
# Probability

EECS 492 Lecture 16
March 8, 2011

## Course Overview: Where We Are

- Logic
  - Languages: PL, FOL
  - Inference (model checking, chaining, resolution)
- Logical Planning
  - Deterministic
  - Non-Deterministic: dealing with unknown propositions
- Probability
  - Language
  - Inference

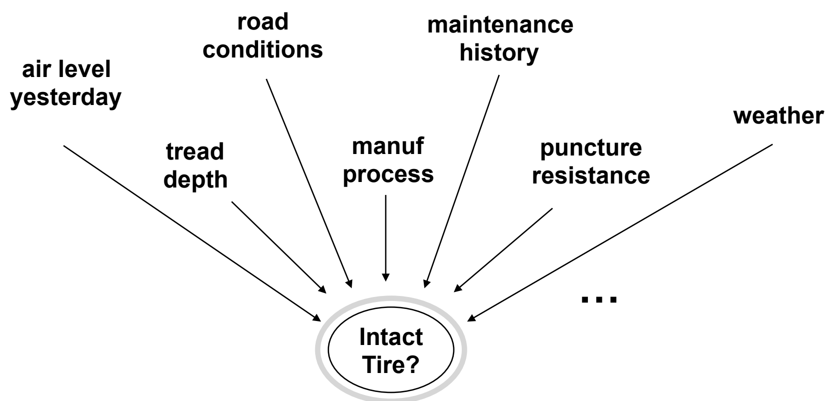
## Today

- How does probability make for better agents?
- Decision Theory
  - What does rationality mean for a probabilistic agent?
  - Maximizing *expected* utility
- The language of probabilities
  - Joint, Marginal, Conditional Distributions
  - Bayes' rule
  - Simple methods of probabilistic inference

## What do probabilities *mean*?

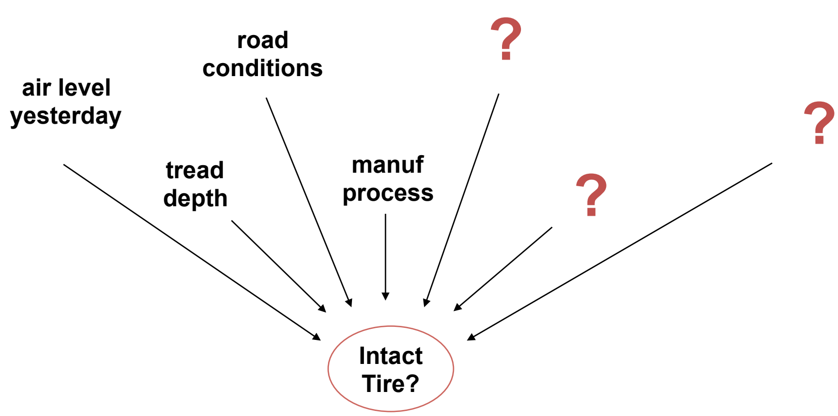

- And where do they come from?
  - Pr(Head) = ?
- Interpretations:
  - A) The coin has an intrinsic property of coming up heads at a particular rate
  - B) Given a large number of trials, the fraction of heads approaches Pr(Head)
  - C) I am uncertain about Pr(Head), but have some prior belief which can be refined through observation.
  - D) The coin has qualities of being both heads and tails, and the "headness" of it is Pr(Head).

## Certainty through Detail



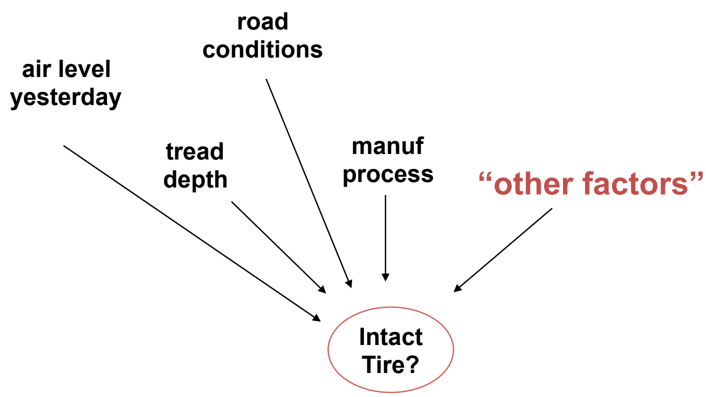

- Can approach certainty in the proposition of interest, given enough detail in causal factors.

## Uncertainty and Abstraction




- Conversely, leaving out causal factors induces *uncertainty* in the proposition of interest.

## Uncertainty as Summarization




Degrees of belief are summary measures of the uncertainty induced by leaving out model details.

