Artificial Intelligence:

Quantifying Uncertainty

Muhammad Abul Hasan, Ph.D.

Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
muhammad.hasan@ulab.edu.bd

August 25, 2021

Outline

1. Probability Theory
   - Probability Basics
   - Prior, Joint and Conditional Probabilities
2. Probability Distributions
   - Probability Distribution Basics
   - Inference from Joint Distributions
3. Using Independence
   - Independence
   - Bayes' Rule

“As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

– Albert Einstein”

Motivation

Goal

Look out window - is it warm?

Rules

$\forall d \; Sunny(d) \Rightarrow Warm(d)$
Wrong; it can be sunny but cold
$\forall d \; Warm(d) \Rightarrow Sunny(d)$
Wrong; it can be warm and cloudy

Problem

Can't capture that sunny days are usually warm

Probability Theory

Key Idea

Measure degrees of belief in propositions

Random Variable An interesting part of the world
e.g. *Sky* or *Temp*

Domain Possible values of a random variable
e.g. $domain(Temp) = \langle warm, cold, \ldots \rangle$

Proposition A statement, like propositional logic
e.g. $Sky = sunny \wedge Temp = warm$

Probability The degree of belief in a proposition
e.g. $P(Temp = warm) = 0.7$

Aside: Fuzzy Logic

Probability Theory

- Propositions are believed to a certain degree
- Belief values range from [0, 1]

Fuzzy Logic

- Propositions are *true* to a certain degree
- *Truth* values range from [0, 1]

For example:

- $T(Tall(Steve)) = 0.5$
- $T(Fat(Steve)) = 0.1$

Better for describing indefinite classes than for reasoning

Probability Rules

Range of Probabilities

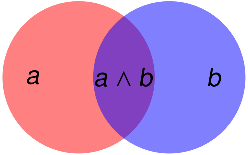
$0 \leq P(a) \leq 1$

Propositions Known to be True or False

$$
\begin{aligned}
P(true) &= 1 \\
P(false) &= 0
\end{aligned}
$$

Probability of Disjunctions

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



Random Variables

Boolean Random Variables

Have the domain $\langle \mathit{true}, \mathit{false} \rangle$, e.g.

- *IsSunny*

Discrete Random Variables

Have a countable domain, e.g.

- $\mathit{domain}(\mathit{Sky}) = \langle \mathit{sunny}, \mathit{cloudy}, \ldots \rangle$
- $\mathit{domain}(\mathit{DieRoll}) = \langle$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\rangle$

Continuous Random Variables

Have a real-valued domain, e.g.

- $\mathit{domain}(\mathit{Temperature}) = \langle \ldots, -40^\circ, \ldots, 98.6^\circ, \ldots \rangle$

Prior Probability

Definition

$P(a)$ The unconditional or prior probability of a proposition a is the degree of belief in that proposition *given no other information*

Examples

$$
\begin{aligned}
P(DieRoll = ⚄) &= 1/6 \\
P(CardDrawn = \text{A}\spadesuit) &= 1/52 \\
P(SkyInDhaka = sunny) &= 210/365
\end{aligned}
$$

Joint Probability

Definition

$P(a_1, \ldots, a_n)$ The joint probability of propositions $a_1, \ldots, a_n$ is the degree of belief in the proposition $a_1 \wedge \ldots \wedge a_n$

Example

$$
\begin{aligned}
& P(DieRoll_1 = ⚃, DieRoll_2 = ⚅) \\
= \quad & P(DieRoll_1 = ⚃ \wedge DieRoll_2 = ⚅) \\
= \quad & 1/36
\end{aligned}
$$

Conditional Probability

Definition

$P(a \mid b)$ The posterior or conditional probability of a proposition a given a proposition b is the degree of belief in a, *given that we know only* b

Examples

$$
\begin{aligned}
P(Card = A\spadesuit \mid CardSuit = \spadesuit) &= 1/13 \\
P(DieRoll_2 = \text{⚄} \mid DieRoll_1 = \text{⚄}) &= 1/6
\end{aligned}
$$

Prior, Joint or Conditional?

