STAT 712 fa 2022 Lec 1 slides

Set theory and basics of probability theory

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Experiment

An *experiment* is a process which generates an outcome such that there is

(i) more than one possible outcome
(ii) the set of possible outcomes is known
(iii) the outcome is not known in advance

Sample space

The *sample space* Ω of an experiment is the set of possible outcomes.

Call each point $\omega \in \Omega$ a *sample point*.

The symbol $\in$ denotes membership; $\omega \in \Omega$ means ω belongs to/is a member of Ω.

Exercise: Give the sample space Ω for the following experiments:

1. Roll of a 6-sided die.
2. Number of bacteria counted in sample.
3. Blood type of a randomly selected student.
4. Time until you drop your new phone and crack the screen.
5. Proportion of people in a sample with antibodies to a virus.
6. Deviation of today's temperature from the historical average.

Finite/infinite and countable/uncountable sets

- The *cardinality* of a set is the number of elements in the set.
- A set is *finite* if its cardinality is some positive integer.
- The empty set $\emptyset$, the set containing no elements, is finite with cardinality 0.
- Sets that do not have finite cardinality are *infinite* sets.
- A set is *countable* if it has the same cardinality as some subset of the positive integers.
- Sets that are not countable are called *uncountable*.

If a set is countable, we can list (or begin to list) its elements.

Event

An *event* is a collection of possible outcomes of an experiment, that is any subset of Ω (including Ω itself).

- Usually represent events with capital letters A, B, C,...
- Say an event A occurs if the outcome is in the set A.
- So events are equivalent to sets. Can refer to events as sets, to sets as events.

Example: Rolling a die has sample space $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$.

- $A = \{$⚀, ⚂, ⚄$\}$ is the event "you roll an odd number".
- $B = \{$⚅$\}$ is the event "you roll a 6".

Let A and B be events/sets in a sample space Ω.

Relationships between sets

- *Containment*:
- *Equality*:

The symbol $\subset$ means "is a subset of". Note the difference between $\subset$ and $\in$.

Elementary set operations

Let A and B be subsets of Ω.

- *Union set*: $A \cup B$ is the set of elements in A or B or in both.
- *Intersection set*: $A \cap B$ is the set of elements in both A and B.
- *Set subtraction*: $B \setminus A$ is the set of elements in B but not in A.
- *Complement set*: A^c is the set of elements in Ω but not in A, i.e. $A^c = \Omega \setminus A$.

Exercise: Write each of the above as a set $\{\omega \in \Omega : \dots\}$.

Unions and intersections of infinite collections of sets

For a collection of sets $A_1, A_2, \dots \subset \Omega$ we define

$\bigcup_{n=1}^{\infty} A_n =$

$\bigcap_{n=1}^{\infty} A_n =$

Theorem (Properties of the union and intersection operations)

For any events $A, B, C \subset \Omega$, we have

- *Commutativity:*
- *Associativity:*
- *Distributive laws:*
- *De Morgan's Laws:*

Exercise: Give proof of De Morgan's laws. In addition, verify the identities:

1. $A \setminus B = A \cap B^c$
2. $A = (A \cap B) \cup (A \cap B^c)$
3. $A \cup B = A \cup (B \cap A^c)$

Mutual exclusivity/disjoint-ness, partition

- Two events A and B are called *mutually exclusive* or *disjoint* if $A \cap B = \emptyset$.
- The events $A_1, A_2, \dots$ are called *mutually exclusive* or *pairwise disjoint* if $A_i \cap A_j = \emptyset$ for all $i \neq j$.
- If $A_1, A_2, \dots \subset \Omega$ are pairwise disjoint and $\bigcup_{n=1}^{\infty} A_n = \Omega$, then the collection of sets $A_1, A_2, \dots$ is called a *partition* of Ω.

Think of:

1. Partitions of sample space for the time until you drop your phone.
2. Partitions of sample space when drawing one card from a 52-card deck.
3. Partitions of the set $[0, 1)$.

We want to be able to assign probabilities to events. To which events?

σ-algebra

Let $\mathcal{B}$ be a collection of subsets of Ω. We call $\mathcal{B}$ a σ-algebra on Ω if it has the three properties:

1. $\emptyset \in \mathcal{B}$.
2. If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$.
3. If $A_1, A_2, \dots \in \mathcal{B}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Exercise: Determine which of the following are σ-algebras of $\Omega = \{a, b, c, d\}$:

- $\mathcal{B}_1 = \{\emptyset, \{a, b, c, d\}\}$
- $\mathcal{B}_2 = \{\{a\}, \{b, c, d\}, \emptyset, \{a, b, c, d\}\}$
- $\mathcal{B}_3 = \{\{a, b\}, \{a, b, c\}, \{d\}, \emptyset, \{a, b, c, d\}\}$
- $\mathcal{B}_4 = \{\text{all subsets of } \Omega, \text{ including } \Omega \text{ itself}\}$

Exercise: Consider the sample space

$$\Omega = [0, 1]$$

Add only the subsets required to make these collections into σ-algebras on Ω:

1. $\mathcal{B}_1 = \{\emptyset, \{1\}, [0, 1/2), \dots\}$
2. $\mathcal{B}_2 = \{[0, 1/2), [1/2, 3/4), \dots\}$

Theorem (Derived properties of σ-algebras)

Let $\mathcal{B}$ be a σ-algebra on a sample space Ω.

