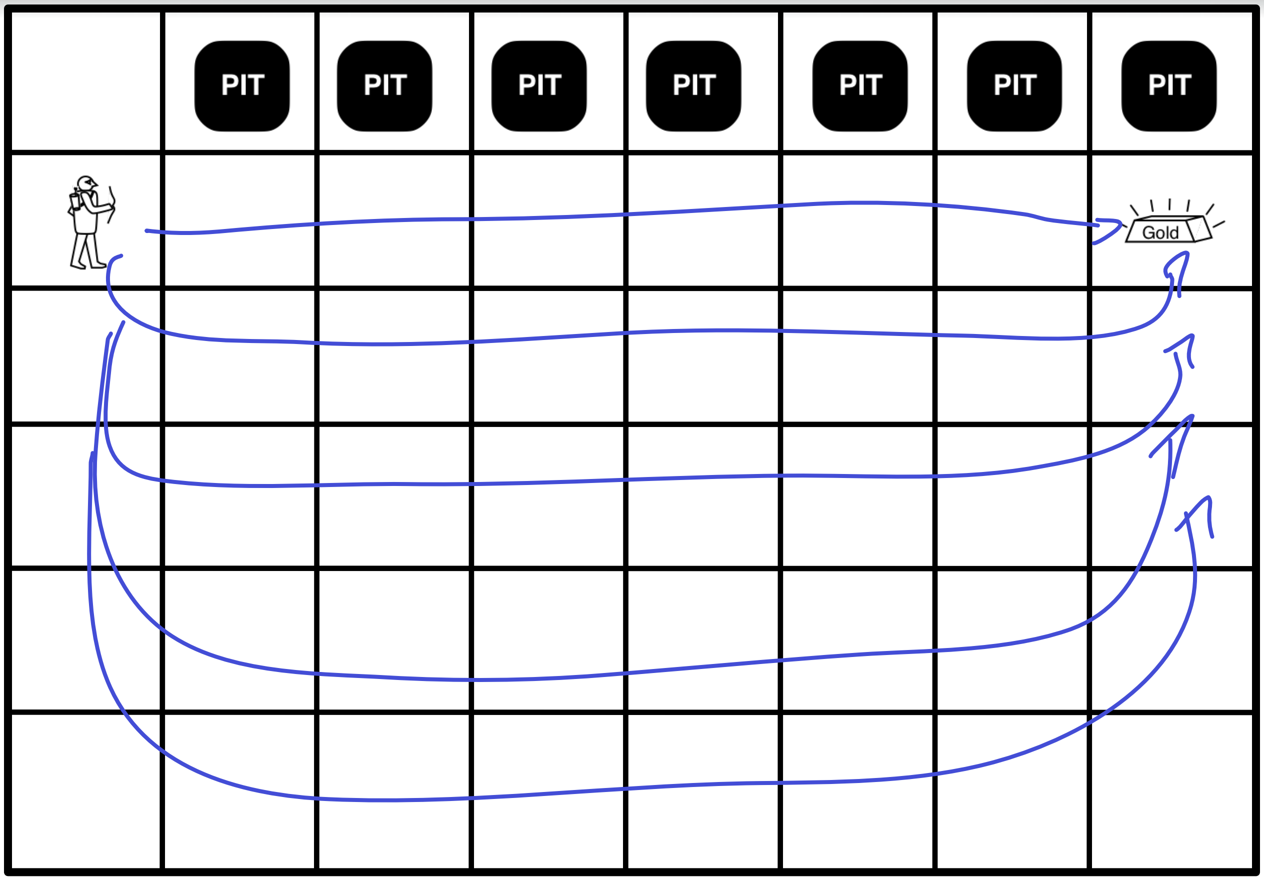
- One approach: Nondeterministic actions / observations
  - Not specified which outcome is more likely
  - Purely *qualitative* model of uncertainty

Problems with Nondeterminism



- Motion model: sometimes, actual direction is off by 45 degrees of intended direction
- Nondeterministic planning finds no feasible solution
- Suppose, error occurs with at most 20% chance.. What should we do?

Decision making under uncertainty



Cost (Pit) = −1000

Cost (1 step) = −1

Path	P(Gold)	length
1	.01	7
2	.3	9
3	.9	11
⋮	.99	13
	.999	10000

Choose path that minimises expected cost

- Which path should we choose?

