Probability, Markov Models

Andrea Danyluk
March 15, 2017

With thanks to CS188 slides, as well as content from University of Washington CSE515, Penn State Stats, Yale University Stats, and others.

Announcements

- Programming Assignment 3: RL
  - Code reviews today

Today's Lecture

- Probability
  - Reminders
  - Exercises
  - Preparation for Markov Models

Uncertainty

- We don't always have complete/perfect information – yet we need to act/make decisions

Many ways to handle uncertainty

- Ignore it – pretend that the world is completely and correctly specified and that no contingencies will arise
- Ad hoc solutions – for example, in MAX
  - If mechanized loop test results appear "funny", perform retest
  - Assume that all data that look reasonable are accurate
  - Allow multiple conclusions to be reached
  - Attach rankings to rules to enable choice of one conclusion over another
- Incorporate probabilities into **knowledge representation and reasoning**

Probabilistic Reasoning

- General situation:
  - **Observed variables** (evidence): Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - **Unobserved variables**: Agent needs to reason about other aspects (e.g., where an object is or what disease is present)
  - **Model**: Agent knows something about how the known variables related to the unknown variables
- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge

[CS 188 Berkeley]

Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - Voltage of a (wired) phone line?
  - Will it snow today?
  - Where is Blinky?
  - What number will come up in the roll of a die?
- A random variable has a domain, i.e., a sample space of possible outcomes Ω
  - V in [0, 57]
  - S in {True, False}
  - B in {(0,1), (0,2), (1, 1), (1, 2)}
  - D in {1, 2, 3, 4, 5, 6}

Random Variables

- If X is a random variable over sample space Ω, probability measure P defined on Ω induces a probability distribution on X

Probability Distributions

Unobserved random variables have distributions.
A probability distribution = a table of probability values.

Sex	P
female	0.5
male	0.5

P(Sex=female) = 0.5

P(Sex=female) is an unconditional or prior probability.

Note: will abbreviate as P(female) when it is unambiguous.

Must have:

$$\forall x \;\; P(X=x) \geq 0 \quad \text{and} \quad \sum_x P(X=x) = 1$$

Joint Distributions

A joint distribution over a set of random variables X_1, X_2,… X_n specifies a probability for each possible outcome (i.e., assignment).

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
other	0.43	.34

P(kim ∧ male) = 0.01

P(kim, male) = 0.01

If these are all the random variables in the "world", then this table gives the **full joint** probability distribution. All entries sum to 1.

Probabilistic Models

A joint distribution over a set of random variables X_1, X_2,… X_n specifies a probability for each possible outcome (i.e., assignment).

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
other	0.43	.34

P(kim ∧ male) = 0.01

P(kim, male) = 0.01

A probabilistic model is a joint distribution over a set of random variables.

Events

An event E is a set of outcomes.

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

From a joint distribution, can calculate the probability of any event.

This process of summing over entries in the joint distribution is called inference by enumeration.

Events

- From a joint probability distribution, can calculate the probability of any event
  - P(kate)? P(male)?
  - P(michael ∧ female)?
  - P(kate ∨ kim)?

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
Other	0.43	.34

Events

- From a joint probability distribution, can calculate the probability of any event
  - P(kate)? P(male)?
  - P(michael ∧ female)?
  - P(kate ∨ kim)?

Name	Sex	P(N, S)
kate	female	0.04
kate	male	0.0
kim	female	0.02
kim	male	0.01
michael	female	0.01
michael	male	0.1
other	female	0.43
other	male	0.39

Atomic Events

- An atomic event is a complete specification of the state of the world
  - An assignment of values to all random variables
- Joint probability distribution assigns probabilities to all possible atomic events.

Marginal Distributions

- Marginal distributions are subtables that eliminate variables.
- Marginalization is the process of combining collapsed rows/columns by summing them.

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
Other	0.43	.34

kate	0.04
kim	0.03
michael	0.11
tom	0.05
Other	0.77

female	male
0.5	0.5

Marginal Distributions

Name	Sex	P(N, S)
kate	female	0.04
kate	male	0.0
kim	female	0.02
kim	male	0.01
michael	female	0.01
michael	male	0.1
other	female	0.43
other	male	0.39

S	P(S)
female	0.5
male	0.5

N	P(N)
kate	?
kim	?
michael	?
other	?