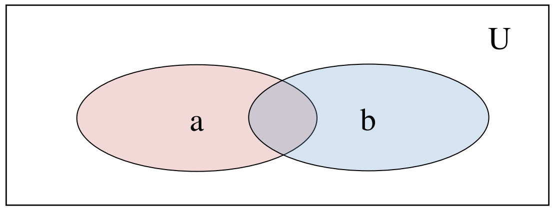
## Probability Theory

- Probability function

$$Pr: S \rightarrow [0,1]$$

  - $S$ is a sentence in a logic (typically propositional)
- Random variables (boolean, discrete, continuous)
  - Analogous to a propositional symbol
- Axioms
  1. $0 \le Pr(a) \le 1$
  2. $Pr(true) = 1$ *and* $Pr(false) = 0$
  3. $Pr(a \vee b) = Pr(a) + Pr(b) - Pr(a \wedge b)$

## Your turn!




Given:

$$\Pr(a \vee b) = \Pr(a) + \Pr(b) - \Pr(a \wedge b)$$

Show that:

$$\Pr(\neg a) = 1 - \Pr(a)$$

Hint: Consider $\Pr(a \vee \neg a)$

## Justifying the Axioms

- Axioms of probability restrict the set of probabilistic beliefs an agent can hold
- Why are these beliefs irrational?
  - $\Pr(a) = 0.4$, $\Pr(b) = 0.3$, $\Pr(a \vee b) = 0.8$
- de Finetti's argument
  - Agent should be willing to bet based on beliefs
  - If Pr(a) = 0.4, then agent should be *indifferent* to [ \$6 if a; \$4 if ¬a ]
  - Any agent violating axioms can be turned into a money machine (!) via a Dutch Book

## Dutch Book Example

| Agent1 | | | Agent2 bets on | Outcome for Agent1 | | | |
|---|---|---|---|---|---|---|---|
| Event | Belief | Odds | --- | a∧b | a∧¬b | ¬a∧b | ¬a∧¬b |
| a | 0.4 | 4 : 6 | a | -6 | -6 | 4 | 4 |
| b | 0.3 | 3 : 7 | b | -7 | 3 | -7 | 3 |
| a V b | 0.8 | 2 : 8 | ¬(a V b) | 2 | 2 | 2 | -8 |
| | | | | -11 | -1 | -1 | -1 |

## Joint Probability

- Probability of multiple propositions, considered simultaneously.
  - P(H1 ^ ~H2) = 0.25 (joint probability)
- Specifying joint probability over all atomic events = full joint distribution = complete probabilistic description of the world
  - In discrete case, could use a big table
  - How many entries in the table?
    - Assume N binary random variables

| Two crooked coins (p=0.7) | | |
|---|---|---|
| H1 | H2 | P(H1 ^ H2) |
| False | False | 0.09 |
| False | True | 0.21 |
| True | False | 0.21 |
| True | True | 0.49 |

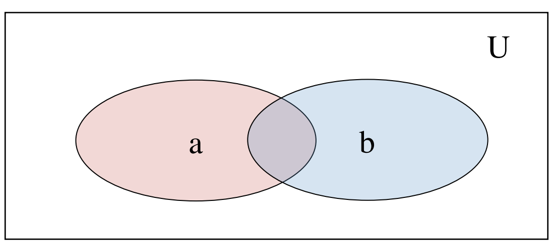
# Marginal Probability

- Start with a joint probability, ignore some random variables
- Crooked coin
  - Given full joint distribution
  - Suppose we can't see the second flip
    - Can we still characterize P(H1)?
  - P(H1) = P(H1, H2) + P(H1, ~H2)
    = 0.7 (whew!)

Marginalization

| Two crooked coins (p=0.7) | | |
|---|---|---|
| H1 | H2 | P(H1 ^ H2) |
| False | False | 0.09 |
| False | True | 0.21 |
| True | False | 0.21 |
| True | True | 0.49 |

# Conditional Probability




$$\Pr(a \mid b) = \frac{\Pr(a \wedge b)}{\Pr(b)}$$

- Undefined if Pr(b) = 0
- Means probability of *a* given all we know is *b*
- Often: P( a | KB)

## Your turn: Marginals

| #Legs | Species | P(Legs=#Legs, Species =Species) |
|---|---|---|
| 2 | Dog | .001 |
| | Cat | .001 |
| | Bird | .2 |
| 3 | Dog | .057 |
| | Cat | .04 |
| | Bird | .001 |
| 4 | Dog | .4 |
| | Cat | .3 |
| | Bird | 0 |

1. P(#legs=2 v #legs=3 v #legs=4)
2. P(Dog v Cat v Bird)
3. P(Bird)
4. P(Bird, #legs = 2)
5. P(Bird | #legs = 2)
6. P(#legs = 3 | Cat)

## Evidence Evaluation Example

- Disease testing (hypothetical):
  - Prior probability (prevalence)
    - Pr(disease) = .0005
  - Conditionals (test accuracy)
    - Pr(pos test | disease) = 1
    - Pr(neg test | ¬ disease) = .995
  - Posterior
    - Pr(disease | pos test) = ?