Review: Probability

- Describe probability of events
  - P(Pit at [2,2])
  - P(Wumpus dead)
  - P(A v B)
- Formally: Probability Space $(\Omega, \mathcal{F}, P)$
  - Set of "atomic events": Ω Eg. $\Omega = \{1, 2, \ldots, 6\}$
  - Set of all non-atomic events: $\mathcal{F} \subseteq 2^\Omega$

  $\mathcal{F}$ is a σ-Algebra (closed under complements and countable unions)

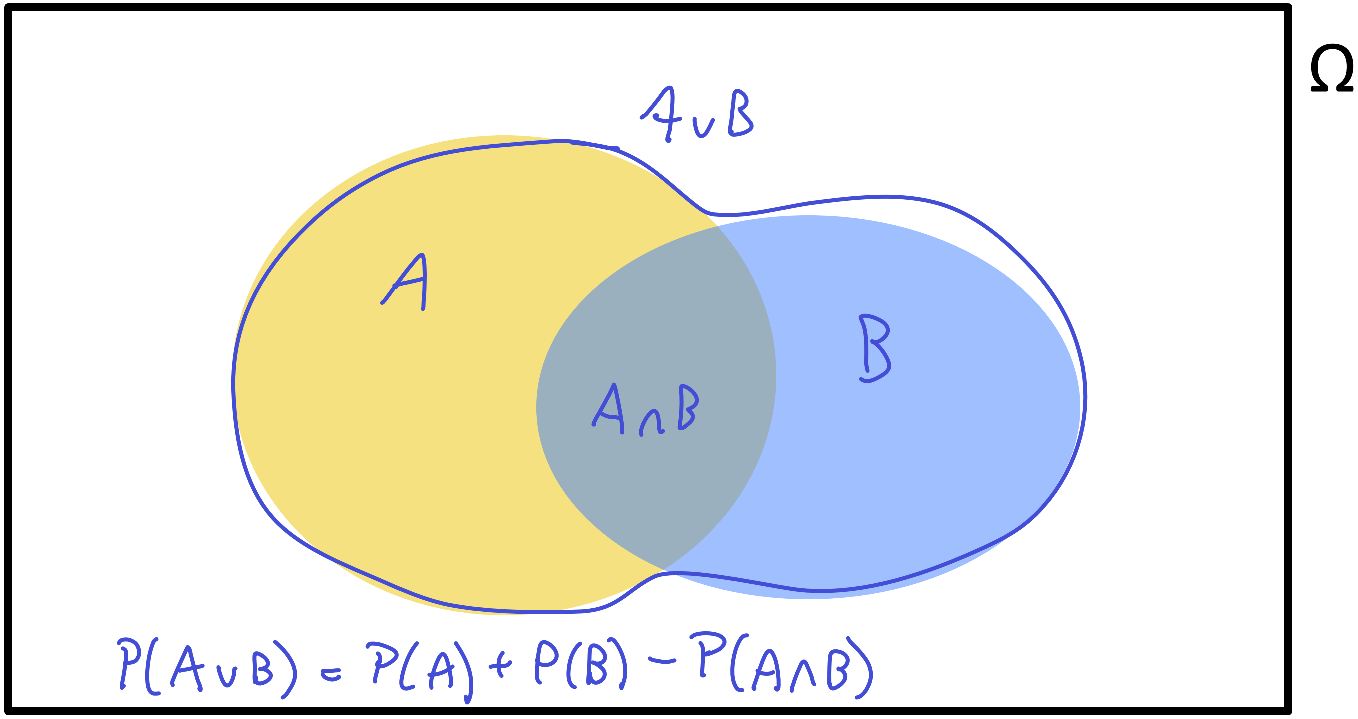
  even $= \{2, 4, 6\}$, odd $= \{1, 3, 5\}$
  - Probability measure $P : \mathcal{F} \to [0, 1]$

    For $\omega \in \mathcal{F}$, $P(\omega)$ is the probability that event ω happens

    $P(\{1\}) = P(\{2\}) = \ldots P(\{6\}) = \frac{1}{6}$; $P(\text{even}) = \frac{1}{2}$

Why use probabilities?

