Lecture 2,

Tuesday, August 27

Introduction to Probability

Set Theory: Algebra (without proof)

- Algebraic Laws
  - commutativity:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

  - associativity:

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

  - distributive law:

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

  - DeMorgan's laws:

$x \in (A \cup B)^c \Leftrightarrow x \notin A$ & $x \notin B$

$\Leftrightarrow x \in A^c$ & $x \in B^c$

$\Leftrightarrow x \in A^c \cap B^c$

$\Rightarrow \begin{cases} (A \cup B)^c \subset A^c \cap B^c \\ A^c \cap B^c \subset (A \cup B)^c \end{cases}$

$$\boxed{(A \cup B)^c = A^c \cap B^c}$$
$$(A \cap B)^c = A^c \cup B^c$$

Introduction to Probability

Set Theory: Algebra (without proof), cont'd

- Extending the previous set laws:

  Let $A_1, A_2, \ldots$ be sets defined on a sample space S.

  The previous set results (last slide) can be generalized to finite unions $\bigcup_{i=1}^{n} A_i$, countably infinite unions $\bigcup_{i=1}^{\infty} A_i$, and unions $\bigcup_{b\in\Gamma} A_b$ over a possibly continuous index set Γ and similarly defined intersections.

e.g. define sets for real numbers $A_i = [i, i+1)$, $\forall i \geq 1$.

$$\bigcup_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ for some } i \geq 1\} = [1, \infty)$$

Define $B_i := (0, \frac{1}{i})$, $i \geq 1$.

$$\bigcap_{i=1}^{\infty} B_i = \{x : x \in B_i, \forall i, \text{ for any } i\}$$

$$= \{x : 0 < x < \tfrac{1}{i}, \forall i\} = \Phi$$

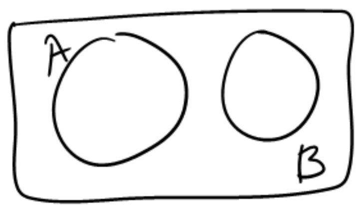
$$\bigcup_{i=1}^{\infty} B_i = (0, 1) \quad \text{(check!)}$$

Introduction to Probability

Set Theory: Partitions

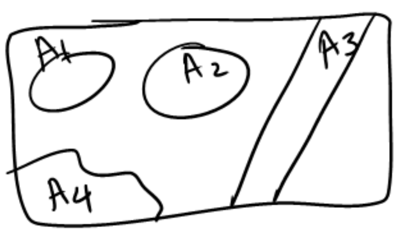
Definitions: Disjoint sets and partitions

- events A and B are **disjoint** (mutually exclusive) if $A \cap B = \emptyset$



- For a sequence $A_1, A_2, \ldots$ of events, we say $A_1, A_2, \ldots$ are **pairwise disjoint** if

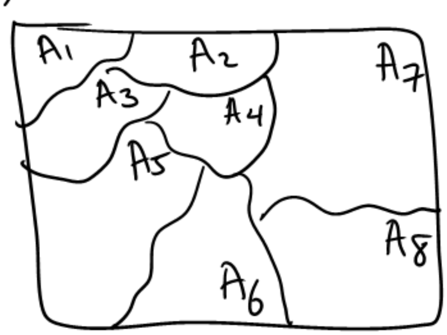
$$A_i \cap A_j = \emptyset \quad \text{for all } i \neq j$$



- $A_1, A_2, \ldots$ is a **partition** of S if the A_i's are pairwise ~~joint~~ disjoint and ~~exhaust S~~, that is

$$\underbrace{\bigcup_{i=1}^{\infty}}A_i = S \quad \& \quad A_i \cap A_j = \emptyset \quad \text{for all } i \neq j$$

ex:



Introduction to Probability

Thinking about probability

- Want to assign probabilities to events $A \subset S$ Sample space

- Interpretations of probability

  – limiting relative frequency

  – subjective belief

- For now, we'll ignore interpretation issues & focus on probability as a *set function* (no interpretation necessary)

ASIDE $P:(\Omega, \mathcal{F}, p) \to \mathbb{R}$

- A technical difficulty: It is *not* generally possible to assign probabilities to *every* possible subset of S.

