Lecture 2: Multivariate Probability

UNIVERSITY OF SAN FRANCISCO

James D. Wilson

MSAN 628

Computational Statistics

Plan for this Lecture

- Probability Models
- Axioms of Probability
- Law of Total Probability and Bayes Rule

Introduction

- Today we will start by going on a formal excursions into probability.
- First need to fix our ideas regarding possible outcomes in the context of uncertainty and mathematical operations in the light of such uncertainty.
- We'll go a little quickly through this material as this should mostly be review. I want to make sure we are all on the same page!

Introduction

Real world motivation: Cognitive biases

- Make it very hard for human brain to reason and make decisions in the light of uncertainty
- People use simple heuristic principles which lead to severe and systematic errors.
- Two ways of think:
  - **I. Fast** (intuitive)
  - **II. Slow** (reasoning)
- We think we are in state II most of the time (at least in MSAN!) but most of the time we are in state I.

Example

Example

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

Consider the following statements about Linda:

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

Which do you think is more probable?

Probability Models

Setting: An experiment is performed or an observation is made where outcomes are *uncertain* or *non-deterministic*

Examples:

- Daily precipitation in San Francisco
- How many students in the class are paying attention right now
- The sex of a newborn child
- Whether or not you'll like this class
- Whether or not you'll receive an email from David "reinforcing" your attendance, participation, and interviewing habits over the next 3 days.

Probability Models

A probability model is used to quantify the likelihood of such an experimental or observational study. Consists of two main components:

1. Sample space - enumerates all possible outcomes
2. Probability distribution / measure - quantifies the likelihood of subsets of possible outcomes

Sample space

The sample space associated with a family of experiments or observations is the set of all possible outcomes.

Notation: S.

Sample Space Examples

Examples

Ex. 1: Experiment = determining the sex of a newborn

☼ $S =$

Ex. 2: Experiment = flipping 2 coins

☼ $S =$

Ex. 3: Experiment = observe coin flips until the first H appears

☼ $S =$

Ex. 4: Experiment = measuring lifetime (in hours) of a bulb

☼ $S =$

Events

Event

An event is any subset of the sample space S. These are often described in words.

Notation: A, B, C, D, E, etc. (capitals)

Examples

Ex. 1: $E = \{$1st coin of three tosses is $H\}$ =

Ex. 2: E = amongst 4 runners (A, B, C, D), B won the race =

Ex. 3: E = number of coin flips to observe a head is at least 3 =

Event Operations

Given two events E, F.

Union of events

The union of E and F = event consisting of all outcomes that are either in E, or in F, or in both E and F.

Notation: $E \cup F$.

Terminology: $E \cup F$ occurs = E or F occurs.

Example

$E = \{$1st coin of two tosses is $H\} = \{(H,T),(H,H)\}$

$F = \{$outcomes of 1st and 2nd coins are different$\} = \{(H,T),(T,H)\}$

$E \cup F =$

Event Operations

Intersection

The intersection of E and F = event consisting of all outcomes that are both in E **and** F.

Notation: $E \cap F$.

Terminology: $E \cap F$ occurs = E and F occur.

Continuing the previous example

$E \cap F =$

Null Events and Mutually Exclusive

Null events

A null event: event that cannot occur on the given sample space.

Notation: $\varnothing$.

If $E \cap F = \varnothing$, then E and F are called mutually exclusive (disjoint) and cannot occur together.

Event Operations

Extensions to finitely many events

Events $E_1, E_2, \ldots, E_n$.

Union of E_n's, $1 \leq j \leq n$ = event of outcomes that are in E_j for at least one of the $1 \leq j \leq n$

Notation: $\cup_{j=1}^{n} E_j$

Intersection of E_n's, $n \geq 1$ = event of outcomes that are in E_n for all $1 \leq j \leq n$

Notation: $\cap_{j=1}^{n} E_j$

Event Operations

Extension to countably many events:

Consider events $E_1, E_2, \ldots$

Union of E_n's, $n \geq 1$ = event of outcomes that are in E_n for at least one $n \geq 1$

Notation: $\cup_{n=1}^{\infty} E_n$

Intersection of E_n's, $n \geq 1$ = event of outcomes that are in E_n for all $n \geq 1$

Notation: $\cap_{n=1}^{\infty} E_n$

Complements

Complement

The complement of E = event consisting of all outcomes of S that are not in E.

Notation: E^c

Terminology: E^c occurs = E does not occur.

Note: $S^c = \varnothing$.

Example continued

$E^c =$

Event Comparison

Comparison

"E is contained in F" = all outcomes in E are also in F.

Notation: $E \subset F$ or $F \supset E$

Terminology: If E occurs, then F occurs as well.