1. Probability of a having a cavity?
   Prior, $P(Cavity = true)$
2. Probability of it being warm and cloudy?
   Joint, $P(Temp = warm, Sky = cloudy)$
3. Probability of car being stolen?
   Prior, $P(CarStolen = true)$
4. Probability of car being stolen and being in Dhaka?
   Joint, $P(CarStolen = true, InDhaka = true)$
5. The car is in Dhaka. Probability of car being stolen?
   Conditional, $P(CarStolen = true \mid InDhaka = true)$
6. It's cloudy. Probability of it being warm?
   Conditional, $P(Temp = warm \mid Sky = cloudy)$

Relation between Joint and Conditional

Product Rule

$P(a \wedge b) = P(a \mid b)P(b)$ or $P(a \mid b) = P(a \wedge b)/P(b)$

Intuition

To have $a \wedge b$ true, we need b true, and a true given b

Example

$$
\begin{aligned}
P(A \wedge \spadesuit) &= 1/52 \\
P(A \mid \spadesuit) &= 1/13 \\
P(\spadesuit) &= 1/4 \\
P(A \mid \spadesuit)P(\spadesuit) &= 1/13 \cdot 1/4 = 1/52
\end{aligned}
$$

Chain Rule

Key Ideas

- Repeatedly apply product rule, $P(a, b) = P(a \mid b)P(b)$
- Joint probability $\rightarrow$ conditional probabilities

Example

$$
\begin{aligned}
&P(sunny, dry, warm) \\
&\quad = P(sunny \mid dry, warm)P(dry, warm) \\
&\quad = P(sunny \mid dry, warm)P(dry \mid warm)P(warm)
\end{aligned}
$$

Probability Distributions

Definition

$\mathbf{P}(X)$ The probability distribution of a random variable X is a list of probabilities for each domain value

Example

$\mathbf{P}(SkyInDhaka) = \langle 210/365, 155/365 \rangle$ means

$$\begin{aligned} P(SkyInDhaka = sunny) &= 210/365 \\ P(SkyInDhaka = cloudy) &= 155/365 \end{aligned}$$

Notation Warning

- $P(a)$ or $P(X = a)$ means prior probability
- $\mathbf{P}(X)$ means probability distribution

Probability Distributions

Key Idea

$$\sum_i^d P(X = X_i) = 1$$

The sum of the probabilities for all possible value assignments of the random variable is always 1

Example

$$\begin{aligned}
\sum_x P(Die = x) \\
&= P(Die = \text{⚀}) + P(Die = \text{⚁}) + P(Die = \text{⚂}) + \\
&\quad P(Die = \text{⚃}) + P(Die = \text{⚄}) + P(Die = \text{⚅}) \\
&= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\
&= 1
\end{aligned}$$

Continuous Variable Probability Distributions

Definition

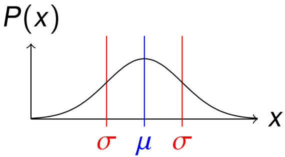
A probability density function is a probability distribution over a continuous variable

Key Idea

Function assigns a probability to all possible values

Example

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



Joint Probability Distributions

Definition

$\mathbf{P}(X_1, \ldots, X_n)$ The joint probability distribution of random variables $X_1, \ldots, X_n$ is a table of probabilities for each combination of values in the variable domains

Example

$\mathbf{P}(Gender, Smoker) =$

	$Gender = male$	$Gender = female$
$Smoker = true$	0.113	0.107
$Smoker = false$	0.377	0.403

