Lecture Notes

Kyle Chui

2022-01-04

Contents

1 Lecture 1

The goal of this class is to *quantify randomness*. The main topics for the term are:

1. The fundamentals of probability theory, including conditional probability and enumeration arguments.
2. Discrete and continuous random variables.
3. Sequences of i.i.d. random variables, including the Weak Law of Large Numbers and the Central Limit Theorem.

1.1 Properties of Probability

Probability theory takes place inside a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition. *Probability Space*
A *probability space* is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying:

1. A non-empty set Ω, called the *sample space*.
2. A set $\mathcal{F}$ of subsets of Ω satisfying certain properties:
   - Elements of $\mathcal{F}$ are called *events*.
   - Events $A_1, A_2, \ldots, A_k$ are called *mutually exclusive* if they are *pairwise disjoint*, i.e. if $i \neq j$ then $A_i \cap A_j = \varnothing$.
   - Events $A_1, A_2, \ldots, A_k$ are called *exhaustive* if their union is the sample space, i.e.
   $$\bigcup_{j=1}^{k} A_j = \Omega.$$
   - For this class you may ignore $\mathcal{F}$ and assume that all subsets of Ω are events.
3. A function $\mathbb{P}\colon \mathcal{F} \to [0, 1]$, called a *probability measure*, which satisfies:
   - $\mathbb{P}[\Omega] = 1$, or "the probability that something happens is 1".
   - If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then
   $$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] = \sum_{j=1}^{n} \mathbb{P}[A_j].$$
   - If $A_1, A_2, \ldots$ are mutually exclusive events, then
   $$\mathbb{P}\left[\bigcup_{j=1}^{\infty} A_j\right] = \sum_{j=1}^{\infty} \mathbb{P}[A_j].$$

Example. Suppose I flip two fair coins. Then the sample space can be written as $\Omega = \{HH, HT, TH, TT\}$. The probability measure should be defined as

$$\begin{aligned} P[HH] &= \frac{1}{4} \\ P[HT] &= \frac{1}{4} \\ P[TH] &= \frac{1}{4} \\ P[TT] &= \frac{1}{4}. \end{aligned}$$

The probability of getting exactly one head is hence $\mathbb{P}[\{HT, TH\}] = \mathbb{P}[HT] + \mathbb{P}[TH] = \frac{1}{2}$.

Theorem. $\mathbb{P}[\varnothing] = 0$.

Proof. We know that Ω and $\varnothing$ are mutually exclusive, since, $\Omega \cap \varnothing = \varnothing$. Thus

$$\begin{aligned} \mathbb{P}[\Omega] &= \mathbb{P}[\Omega \cup \varnothing] \\ &= \mathbb{P}[\Omega] + \mathbb{P}[\varnothing], \end{aligned}$$

and so $\mathbb{P}[\varnothing] = 0$. $\square$

Theorem. If $A \subseteq \Omega$ is an event and $A' = \Omega \setminus A$ then

$$\mathbb{P}[A] = 1 - \mathbb{P}[A'].$$

Proof. Since we have that $A' = \Omega \setminus A$, we know that $A' \cap A = \varnothing$, so they are mutually exclusive. Thus we have

$$\begin{aligned} \mathbb{P}[\Omega] &= \mathbb{P}[A \cup A'] \\ 1 &= \mathbb{P}[A] + \mathbb{P}[A'] \\ \mathbb{P}[A] &= 1 - \mathbb{P}[A']. \end{aligned}$$

$\square$

Theorem. If $A \subseteq B$ then

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

Proof. We know that $B = A \cup (B \setminus A)$ and $A \cap (B \setminus A) = \varnothing$. Hence

$$\mathbb{P}[B] = \mathbb{P}[A \cup (B \setminus A)] = \mathbb{P}[A] + \mathbb{P}[B \setminus A],$$

and the result follows. $\square$

Theorem. If $A \subseteq B$ then $\mathbb{P}[A] \leq \mathbb{P}[B]$.