  In other words, when one develops rules or ways of assigning probabilities to sets, technically one has to be careful to work with special collections of sets (or events), called σ-algebras or Borel fields.

  Aside from a brief description (next), we will not concern ourselves with the theory involved (measure theory or measure-theoretic probability, STAT 642).

$S = \{1,2,3,4,5,6\}$

subset/event: all outcomes where the result is a Prime number, and If you Concatenate the outcome with the next highest number, the result is divisible by 3.

2: Next number 3 $\to$ $23 \div 3$ $\to$ leaves a reminder, So do not include

3: Next number 4 $\to$ $34 \div 3$ $\to$ 〃 〃 〃 〃 〃

5: Next number 6 $\to$ $56 \div 3$ $\to$ 〃

ASIDE
STAT 641

Introduction to Probability

Borel fields

- *Definition:* A collection $\mathcal{B}$ of subsets of S is a **σ-algebra** or **Borel field** if its satisfies

  1. $\emptyset \in \mathcal{B}$.
  2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$.
  3. If a sequence of sets $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

- Want to assign a probability to each set $A \in \mathcal{B}$ (in a logically consistent way)

- For a given S, there can be many possible Borel fields.

  Suppose $S = \{a, b, c, d\}$ then here are three Borel fields associated with S

  - $\mathcal{B}_1 = \{\emptyset, S\}$ (trivial Borel field)
  - $\mathcal{B}_2 = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{d, c\}, \{b, c, d\}, \{a, c, d\}, S\}$
    (Borel field "generated" by $\{a\}$ and $\{b\}$)
  - $\mathcal{B}_3 = \{$all possible subsets of $S\}$

- For probability applications, choose $\mathcal{B}$ to contain all sets of practical interest

  1. for countable S: take $\mathcal{B}$ to contain all possible subsets
  2. for uncountable S: take $\mathcal{B}$ to contain intervals and everything derived from intervals (i.e., the Borel field "generated" by intervals)

Introduction to Probability

Historical side note on Borel fields & measure theory

- Probability was inspired by studies of gambling in the 1700's (e.g., Laplace, 1800's), but wasn't considered a branch of mathematics at first

- In 1920's, the *time waiting paradox* **threatened** the notion of "probability"
  - Problem: A person shows up a bus-stop at 9am and inter-arrival times between buses are random, independent, and on average 1 hour. How long does the person have to wait on average?
  - By *the standards of probability logic at that time*, there were two correct (but contradictory) answers: 30 minutes & 1 hour.

- This was a big existential dilemma... did "probability" have *any* foundation?

- Kolmogorov (1930's) rescued by probability theory by formulating it based upon ideas of "measure theory" → STAT 641

  i.e., probability is a measuring device (a function) on special sets of events (Borel fields)

  This gave a mathematical legitimacy and new clarity/rigor to probability, in ways that we take for granted.

- To re-iterate, we don't evoke measure theory and simply skip technical ideas like Borel fields. But, studying these later (e.g., STAT 642) can improve your understanding of probability and statistics (e.g., expand probability models)

Introduction to Probability

Axiomatic definition of probability (due to Kolmogorov)

- A **probability function** is a function P defined on a Borel field $\mathcal{B}$ of the sample space S that satisfies:

Axiom 1 1. $P(A) \geq 0$ for all $A \in \mathcal{B}$

Axiom 2 2. $P(S) = 1$

Axiom 3 3. If $A_1, A_2, \ldots \in \mathcal{B}$ are *pairwise disjoint* then

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

- Any function satisfying the above is a legitimate probability function.