Inference with Joint Probability Distributions

Key Idea

Sum entries in joint distribution where proposition is true

Example: $P(Gender = female \vee Smoker = false)$

	$Gender = male$	$Gender = female$
$Smoker = true$	0.113	0.107
$Smoker = false$	0.377	0.403

$$
\begin{aligned}
&P(Gender = female \vee Smoker = false) \\
&= P(Gender = male, Smoker = false) + \\
&\quad P(Gender = female, Smoker = true) + \\
&\quad P(Gender = female, Smoker = false) \\
&= 0.377 + 0.107 + 0.403 = 0.887
\end{aligned}
$$

Marginalization

Key Idea

$$\mathbf{P}(Y) = \sum_z \mathbf{P}(Y, z)$$

Marginalization removes all variables but **Y** by summing over the values of the other variables

Example: $P(Gender = female)$

The other variable is *Smoker*, so:

$$\begin{aligned} & P(Gender = female) \\ = \quad & P(Gender = female, Smoker = true) + \\ & P(Gender = female, Smoker = false) \\ = \quad & 0.107 + 0.403 = 0.51 \end{aligned}$$

Normalization

Key Idea

$$\mathbf{P}(Y \mid z) = \frac{\mathbf{P}(Y, z)}{P(z)} = \frac{\mathbf{P}(Y, z)}{\sum_y P(y, z)} = \alpha \mathbf{P}(Y, z)$$

Calculating a Normalizing Constant

$$\begin{aligned} P(\spadesuit, A) &= \frac{1}{52} \\ P(\clubsuit, A) &= \frac{1}{52} \\ P(\diamondsuit, A) &= \frac{1}{52} \\ P(\heartsuit, A) &= \frac{1}{52} \end{aligned}$$

$$\alpha = \frac{1}{\frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52}} = 13$$

$$\mathbf{P}(Suit \mid A) = \langle \frac{13}{52}, \frac{13}{52}, \frac{13}{52}, \frac{13}{52} \rangle = \langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \rangle$$

Inference Exercises

$sunny$	$warm$	$snow$	0.01
$sunny$	$warm$	$\neg snow$	0.59
$sunny$	$cold$	$snow$	0.08
$sunny$	$cold$	$\neg snow$	0.14
$cloudy$	$warm$	$snow$	0.03
$cloudy$	$warm$	$\neg snow$	0.07
$cloudy$	$cold$	$snow$	0.06
$cloudy$	$cold$	$\neg snow$	0.02

1. $P(sunny) = 0.82$
2. $P(warm) = 0.70$
3. $P(snow) = 0.18$
4. $P(sunny \vee \neg snow) = 0.91$
5. $P(sunny \mid snow) = 0.50$
6. $P(snow \mid sunny, cold) = 0.36$

Joint Distribution Inference

Properties

Given n random variables with maximum domain size d:

Time Complexity? $O(d^n)$

Space Complexity? $O(d^n)$

Biggest Problem

How do you fill in a table of $O(d^n)$ probabilities?

Independence

Key Ideas

- Sometimes no connection exists between variables
- Such independence determined by world knowledge

Formal Independence

$$P(a, b) = P(a)P(b) \quad \text{or} \quad P(a \mid b) = P(a)$$

Examples

$$\begin{aligned}
P(EyeColor, Gender) &= P(EyeColor)P(Gender) \\
P(Cavity, BlazersWon) &= P(Cavity)P(BlazersWon) \\
P(DieRoll_1, DieRoll_2) &= P(DieRoll_1)P(DieRoll_2)
\end{aligned}$$

Independence

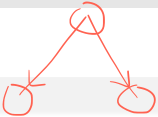
Using Independence

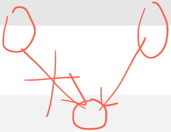
- No need to store entire joint probability table
- Can store several smaller independent tables

