Lecture Notes for Math 372:
Elementary Probability and Statistics

Last updated: January 11, 2026

Contents

0 Tentative course outline

This course is a problem-oriented introduction to the basic concepts of probability and statistics, providing a foundation for applications and further study.

1. **Weeks 1-2.** Sampling distributions (4 lessons).

   chi-squared, t, and F distributions, distributions of sample mean and variance

2. **Weeks 3-4.** Point estimation (5 lessons)

   properties and methods of point estimation

3. **Weeks 5-6.** Interval estimation (4 lessons)

   Confidence intervals for means, variances, proportions and differences

4. **Weeks 7-12.** Hypothesis Testing (19 lessons)

   Neyman-Pearson lemma, likelihood ratio test; tests concerning means and variances, tests based on count data, nonparametric tests, analysis of variance

5. **Weeks 13-14.** Regression and correlation (6 lessons)

   regression, bivariate normal distributions, method of least squares

1 2025-01-12 | Week 01 | Class 01

- give syllabus
- do activity with why you're in this course

> ***The nexus question of this lecture:*** *What is a probability?*

Reading assignment: Sections 1.1, 1.2, 1.3, 2.1, 2.4 of the textbook.

1.1 What is probability?

1.1.1 A general framework: sample space, events, etc

We begin with a general framework and some terminology to formalize the notions of probability. This is based on section 2.4 in the textbook.

- An ***experiment*** is an activity or process whose outcome is subject to uncertainty, and about which an observation is made.

  Examples include flipping a coin, rolling a dice, measuring the size of a wave, or the amount of rainfall, conducting a poll, performing a diagnostic test, opening a pack of Pokemon cards, etc.

- The ***sample space*** S of an experiment is the set of all possible outcomes. The elements of the sample space are called ***sample points***.

  We think of each sample point as representing a unique outcome of the experiment. In the case of rolling a dice, the sample points are $1, 2, 3, 4, 5$ and 6, and the sample space is $S = \{1, 2, 3, 4, 5, 6\}$.

- We use the term ***event*** to refer to a collection of outcomes, i.e., a subset of S.

  Example: if our experiment is rolling a 6-sided dice, here are some events

$$
\begin{aligned}
A &= [\text{observe an odd number}] & E_2 &= [\text{observe a } 2] \\
B &= [\text{observe an even number}] & E_3 &= [\text{observe a } 3] \\
C &= [\text{observe a number less than } 5] & E_4 &= [\text{observe a } 4] \\
D &= [\text{observe a } 2 \text{ or a } 3] & E_5 &= [\text{observe a } 5] \\
E_1 &= [\text{observe a } 1] & E_6 &= [\text{observe a } 6]
\end{aligned}
$$

- There are two types of events: ***compound events***, which can be decomposed into other events, and ***simple events***, which cannot.

  In the above example, the events A, B, C and D are compound events. $E_1, \ldots, E_6$ are simple events.

- A sample space is ***discrete*** if it is countable (i.e., finite or countably infinite). In a discrete sample space S, the set of all possible events is the *power set* of S.[1]

  In the dice-rolling example, the set of all possible events is $\{E : E \subseteq \{1, 2, 3, 4, 5, 6\}\}$.

$$
\begin{aligned}
A &= [\text{observe an odd number}] = \{1, 3, 5\} & E_2 &= [\text{observe a } 2] = \{2\} \\
B &= [\text{observe an even number}] = \{2, 4, 6\} & E_3 &= [\text{observe a } 3] = \{3\} \\
C &= [\text{observe a number less than } 5] = \{1, 2, 3, 4\} & E_4 &= [\text{observe a } 4] = \{4\} \\
D &= [\text{observe a } 2 \text{ or a } 3] = \{2, 3\} & E_5 &= [\text{observe a } 5] = \{5\} \\
E_1 &= [\text{observe a } 1] = \{1\} & E_6 &= [\text{observe a } 6] = \{6\}
\end{aligned}
$$

[1] If S is not discrete, a complication arises: in that case, some subsets of S are too wild and untameable for us to treat them mathematically as "events". Resolving that issue requires introducing measure theory, which is beyond the scope of this class, so we will ignore it and simply steer clear of any setting where any issues might arise.