Related events must have related probabilities



Agents that bet according to beliefs that violate probability axioms can be forced to lose in expectation

(de Finetti 1931)

Independent events

- Two random events *A, B* are independent iff

$$P(A \cap B) = P(A) \cdot P(B)$$

Eg. A = Dice 1 comes up 3 $\quad P(A) = P(B) = \frac{1}{6}$
B = Dice 2 " 2

$P($Dice 1 comes up 3, Dice 2 comes up 2$) = P(A)P(B) = \frac{1}{36}$

- Events $A_1, A_2, \ldots, A_n$ are independent iff

For all subsets $A_{i_1}, \ldots A_{i_k}$ it holds that

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$$

Interpretation of probabilities

- Philosophical debate..
- Frequentist interpretation
  - P(ω) is relative frequency of ω in repeated experiments
  - Often difficult to assess with limited data
- Bayesian interpretation
  - P(ω) is "degree of belief" that ω will occur
  - Where does this belief come from?
  - Many different flavors (subjective, objective, pragmatic, …)
- For now assume probabilities are known

Random variables

- Events are cumbersome to work with.

- Let D be some set (e.g., the integers)
- A random variable X is a mapping $X : \Omega \rightarrow D$
- For some $x \in D$, we say

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

"probability that variable X assumes state x"

$\Omega = \{1, \ldots, 6\}$, $\mathrm{Even}(\omega) = \begin{cases} 1 & \text{if } \omega \in \{2,4,6\} \\ 0 & \text{if } \omega \in \{1,3,5\} \end{cases}$

$P(\mathrm{Even} = 1) = P(\{2,4,6\})$

Examples

- Bernoulli distribution: "(biased) coin flips"

  D = {H,T}

  Specify P(X = H) = p. Then P(X = T) = 1-p.

  Note: can identify atomic events ω with {X=H}, {X=T}
- Binomial distribution counts the number of heads S

  $p(S=k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Categorical distribution: "(biased) m-sided dice"

  D = {1,…,m}

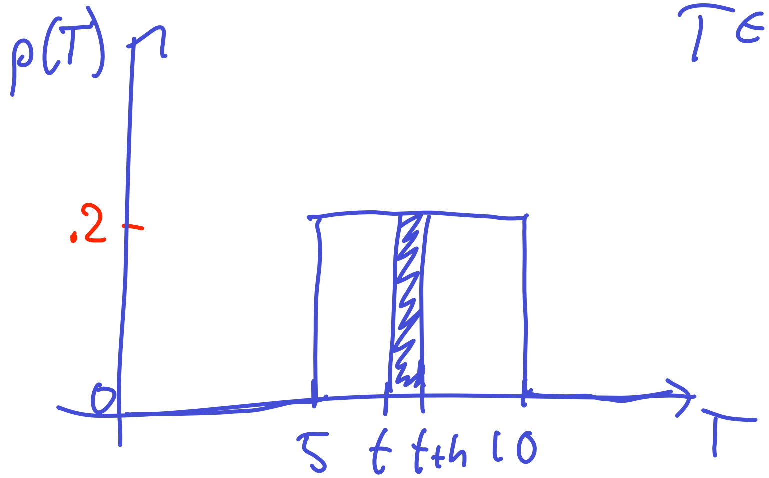
  Specify P(X = i) = p_i, s.t. $\sum_i p_i = 1$
- Multinomial distribution counts the number of outcomes for each side