## Product Rule

- P(A | B) P(B) = P(A, B)
- Of course P(A, B) = P(B, A), so:
  - P(A | B) P(B) = P(B | A) P(A) = P(A, B)
- If we rearrange a bit, we arrive at one of the most important probabilistic theorems:

## Bayes's Theorem

$$\Pr(h \mid e) = \frac{\Pr(h \wedge e)}{\Pr(e)}$$

$$= \frac{\Pr(e \mid h)\Pr(h)}{\Pr(e)}$$

## Evidence Evaluation Example

- Disease testing (hypothetical):
  - Prior probability (prevalence)
    - Pr(disease) = .0005
  - Conditionals (test accuracy)
    - Pr(pos test | disease) = 1
    - Pr(neg test | ¬ disease) = .995
  - Posterior
    - Pr(disease | pos test) =
      Pr(pos test | disease) Pr(disease) / Pr(pos test) =
      1 * .0005 / ??

## Evidence Evaluation Example

Pr(pos test) = Pr(pos-test ^ disease) +
Pr(pos-test ^ ~disease)
= Pr(pos-test|disease)Pr(disease) +
Pr(pos-test|~disease)Pr(~disease)
= (1 * .0005) +
((1-Pr(neg|~dis))*(1-Pr(disease))
= .0005 + (.005 * .9995)
= .0054975

## Evidence Evaluation Example

- Disease testing (hypothetical):
  - Prior probability (prevalence)
    - Pr(disease) = .0005
  - Conditionals (test accuracy)
    - Pr(pos test | disease) = 1
    - Pr(neg test | ¬ disease) = .995
  - Posterior
    - Pr(disease | pos test) =
      Pr(pos test | disease) Pr(disease) / Pr(pos test) =
      1 * .0005 / .0054975 =
      0.09095

## Causal versus Diagnostic Information

- P(funny engine noise | loose hose)
  - Causal or Diagnostic?
- P(loose hose | funny engine noise)
  - Causal or Diagnostic?

- Bayes' rule allows us to go back and forth
  - Which fact is more useful?
  - News report: "Police have identified the notorious hose loosener, who has doubled the prevalence of loose hoses. This hose loosener is still on the loose!"
    - What is P(funny engine noise | loose hose) now?
    - What is P(loose hose | funny engine noise) now?

## Independence

- a and b are *independent* iff:
  - Pr(a|b) = Pr(a)
- Independence implies
  - Pr(a ∧ b) = Pr(a)Pr(b)
- a and b are *conditionally* independent given c iff:
  - Pr(a|b c) = Pr(a|c)
  - Equiv: P(a,b | c) = P(a|c) P(b|c)

## Conditional Independence

- Consider:
  - HIV → Infection → Fever
- If we don't know "infection", then HIV and Fever are dependent.
- But if we *do* know "infection", HIV and Fever become independent
  - The value of "Infection" conveys all of the relevant information of HIV to fever.

## Combining Conditions

- How to calculate
  - Pr( Intact | Flat, Glass )
- Given
  - Pr ( Flat | Intact ), Pr( Flat | ~Intact )
  - Pr ( Intact | Glass ), Pr(~Intact | Glass )
  - Flat (looks flat) is conditionally independent of Glass (glass in road) given Intact
- *Hint*: use the conditional independence, normalization

## Continuous-valued Probabilities

- So far, we've only described discrete-valued probabilities
- Many real-world quantities are continuous
  - Tire pressure
  - GPS coordinates of car
- Very similar to discrete-valued probabilities…

# Continuous-valued Probabilities

| Discrete | Continuous |
| --- | --- |
| Probability functions<br>P(S) = [0, 1] | Probability *Density* Functions<br>P(x) >= 0<br>Prob(x) = 0 |
| ∑ P(S) = 1 | ∫ P(x) dx = 1 |

- Despite differences, notation for discrete probability distribution and continuous probability density function is (usually) the same!
- Common continuous distributions
  - Uniform: U(0,5)
  - Gaussian: $N(\mu,\sigma^2)$

# Next Time

- Bayesian Networks
  - Full joint distributions can be very big
  - The world has structure: not every proposition is correlated with very other proposition!
  - Exploit conditional independence to reduce problem size
  - Faster inference