$E = F$ means $E \subset F$ and $F \subset E$ (important proof technique).

Laws for operations

Set Operations

(a) Commutative laws:

- $E \cup F = F \cup E$
- $E \cap F = F \cap E$

(b) Associative laws:

- $(E \cup F) \cup G = E \cup (F \cup G)$
- $(E \cap F) \cap G = E \cap (F \cap G)$

(c) Distributive laws:

- $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$
- $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

De Morgan's laws[a]

[a] Augustus De Morgan (1806-1871), British mathematician and logician

$$\left(\bigcup_{n=1}^{k} E_n\right)^c = \bigcap_{n=1}^{k} E_n^c, \quad \left(\bigcap_{n=1}^{k} E_n\right)^c = \bigcup_{n=1}^{k} E_n^c$$

De Morgan's laws in words

Example: E = Superman saved Metropolis, F = Batman saved Gotham city.

(i) $E \cup F$ in words =

(ii) $(E \cup F)^c$ in words =

(iii) $(E \cup F)^c$ =

How can we introduce and think of probabilities of events?

Natural to think: repeat the experiment n times under same conditions; let $n(E)$ be the number of times event E occurs in these n repetitions; think

$$\lim_{n\to\infty} \underbrace{\frac{n(E)}{n}}_{\text{frequency of } E \text{ in } n \text{ rep.}} = \underbrace{\mathbb{P}(E)}_{\text{probability of } E} \qquad (*)$$

Main idea: The probability of an event E measures the likelihood of E occurring in the *long* run. **Note**: $(*)$ is not a definition.

Axiomatic approach to Probability Theory

Start with simpler, more evident assumptions; then show $(*)$. In particular, one can continue thinking of probability through $(*)$.

Axioms of Probability

Let S be a sample space, and $\mathbb{P}(E)$ the probability of event $E \subset S$.

Assign $\mathbb{P}(E)$'s so that the following axioms hold:[a]

[a] Andrey Kolmogorov (1903-1987), Russian mathematician, 1930's.

Axiom 1: $0 \leq \mathbb{P}(E) \leq 1$ for any event E.

Axiom 2: $\mathbb{P}(S) = 1$.

Axiom 3: for mutually exclusive (disjoint) events $E_1, E_2, \ldots$:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

Axioms of Probability

Consequence 1:

$\mathbb{P}(\varnothing) = 0.$

Consequence 2:

If $E_1, E_2, \ldots, E_n$ are mutually exclusive (disjoint), then

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{i=1}^{n} \mathbb{P}(E_i).$$

In particular, for mutually exclusive A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Axioms of Probability

Ex. Constructing a Probability Model

(i) $S = \{x_1, x_2, \ldots, x_N\}$.

e.g. $S = \{1, 2, 3, 4, 5, 6\}$ = {outcomes in rolling a die}.

(ii) Assign $\mathbb{P}(\{x_i\}) = p_i$, $p_i \in (0, 1)$ with $p_1 + p_2 + \ldots + p_N = 1$.

e.g. $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$ above.

(iii) For any event E, assign

$$\mathbb{P}(E) = \sum_{i:x_i \in E} \mathbb{P}(\{x_i\}) = \sum_{i:x_i \in E} p_i.$$

One can show that all three axioms satisfied with this model.

Axioms of Probability

Example

You have an unfair dice such that

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = 1/4$$

and

$$\mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) = 1/12$$

What is the chance that when you throw the dice, you get an even number?

Axioms of Probability

Probability Model

Terminology: $(S, \text{events}, \mathbb{P})$ or $(S, \mathbb{P})$ (with $\mathbb{P}$ satisfying Axioms 1, 2, and 3) is called a probability space or probability model).

Notes:

1. Axioms will often be taken as rules, especially for equally likely outcomes.
2. In the context of Venn diagrams, one can think of probability as area (or mass).

Models for Equally Likely Events

Probability model where all outcomes equally likely

(i) Sample space $S = \{x_1, x_2, \ldots, x_N\}$.

(ii) $p_1 = p_2 = \ldots = p_N = \frac{1}{N}$ (equally likely outcomes), in which case

$$\begin{aligned}\mathbb{P}(E) &= \frac{\text{Number of outcomes } x_i \text{ in } E}{\text{Total number of outcomes } x_i \text{ in } S} \\ &= \frac{\text{Number of outcomes } x_i \text{ in } E}{N}.\end{aligned}$$

(iii) **Note**: We will often use combinatorics to compute the numbers in the numerator and denominator.