Continuous distributions

- Probability density

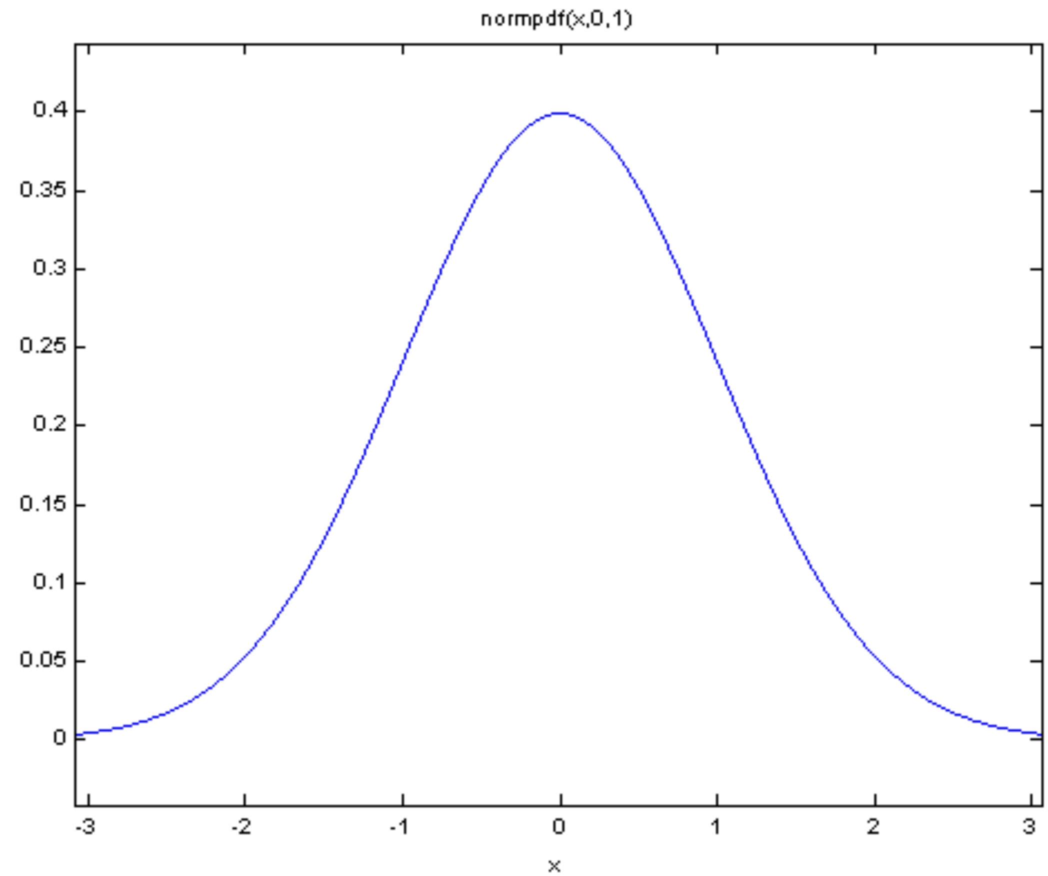
T = time it takes to take an action
$T \in \mathbb{R}$

$p(T = 7.32198\ldots)$



$p(t) = \lim_{h \to 0} P(T \in [t, t+h]) / h$

Example: Gaussian distribution



- σ = Standard deviation
- μ = mean

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Joint distributions

- Instead of random variable, have random vector
$$\mathbf{X}(\omega) = [X_1(\omega), \ldots, X_n(\omega)] \in D^n$$
- Can specify $P(X_1=x_1,\ldots,X_n=x_n)$ directly
(atomic events are assignments $x_1,\ldots,x_n$)
- Joint distribution describes relationship among all variables
- Example: X_1, X_2 ideal 6 sided dice, $X_3 = X_1 + X_2$

