STOR 565 Machine Learning

Probability Overview

Andrew Nobel

January 2016

Basics of Set Theory

Definition (informal): A set is an unordered collection of objects.

Notation: A, B, C denote sets and x, y, z denote objects.

Basic relations between sets and objects

1. Containment

- Notation $x \in A$
- Meaning: Object x is a member/element of the set A

Basics of Set Theory, cont.

2. Subset

- Notation $A \subseteq B$
- Set A is a subset of B. Formally, if $x \in A$ then $x \in B$.

3. Equality

- Notation $A = B$
- Sets A and B are equal if they have the same members.

Note: $A = B$ is logically equivalent to $A \subseteq B$ and $B \subseteq A$.

Definition: The **Cardinality** of a set A, written $|A|$, is the number of objects in A, which may be infinite.

Sets Specified by *Lists*

Template: If $x_1, x_2, \ldots$ is a finite or infinite family of objects, then

$$\{x_1, x_2, \ldots\}$$

is the set containing these objects, without regards to order or repetition.

Examples:

- $A = \{1, 2, 5, 9\}, \quad |A| = 4$
- $B = \{$Ziggy, JoJo, Buddy$\}, \quad |B| = 3$
- $C = \{$All freshman at UNC$\}, \quad |C|$ is finite
- Natural numbers $\mathbb{N} = \{0, 1, 2, 3, \ldots\}, \quad |\mathbb{N}|$ is (countably) infinite
- Real numbers $\mathbb{R} = (-\infty, +\infty), \quad |\mathbb{R}|$ is (uncountably) infinite

Note: $A = \{5, 2, 1, 9\}$ and $B = \{$JoJo, Ziggy, Buddy, Ziggy$\}$

The Empty Set and Recursive Bracing

Definition: The **empty set**, denoted $\emptyset$, is the set with no elements.

Fact: For any set A, it is vacuously true that $\emptyset \subseteq A$.

Note: If B is a set then $\{B\}$ is the set that contains A as its sole element. Applying this rule recursively to the empty set $\emptyset$, we can define

$$A = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$$

Questions:

- What is $|A|$?
- Is $\emptyset \subseteq A$? Is $\emptyset \in A$?
- Is $\{\emptyset\} \subseteq A$? Is $\{\emptyset\} \in A$?

Sets Specified by *Conditions*

Template: Wish to specify a subset of some universal/background set that satisfy some conditions. General format

$$\{x : \text{conditions}(x)\} = \text{The set of } x \text{ satisfying the conditions}$$

Examples

- $\{x : x \text{ is prime }\} \subseteq \mathbb{N}$
- Rational numbers $\{x : x = a/b \text{ where } a, b \in \mathbb{N} \text{ and } b > 0\} \subseteq \mathbb{R}$
- Unit interval $[0, 1] = \{x : 0 \leq x \leq 1\} \subseteq \mathbb{R}$

Basic Operations on Sets

1. **Complement** $A^c = \{x : x \notin A\} =$ Objects *not* in A

2. **Union** $A \cup B = \{x : x \in A \text{ or } x \in B\} =$ Objects in A *or* B

3. **Intersection** $A \cap B = \{x : x \in A \text{ and } x \in B\} =$ Objects in A *and* B

Note: Unions and intersections of three or more sets can be defined similarly

- $A \cup B \cup C$
- $A \cap B \cap C$.

Definition: Sets A and B are **disjoint** if $A \cap B = \emptyset$, that is, they have no elements in common.

Set Operations, cont.

Basic containment relations: $A \cap B \subseteq A, B \subseteq A \cup B$

DeMorgan's Laws

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

Distributive Laws

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Random Experiments

An **experiment** is an idealized procedure that can be repeated and that has a well-defined set of possible outcomes.

An experiment is **random** if its outcome is uncertain in advance.

Examples

- Flip a coin three times
- Roll a die two times
- Choose three cards at random from a standard deck
- Count number of defective items in shipment of electronic components
- Measure the lifetime of a particular brand of lightbulb

Probability Model

Study of probability begins with a basic model for random experiments. The first two components of the model are:

1. **Sample space** Denoted Ω, this is just the set of possible outcomes

2. **Events** An event is a (sometimes any) subset of the sample space

Terminology: An event $A \subseteq \Omega$ **occurs** if outcome of the experiment is in A.

Ex: Three Flips

1. **Sample space:** Let H $=$ Heads, T $=$ Tails. Then

$$\Omega = \{\text{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH}\}$$

Note that $|\Omega| = 2 \times 2 \times 2 = 8$.

