Lecture Notes for Math 372:
Elementary Probability and Statistics

Last updated: January 16, 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 | Lecture 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

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

- 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 $\{E : E \subseteq \{1, 2, 3, 4, 5, 6\}\}$.

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

[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 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 propabilities 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,* $\mathbb{P}[E] \in [0,1]$.

2 2026-01-14 | Week 01 | Lecture 02

The topic of this lecture: *independent events, conditional probabilities, random variables*

2.1 Independent events and conditional probabilities

This section is based on section 2.7 in the textbook.

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

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

Interpretation: $\mathbb{P}[A \mid B]$ is the probability of A when we know that event B happened.

Definition 5. We say that there exists a ***positive relationship*** between events A and B if

$$\mathbb{P}[A \mid B] > \mathbb{P}[A],$$

and a ***negative relationship*** if

$$\mathbb{P}[A \mid B] < \mathbb{P}[A].$$

Remark 6. Note the the conditions of Definition 5 are symmetric in the sense that

$$\mathbb{P}[A \mid B] > \mathbb{P}[A] \iff \mathbb{P}[B \mid A] > \mathbb{P}[B],$$

provided that both A and B have positive probability.

Example 7. Roll a 6-sided dice. Let A be the event that a '2' was rolled, and B be the event that an even number was observed.

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

Since $\frac{1}{3} > \frac{1}{6}$, we conclude there is a positive relationship between rolling a '2' and rolling an even number.

End of Example 7. □

The notion of independence formalizes the idea of "no relationship".

Proposition 8. *If A, B are events with positive probabilities then the following are equivalent:*

(i.) *A and B are independent.*

(ii.) $\mathbb{P}[A \mid B] = \mathbb{P}[A]$ *and* $\mathbb{P}[B \mid A] = \mathbb{P}[B]$.

In words, independence means that the probabilities of each event are unaffected by whether or not the other event occurs. Proposition 8 simply formalizes this idea using conditional probabilities.

2.2 Random variables

Based on Sections 2.11, 4.2 in the textbook

Definition 9 (Random variable)**.** A ***random variable*** (or ***rv***) 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 (the sample point). Random variables are usually denoted with capital letters, like X, Y, Z.

Example 10 (Sum 2d4)**.** Roll a 4-sided dice twice (this is the **experiment**). There are 16 possible **outcomes**. The **sample space** is

$$S = \{(x, y) : x, y \in \{1, 2, 3, 4\}\} .$$

Let X be the sum of the two rolls. We can represent X by the following table:

		Dice 2			
		1	2	3	4
	1	2	3	4	5
Dice 1	2	3	4	5	6
	3	4	5	6	7
	4	5	6	7	8

Events are often defined using preimages of random variables. For example, the event that $X = 6$ is:

$$\begin{aligned}[X = 6] &= \{\omega \in S : X(\omega) = 6\} \\ &= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} .\end{aligned}$$

The textbook uses the notation $\{X = 6\}$ instead of $[X = 6]$.

Here's another example of an event. Let $E = \{2, 4, 6, 8\}$. Then

$$\begin{aligned}[X \text{ is even}] &= [X \in E] \\ &= \{\omega \in S : X(\omega) \in E\} \\ &= \{(1, 1), (1, 3), (2, 2), (2, 4), (3, 1), (3, 3), (4, 2), (4, 4)\} .\end{aligned}$$

When writing random variables, we usually suppress the arguments, e.g., writing X rather than $X(\omega)$.

End of Example 10. □

3 2026-01-16 | Week 01 | Lecture 03

3.1 Random variables

3.1.1 Discrete vs continuous

Definition 11 (Discrete random variable)**.** We say that a random variable X is a ***discrete random variable*** if it can assume only a finite or countably infinite number of distinct values.

Definition 12 (Probability mass function, pmf)**.** Let X be a discrete random variable. The ***probability mass function*** (or ***pmf***) of X is the function

$$p(x) = \mathbb{P}[X = x],$$

defined for every $x \in \mathbb{R}$.