- $\Omega \in \mathcal{B}$.
- *If* $A_1, \dots, A_n \in \mathcal{B}$ *then* $\cup_{i=1}^{n} A_i \in \mathcal{B}$.
- *If* $A_1, \dots, A_n \in \mathcal{B}$ *then* $\cap_{i=1}^{n} A_i \in \mathcal{B}$.
- *If* $A_1, A_2, \dots \in \mathcal{B}$ *then* $\cap_{n=1}^{\infty} A_n \in \mathcal{B}$.
- *If* $A, B \in \mathcal{B}$ *then* $A \setminus B \in \mathcal{B}$.

Exercise: Prove each of the above.

Exercise: We sometimes restrict our attention to a subset of the sample space. Let $\mathcal{B}$ be a σ-algebra on a sample space Ω and let $C \in \mathcal{B}$. Show that the collection of sets $\mathcal{B}_C = \{C \cap A : A \in \mathcal{B}\}$ is a σ-algebra on C.

Borel σ-algebra on $\mathbb{R}$

- For $\Omega = (-\infty, \infty)$, we often consider the σ-algebra consisting of all countable unions and intersections of open intervals (a, b), $-\infty < a < b < \infty$.
- We call this the Borel σ-algebra on $\mathbb{R}$ and denote it by $\mathcal{B}(\mathbb{R})$.
- $\mathcal{B}(\mathbb{R})$ contains all sets of the form

$$[a, b], \quad [a, b), \quad (a, b], \quad (a, b), \quad -\infty \leq a < b \leq \infty$$

and any countable unions and intersections of such sets.

You will become better acquainted with $\mathcal{B}(\mathbb{R})$ in STAT 810 and STAT 811.

Probability function (Колмогоров axioms)

Given a sample space Ω and a σ-algebra $\mathcal{B}$ on Ω, a *probability function* is a function P with domain $\mathcal{B}$ that satisfies

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.
2. $P(\Omega) = 1$.
3. If $A_1, A_2, \dots \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

These three properties are known as the Колмогоров axioms.

Probability space

We often introduce a sample space Ω, a σ-algebra $\mathcal{B}$ on Ω, and a probability function P on $\mathcal{B}$ together as a *probability space*, which we write as $(\Omega, \mathcal{B}, P)$.

School of deFinetti rejects the axiom of countable additivity. Instead asserts:

Axiom of finite additivity

For $n \geq 1$, if $A_1, \ldots, A_n \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Simplified: If $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then $P(A \cup B) = P(A) + P(B)$.

Exercise: Show that the axiom of countable additivity implies finite additivity.

Theorem (Probability function for a finite sample space)

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite sample space and let $\mathcal{B}$ be any σ-algebra on Ω. In addition, let $p_1, \ldots, p_n \geq 0$ such that $\sum_{j=1}^n p_j = 1$. Then the function given by

$$P(A) = \sum_{\{i \,:\, \omega_i \in A\}} p_i \quad \text{for any } A \in \mathcal{B}$$

is a probability function on $\mathcal{B}$ (satisfies the Колмогоров axioms).

The above remains true if Ω is a countable set.

Exercise: Prove the theorem.

Exercise: Consider rolling two dice.

1. Write down the points in the sample space.
2. What probabilities p_j should be assigned to the sample points?
3. Compute the probabilities of the following events:
   - You roll doubles.
   - The sum of the rolls is equal to 7.
   - The sum of the rolls is greater than 10.
   - The absolute value of the difference between the rolls is less than 2.

Sample space for rolling two dice, tabulated as $(\text{roll } 1, \text{roll } 2)$:

$$\Omega = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}.$$

Exercise: Let $\mathcal{B}$ be a σ-algebra on a sample space Ω and let P_1 and P_2 be probability functions on $\mathcal{B}$. Show that the function given by

$$P(A) = \alpha P_1(A) + (1-\alpha)P_2(A) \quad \text{for all } A \in \mathcal{B},$$

where $\alpha \in [0,1]$, is a probability function on $\mathcal{B}$ (satisfies the K. axioms).