2. **Some events**

- No Heads $= \{\text{TTT}\}$
- Number of Heads equals Number of Tails $= \emptyset$
- First flip is Heads $= \{\text{HHH, HTH, HHT, HTT}\}$
- At least two Heads $= \{\text{HHT, HTH, THH, HHH}\}$
- All flips the same $= \{\text{HHH, TTT}\}$

Ex: Two Rolls

1. Sample space: Each roll shows $1, 2, 3, 4, 5$, or 6 dots. Then

$$\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$$

Here $i =$ outcome of first roll, $j =$ outcome of second roll. $|\Omega| = 6 \times 6 = 36$.

Some events

- First roll is 1 $= \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$
- Both rolls the same $= \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$
- Sum of rolls is 9 $= \{(3, 6), (6, 3), (4, 5), (5, 4)\}$
- Product of rolls is 8 $= \{(2, 4), (4, 2)\}$

Final Component of Probability Model

Probability measure $P(\cdot)$: For each event $A \subseteq \Omega$

$$P(A) \;=\; \text{probability that the outcome of the experiment lies in } A$$

Probability measures satisfy the following **axioms**

1. $P(A) \geq 0$ for every event $A \subseteq \Omega$ (non-negative)
2. $P(\Omega) = 1$ (sample space has probability one)
3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$ (additivity)

Basic Idea: If we think of Ω as a figure in the plane with unit area, then $P(A)$ is like the area of A.

Elementary Properties of Probabilities

1. If $A \subseteq B$ then $P(A) \leq P(B)$.

2. For every event A, the probability $P(A)$ is between 0 and 1.

3. $P(A^c) = 1 - P(A)$

4. $P(A \cup B) \; = \; P(A) + P(B) - P(A \cap B) \; \leq \; P(A) + P(B)$

5. $P(\cup_{i=1}^n A_i) \; \leq \; \sum_{i=1}^n P(A_i)$

Meaning/Interpretations of Probability

1. Objective/Frequentist: Long run relative frequency

If we repeat the basic experiment many times then for any event A

$$\frac{\text{\# times } A \text{ occurs}}{\text{total \# repetitions}} \approx P(A)$$

Problem: How to interpret probability of singular events, e.g.,

- The Buffalo Bills win next year's Superbowl
- The next president of the U.S. is a woman

2. Subjective/Bayesian: Degree of belief or evidentiary support

Important: Mathematical truth of theorems of probability theory does not depend on interpretation.

Uniform Probability Model

Consider: experiment with finite sample space $\Omega = \{\omega_1, \ldots, \omega_m\}$

Uniform probability model: All outcomes are equally likely, i.e.,

$$P(\omega_j) = \frac{1}{|\Omega|} \text{ for } 1 \leq j \leq m$$

In this case, the axioms ensure that for each event $A \subseteq \Omega$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{\# outcomes in } A}{\text{total \# outcomes}}$$

so we can find probabilities by counting outcomes.

Ex: Three Flips

Recall $\Omega = \{$TTT, TTH, THT, HTT, THH, HTH, HHT, HHH$\}$ and $|\Omega| = 8$

Consider events

$A =$ First flip is Heads $= \{$HHH, HTH, HHT, HTT$\}$

$B =$ Second flip is Heads $= \{$HHH, THH, HHT, THT$\}$

$A \cap B =$ First two flips are Heads $= \{$HHH, HHT$\}$

Under the uniform probability model $P(\cdot)$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{8} = \frac{1}{2} \qquad P(B) = \frac{|B|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}$$

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{2}{8} = \frac{1}{4}$$

Independence

Definition: Events A and B in a random experiment are **independent** if

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

Note:

- Independence is different from disjointness
  - A and B are disjoint if they *don't* overlap
  - A and B are independent if they overlap *just the right amount*
- Independence *depends on the probability* $P(\cdot)$.

Ex: Three Flips

$A =$ First flip is Heads $=$ {HHH, HTH, HHT, HTT}

$B =$ Second flip is Heads $=$ {HHH, THH, HHT, THT}

$C =$ At least two Heads $=$ {HHT, HTH, THH, HHH}

Under uniform probability $P()$,

$$P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A \cap B)$$

Thus A and B are independent. By contrast,

$$P(A \cap C) = \frac{|A \cap C|}{8} = \frac{3}{8} \neq P(A) \cdot P(C) = \frac{1}{4}$$

Thus A and C are not independent.

Ex: Three Flips, continue

Let events A, B, C be as before, and consider the probability P' such that

$$P'(HHH) = P'(TTT) = 1/2$$

and all other outcomes have probability zero.

The probability $P'()$ models an experiment in which the outcomes are **completely dependent**. Either all flips are H or all flips are T.

Under $P'()$ the events A and C are not independent. In fact,

$$P(A) = P(B) = P(A \cap B) = 1/2$$

Conditional Probability

Definition: Let A and B be events with $P(B) > 0$. The **conditional probability** of A *given* B is defined by

$$P(A \mid B) \;=\; \frac{P(A \cap B)}{P(B)}$$

Notes:

- The vertical bar '$|$' is read as "given"
- B is the *conditioning event*; its probability goes in the denominator
- Interpretation of $P(A \mid B)$
  - What is the probability of A if we are told that B occurred?
  - The proportion of B occupied by A
  - The probability of A if Ω were replaced by B

Conditional Probability, cont.