Example 13. The pmf of X in Example 10 is

$$p(2) = 1/16 \quad p(3) = 2/16, \quad p(4) = 3/16, \quad p(5) = 4/16, \quad p(6) = 3/16, \quad p(7) = 2/16, \quad p(8) = 1/16$$

and $p(x) = 0$ for all other $x \in \mathbb{R}$.

End of Example 13. □

Definition 14 (Distribution function - section 4.2)**.** Let X be any random variable. The ***cumulative distribution function*** (or ***cdf***) of X is the function

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

defined for all $x \in \mathbb{R}$.

Remark 15. The domain of a cdf is always $\mathbb{R}$, and it is always a nondecreasing function with $F(-\infty) = 0$ and $F(+\infty) = 1$. The cdf of a discrete random variable is always a step function.

Definition 16 (Continuous rv)**.** Let Y be a random variable with distribution function F. We say that Y is a ***continuous random variable*** if there exists a nonnegative function f such that

$$F(y) = \int_{-\infty}^{y} f(t)dt \tag{1}$$

for all $y \in \mathbb{R}$. The function f is called the ***probability density function*** (or ***pdf***) of Y.

Remark 17. For continuous random variables, the distribution function F is always continuous. Moreover, for a continuous random variable Y, $\mathbb{P}[Y = b] = 0$ for all $b \in \mathbb{R}$.

Theorem 18 (Theorem 4.3 in textbook)**.** *If Y is a continuous random variable with pdf f, then*

$$\mathbb{P}[a \leq Y \leq b] = \int_{a}^{b} f(t)dt$$

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

3.1.2 Expected value

Definition 19 (Expectation of a continuous random variable)**.** If Y is a random variable with pdf f, then the ***expected value*** of Y, denoted $\mathbb{E}[Y]$, is the quantity

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f(y)dy,$$

provided that $\int_{-\infty}^{\infty} |y| f(y)dy < \infty$.

Remark 20. $\mathbb{E}[Y]$ is the long-run average of Y, if we were to repeat the experiment many times.

Theorem 21 (LOTUS - single variable case)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be a function.*

(i.) *If X has pmf p, then*

$$\mathbb{E}[g(X)] = \sum_{x \in \mathbb{R}:\ p(x)>0} g(x)p(x).$$

(ii.) *If Y has pdf f, then*

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy.$$

Remark 22. Often we wish to compute probabilities of functions of multiple random variables, for example:

- What is the probability that $\frac{X_1+\ldots+X_n}{n} \in (0,1)$? Here, the function is $g(x_1,\ldots,x_n) = \frac{x_1+\ldots+x_n}{n}$.
- What is the probability that $\max(X,Y) \leq 10$? Here the function is $g(x,y) = \max(x,y)$.
- Suppose we roll two dice and take the maximum. What is the expected value? In this case, our dice rolls are X, Y and we want to compute $\mathbb{E}[g(X,Y)]$, where $g(x,y) = \max(x,y)$.

To answer these sorts of questions, we need the notion of a "joint distribution".

3.1.3 Joint distributions

This subsection is based on section 5.4 in the textbook. Everything in this section generalizes naturally to n variables, but the results are simpler to state for just 2 random variables.

Definition 23 (Joint pmf)**.** Let X_1 and X_2 be discrete random variables. The ***joint probability mass function*** for X_1 and X_2 is the function

$$p(x_1,x_2) = \mathbb{P}[X_1 = x_1, X_2 = x_2],$$

defined for all $x_1, x_2 \in \mathbb{R}$.

Definition 24 (Joint pdf)**.** Let Y_1 and Y_2 be continuous random variables. We say that Y_1 and Y_2 are ***jointly continuous*** if there exists a function $f : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ such that

$$\mathbb{P}[Y_1 \leq y_1, Y_2 \leq y_2] = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f(t_1,t_2)dt_2dt_1.$$

for all $y_1, y_2 \in \mathbb{R}$. The function f is called the ***joint probability density function*** for Y_1 and Y_2.