$P(X_1=1, X_2=3, X_3=4) = \frac{1}{36}$

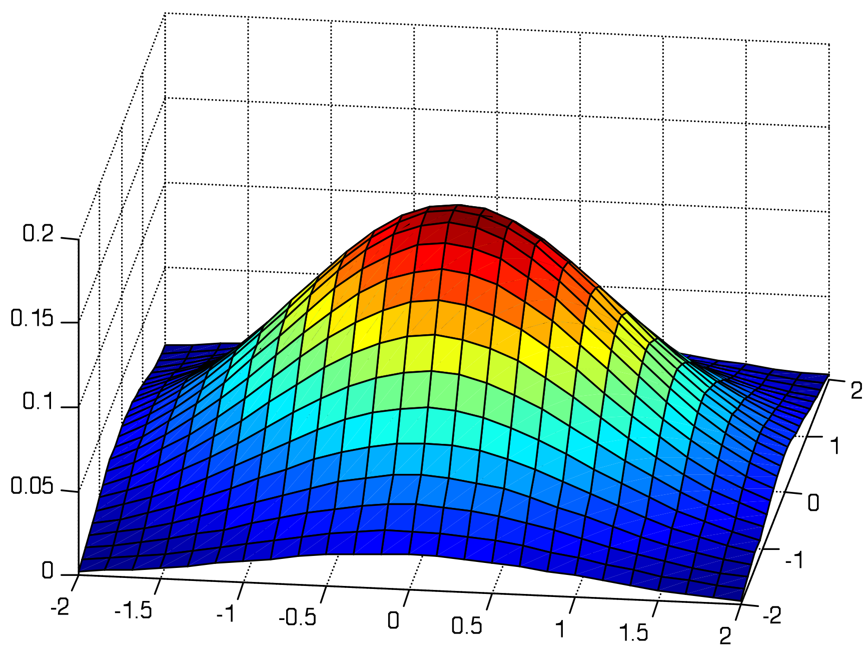
$P(X_1=1, X_2=3, X_3=5) = 0$

$\Omega = \{(1,1,2), (1,2,3), \ldots (6,6,12)\}$

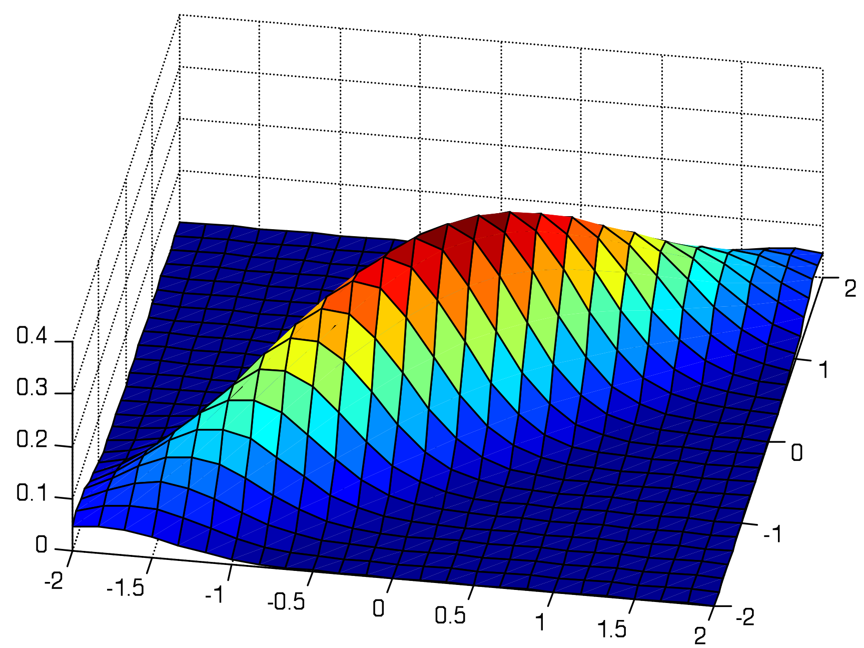
$P(\omega) = \frac{1}{|\Omega|} = \frac{1}{36}$

Example: Multivariate Gaussian

$$\frac{1}{2\pi\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \qquad \Sigma = \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{pmatrix} \qquad \mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$$



$$\Sigma = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$



$$\Sigma = \begin{pmatrix}1 & 0.9\\ 0.9 & 1\end{pmatrix}$$

Probabilistic propositional logic

- Suppose we would like to express uncertainty about *logical propositions*
- Birds can typically fly $P(Bird \Rightarrow CanFly) = .95$
- Propositional symbols ➔ Bernoulli random variables
  - Specify $P(Bird = b, CanFly = f)$ for all $b, f \in \{true, false\}$
  - Events ω encode assignments to all propositional symbols

	bird	¬bird
canfly	.4	.3
¬canfly	.01	.29

$P(Bird \Rightarrow Canfly) =$
$P(\neg Bird \vee Canfly) = .4 + .3 + .29$
$= .99$

- Probability of a proposition ϕ is the probability mass of all models of ϕ (i.e., all ω that make ϕ true)

Marginal distributions

What is P(Toothache)?

$= .108 + .012 + .016 + .064$

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Joint distribution

$P(X_1, \ldots X_n)$

$$P(X_i) = \sum_{x_1 \ldots x_{i-1}, x_{i+1} \ldots x_n} P(x_1 \ldots x_{i-1}, X_i, x_{i+1} \ldots x_n)$$

Conditional distributions

- Conditional (or posterior) probabilities. E.g.,

$$P(cavity \mid toothache) = .8$$

$$P(cavity \mid \neg toothache) = .1$$

- Conditional distributions $P(Cavity \mid Toothache)$ specify values for all state combinations

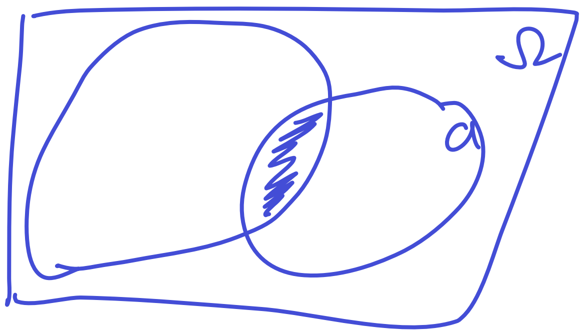
- New evidence can change posterior belief

$$P(cavity \mid toothache, cavity) = 1$$

$$P(cavity \mid toothache, sunny) = P(cavity \mid toothache)$$

Conditional distributions

- Formal definition: $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$



- Product rule $P(a \wedge b) = P(a \mid b) P(b)$
- For distributions: $P(A, B) = P(A \mid B) P(B)$
  (set of equations, one for each instantiation of A,B)

- Chain rule: $P(X_1, \ldots, X_n) = P(X_1) \cdot P(X_2 \mid X_1) \cdot P(X_3 \mid X_1, X_2) \cdot \ldots \cdot P(X_n \mid X_1, \ldots, X_{n-1})$

Example: Conditional distributions

What is P(Toothache|cavity)?

