Math Camp

Justin Grimmer

Professor
Department of Political Science
Stanford University

September 13th, 2018

Where are we going?

Probability Theory:

1) Mathematical model of uncertainty
2) Foundation for statistical inference
3) Continues our development of key skills
   - Proofs [precision in thinking, useful for formulating arguments]
   - Statistical computing [basis for much of what you'll do in graduate school]

Model of Probability

Three parts to our probability model

1) Sample space: set of all things that could happen
2) Events: subsets of the sample space
3) Probability: chance of an event

Sample Spaces: All Things that Can Happen

Definition

The *sample space* *as the set of all things that can occur. We will collect all distinct outcomes into the set* S

Known perfectly

Examples:

1) House of Representatives: Elections Every 2 Years
   - One incumbent: $S = \{W, N\}$
   - Two incumbents: $S = \{(W, W), (W, N), (N, W), (N, N)\}$
   - 435 incumbents: $S = 2^{435}$ possible outcomes
2) Number of countries signing treaties
   - $S = \{0, 1, 2, \ldots, 194\}$
3) Duration of cabinets
   - All non-negative real numbers: $[0, \infty)$
   - $S = \{x : 0 \leq x < \infty\}$

Key point: this defines all possible realizations

Events: Subsets of Sample Space

Definition

An event, E is a subset of the sample space.
$E \subset S$

Plain English: Outcomes from the sample space, collected in set

Congressional Election Example

- One incumbent:
  - $E =$ W
  - $F =$ N
- Two Incumbents:
  - $E = \{(W, N), (W, W)\}$
  - $F = \{(N, N)\}$
- 435 Incumbents:
  - Outcome of 2010 election: one event
  - All outcomes where Dems retain control of House: one event

Notation: x is an "element" of a set E:
$x \in E$

Events: Subsets of Sample Space

E is a set: collection of distinct objects.
Recall three operations on sets (like E) to create new sets:
Consider two example sets (from two incumbent example):

$$E = \{(W, W), (W, N)\}$$
$$F = \{(N, N), (W, N)\}$$
$$S = \{(W, W), (W, N), (N, W), (N, N)\}$$

Operations determine what lies in new set E^{new}

1) Union: $\cup$
   - All objects that appear in either set
   - $E^{\text{new}} = E \cup F = \{(W, W), (W, N), (N, N)\}$
2) Intersection: $\cap$
   - All objects that appear in both sets
   - $E^{\text{new}} = E \cap F = \{(W, N)\}$
   - Sometimes written as EF

3) Complement of set E: E^c

- All objects in S that aren't in E
- $E^c = \{(N, W), (N, N)\}$
- $F^c = \{(N, W), (W, W)\}$
- $S = \Re$ and $E = [0, 1]$. What is E^c?
- What is S^c? $\emptyset$

Suppose $E = W$, $F = N$. Then $E \cap F = \emptyset$ (there is nothing that lies in both sets)

Events: Subsets of Sample Space

Definition

Suppose E and F are events. If $E \cap F = \emptyset$ then we'll say E and F are mutually exclusive

- Mutual exclusivity $\neq$ independence
- E and E^c are mutually exclusive events

Examples:

- Suppose $S = \{H, T\}$. Then $E = H$ and $F = T$, then $E \cap F = \emptyset$
- Suppose $S = \{(H, H), (H, T), (T, H), (T, T)\}$. $E = \{(H, H)\}$, $F = \{(H, H), (T, H)\}$, and $G = \{(H, T), (T, T)\}$
  - $E \cap F = (H, H)$
  - $E \cap G = \emptyset$
  - $F \cap G = \emptyset$
- Suppose $S = \Re_{+}$. $E = \{x : x > 10\}$ and $F = \{x : x < 5\}$. Then $E \cap F = \emptyset$.