Exercise: Let $\Omega = (0,1)$ and let $\mathcal{B} = \{\Omega \cap B : B \in \mathcal{B}(\mathbb{R})\}$ be the Borel σ-algebra on Ω. Show that the function given by

$$P(A) = \int_A 2x dx \quad \text{for all } A \in \mathcal{B}$$

is a probability function on $\mathcal{B}$ (satisfies the K. axioms).

Theorem (Derived properties of a probability function)

Given a probability space $(\Omega, \mathcal{B}, P)$ *and a set* $A \in \mathcal{B}$*, we have*

1. $P(\emptyset) = 0$
2. $P(A) \leq 1$
3. $P(A^c) = 1 - P(A)$

Exercise: Prove (3), (2), and then (1).

Theorem (First probability results)

Given a probability space $(\Omega, \mathcal{B}, P)$ *and sets* $A, B \in \mathcal{B}$, *we have*

1. $P(B \cap A^c) = P(B) - P(A \cap B)$.
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. *If* $A \subset B$, *then* $P(A) \leq P(B)$.

Exercise: Prove (1), (2), and then (3).

Theorem (Law of total probability and union bound)

Given a probability space $(\Omega, \mathcal{B}, P)$, *we have*

1. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ *for any* $A \in \mathcal{B}$ *and partition* $C_1, C_2, \dots \in \mathcal{B}$ *of* Ω.
2. $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ *for any sets* $A_1, A_2, \dots \in \mathcal{B}$

Exercise: Prove (1) and (2) of the theorem.

Theorem (Inclusion-exclusion principle)

Given a probability space $(\Omega, \mathcal{B}, P)$ *and any events* $A_1, \ldots, A_n \in \mathcal{B}$, *we have*

$$P\left(\cup_{i=1}^n A_i\right) = S_1 - S_2 + S_3 - \cdots + (-1)^{n-1} S_n,$$

where $S_k = \sum_{1 \le i_1 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k})$ *for* $k = 1, \ldots, n$.

Exercise:

1. Write out inclusion-exclusion formula for $n = 2, 3$.
2. Give heuristics of counting proof.

Theorem (Bonferroni inequalities)

Under the inclusion-exclusion setup, for any $m = 1, \dots, n$*, we have*

$$P(\cup_{i=1}^n A_i) \leq \sum_{k=1}^m (-1)^{k-1} S_k \quad \text{for } m \text{ odd}$$
$$P(\cup_{i=1}^n A_i) \geq \sum_{k=1}^m (-1)^{k-1} S_k \quad \text{for } m \text{ even.}$$

Exercise: Write out Bonferroni inequalities for $m = 1, 2$.

Theorem (Statisticians' variant of Bonferroni's inequality)

Given a probability space $(\Omega, \mathcal{B}, P)$ *and any events* $A_1, \ldots, A_n \in \mathcal{B}$, *we have*

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

Exercise: Show that the above follows from the union bound (Boole's).

Exercise: Consider constructing $(1-\alpha) \times 100\%$ C.I.s for parameters $\theta_1, \ldots, \theta_n$. Let A_i be the event that C.I. i contains θ_i, $i = 1, \ldots, n$. Give a way to choose α to ensure all C.I.s simultaneously contain their targets with probability $1 - \alpha^*$.

The limit of a sequence of sets

If $\{A_n\}_{n\geq 1}$ is a sequence of sets, $\lim_{n\to\infty} A_n$ exists and is equal to A if $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$ exist and are both equal to A, where

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

Exercise:

1. For $A_n = [1/n, 1 - 1/n]$, $n \geq 1$, check if $\lim_{n\to\infty} A_n$ exists; if so, find it.
2. Make sense of $\omega \in \limsup_{n\to\infty} A_n \iff \omega \in \{A_n\}_{n\geq 1}$ i.o. (infinitely often).
3. Make sense of $\omega \in \liminf_{n\to\infty} A_n \iff \omega \in A_n$ ev. (eventually).

Increasing and decreasing sequences of sets

A sequence of sets $A_1, A_1, \dots$ is called

- *increasing* if $A_1 \subset A_2 \subset \dots$
- *decreasing* if $A_1 \supset A_1 \supset \dots$

Exercise: Verify the following:

1. If $\{A_k\}_{k=1}^{\infty}$ is an increasing sequence of sets we have $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i$.
2. If $\{A_k\}_{k=1}^{\infty}$ is a decreasing sequence of sets we have $\lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i$.

Theorem (Continuity of probability function)

Let $(\Omega, \mathcal{B}, P)$ be a probability space and let $\{A_n\}_{n\geq 1}$ a sequence of sets in $\mathcal{B}$.

- If $\lim_{n\to\infty} A_n = A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.
- Easier to prove:
  1. If $\{A_n\}_{n\geq 1}$ is increasing and $\cup_{n=1}^{\infty} A_n = A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.
  2. If $\{A_n\}_{n\geq 1}$ is decreasing and $\cap_{n=1}^{\infty} A_n = A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.

Exercise: Prove (1).