$P(X_1 = x1) = \Sigma_{x2} P(X_1 = x1, X_2 = x2)$

Conditional (Posterior) Probabilities

A simple relationship between joint and conditional probabilities

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
Other	0.43	.34

P(female|michael)

= (female, michael) / P(michael)

= 0.01 / 0.11

= 0.09

"given that *michael* is all I know, what is the probability that you're female"

Conditional (Posterior) Probabilities

Probabilities change with new evidence:

P(female|michael) = 0.09

P(female|michael, bass-register voice) = 0.001

Conditional Distributions

Conditional distributions are probability distributions over some variables given fixed values of others.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

P(Name|Sex)

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
Other	0.43	.34

	S=f
kate	0.08
kim	0.04
michael	0.02
tom	0.0
Other	0.86

	S=m
kate	0.0
kim	0.02
michael	0.2
tom	0.1
Other	.68

Conditional Distributions: Normalization

Conditional distributions are probability distributions over some variables given fixed values of others.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

1. Add the entries.
2. Divide by the sum.

	female	male
kate	0.04	0.0
kim	0.02	0.01
michael	0.01	0.1
tom	0.0	0.05
Other	0.43	.34

	S=f
kate	0.04
kim	0.02
michael	0.01
tom	0.0
Other	0.43

	S=f
kate	0.08
kim	0.04
michael	0.02
tom	0.0
Other	0.86

Conditional to Joint Distributions

Can also compute joint distributions from conditional distributions.

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

P(a,b) = P(a|b)P(b)

Or

P(a,b) = P(b|a)P(a)

Conditional to Joint Distributions

Name	Sex	P(N | S)
kim	female	0.04
kim	male	0.02
kate	female	0.08
kate	male	0.0
michael	female	.02
michael	male	.2
other	female	.86
other	male	.78

Sex	P(S)
female	0.5
male	0.5

Name	Sex	P(N, G)
kate	female	0.04
kate	male	0.0
kim	female	0.02
kim	male	0.01
michael	female	0.01
michael	male	0.1
other	female	0.43
other	male	0.39

P(n , s) = P(n | s) P(s)

Useful Axioms, Definitions, and Rules

- All probabilities are between 0 and 1.
  $0 \leq P(a) \leq 1$
  and the total probability of the set of possible worlds is 1
- The probability associated with a proposition is the sum of the probabilities of the worlds in which it holds.
  - Necessarily true propositions have value 1.
  - Necessarily false propositions have value 0.
- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

Useful Axioms, Definitions, and Rules

Definition of conditional probability:

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

Product rule:

P(a,b) = P(a|b)P(b) = P(b|a)P(a)

Chain rule: derived by successive applications of the product rule

$P(x_1,x_2,x_3)=P(x_1)P(x_2|x_1)P(x_3|x_1,x_2)$
$P(x_1,x_2,..x_n)=P(x_n|x_1...x_{n-1})P(x_{n-1}|x_1...x_{n-2})...P(x_2|x_1)P(x_1)$

Bayes rule: P(a|b) = P(b|a)P(a) / P(b)

Diagnostic from Causal Probability

P(Cause|Effect) = P(Effect|Cause)P(Cause) / P(Effect)

[Adapted from CS 188]

Inference in Ghostbusters

- A ghost is in the grid somewhere
- Sensor readings tell how close a square is to the ghost
- Sensors noisy, but we know P(Color|Distance)

Ghostbusters

- Say we have two distributions
  - P(G): say it's uniform
  - Sensor reading model P(R|G)
- Say we get a reading at (1,1)
- Can calculate the posterior distribution P(G|r) over all locations given the reading at (1,1)

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

- Sensor readings
  - On the ghost (1 location): red
  - 1 or 2 away (5 locations): orange
  - 3 or 4 away (3 locations): yellow
- Sensors noisy

P(red | 0)	P(orange | 0)	P(yellow | 0)
0.7	0.2	0.1

P(red | 1 or 2)	P(orange | 1 or 2)	P(yellow | 1 or 2)
0.15	0.7	0.15

P(red | 3 or 4)	P(orange | 3 or 4)	P(yellow | 3 or 4)
0.1	0.2	0.7

[Adapted from CS 188 Berkeley]

P(g|yellow)
P(0 away|yellow)
P(1-2 away|yellow)
P(3-4 away|yellow)
P(yellow)