Events: Subsets of the Sample Space

Definition

Suppose we have events $E_1, E_2, \ldots, E_N$*.*
Define:

$$\cup_{i=1}^{N} E_i \quad = \quad E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_N$$

$\cup_{i=1}^{N} E_i$ *is the set of outcomes that occur at least once in* $E_1, \ldots, E_N$*.*
Define:

$$\cap_{i=1}^{N} E_i \quad = \quad E_1 \cap E_2 \cap \ldots \cap E_N$$

$\cap_{i=1}^{N} E_i$ *is the set of outcomes that occur in each* E_i

Probability

1) Sample Space: set of all things that could happen
2) Events: subsets of sample space
3) Probability: chance of event
    - P is a function
    - Domain: all events E

Probability

Definition

All probability functions, P, satisfy three axioms:

1) *For all events E,*
   $0 \leq P(E) \leq 1$
2) $P(S) = 1$
3) *For all sequences of mutually exclusive events $E_1, E_2, \ldots, E_N$ (where N can go to infinity)*
   $P\left(\cup_{i=1}^N E_i\right) = \sum_{i=1}^N P(E_i)$

Probability

- Suppose we are flipping a fair coin. Then $P(H) = P(T) = 1/2$
- Suppose we are rolling a six-sided die. Then $P(1) = 1/6$
- Suppose we are flipping a pair of fair coins. Then $P(H, H) = 1/4$

Example: Congressional Elections

One candidate example:

- $P(W)$: probability incumbent wins
- $P(N)$: probability incumbent loses

Two candidate example:

- $P(\{W, W\})$: probability both incumbents win
- $P(\{W, W\}, \{W, N\})$: probability incumbent 1 wins

Full House example:

- $P(\{\text{All Democrats Win}\})$ (Cox, McCubbins (1993, 2005), Party Brand Argument)

We'll use data to infer these things

Properties of Probability

We can derive intuitive properties of probability theory. Using just the axioms

Proposition

$P(\emptyset) = 0$

Proof.

Define $E_1 = S$ and $E_2 = \emptyset$,

$$
\begin{aligned}
1 = P(S) = P(S \cup \emptyset) &= P(E_1 \cup E_2) \\
1 &= P(E_1) + P(E_2) \\
1 &= P(S) + P(\emptyset) \\
1 &= 1 + P(\emptyset) \\
0 &= P(\emptyset)
\end{aligned}
$$

$\square$

Properties of Probability

Proposition

$P(E) = 1 - P(E^c)$

Proof.

Note that, $S = E \cup E^c$. And that $E \cap E^c = \emptyset$. Therefore,

$$\begin{aligned} 1 = P(S) &= P(E \cup E^c) \\ 1 &= P(E) + P(E^c) \\ 1 - P(E^c) &= P(E) \end{aligned}$$

□

In words: Probability an outcome in E happens is $1-$ probability an outcome in E doesn't.

Properties of Probability

Proposition

If $E \subset F$ then $P(E) \leq P(F)$.

Proof.

We can write $F = E \cup (E^c \cap F)$. (Why?)
Further, $(E^c \cap F) \cap E = \emptyset$
Then
$P(F) = P(E) + P(E^c \cap F)$ (Done!) $\square$

As you add more "outcomes" to a set, it can't reduce the probability.

Examples in R

Simulation: use pseudo-random numbers, computers to gain evidence for claim

Tradeoffs:

Pro Deep understanding of problem, easier than proofs

Con Never as general, can be deceiving if not done carefully (also, never a monte carlo study that shows a new method is wrong)

Walk through `R` code to simulate these two results

To the `R` code!