Proof. From the previous theorem we have

$$\mathbb{P}[A] \leq \mathbb{P}[A] + \mathbb{P}[B \setminus A] = \mathbb{P}[B].$$

$\square$

2 Lecture 2

2.1 Inclusion-Exclusion Principle

Theorem — *Inclusion-Exclusion Principle*
If $A, B \subseteq \Omega$ are events, then

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

Proof. Observe that we may write

$$A \cup B = [A \setminus (A \cap B)] \cup [A \cap B] \cup [B \setminus (A \cap B)],$$

where $A \cap B$, $B \setminus (A \cap B)$, and $A \setminus (A \cap B)$ are mutually exclusive. Hence

$$\begin{aligned}\mathbb{P}[A \cup B] &= \mathbb{P}[A \setminus (A \cap B)] + \mathbb{P}[B \setminus (A \cap B)] + \mathbb{P}[A \cap B] \\ &= (\mathbb{P}[A] - \mathbb{P}[A \cap B]) + (\mathbb{P}[B] - \mathbb{P}[A \cap B]) + \mathbb{P}[A \cap B] \\ &= \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].\end{aligned}$$

$\square$

Theorem — *Union Bound*
If $A_1, A_2, \ldots, A_n \subseteq \Omega$ are events, then

$$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] \leq \sum_{j=1}^{n} \mathbb{P}[A_j].$$

Proof. We proceed via proof by induction. Observe that for $n = 1$, we have $\mathbb{P}[A_1] \leq \mathbb{P}[A_1]$, which is obviously true. Suppose that this statements holds for some $k \geq 1$. Then

$$\begin{aligned}\mathbb{P}\left[\bigcup_{j=1}^{k+1} A_j\right] &= \mathbb{P}\left[\left(\bigcup_{j=1}^{k} A_j\right) \cup A_{k+1}\right] \\ &= \mathbb{P}\left[\bigcup_{j=1}^{k} A_j\right] + \mathbb{P}[A_{k+1}] - \mathbb{P}\left[\left(\bigcup_{j=1}^{k} A_j\right) \cap A_{k+1}\right] \\ &\leq \mathbb{P}\left[\bigcup_{j=1}^{k} A_j\right] + \mathbb{P}[A_{k+1}] \\ &\leq \sum_{j=1}^{k} \mathbb{P}[A_j] + \mathbb{P}[A_{k+1}] \\ &= \sum_{j=1}^{k+1} \mathbb{P}[A_j].\end{aligned}$$

Hence the statement holds for $k + 1$, and so holds for all natural numbers n. $\square$

2.2 Mutual Independence

Definition. *Independence*
We say that two events $A, B \subseteq \Omega$ are *independent* if

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

If two events are not independent, then we say that they are *dependent*.

Definition. *Mutual Independence*
We say that events $A_1, \ldots, A_n \subseteq \Omega$ are *mutually independent* if, given any $1 \leq k \leq n$ and $1 \leq j_1 < j_2 \cdots < j_k \leq n$ we have

$$\mathbb{P}\left[\bigcap_{\ell=1}^{k} A_{j_\ell}\right] = \prod_{\ell=1}^{k} \mathbb{P}[A_{j_\ell}].$$

3 Lecture 3

Theorem — *Multiplication Principle*
Suppose I run r *mutually independent* experiments so that

- The 1^{st} experiment has n_1 possible outcomes.
- The 2^{nd} experiment has n_2 possible outcomes.
- ...
- The r^{th} experiment has n_r possible outcomes.

The composite experiment then has $n_1 \cdot n_2 \cdot \dots \cdot n_r$ outcomes.

In some experiments we care about taking samples of size r from a set of n objects.

- We can seek *ordered* or *unordered* samples.
- We can do this *with* or *without* replacement.

Theorem. There are n^r possible choices of an *ordered* sample of size r from a set of n objects *with replacement*.

Proof. We run r experiments corresponding to each choice. For each choice, we have n possible outcomes because we are performing the choices with replacement. The Multiplication Principle tells us that there are $n \cdot n \cdot \dots \cdot n = n^r$ outcomes. □

Theorem. There are
$$_nP_r = \frac{n!}{(n-r)!}$$
ordered samples of size r from a set of n objects *without* replacement. The number $_nPr$ is known as the number of *permutations* of n objects, taken r at a time.

Proof. Each choice is an independent experiment:

- 1^{st} choice: n outcomes
- 2^{nd} choice: $n-1$ outcomes
- 3^{rd} choice: $n-2$ outcomes
  $\vdots$
- r^{th}: $n-(r-1)$ outcomes

Hence the composite experiment has
$$n \cdot (n-1) \cdot \dots \cdot (n-r+1) = {}_nP_r$$
outcomes. □

Theorem. There are
$$_nC_r = \frac{n!}{(n-r)!r!}$$
unordered samples of size r from a set of n objects *without* replacement.