Examples

	Table Size
$\mathbf{P}(DieRoll_1, DieRoll_2)$	$6 \cdot 6 = 36$
$\mathbf{P}(DieRoll_1)\mathbf{P}(DieRoll_2)$	$6 + 6 = 12$
$\mathbf{P}(Age, Gender, BlazersWon)$	$125 \cdot 2 \cdot 2 = 500$
$\mathbf{P}(Age, Gender)\mathbf{P}(BlazersWon)$	$125 \cdot 2 + 2 = 252$

Conditional Independence





Key Idea

Two variables can sometimes become independent after the value of a third variable is observed

Definition

A random variable X is conditionally independent of random variable Y given the random variable Z if:

$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z)$$

or

$$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z)$$

Conditional Independence Example

Is *GrayHair* indpendent of *Bifocals*?

No, we expect the two to often come together, e.g.:

$$P(\textit{gray-hair}, \textit{bifocals}) > P(\textit{gray-hair})P(\textit{bifocals})$$

Is *GreyHair* independent of *Bifocals* given *Age*?

Yes, the bifocals add nothing if we know the age, e.g.:

$$P(\textit{gray-hair} \mid \cancel{\textit{bifocals}}, \textit{Age}=x) = P(\textit{gray-hair} \mid \textit{Age}=x)$$

Conditional Independence Example

Noisy Phone

- Adam calls Betty and Charlie and says a number N_A
- Betty hears N_B and Charlie hears N_C

Are N_B and N_C independent?

No, we expect the numbers to be similar, e.g.:

$$P(N_B = 1, N_C = 1) > P(N_B = 1)P(N_C = 1)$$

Are N_B and N_C independent given N_A?

Yes, Betty's number adds nothing if we know Adam's, e.g.:

$$P(N_C = 1 \mid N_B = 2, N_A = 1) = P(N_C = 1 \mid N_A = 1)$$

Thomas Bayes

Definition

Thomas Bayes, (born 1702, London, Englanddied April 17, 1761, Tunbridge Wells, Kent), English Nonconformist theologian and mathematician who was the first to use probability inductively and who established a mathematical basis for probability inference (a means of calculating, from the frequency with which an event has occurred in prior trials, the probability that it will occur in future trials.

P(b) Prior Probability

P(b | some given information) Posterior Probability

We can consider the given information as an evidence

Bayes' Rule

Key Idea

Swap the conditioned and conditioning variables

Bayes' Rule

Casual probability/Likelihood

$$P(b \mid a) = \frac{P(a \mid b)P(b) \rightarrow \text{Prior}}{P(a) \rightarrow \text{Marginal Probability}}$$

P(a) = P(a | b) P(b) + P(a | not b) P(not b)

Derivation

$$\begin{aligned} P(b \mid a) &= \frac{P(a \wedge b)}{P(a)} && \text{Definition} \\ &= \frac{P(a \mid b)P(b)}{P(a)} && \text{Product Rule} \end{aligned}$$

Why Use Bayes' Rule?

Making Diagnoses Based on Causal Knowledge

$$P(Cause \mid Effect) = \frac{\boxed{P(Effect \mid Cause)}\boxed{P(Cause)}}{P(Effect)}$$

100000/160000000

Example

$$P(meningitis \mid stiff\text{-}neck) = \frac{P(stiff\text{-}neck \mid meningitis)P(meningitis)}{P(stiff\text{-}neck)}$$

In an epidemic, where $P(meningitis)$ rises:

+Bayes $P(meningitis \mid stiff\text{-}neck)$ rises proportionally

-Bayes Collect data, re-estimate $P(meningitis \mid stiff\text{-}neck)$

Bayes' Rule & Conditional Independence

Deriving More Manageable Models

$$\begin{aligned} & P(\mathit{Cavity} \mid \mathit{toothache}, \mathit{catch}) \\ & = \alpha P(\mathit{toothache}, \mathit{catch} \mid \mathit{Cavity}) P(\mathit{Cavity}) \\ & = \alpha P(\mathit{toothache} \mid \mathit{Cavity}) P(\mathit{catch} \mid \mathit{Cavity}) P(\mathit{Cavity}) \end{aligned}$$