Axioms of Probability

Example

In a small town with 900 adults, there are 600 Democrats (say $D_1, D_2, \ldots, D_{600}$) and 300 Republicans (say $R_1, R_2, \ldots R_{300}$). You want to sample 2 people from this population and record the political affiliation of each. To do this you fill a box with the names of the people and then pick 2 pick names at random without replacement. What is the probability model?

Rules of Probability

Proposition 1:

$\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.

Proposition 2:

If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.

Proposition 3

For any two events E, F,

$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

Conditional Probability and Bayesian Thinking

Basic point

- Think of probability as the amount of belief we have in an outcome.
- If we are rational, we update our prior beliefs in light of this data.
- In the world of probability, *updating* your beliefs refers to *conditioning* your current belief on previously observed outcomes.
- This is the crux of Bayesian analysis.

Extreme Example

Suppose your friend fills up a box with 10 balls of two type: Red and White. He/she does not tell you the colors but says they are either 2 Red and 8 Black or 8 Red and 2 black.

- Since you do not quite know which she/he did, your a natural choice for the “prior probability” is $1/2$ and $1/2$
- You pick three balls **without replacement** and find all 3 of them are red.
- What is your best guess of what your friend? With what certainty can you say this?

History Time: Bayes Theorem and Conditioning

- Goes back to Reverend Thomas Bayes in the 1740s.
- Amateur mathematician (owing to religious persecution was barred from English universities).
- His motivation: “Theological”.
- Discovered independently and developed in a tremendous number of applications by Laplace, the “Newton of France”.
- Subsequent uses of just Bayes theorem:
    1. Alan Turing and cracking of the “ENIGMA” (German navy secret code).
    2. Cold War: Finding missing H-bomb, nuclear subs etc.
    3. Bayesian spam filters, computer science, Google, neuroscience (adaptive brains).

History

Figure: Thomas Bayes

- Bayes never actually published his results.
- Famous amateur mathematician (Price) published his mathematical work posthumously.
- The form we use formulated by Laplace one of the most powerful mathematicians in the history of science.

Bayesian Thinking

Punch line

$$\text{Use: } \underbrace{\text{Initial Belief}}_{\text{Prior probability}} \text{ and } \text{Observed Data} \Longrightarrow \underbrace{\text{New improved Belief}}_{\text{Posterior Probability}}$$

Major road blocks initially

1. No particular reason to assume some specific initial belief then Laplace/Bayes suggested all possibilities equally likely.
2. Caused major complaints: “subjectivity”

Conditional Probability

Conditional Probability

Probability that E occurs given that F has occurred.

Name: conditional probability of E given F.

Notation: $\mathbb{P}(E|F)$.

Example

Ex. 1: Toss 2 dice; $E = \{\text{first die is } 3\}$ and $F = \{\text{sum is } 8\}$. What is $\mathbb{P}(E|F)$?.

Ex. 2: blood test for a disease; false positive can happen; $E = \{\text{person has disease}\}$ and $F = \{\text{test is positive}\}$. What is $P(E|F)$?.

Conditional Probability

Interpretation:

Repeat experiment n times; treat $\mathbb{P}(E|F)$ as the long-run proportion of times that E occurs when F occurs, that is,

$$\mathbb{P}(E|F) = \frac{\text{number of times } E \cap F \text{ occurs in } n \text{ rep.}}{\text{number of times } F \text{ occurs in } n \text{ rep.}}$$

For large n, this can be thought as

$$\frac{n\mathbb{P}(E \cap F)}{n\mathbb{P}(F)} = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Conditional Probability

If $\mathbb{P}(F) > 0$,

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Conditional Probability

Example

Two cards are randomly chosen without replacement from an ordinary deck of 52 cards. Let B be the event that both cards are aces and let A be the event that at least one ace is chosen. Find $\mathbb{P}(B|A)$.

☼

Conditional Probability

Computing probabilities via conditioning

$$\mathbb{P}(E \cap F) = \mathbb{P}(F)\frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \mathbb{P}(F)\mathbb{P}(E|F)$$

More generally:

The multiplication rule

$$\mathbb{P}(\bigcap_{j=1}^{n} E_j) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_1 \cap E_2)\cdots\mathbb{P}(E_n|E_1 \cap \ldots \cap E_{n-1})$$

Conditional Probability

Example

A recent college graduate is planning to write her actuarial exams. She will write the first exam in June (which she has a .9 probability of passing) and then will write the second exam in July. Given that she passed the 1st exam, the probability that she passes the second exam is .8. What is the probability that she passes both exams?

Motivation: Conditional probability and medicine

http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Doctors were asked: “The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram. Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?”

“When Gigerenzer asked 24 other German doctors the same question, their estimates whipsawed from 1 percent to 90 percent. Eight of them thought the chances were 10 percent or less, 8 more said 90 percent, and the remaining 8 guessed somewhere between 50 and 80 percent. Imagine how upsetting it would be as a patient to hear such divergent opinions.”