Proof. From the previous theorem, there are $_nP_r$ ordered samples of size r from n objects without replacement. However, we have over counted because our sample will show up $r!$ times (in every possible permutation). Hence we divide by $r!$ to get

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!r!}. $$

$\square$

Note. ${}_nC_r = {}_nC_{n-r}$.

4 Lecture 4

- There are n^r *ordered* samples of size r from n objects *with replacement*.
- There are ${}_nP_r$ *ordered* samples of size r from n objects *without replacement*.
- There are ${}_nC_r = \frac{n!}{r!(n-r)!}$ *unordered* samples of size r from n objects *without replacement*.
- There are ${}_{n+r-1}C_r = \frac{(n+r-1)!}{r!(n-1)!}$ *unordered* samples of size r from n objects *with replacement*.

4.1 Distinguishable Permutations

- Suppose we are given n objects, but some of them are identical.
- How many *distinguishable* permutations of the n objects are there?

Theorem. Suppose you have:

- n_1 objects of type 1,
- n_2 objects of type 2,
- ...
- n_r objects of type r.

Let $n = n_1 + n_2 + \cdots + n_r$. Then the number of distinguishable permutations is

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}.$$

Proof. We have n locations.

- First choose n_1 locations for type 1 objects: ${}_nC_{n_1}$ choices.
- Then choose n_2 locations for type 2 objects: ${}_{n-n_1}C_{n_2}$ choices.
- $\cdots$
- Finally we choose n_r locations for type r objects: ${}_{n-n_1-\cdots-n_{r-1}}C_{n_r}$ choices.

Using the multiplication principle to take the product of all of these combinations, we have

$$\binom{n}{n_1, \dots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}.$$

□

Note. An alternate way to think about this theorem is to first consider how many regular permutations of n objects there are ($n!$), and then divide by how many possible times we over count ($n_k!$ for each $1 \leq k \leq r$).

4.2 The Binomial Theorem

Theorem — *Binomial Theorem*

If $n \geq 0$ then

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r},$$

where the *binomial coefficient* is

$$\binom{n}{r} = {}_nC_r.$$

Proof. If we multiply out $(x+y)^n = \underbrace{(x+y)\cdots(x+y)}_{n\text{ times}}$ without using the fact that multiplication is commutative, we see that the number of times $x^r y^{n-r}$ appears is equal to how many different ways there are to rearrange r "x" terms in n total terms. $\square$

Theorem. We have

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n.$$

Theorem. If $n, r \geq 0$ then

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{n_1 + \cdots + n_r = n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}.$$

Proof. Similar to the binomial theorem, we have that to get each term we just need to find the number of distinguishable permutations of n_1 terms of x_1, ..., n_r terms of x_r, which is our multinomial coefficient from before. $\square$

5 Lecture 5

5.1 Conditional Probability

Suppose that $A, B \subseteq \Omega$ are events. If we know that the event B occurs, how does this affect the probability that A occurs?

We write $\mathbb{P}[A]$ for the probability of A, and $\mathbb{P}[A \mid B]$ for the probability of A *conditioned on* B.

Example.

- Suppose that 5% of UCLA students take a math class and 6% take a physics class.
- Suppose that 80% of UCLA students that take a math class also take a physics class.
- If we know that a randomly chosen student takes a math class, what is the probability they take a physics class?

Let us define

$$\begin{aligned}\Omega &= \{\text{All UCLA students}\}, \\ A &= \{\text{Students taking a physics class}\}, \\ B &= \{\text{Students taking a math class}\}.\end{aligned}$$

Then we want to find $\mathbb{P}[A \mid B]$. Thus we have

$$\begin{aligned}\mathbb{P}[A \mid B] &= \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \\ &= \frac{0.05 \cdot 0.8}{0.05} \\ &= 0.8.\end{aligned}$$

Definition. *Conditional Probability*
Suppose A, B are events. Then the probability of A *conditioned on* B is given by

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

Theorem. If $B \subseteq \Omega$ is an event so that $\mathbb{P}[B] \neq 0$ then $\mathbb{P}[\cdot \mid B]$ is a probability measure. Precisely:

- $\mathbb{P}[\Omega \mid B] = 1$.