Introduction to Probability

Axiomatic definition of probability, (cont'd)

- More than one probability function can be defined on a sample space S

e.g. rain "R" or not rain "NR" over next 2 days.

$S = \{RR, RN, NR, NN\}$

	Model 1	Model
P(RR)	1/4	0
P(NR)	1/4	3/5
P(RN)	1/4	0
P(NN)	1/4	2/5

- How should we decide which P to use?
    - physical considerations (e.g., model failure probabilities in engineering)
    - mathematical considerations (e.g., central limit theorem)
    - plausible assumptions (e.g., observations are dependent or independent)

Introduction to Probability

Properties of probability functions

Theorem 1.2.8. If P is a probability function and A is any set in $\mathcal{B}$, then

(a) $P(\emptyset) = 0$ $\quad P(\phi) \underset{\text{part c}}{=} 1 - P(\phi^c) = 1 - \underbrace{P(S)}_{} \underset{\text{axiom 2}}{=} 1 - 1 = 0$

(b) $P(A) \leq 1$ $\quad 0 \underset{\text{axiom 1}}{\leq} P(A^c) \underset{\text{part c}}{=} 1 - P(A) \Rightarrow \boxed{P(A) \leq 1}$

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

Proof of (c) (parts a, b follow from c and the axioms)

$$S = A \cup A^c, \quad 1 \underset{\text{axiom 2}}{=} P(S) = P(\underbrace{A \cup A^c}_{\text{disjoint}}) \underset{\text{axiom 3}}{=} P(A) + P(A^c)$$

$$\boxed{1 = P(A) + P(A^c)}$$

Theorem 1.2.9. If P is a probability function and A, B are sets in $\mathcal{B}$, then

(a) $P(B \cap A^c) = \boxed{P(B) - P(B \cap A)}$

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(c) if $A \subset B$, then $P(A) \leq P(B)$

Proof of (a) (parts b, c follow from a)

$$(a) \quad B = \underbrace{(B \cap A) \cup (B \cap A^c)}_{\text{disjoint}} \Rightarrow P(B) \overset{\text{axiom 3}}{=} P(B \cap A) + P(B \cap A^c)$$

$$\Rightarrow P(B \cap A^c) = P(B) - P(B \cap A).$$

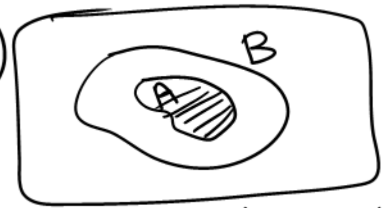
$$(b) \quad \underbrace{A \cup B}_{} = \underbrace{A \cup (B \cap A^c)}_{\text{disjoint}} \Rightarrow P(A \cup B) \underset{\text{axiom 2}}{=} P(A) + \underbrace{P(B \cap A^c)}_{}$$

$$\underset{\text{Part (a)}}{=} P(A) + P(B) - P(B \cap A).$$

$$A \cup (B \cap A^c) = (A \cup B) \cap \underbrace{(A \cup A^c)}_{S}$$

(c) $A \subset B \Rightarrow P(A) \leq P(B)$

$B = A \cup (B \cap A^c) \Rightarrow P(B) = P(A) + P(\underbrace{B \cap A^c}_{\geq 0})$



$\Rightarrow P(B) \geq P(A)$.

Introduction to Probability

Properties of probability functions (cont'd)

Bonferroni's Inequality:

Theorem 1.2.11. If P is a probability function, then

(a) $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \ldots \in \mathcal{B}$ (i.e., disjoint C_i's & $\bigcup_{i=1}^{\infty} C_i = S$)

(b) $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets $A_1, A_2, \ldots \in \mathcal{B}$

Proof of (b)

Principle of Inclusion-Exclusion: For any sets $A_1, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right)$$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n}^{n} P(A_i \cap A_j) + \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

This generalizes $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and is proven by induction.