MATH 205: Statistical methods

September 15th, 2021

Lecture 4: Sample space, events and probability

Tentative schedule

Date	Theme/Topic	Labs	Assignments
Sep 1	Syllabus		
Sep 8	Chapter 1: Describing dataset	Section 2: Handling data	
Sep 13 - 15	Chapter 2: Looking at Relationships	Section 3: Univariate data	
Sep 20-22	Chapter 3: Basic Ideas in Probability	Section 4: Bivariate Data	Homework 1 (due 09/22)
Sep 27-29	Chapters 3-4	Section 4: Correlation	
Oct 4-6	Chapter 4: Random variables and expectations	Section 6: Random data	Homework 2 (due 10/06)
Oct 11-13	Chapter 5: Useful distributions	Section 7: The central limit theorem	
Oct 18-20	Chapter 6: Samples and populations	Section 9: Confidence interval estimation	Homework 3 (due 10/20)
Oct 25-27	Review and midterm exam		Midterm: Oct 27 (lecture), Oct 25-27 (labs)
Nov 1-3	Chapter 7: The significance of evidence	Section 10: Hypothesis testing	
Nov 8-10	Goodness of Fit	Section 12: Goodness of Fit	Homework 4 (due 11/10)
Nov 15-17	Linear Regression	Section 13: Linear regression	
Nov 22-24	Thanksgiving break		
Nov 29 - Dec 1	One-Way Analysis of Variance	Section 15: Analysis of variance	Homework 5 (due 12/01)
Dec 6-8	Selected topics + Review		
Exam week			

Topics

- Sample space and events
- Basic properties of probability
- Advanced properties of probability
- Compute probability
    - Computing event probabilities by counting outcomes
    - Computing probabilities by reasoning about sets

Sample space and events

1. An experiment: is any action, process, or phenomenon whose outcome is subject to uncertainty
2. An outcome: is a result of an experiment
   Each run of the experiment results in one outcome
3. A sample space: is the set of all possible outcomes of an experiment
4. An event: is a subset of the sample space.
   An event occurs when one of the outcomes that belong to it occurs

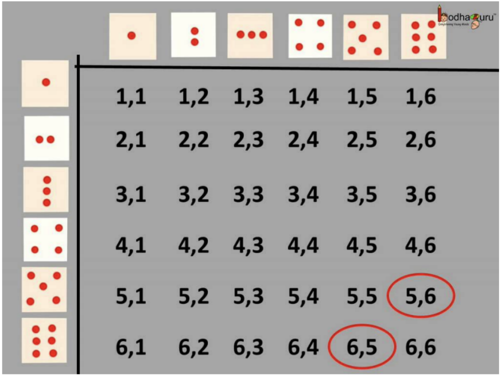
Different views of probability

- Frequentist: The probability of an outcome is the frequency of that outcome in a very large number of repeated experiments
- Bayesian: Probability is a quantification of a belief about how often an outcomes occurs

Bottom line: Probability is a function defined on the set of events of an experiment

1. Experiment: Toss two regular dice
2. Event E_1 = the summation of the two dice is 11

$$P[E_1] = 1/18$$



Basic set operators

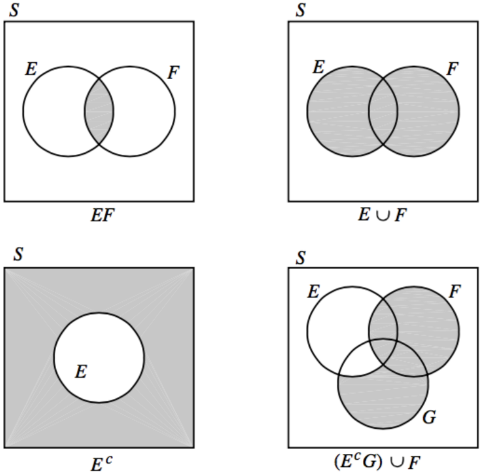


Figure 1.1 Venn diagrams of the events specified.

What conditions should we impose to define probability?

- P[the sample space] $= 1$
- $0 \leq P[E] \leq 1$ for all events E
- $P[\emptyset] = 0$
- If E_1 and E_2 are disjoint then $P[E_1 \cup E_2] = P[E_1] + P[E_2]$
- If E_1, E_2 and E_3 are mutually disjoint then

$$P[E_1 \cup E_2 \cup E_3] = P[E_1] + P[E_2] + P[E_3]$$

Basic properties of probability

Useful Facts 3.1 (Basic Properties of the Probability Events)
We have

- The probability of every event is between zero and one; in equations
$$0 \leq P(\mathcal{A}) \leq 1$$
for any event $\mathcal{A}$.
- Every experiment has an outcome; in equations,
$$P(\Omega) = 1.$$
- The probability of disjoint events is additive; writing this in equations requires some notation. Assume that we have a collection of events $\mathcal{A}_i$, indexed by i. We require that these have the property $\mathcal{A}_i \cap \mathcal{A}_j = \varnothing$ when $i \neq j$. This means that there is no outcome that appears in more than one $\mathcal{A}_i$. In turn, if we interpret probability as relative frequency, we must have that
$$P(\cup_i \mathcal{A}_i) = \sum_i P(\mathcal{A}_i)$$

Useful Facts 3.2 (Properties of the Probability of Events)

- $P(\mathcal{A}^c) = 1 - P(\mathcal{A})$
- $P(\varnothing) = 0$
- $P(\mathcal{A} - \mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B})$
- $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$

Others

Problem

If $A \subset B$, *then* $P(A) \leq P(B)$.

Others

Problem

For any events A, B

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

Computing event probabilities by counting outcomes

In some problems, you can compute the probabilities of events by counting outcomes.

Problem

We throw a fair (each number has the same probability) six-sided die twice, then add the two numbers. What is the probability of getting a number divisible by five?

Computing probabilities by reasoning about sets

Problem

In a community, 32% of the population smokers who like math; 27% are smokers who don't. What percentage of the population of this community smoke?

Computing probabilities by reasoning about sets

Problem

You flip two fair six-sided dice, and add the number of spots. What is the probability of getting a number divisible by 2, but not by 5?

Problem

You flip two fair six-sided dice, and add the number of spots. What is the probability of getting a number divisible by either 2 or 5, or both?

Computing probabilities by reasoning about sets

Problem

Suppose that in a community of 400 adults, 300 bike or swim or do both, 160 swim, and 120 swim and bike. What is the probability that an adult, selected at random from this community, bikes?