  Proof. Observe that

  $$\mathbb{P}[\Omega \mid B] = \frac{\mathbb{P}[\Omega \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B]}{\mathbb{P}[B]} = 1.$$

  □

- If $A_1, A_2, \ldots, A_n$ are mutually exclusive events then

  $$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j \,\middle|\, B\right] = \sum_{j=1}^{n} \mathbb{P}[A_j \mid B].$$

Proof. If $A_1, \dots, A_n$ are mutually exclusive, so are $A_1 \cap B, A_2 \cap B, \dots, A_n \cap B$. Then

$$\begin{aligned}\mathbb{P}\left[\bigcup_{j=1}^{n} A_j \,\middle|\, B\right] &= \frac{\mathbb{P}\left[\left(\bigcup_{j=1}^{n} A_j\right) \cap B\right]}{\mathbb{P}[B]} \\ &= \frac{\mathbb{P}[\bigcup_{j=1}^{n}(A_j \cap B)]}{\mathbb{P}[B]} \\ &= \sum_{j=1}^{n} \frac{\mathbb{P}[A_j \cap B]}{\mathbb{P}[B]} \\ &= \sum_{j=1}^{n} \mathbb{P}[A_j \mid B].\end{aligned}$$

□

- If $A_1, A_2, \dots, A_n$ are mutually exclusive events then

$$\mathbb{P}\left[\bigcup_{j=1}^{\infty} A_j \,\middle|\, B\right] = \sum_{j=1}^{\infty} \mathbb{P}[A_j \mid B].$$

Proof. The proof for this is very similar to the one above. □

Theorem. If A and B are independent events so that $\mathbb{P}[B] \neq 0$ then

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

Proof. Observe that

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

□

6 Lecture 6

Theorem — *The Law of Total Probability*
Let $A \subseteq \Omega$ be an event and $B_1, B_2 \dots, B_n \subseteq \Omega$ be mutually exclusive events so that $\mathbb{P}[B_j] \neq 0$ and

$$A \subseteq \bigcup_{j=1}^{n} B_j.$$

Then

$$\mathbb{P}[A] = \sum_{j=1}^{n} \mathbb{P}[A \mid B_j]\mathbb{P}[B_j].$$

Proof. As $A \subseteq \bigcup_{j=1}^{n} B_j$ we have

$$A = A \cap \left[\bigcup_{j=1}^{n} B_j\right] = \bigcup_{j=1}^{n} (A \cap B_j).$$

As all of the B_j are mutually exclusive, so are $A \cap B_j$. Hence

$$\begin{aligned}\mathbb{P}[A] &= \mathbb{P}\left[\bigcup_{j=1}^{n} (A \cap B_j)\right] \\ &= \sum_{j=1}^{n} \mathbb{P}[A \cap B_j] \\ &= \sum_{j=1}^{n} \mathbb{P}[A \mid B_j]\mathbb{P}[B_j].\end{aligned}$$

□

Note. The same result is true if we have a countable number of events $B_1, B_2, \dots$.

6.1 Bayes' Theorem

Theorem — *Bayes' Theorem*
If $A, B \subseteq \Omega$ are events so that $\mathbb{P}[A], \mathbb{P}[B] \neq 0$ then

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

Proof. By definition, we have

$$\begin{aligned}\mathbb{P}[B \mid A] &= \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]} \\ &= \frac{\mathbb{P}[A \mid B]\mathbb{P}[B]}{\mathbb{P}[A]}.\end{aligned}$$

□

Theorem — *Bayes' Theorem (Improved Version)*
If $A \subseteq \Omega$ is an event and $B_1, B_2, \ldots, B_n \subseteq \Omega$ are mutually exclusive events so that $\mathbb{P}[A], \mathbb{P}[B_j] \neq 0$ and
$$A \subseteq \bigcup_{j=1}^{n} B_j,$$
then for any $1 \leq k \leq n$ we have
$$\mathbb{P}[B_k \mid A] = \frac{\mathbb{P}[A \mid B_k]\mathbb{P}[B_k]}{\sum_{j=1}^{n} \mathbb{P}[A \mid B_j]\mathbb{P}[B_j]}.$$

Proof. The Law of Total Probability tells us that
$$\mathbb{P}[A] = \sum_{j=1}^{n} \mathbb{P}[A \mid B_j]\mathbb{P}[B_j].$$
Hence by Bayes' Theorem we have
$$\begin{aligned}\mathbb{P}[B_k \mid A] &= \frac{\mathbb{P}[A \mid B_k]\mathbb{P}[B_k]}{\mathbb{P}[A]} \\ &= \frac{\mathbb{P}[A \mid B_k]\mathbb{P}[B_k]}{\sum_{j=1}^{n} \mathbb{P}[A \mid B_j]\mathbb{P}[B_j]}.\end{aligned}$$
$\square$