Note: If A and B are independent, then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\,P(B)}{P(B)} = P(A)$$

and similarly, $P(B \mid A) = P(B)$.

Interpretation: Knowing that B occurred tells us nothing about the probability of A, and vice-versa.

In general, $P(A \mid B)$ can be less than, equal to, or greater than $P(A)$.

Example: Three Flips

Consider events

$A =$ First flip is H

$B =$ Second flip is T

$C =$ At least two Hs

Note: Under uniform probability P, events A, B are independent, so

$$P(A \mid B) = P(A)$$

But events A, C and events B, C are **not** independent. In fact,

$$P(C \mid A) > P(C) \qquad P(C \mid B) < P(C)$$

Independence in Inference

Problem: Estimate (unknown) number N of people killed in a conflict

Available data: Two lists of individuals killed in the conflict, prepared by independent observers.

- N_1 = number of individuals on first list
- N_2 = number of individuals on second list
- $N_{1,2}$ = number of individuals on both lists

Consider **proportions**

$$\hat{p}_1 = \frac{N_1}{N} = \text{proportion of casualties on first list}$$

$$\hat{p}_2 = \frac{N_2}{N} = \text{proportion of casualties on second list}$$

$$\hat{p}_{1,2} = \frac{N_{1,2}}{N} = \text{proportion of casualties on both lists}$$

Independence in Inference, cont.

Key Ideas

- Proportions $\hat{p}$ are estimates of true probabilities p
- If the lists are collected independently, then $p_{1,2} = p_1 \cdot p_2$.

Substituting estimates $\hat{p}$ for true probabilities p we find

$$\frac{N_{1,2}}{N} = \hat{p}_{1,2} = \hat{p}_1 \cdot \hat{p}_1 = \frac{N_1}{N} \cdot \frac{N_2}{N}$$

Rearranging terms gives estimate of number of fatalities

$$N \approx \frac{N_1\, N_2}{N_{1,2}}$$

Conditional Probability, cont.

Chain Rule: Rewriting definition of conditional probability gives

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

Law of Total Probability: Note that $B = (B \cap A) \cup (B \cap A^c)$. Then we have

$$\begin{aligned} P(B) &= P((B \cap A) \cup (B \cap A^c)) \quad \text{[by the identity above]} \\ &= P(B \cap A) + P(B \cap A^c) \quad \text{[as } B \cap A \text{ and } B \cap A^c \text{ are disjoint]} \\ &= P(B \mid A)\, P(A) + P(B \mid A^c)\, P(A^c) \quad \text{[by the chain rule]} \end{aligned}$$

Bayes's Formula

Fact: If A and B are events with $P(A), P(B) > 0$ then

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B \mid A)\, P(A) + P(B \mid A^c)\, P(A^c)}$$

Proof: By definition

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Apply chain rule to the numerator, law of total probability to the denominator.

Interpretation

- $P(A)$ is the **prior** probability of A
- $P(A \mid B)$ is the **posterior** probability of A *after* we know B occurred
- Think of A as unobserved "cause" and B as observed "effect"
- Formula tells us how to update prior probability of A given B

Example: Guessing on Multiple Choice Question

Multiple choice question with 5 possible answers.

If a student knows the correct answer, they choose it; if not, they choose one of the answers at random (with equal probability).

Suppose that a student knows the correct answer with probability .70. If they answer the question correctly, how likely is it that they were guessing?

Cause (unobserved) $A =$ student knows the correct answer

Effect (observed) $B =$ student answers question correctly

Multiple Choice Question, cont.

We are given

- $P(A) = .70$ so $P(A^c) = 1 - P(A) = .30$
- $P(B \mid A) = 1$ and $P(B \mid A^c) = 1/5 = .20$

Applying Bayes formula gives

$$\begin{aligned} P(A \mid B) &= \frac{P(B \mid A)\, P(A)}{P(B \mid A)\, P(A) + P(B \mid A^c)\, P(A^c)} \\ &= \frac{1 \times .70}{1 \times .70 + .20 \times .30} \\ &\approx .92 \end{aligned}$$

Boys and Girls

Family with two children: possible sexes of first, second children are

$$\{(M, M), (M, F), (F, M), (F, F)\}$$

Define events

$A =$ First child a girl

$B =$ Second child a girl

$C =$ At least one girl

Suppose all four possibilities are equally likely. Find

$$P(B \mid A)$$

$$P(B \mid C)$$

Note that $P(C \mid B) = 1$ so $P(C \mid B) \neq P(B \mid C)$. Order of events matters!