Correct answer: 9%

Conditional Probability

Important result:

$$\mathbb{P}(E) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)\mathbb{P}(F^c)$$

More generally,

The Law of Total Probability

Suppose $S = \cup_{i=1}^{n} F_i$, where F_i are mutually disjoint. Then,

$$\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E \cap F_i) = \sum_{i=1}^{n} \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

Bayes Rule

Expressing $\mathbb{P}(F|E)$ in terms of $\mathbb{P}(E|F)$: Note that

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(F \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F)\mathbb{P}(F)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F)\mathbb{P}(F)}{\mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)\mathbb{P}(F^c)}.$$

More generally:[1]

Bayes Rule

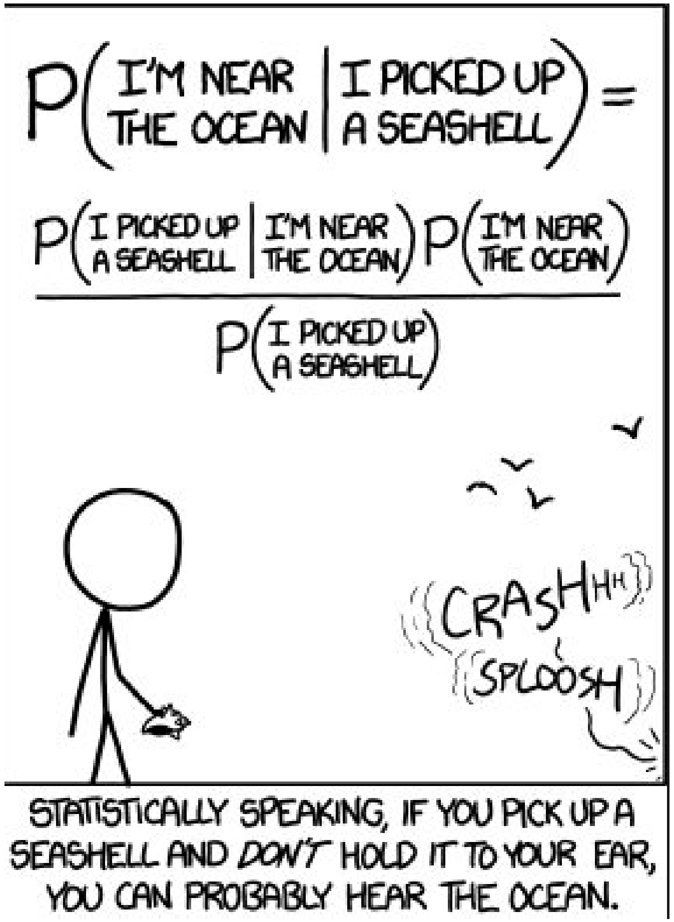
Suppose $S = \cup_{i=1}^{n} F_i$, where F_i are mutually disjoint. Then,

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_{i=1}^{n} \mathbb{P}(E|F_i)\mathbb{P}(F_i)}.$$

So, Bayes Rule is just a direct consequence of the law of total probability!

[1] Thomas Bayes (1701-1761), English mathematician

Exactly...


P(I'M NEAR THE OCEAN | I PICKED UP A SEASHELL) =
P(I PICKED UP A SEASHELL | I'M NEAR THE OCEAN) P(I'M NEAR THE OCEAN)
P(I PICKED UP A SEASHELL)
CRASHHH
SPLOOSH
STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

Independence

Independent events I:

Events E and F are independent if

$$\mathbb{P}(E|F) = \mathbb{P}(E), \quad \mathbb{P}(F|E) = \mathbb{P}(F). \tag{1}$$

In other words, *knowing that one of them occurs does not change the probability that the other occurs.*

Independent events II:

Equation (1) gave one definition of independence. It turns out this is equivalent to the following: Events E and F are independent if

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F).$$

Why is this true?

Independence

Example

You select a card randomly from a deck. Let E be the event that it is a ♣ and F be the event it is a 6. Are these two events independent?

☼

Independence and Mutual Independence

Independence of n events

Events $E_1, E_2, \ldots, E_n$ are independent if, for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r' \leq n$:

$$\mathbb{P}(E_{1'} \cap E_{2'} \cap \ldots \cap E_{r'}) = \mathbb{P}(E_{1'})\mathbb{P}(E_{2'}) \ldots \mathbb{P}(E_{r'}).$$

Mutual Independence of n events

Events $E_1, E_2, \ldots, E_n$ are mutually independent if

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\mathbb{P}(E_j)$$

for all $i \neq j$.