7 Lecture 7

7.1 Discrete Random Variables

Definition. *Random Variable*
Given a set S, a *random variable* is a function $X\colon \Omega \to S$ satisfying certain properties. For the sake of this class, we will assume that all functions $X\colon \Omega \to S$ are random variables. Some notation follows:

$$\mathbb{P}[X = x] = \mathbb{P}\{\omega \in \Omega \mid X(\omega) = x\},$$
$$\mathbb{P}[X \in A] = \mathbb{P}\{\omega \in \Omega \mid X(\omega) \in A\}.$$

Definition. *Discrete Random Variable*
Let $X\colon \Omega \to S$ be a random variable. We say that X is a *discrete random variable* if $S \subseteq \mathbb{R}$ is a countable set. We define the *probability mass function* (PMF) of X to be the function $p_X\colon S \to [0,1]$ given by

$$p_X(x) = \mathbb{P}[X = x].$$

We define the *cumulative distribution function* (CDF) of X to be the function $F_X\colon \mathbb{R} \to [0,1]$ given by

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

We say that two random variables X, Y are *identically distributed* if they have the same CDF, and we write $X \sim Y$.

Definition. *Uniform Distribution*
Let $m \geq 1$. We say that a discrete random variable X is *uniformly distributed* on $\{1, 2, \dots, m\}$ and write $X \sim \text{Uniform}(\{1, 2, \dots, m\})$ if it has PMF

$$p_X(x) = \frac{1}{m} \quad \text{if } x \in \{1, 2, \dots, m\}.$$

If $X \sim \text{Uniform}(\{1, 2, \dots, m\})$, then it has CDF

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1, \\ \frac{k}{m} & \text{if } k \leq x < k + 1 \text{ and } k \in \{1, 2, \dots, m-1\}, \\ 1 & \text{if } x \geq m. \end{cases}$$

If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ and $A \subseteq \mathbb{R}$ is any set then

$$\mathbb{P}[X \in A] = \sum_{x \in A \cap S} p_X(x).$$

Proof. Since S is countable, we know that $A \cap S = \{a_1, a_2, \ldots, a_n\}$ (or $\{a_1, a_2, \ldots\}$). Hence

$$\begin{aligned}
\mathbb{P}[X \in A] &= \mathbb{P}[X \in \{a_1, \ldots, a_n\}] \\
&= \mathbb{P}\left[\bigcup_{j=1}^{n} \{X = a_j\}\right] \\
&= \sum_{j=1}^{n} \mathbb{P}[X = a_j] \\
&= \sum_{j=1}^{n} p_X(a_j) \\
&= \sum_{x \in A \cap S} p_X(x).
\end{aligned}$$

$\square$

If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ then

$$F_X(x) = \sum_{\substack{y \leq x \\ y \in S}} p_X(y).$$

Proof. Observe that

$$\begin{aligned}
F_X(x) &= \mathbb{P}[X \leq x] \\
&= \mathbb{P}[X \in (-\infty, x]] \\
&= \sum_{y \in (-\infty, x] \cap S} p_X(y) \\
&= \sum_{\substack{y \leq x \\ y \in S}} p_X(y).
\end{aligned}$$

$\square$

If X is a discrete random variable and $a < b$ then

$$\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a).$$

Proof. Observe that

$$\begin{aligned}
\mathbb{P}[a < X \leq b] &= \mathbb{P}[X \in (a, b]] \\
&= \sum_{x \in (a, b] \cap S} p_X(x) \\
&= \sum_{\substack{x \leq b \\ x \in S}} p_X(x) - \sum_{\substack{x \leq a \\ x \in S}} p_X(x) \\
&= F_X(b) - F_X(a).
\end{aligned}$$

$\square$

If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ then

$$\sum_{x \in S} p_X(x) = 1.$$

Proof. Observe that

$$\begin{aligned}\sum_{x\in S} p_X(x) &= \sum_{x\in \mathbb{R}\cap S} \\ &= \mathbb{P}[X \in R] \\ &= \mathbb{P}[\Omega] \\ &= 1.\end{aligned}$$

$\square$