- Some observations about events:
  - The sample points are *elements* of S. The simple events are *singleton subsets* of S. In the dice example, we have:
    - Sample points: 1,2,3,4,5,6.
    - Simple events: $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$.
  - The empty set $\emptyset$ and the whole sample space S are always both events: $\emptyset$ is the event "nothing happens" and S is the event "something happens".
  - Events satisfy a the properties of a boolean algebra:
    - **"And":** If E and F are events, then $E \cap F$ is the event that E and F occur.
    - **"Or":** If E and F are events, then $E \cup F$ is the event that E or F occurs.
    - **"Not":** If E is an event, then $E^c = S \backslash E$ is the event that E does not occur.
- Two events E and F are ***mutually exclusive*** if $E \cap F = \emptyset$. This means that E and F cannot both happen at the same time.

  In the dice example, the events A and B are mutually exclusive, since the dice roll cannot be both even and odd. But A and C are not mutually exclusive because $A \cap C = \{1, 3\} \neq \emptyset$. If a 1 or a 3 is rolled, then both A and C occur.

1.1.2 Definition of probability measure

Definition 1 (Probability measure)**.** Let S be a sample space associated with an experiment. A function $\mathbb{P}$ is said to be a ***probability measure*** on S if it satisfies the following three axioms:

A.1 (Nonnegativity) For every event $E \subseteq S$,

$$\mathbb{P}[E] \geq 0.$$

A.2 (Total mass one) $\mathbb{P}[S] = 1$.

A.3 (Countable additivity) If $E_1, E_2, \ldots$ is a sequence of events which are pairwise mutually exclusive (meaning $E_i \cap E_j = \emptyset$ if $i \neq j$), then

$$\mathbb{P}[E_1 \cup E_2 \cup \ldots] = \sum_{i=1}^{\infty} \mathbb{P}[E_i].$$

If $\mathbb{P}$ is a probability measure, then for every event $E \subseteq S$, the number $\mathbb{P}[E]$ is called the ***probability*** of E.

The above definition only tells us the conditions an assignment of probabilities must satisfy; it doesn't tell us how to assign specific propbabilities to events.

Probability measures satisfy some basic properties:

Proposition 2 (Basic properties of probability measure)**.** *If $\mathbb{P}$ is a probability measure, then the following properties hold:*

(i.) *(The null event has probability zero)* $\mathbb{P}[\emptyset] = 0$.

(ii.) *(Finite additivity) Let $\{E_1, \ldots, E_n\}$ be a finite sequence of events. If the sequence is pairwise disjoint, then*

$$\mathbb{P}[E_1 \cup E_2 \cup \ldots \cup E_n] = \mathbb{P}[E_1] + \mathbb{P}[E_2] + \ldots + \mathbb{P}[E_n].$$

(iii.) *("With probability one, an event E either does occur or doesn't")* $\mathbb{P}[E^c] = 1 - \mathbb{P}[E]$.

(iv.) *(Excision Property) If A, B are events and $A \subseteq B$, then*

$$\mathbb{P}[B \backslash A] = \mathbb{P}[B] - \mathbb{P}[A].$$

(v.) *("The particular is less likely than the general") If A, B are events and $A \subseteq B$, then $\mathbb{P}[A] \leq \mathbb{P}[B]$.*

(vi.) *("Probabilities are between 0 and 1") For any event E, $\mathbb{P}[E] \in [0, 1]$.*

1.2 Conditional probabilities and independence

Definition 3 (Independence)**.** Two events A and B are said to be ***independent*** if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$. Otherwise, the events are said to be dependent.

Definition 4 (Conditional probability)**.** Let A, B be events, and assume that $\mathbb{P}[B] > 0$. Then the ***conditional probability of A, given B***, denoted $\mathbb{P}[A \mid B]$, is given by the formula

$$\mathbb{P}[A \mid B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

(Intuitively, $\mathbb{P}[A \mid B]$ is the probability of A when we know that event B happened.)

Example 5. Roll a dice. Let B be the event that an odd number was observed, and A the event that a '1' was rolled.

- The unconditional probability: $\mathbb{P}[A] = 1/6$.
- The conditional probability: $\mathbb{P}[A \mid B] = 1/3$

End of Example 5. □

Proposition 6. *If A, B are independent events with positive probabilities, then*

$$\mathbb{P}[A \mid B] = \mathbb{P}[A] \quad \textit{and} \quad \mathbb{P}[B \mid A] = \mathbb{P}[B]$$

This proposition provides the basis for the following intuition about independence: the occurance of one of the events is unaffected by the occurrence or nonoccurence of the other.

1.3 Random variables

Section 2.11

Definition 7 (Random variable)**.** A ***random variable*** is a real-valued function whose domain is a sample space.

The value of a random variable is thought of as varying depending on the outcome of the experiment.