Naive Bayes Models

$$P(\mathit{Cause}, \mathit{Effect}_1, \ldots, \mathit{Effect}_n) = P(\mathit{Cause}) \prod_i P(\mathit{Effect}_i \mid \mathit{Cause})$$

- All effects conditionally independent given cause
- Common class of machine learning models

Probability Rules

Product Rule

$$P(a, b) = P(a \mid b)P(b)$$

Independence

$$P(a, b) = P(a)P(b)$$

Conditional Independence

$$P(a, b \mid c) = P(a \mid c)P(b \mid c)$$
$$P(a \mid b, c) = P(a \mid c)$$

Given: w indep. x, y, z
x indep. z given y

Show:

$$P(w, x, y, z) = P(w)P(x \mid y)P(z, y)$$

Proof:

$$\begin{aligned} &P(w, x, y, z) \\ &= P(w)P(x, y, z) \\ &= P(w)P(x, z \mid y)P(y) \\ &= P(w)P(x \mid y)P(z \mid y)P(y) \\ &= P(w)P(x \mid y)P(z, y) \end{aligned}$$

Probability Rule Exercises

Product Rule

$$P(a, b) = P(a \mid b)P(b)$$

Independence

$$P(a, b) = P(a)P(b)$$

Conditional Independence

$$P(a, b \mid c) = P(a \mid c)P(b \mid c)$$
$$P(a \mid b, c) = P(a \mid c)$$

Given:
- z indep. w, x, y
- w indep. x given y

Show:

$$P(w, z \mid x, y) = P(w \mid y)P(z)$$

Given:
- w indep. y, z given x
- x indep. z given y
- y indep. z

Show:

$$P(w, x, y, z) = P(w \mid x)P(x \mid y)P(y)P(z)$$

Normalization Rules

Complements

$$P(\neg a) + P(a) = 1$$

Summing Out

$$P(a, b) + P(\neg a, b) = P(b)$$

Distributions

$$\sum_i^d P(X = x_i) = 1$$

Given: $\mathbf{P}(X \mid y, z) = \alpha\langle 0.1, 0.3\rangle$

Show: $\mathbf{P}(X \mid y, z) = \langle 0.25, 0.75\rangle$

$$\begin{aligned}
P(y, z) &= P(x, y, z) + P(\neg x, y, z) \\
1 &= \frac{P(x, y, z) + P(\neg x, y, c)}{P(y, z)} \\
1 &= \frac{P(x, y, z)}{P(y, z)} + \frac{P(\neg x, y, z)}{P(y, z)} \\
1 &= P(x \mid y, z) + P(\neg x \mid y, z) \\
1 &= 0.1\alpha + 0.3\alpha \\
1 &= 0.4\alpha \\
2.5 &= \alpha
\end{aligned}$$

Normalization Rule Practice

Complements

$P(\neg a) + P(a) = 1$
$P(a \mid b) + P(\neg a \mid b) = 1$

Summing Out

$P(a, b) + P(\neg a, b) = P(b)$

Distributions

$$\sum_{i}^{d} P(X = x_i) = 1$$

Given:

$$\begin{aligned} P(y \mid x) &= 0.4 \\ P(y \mid \neg x) &= 0.9 \\ P(x) &= 0.2 \\ P(\neg x) &= 0.8 \end{aligned}$$

Show: $\mathbf{P}(X \mid y) = \langle 0.1, 0.9 \rangle$

Key Ideas

Probability Measures Belief

$P(false) = 0 \quad P(true) = 1 \quad \sum_x P(X = x) = 1$

Types of probabilities

- Prior, $P(X = x)$
- Joint, $P(X = x, Y = y)$
- Conditional, $P(X = x \mid Y = y)$

Inference

- Sums over full joint distribution
- Conditional independence + product and Bayes' rule

Thank You!