Theorem 25 (LOTUS - multivariable case)**.** *Let $g : \mathbb{R}^2 \to \mathbb{R}$.*

- *If X_1, X_2 have joint pmf $p(x_1,x_2)$, then*

$$\mathbb{E}[g(X_1,X_2)] = \sum_{\substack{(x_1,x_2)\in\mathbb{R}^2:\\ p(x_1,x_2)>0}} g(x_1,x_2)p(x_1,x_2).$$

- *If Y_1, Y_2 are jointly continuous random variables with joint pdf $f(y_1,y_2)$, then*

$$\mathbb{E}[g(Y_1,Y_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1,y_2)f(y_1,y_2)dy_1dy_2.$$

Remark 26. Theorem 25 generalizes to n variables. It gives us a way to answer questions like the third question posed in Remark 22.

4 2026-01-21 | Week 02 | Lecture 04

4.1 Independence

4.1.1 Definition and characterization

The aim of this section is to define what it means for random variables to be independent.

Definition 27. We say that the random variables X_1, X_2 are ***independent*** if

$$\mathbb{P}[X_1 \le x_1, X_2 \le x_2] = \mathbb{P}[X_1 \le x_1]\,\mathbb{P}[X_2 \le x_2]$$

for all $x_1, x_2 \in \mathbb{R}$.

Theorem 28 (Factorization theorem – Theorem 5.4 in textbook)**.** *For discrete/continuous random variables, independence is equivalent to factorizability of the joint pmf/pdf. More formally, we have:*

- ***Discrete case:*** *Let X_1, X_2 be discrete random variables with pmfs p_1, p_2 and joint pmf p. Then X_1 and X_2 are independent if and only if*

  $$p(x_1, x_2) = p_1(x_1) \cdot p_2(x_2)$$

  for all $x_1, x_2 \in \mathbb{R}$.

- ***Continuous case:*** *Let Y_1 and Y_2 be continuous random variables with pdfs f_1 and f_2, and joint pdf f. Then Y_1 and Y_2 are independent if and only if*

  $$f(y_1, y_2) = f_1(y_1) \cdot f_2(y_2)$$

  for all $y_1, y_2 \in \mathbb{R}$.

Remark 29. If two random variables are independent, then observing one of them does not give any information about what value the other one takes.

Remark 30. In this course, we will frequently work with "samples" of n independent random variables $X_1, \ldots, X_n$. For that setting, note that Definitions 23, 24 and 27 and Theorem 28 all generalize in the natural way (i.e., with n variables rather than 2). In that case, the intuition is that observing any number of them doesn't give you any information about the others.

4.1.2 Some useful consequences of independence

Theorem 31. *If X_1 and X_2 are independent, then*

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\,\mathbb{E}[X_2].$$

Proof. This can be proven for continuous/discrete random variables using Theorem 25. □

Definition 32 (Variance of a random variable)**.** If X is a random variable and $\mu = \mathbb{E}[X]$, the ***variance*** of X, denoted $V(X)$ or $\mathrm{Var}(X)$, is the quantity

$$V(X) = \mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right]$$

The positive square root of the variance is the ***standard deviation*** of X.

Theorem 33. *If X and Y are independent random variables and a are scalars, then*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

and

$$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X).$$

4.2 What is a statistic?

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 34 (Point Estimator)**.** A **point estimator** is any function $W(X_1, \ldots, X_n)$ 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 35 (Statistic, Sampling Distribution)**.** Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \ldots, X_n)$. Then the random variable or random vector

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

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

Definition 36 (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

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

Theorem 37 (Chebychev's inequality)**.** *Let Y be a random variable with mean μ and finite variance σ^2. Then for any constant $c > 0$,*

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