Example 8. Consider an experiment in which we flip a coin twice. The sample space is consists of four sample points:

$$E_1 = HH, \quad E_2 = HT, \quad E_3 = TH, \quad E_4 = TT$$

Let Y be the number of heads flipped. So Y can take three values: $0, 1$ and 2. These are examples of events:

$$[Y = 0] = \{E_4\}, \quad [Y = 1] = \{E_2, E_3\}, \quad \text{and} \quad [Y = 2] = \{E_1\}.$$

We can compute the probabilities of these:

$$\mathbb{P}[Y = 0] = \mathbb{P}[E_4] = 1/4.$$

and

$$\mathbb{P}[Y = 1] = \mathbb{P}[E_2] + \mathbb{P}[E_3] = 1/4 + 1/4 = 1/2.$$

End of Example 8. □

Definition 9 (Distribution function - section 4.2)**.** Let X be any real-valued random variable. The ***(cumulative) distribution function*** of X is the function

$$F(x) = \mathbb{P}[X \leq x]$$

for $-\infty < x < \infty$.

A distribution function is a nondecreasing function with $F(-\infty) = 0$ and $F(+\infty) = 1$. For discrete random variables, the distribution function is always a step function.

A real-valued random variable X is said to be ***continuous*** if there exists a nonnegative function f_X such that

$$F(x) = \int_{-\infty}^{x} f_X(t)dt$$

for all $x \in \mathbb{R}$, where F is the distribution function of X. The function f_X is called the ***probability density function*** of X (this usually shortened to "pdf" or just "density").

The distribution function F of a continuous random variable is always continuous (because it is differentiable by the fundamental theorem of calculus with derivative $F'(x) = f(x)$, and differentiability implies continuity).

Theorem 10 (Theorem 4.3 in textbook)**.** *If a random variable Y has density f_Y then*

$$\mathbb{P}\left[a \le Y \le b\right] = \int_a^b f_Y(y)dy$$

for all $-\infty \le a \le b \le +\infty$.

If a random variable Y has pdf f_Y, then its ***expectation*** is $\mathbb{E}\left[Y\right] = \int y f(y)dy$, where ther integral ranges over the support of the function f, provided that $\int |y| f(y)dy < \infty$.

If Y_1, Y_2 are discrete random variables, the joint probability mass function for Y_1 and Y_2 is

$$f(y_1, y_2) = \mathbb{P}\left[Y_1 = y_1, Y_2 = y_2\right]$$

for $-\infty < y_1, y_2 < +\infty$.

Add definition 5.2 and 5.3.

Add definitions of marginal pdfs and pms

Definition 11. Two random variabes X and Y are independent if

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y)$$

for all $x, y \in \mathbb{R}$.

Here, $F_{(X,Y)}(x, y) = \mathbb{P}\left[X \le x, Y \le y\right]$ and $F_X(x) = \mathbb{P}\left[X \le x\right]$ and $F_Y(y) = \mathbb{P}\left[Y \le y\right]$.

Add Theorem 5.4, 5.5 – important

1.4 Statistics

Random sample is section 2.12.

A ***population*** is a large body of data that is the target of our interest. The subset collected from it is our ***sample***.

The objective of statistics to is to make an inference about a population based on information contained in a sample from that population and to provide an associated measure of goodness for the inference.

Definition 12 (Point Estimator)**.** A **point estimator** is any function $W(X_1, \dots, X_n)$ of a sample. Thus, any statistic is a point estimator. In general we refer to an *estimator* as a function of the sample, while an *estimate* is the realized value of an estimator (e.g., a number) that is obtained when a sample is actually taken.

Definition 13 (Statistic, Sampling Distribution)**.** Let $X_1, \dots, X_n$ be a random sample of size n from a population and let $T(x_1, \dots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \dots, X_n)$. Then the random variable or random vector

$$Y \triangleq T(X_1, \dots, X_n)$$

is called a **statistic**. The probability distribution of a statistic Y is called the **sampling distribution** of Y.

Definition 14 (Sample Mean, Sample Variance)**.** The **sample mean** is the statistic defined by

$$\overline{X} \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and the **sample variance** is the statistic defined by

$$S^2 \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Chebychev's theorem

Theorem 15 (Chebychev's theorem)**.** *Let Y be a random variable with mean μ and finite variance σ^2. Then for any constant c,*

$$\mathbb{P}\left[|Y-\mu| \geq k\sigma\right] \leq \frac{1}{k^2}$$

Law of large numbers