STAT 511 fa 2019 Lec 01 slides

Set theory and beginning 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.

Table of Contents

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* $\mathcal{S}$ of an experiment is the set of possible outcomes.

Sample points

The outcomes in a sample space $\mathcal{S}$ are called *sample points*.

Exercise: Give the sample space $\mathcal{S}$ for the following experiments:

1. Roll of a 6-sided die.
2. Number of COVID-19 cases confirmed in SC tomorrow.
3. Number of stars counted in a patch of space.
4. Blood type of a randomly selected student.
5. Time until you drop your new phone and crack the screen.
6. Proportion of people in a sample with antibodies to a virus.
7. Deviation of today's temperature from the historical average.

Event

An *event* is a collection of possible outcomes of an experiment, that is any subset of $\mathcal{S}$ (including $\mathcal{S}$ 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.
- Often refer to members of sets as *elements* of the set.

Exercise: Express the following events as subsets of the sample space.

1. Roll of a 6-sided die:
   $A =$ *odd number rolled*.
2. Number of COVID-19 cases confirmed in SC tomorrow:
   $A =$ *no new cases*.
3. Number of stars counted in a patch of space:
   $A =$ *more than 1,000*.
4. Blood type of a randomly selected student:
   $A =$ *has A antigen in RBCs.*
5. Time until you drop your new phone and crack the screen:
   $A =$ *within the first month.*
6. Proportion of people in a sample with antibodies to a virus:
   $A =$ *60% or more with antibodies.*
7. Deviation of today's temperature from the historical average:
   $A =$ *at least 10 degrees cooler.*

Let A and B be events/sets in a sample space $\mathcal{S}$.

Relationships between sets

- *Containment*:
- *Equality*:

Elementary set operations

- The *union set* of A and B, written $A \cup B$, is the set of elements which are either in A or B or in both.
- The *intersection set* of A and B, written $A \cap B$, is the set of elements which are in both A and B.
- The *complement set* of a set A, written A^c, is the set of elements in $\mathcal{S}$ which are not in A.
- *Set subtraction*: The set of elements in B that are not in A is denoted by $B \setminus A = B \cap A^C$.

Draw pictures.

Exercise: Suppose we flip a coin three times and record the sequence of heads and tails.

1. Give the sample space $\mathcal{S}$.
2. Give the points in the event $A =$ at least two heads come up.
3. Give the points in the event $B =$ at exactly two tails come up.
4. Give the points in the event $B \cap A$.
5. Give the points in the event $B^C \cup A^C$.
6. Give the points in the event $(B \cup A)^C$.
7. Give the points in the event $(B \cap A)^C$.
8. Give the points in A^C.

Theorem

For any events A, B, and C in a sample space $\mathcal{S}$, we have

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

Exercise: Suppose the animals you might see on a safari are the wildebeest, the crocodile, and the giraffe. Regarding these as events W, C, and G, write down the following events using elementary set operations. You see...

1. a giraffe but no wildebeest.
2. all three types of animals.
3. not all three types of animals.
4. a giraffe and a wildebeest but no crocodile.
5. not both a giraffe and a wildebeest.
6. a giraffe and a wildebeest without seeing a crocodile or a crocodile and a wildebeest without seeing a giraffe.
7. any two of the three types of animal.
8. exactly one of the three types of animal.
9. at least one of the three types of animal.
10. none of the animals.

Mutual exclusivity/disjoint-ness

Two events A and B are called *mutually exclusive* or *disjoint* if $A \cap B = \emptyset$.

The set $\emptyset$ is the the *empty set*, which is the set containing no elements.

Pairwise disjoint-ness/mutual exclusivity

The events A_1, A_2, A_3, ... are called *mutually exclusive* or *pairwise disjoint* if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Partition

If $A_1, A_2, \dots \subset \mathcal{S}$ are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = \mathcal{S}$, then the collection of sets $A_1, A_2, \dots$ is called a *partition* of $\mathcal{S}$.

Discuss:

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 $\mathbb{R}$, the set of all real numbers.

Any event $A \subset \mathcal{S}$ together with its complement A^C form a partition of $\mathcal{S}$. Why??

Probability function (Колмогоров axioms)

Given a sample space $\mathcal{S}$ and a collection $\mathcal{E}$ of events in $\mathcal{S}$ to which we wish to assign probabilities, a *probability function* P on $\mathcal{E}$ is a function which satisfies

- ax1
- ax2
- ax3

Theorem (Computing probabilities on a finite sample space)

Let $\mathcal{S} = \{s_1, \dots, s_n\}$ be a finite sample space and let $p_1, \dots, p_n$ be nonnegative numbers which sum to one. For any subset A of $\mathcal{S}$, including $\mathcal{S}$ itself, define

$$P(A) = \sum_{\{i : s_i \in A\}} p_i .$$

Then P is a probability function (satisfies K'gorov axioms) on the subsets of $\mathcal{S}$.

The above remains true if $\mathcal{S}$ is a countable set.

Tells how to compute probabilities: sum the probabilities of the points in the event.

Exercise: For sample space of blood types

$$\mathcal{S} = \{\text{O}-,\text{O}+,\text{A}-,\text{A}+,\text{B}-,\text{B}+,\text{AB}-,\text{AB}+\}$$

we have the probabilities (if a person drawn randomly from USA):

$$\begin{aligned} p_{\text{O}-} &= 0.066, & p_{\text{B}-} &= 0.015 \\ p_{\text{O}+} &= 0.374, & p_{\text{B}+} &= 0.085 \\ p_{\text{A}-} &= 0.063, & p_{\text{AB}-} &= 0.006 \\ p_{\text{A}+} &= 0.357, & p_{\text{AB}+} &= 0.034 \end{aligned}$$

Give the probability that

1. a randomly selected person is Rh+.
2. a randomly selected person has the A antigen in the RBCs.

Exercise: Consider rolling two dice.

1. Write down the points in the sample space.
2. What probability should be assigned to the individual sample points?
3. Compute the following probabilities:
   - 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.

Motivation to study counting rules

If all outcomes in $\mathcal{S}$ are equally likely, for any $A \subset \mathcal{S}$, we have

$$P(A) = \frac{\#\{\text{sample points in } A\}}{\#\{\text{sample points in } \mathcal{S}\}}.$$

This leads to an interest in counting rules.

Theorem

For P a probability function and $A \subset \mathcal{S}$, we have

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

Exercise: Prove the identity $P(A^C) = 1 - P(A)$.

Theorem

For P a probability function and $A, B \subset \mathcal{S}$, 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* $P(A) \leq P(B)$.

Draw pictures.

Exercise: Consider the the events

$$A = \text{Oversleep on day 1 of semester}$$
$$B = \text{Oversleep on day 2 of semester}$$

and suppose $P(A) = 1/4$, $P(B) = 1/3$, and $P(A \cap B) = 1/10$.

Find

1. $P(A \cap B^C)$
2. $P(A \cup B)$

Theorem

If P is a probability function then

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

The second result is called *Boole's inequality* or the *union bound*.

Illustrate these results.

Theorem (Bonferroni's inequality)

For events $A_1, \dots, A_n$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq 1 - \sum_{i=1}^{n} P(A_i^C).$$

For two events A *and* B, *the above reduces to*

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

Prove the above.

Exercise: Suppose you make free throws with probability 0.7. Give a lower bound for the probability with which

1. you make 2 out of 2 free throws.
2. you make 3 out of 3 free throws.