4.2. *Three different combination rules were used.* We then tried to identify the rules used to combine individual drug predictions into a combination score. Letting P() indicate probability of sensitivity, the rules used are:

$$
\begin{aligned}
P(TFAC) &= P(T) + P(F) + P(A) + P(C) - P(T)P(F)P(A)P(C), \\
P(TET) &= P(ET) = \max[P(E), P(T)], \text{ and}
\end{aligned}
$$

5 1

Inclusion/Exclusion

Proposition

Suppose $E_1, E_2, \ldots, E_n$ *are events. Then*

$$
\begin{aligned}
P(E_1 \cup E_2 \cup \cdots \cup E_n) &= \sum_{i=1}^{N} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \cdots \\
&+ (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) \\
&+ \cdots + (-1)^{n+1} P(E_1 \cap E_2 \cap \cdots E_n)
\end{aligned}
$$

Proof: Version 1, Intuition

- Suppose that we have an outcome.
- If it isn't in the event sequence, doesn't appear anywhere.
- If it is in the event sequence, appears once in $\cup_{i=1}^{n} E_i$ (contributes once to $P(\cup_{i=1}^{n} E_i)$.
- How many times on the other side? Suppose it appears in m of the E_i $m > 0$

$$
\begin{aligned}
\text{count} &= \binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \cdots + (-1)^{m+1}\binom{m}{m} \\
\text{count} &= \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \\
\text{count} &= -\sum_{i=1}^{m} \binom{m}{i} (-1)^{i}
\end{aligned}
$$

Proof: Version 1, intuition

count $= -\sum_{i=1}^{m} \binom{m}{i}(-1)^i$
Binomial Theorem: $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i}(x)^{n-i}y^i$.

$$\begin{aligned}
0 = (-1+1)^m &= \sum_{i=0}^{m} \binom{m}{i}(-1)^i \\
0 &= 1 + \sum_{i=1}^{m} \binom{m}{i}(-1)^i \\
0 &= 1 - \text{count} \\
1 &= \text{count}
\end{aligned}$$

Inclusion/Exclusion

Corollary

Suppose E_1 and E_2 are events. Then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

`R Code!`

Proposition

Consider events E_1 and E_2. Then

$$P(E_1 \cap E_2) \quad = \quad P(E_1) - P(E_1 \cap E_2^c)$$

Proof.

$$\begin{aligned} E_1 &= (E_1 \cap E_2) \cup (E_1 \cap E_2^c) \\ P(E_1) &= P(E_1 \cap E_2) + P(E_1 \cap E_2^c) \\ P(E_1 \cap E_2) &= P(E_1) - P(E_1 \cap E_2^c) \end{aligned}$$

□

Proposition

Boole's Inequality

$$P(\cup_{i=1}^{N} E_i) \leq \sum_{i=1}^{N} P(E_i)$$

Proof.

Proceed by induction. Trivially true for $n = 1$. Now assume the proposition is true for $n = k$ and consider $n = k + 1$.

$$P(\cup_{i=1}^{k} E_i \cup E_{k+1}) = P(\cup_{i=1}^{k} E_i) + P(E_{k+1}) - P(\cup_{i=1}^{k} E_i \cap E_{k+1})$$

$$P(E_{k+1}) - P(\cup_{i=1}^{k} E_i \cap E_{k+1}) \leq P(E_{k+1})$$

Proof Continued

$$
\begin{aligned}
P(\cup_{i=1}^{k} E_i) &\leq \sum_{i=1}^{k} P(E_i) \\
P(\cup_{i=1}^{k} E_i) + P(E_{k+1}) - P(\cup_{i=1}^{k} E_i \cap E_{k+1}) &\leq \sum_{i=1}^{k} P(E_i) + P(E_{k+1}) \\
P(\cup_{i=1}^{k+1} E_i) &\leq \sum_{i=1}^{k+1} P(E_i)
\end{aligned}
$$

Proposition

Bonferroni's Inequality

$$P(\cap_{i=1}^{n} E_i) \;\; \geq \;\; 1 - \sum_{i=1}^{n} P(E_i^c)$$

Proof.

