# 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 $\Omega$, called the *sample space*.
2. A set $\mathcal{F}$ of subsets of $\Omega$ satisfying certain properties:
   - Elements of $\mathcal{F}$ are called *events*.
   - Events $A_1, A_2, \dots, A_k$ are called *mutually exclusive* if they are *pairwise disjoint*, i.e. if $i \neq j$ then $A_i \cap A_j = \varnothing$.
   - Events $A_1, A_2, \dots, A_k$ 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 $\Omega$ 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 $A_1, A_2, \dots, A_n$ are mutually exclusive events, then
   $$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] = \sum_{j=1}^{n} \mathbb{P}[A_j].$$
   - If $A_1, A_2, \dots$ 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 $\Omega$ 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 $\mathbb{P}[\varnothing] = 0$. □

**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

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

□

**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. □

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

*Proof.* From the previous theorem we have

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

□

# 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

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

□

**Theorem** — *Union Bound*
If $A_1, A_2, \dots, A_n \subseteq \Omega$ 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 $n$. □

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

The composite experiment then has $n_1 \cdot n_2 \cdot \cdots \cdot n_r$ 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 $n \cdot n \cdot \cdots \cdot n = n^r$ 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:

- $1^{\text{st}}$ choice: $n$ outcomes
- $2^{\text{nd}}$ choice: $n-1$ outcomes
- $3^{\text{rd}}$ choice: $n-2$ outcomes

  ⋮
- $r^{\text{th}}$: $n-(r-1)$ outcomes

Hence the composite experiment has

$$ n \cdot (n-1) \cdot \cdots \cdot (n-r+1) = {}_nP_r $$

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

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

□

**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 ${}_{n+r-1}C_r = \frac{n!}{r!(n-r)!}$ *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

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

*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

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

$\square$

**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. $\square$