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Joint distribution

$$P(t|c) = \frac{P(t \wedge c)}{P(c)} = \frac{.12}{.2}$$

$$P(\neg t|c) = \frac{P(\neg t \wedge c)}{P(c)} = \frac{.08}{.2}$$

sum of 4 numbers each, 8 total

$$P(T|c) = \frac{1}{Z} P(T, c) \quad \text{where} \quad Z = P(t, c) + P(\neg t, c)$$

$$= \frac{1}{Z} [.12, .08] \qquad Z = .12 + .08 = .2$$

sum of 2 numbers

$$P(T|c) = [.6, .4]$$

Posterior inference

- Suppose we know:
  - Prior probability P(C)

$cavity$	$\neg cavity$
.1	.9

  - Likelihood P(T | C)

	$toothache$	$\neg toothache$
$cavity$	.9	.1
$\neg cavity$	.01	.99

Bayes' rule

- How do we get $P(cavity \mid toothache)$

$$P(C|t) = \frac{1}{Z} P(C,t) = \frac{1}{Z} P(t|C) \cdot P(C)$$

$$= \frac{1}{Z} [.9 \cdot .1, .01 \cdot .9]$$

$P(\beta \mid \alpha) P(\alpha)$

Problems with high-dim. distributions

- Suppose we have *n* propositional symbols
- How many parameters do we need to specify $P(X_1=x_1,\ldots,X_n=x_n)$?

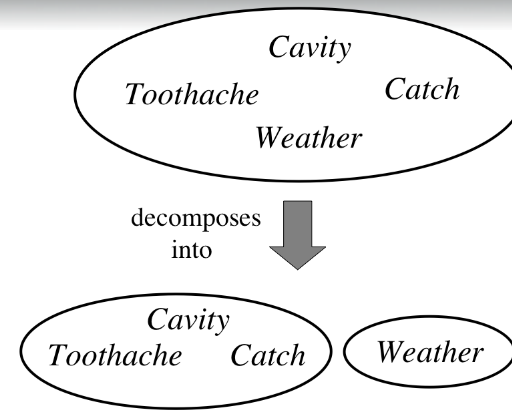
X_1	X_2	$\ldots$		X_n	$P(X_1 \ldots X_n)$
0	0			0	$\vdots$
0	0		0	1	
0			1	0	
				$\vdots$	
1	1	.		1	

$2^n - 1$ params !!

Independent RVs

- What if RVs are independent?

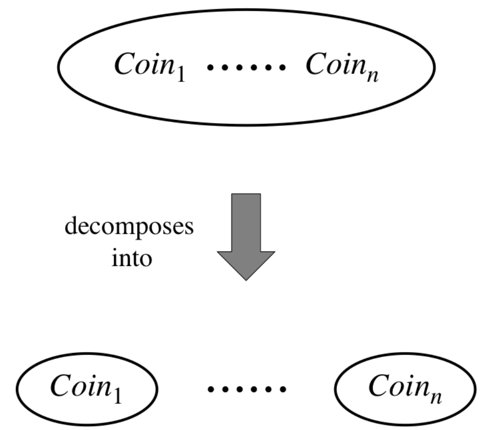
  RVs $X_1,\ldots,X_n$ are independent, if for any assignment $P(X_1=x_1,\ldots,X_n=x_n) = P(x_1)\, P(x_2) \ldots P(x_n)$



- How many parameters are needed in this case?

  n, one for each var $P(X_i=1)=p_i$!

  $n \ll 2^n - 1$!



- Independence too strong assumption… Is there something weaker? Indep: means: $P(X_1 | X_2 \ldots X_n) = P(X_1)$

Key concept: Conditional independence

- How many parameters? $P(Toothache, Cavity, Catch)$
- If I know there's a *cavity*, knowing *toothache* won't help predict whether the probe *catches*

$P(catch \mid cavity, toothache) = P(catch \mid cavity)$

— " — $\neg cavity$ — " — — " — $\neg cavity$

$P(Catch \mid Cavity, Toothache) = P(Catch \mid Cavity)$

Catch and Toothache are conditionally independent given Cavity