$\cup_{i=1}^{n} E_i^c = (\cap_{i=1}^{n} E_i)^c$. So,

$$\begin{aligned}
P(\cup_{i=1}^{N} E_i^c) &\leq \sum_{i=1}^{N} P(E_i^c) \\
P(\cup_{i=1}^{N} E_i^c) &= P((\cap_{i=1}^{n} E_i)^c)) \\
&= 1 - P(\cap_{i=1}^{n} E_i) \\
P(\cap_{i=1}^{n} E_i) &\geq 1 - \sum_{i=1}^{n} P(E_i^c)
\end{aligned}$$

$\square$

Suprising Probability Facts

Formalized Probabilistic Reasoning: helps us to avoid silly reasoning

- "What are the odds" $\rightsquigarrow$ not great, but neither are all the other non-pattens that are missed
- "There is no way a candidate has a 80% chance of winning, the forecasted vote share is only 55%" $\rightsquigarrow$ confuses different events
- "Group A has a higher rate of some behavior, therefore most of the behavior is from group A" $\rightsquigarrow$ confuses two different problems (explain more tomorrow)
- "This is a low probability event, therefore god designed it" $\rightsquigarrow$ (1) Even if we stipulate to a low probability event, intelligent design is an assumption (2) Low probability obviously doesn't imply divine intervention. Take 100 balls and let them sort into an undetermined bins. You'll get a result, but probability of that result $= 1/(10^{29} \times \text{Number of Atoms in Universe})$

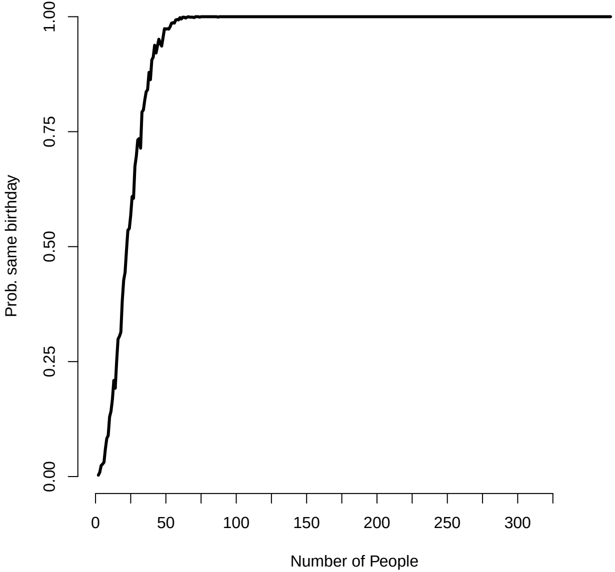
Easy Problems

Surprising Probability Facts:Birthday Problem

Probabilistic reasoning pays off for harder problems

Suppose we have a room full of N people. What is the probability at least 2 people have the same birthday?

- Assuming leap year counts, $N = 367$ guarantees at least two people with same birthday (pigeonhole principle)
- For $N < 367$?
- Examine via simulation


Prob. same birthday
0.00
0.25
0.50
0.75
1.00
0
50
100
150
200
250
300
Number of People

Surprising Probability Facts: the E-Harmony Problem

Curse of dimensionality and on-line dating:

eHarmony matches you based on compatibility in the most important areas of life - like values, character, intellect, sense of humor, and 25 other dimensions.

Suppose (for example) 29 dimensions are binary (0,1):
Suppose dimensions are independent:
Pr(2 people agree) $= 0.5$

$$
\begin{aligned}
\Pr(\text{Exact}) &= \Pr(\text{Agree})_1 \times \Pr(\text{Agree})_2 \times \ldots \times \Pr(\text{Agree})_{29} \\
&= 0.5 \times 0.5 \times \ldots \times 0.5 \\
&= 0.5^{29} \\
&\approx 1.8 \times 10^{-9}
\end{aligned}
$$

1 in 536,870,912 people
Across many "variables" (events) agreement is harder

Probability Theory

- Today: Introducing probability model
- Conditional probability, Bayes' rule, and independence