

Lecture Notes

Kyle Chui

2022-01-04

Contents

1	Lecture 1	1
1.1	Properties of Probability	1
2	Lecture 2	3
2.1	Inclusion-Exclusion Principle	3
2.2	Mutual Independence	4
3	Lecture 3	5
4	Lecture 4	7
4.1	Distinguishable Permutations	7
4.2	The Binomial Theorem	7
5	Lecture 5	9
5.1	Conditional Probability	9
6	Lecture 6	11
6.1	Bayes' Theorem	11
7	Lecture 7	13
7.1	Discrete Random Variables	13
8	Lecture 8	16
8.1	Expected Value	16
8.1.1	Bernoulli Distribution	16
9	Lecture 9	18
9.1	Special Mathematical Expectations	18
9.2	Moment Generating Functions	19
10	Lecture 10	21
10.1	Binomial Random Variables	21
10.2	Geometric Random Variables	22
11	Lecture 11	25
11.1	Negative Binomial Distribution	25
11.2	Poisson Distribution	27

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, \dots, A_k are called *mutually exclusive* if they are *pairwise disjoint*, i.e. if $i \neq j$ then $A_i \cap A_j = \emptyset$.
 - 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 Ω are events.
3. A function $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$, 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. $\mathbb{P}[\emptyset] = 0$.

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

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

and so $\mathbb{P}[\emptyset] = 0$. □

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 $A' \cap A = \emptyset$, 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 $A \cap (B \setminus A) = \emptyset$. 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 $A_1, \dots, A_n \subseteq \Omega$ are *mutually independent* if, given any $1 \leq k \leq n$ and $1 \leq j_1 < j_2 < \dots < j_k \leq n$ 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 1st experiment has n_1 possible outcomes.
- The 2nd 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. \square

Theorem. There are

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

ordered samples of size r from a set of n objects *without* replacement. The number ${}_nP_r$ is known as the number of *permutations* of n objects, taken r at a time.

Proof. Each choice is an independent experiment:

- 1st choice: n outcomes
- 2nd choice: $n - 1$ outcomes
- 3rd 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. \square

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

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

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 $\mathbb{P}[A | B]$ 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 $\mathbb{P}[A | B]$. Thus we have

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

Definition. Conditional Probability

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

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

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

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

Proof. Observe that

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

□

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

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

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 $B_1, B_2, \dots, B_n \subseteq \Omega$ 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

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

Proof. The Law of Total Probability tells us that

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

Hence by Bayes' Theorem we have

$$\begin{aligned} \mathbb{P}[B_k | A] &= \frac{\mathbb{P}[A | B_k] \mathbb{P}[B_k]}{\mathbb{P}[A]} \\ &= \frac{\mathbb{P}[A | B_k] \mathbb{P}[B_k]}{\sum_{j=1}^n \mathbb{P}[A | B_j] \mathbb{P}[B_j]}. \end{aligned}$$

□

7 Lecture 7

7.1 Discrete Random Variables

Definition. *Random Variable*

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

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

Definition. *Discrete Random Variable*

Let $X: \Omega \rightarrow S$ 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 $p_X: S \rightarrow [0, 1]$ given by

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

We define the *cumulative distribution function* (CDF) of X to be the function $F_X: \mathbb{R} \rightarrow [0, 1]$ 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 $A \cap S = \{a_1, a_2, \dots, a_n\}$ (or $\{a_1, a_2, \dots\}$). Hence

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

□

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

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

□

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

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

□

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

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

□

8 Lecture 8

8.1 Expected Value

Definition. *Expected Value*

If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$, we define its *expected value* to be

$$\mathbb{E}[X] = \sum_{x \in S} x \cdot p_X(x)$$

provided the sum converges in a suitable sense.

Note. We often use the notation $\mu_X = \mathbb{E}[X]$ to denote the “mean” or “average” value.

8.1.1 Bernoulli Distribution

Definition. *Bernoulli Random Variable*

Let $p \in (0, 1)$. We say that a discrete random variable X is a *Bernoulli random variable* and write $X \sim \text{Bernoulli}(p)$ if it has PMF

$$p_X(x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

Then by definition we have

$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p.$$

Let X be a discrete random variable. If $a \in \mathbb{R}$ then $\mathbb{E}[a] = a$.

Proof. Let $g(x) = a$. Then

$$\begin{aligned} \mathbb{E}[a] &= \mathbb{E}[g(X)] \\ &= \sum_{x \in S} g(x) p_X(x) \\ &= \sum_{x \in S} a p_X(x) \\ &= a \sum_{x \in S} p_X(x) \\ &= a, \end{aligned}$$

since $\sum_{x \in S} p_X(x) = 1$. □

Let X be a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$. If $a, b \in \mathbb{R}$ and $g, h: S \rightarrow \mathbb{R}$ then

$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)].$$

Proof. By definition, we have

$$\begin{aligned} \mathbb{E}[ag(X) + bh(X)] &= \sum_{x \in S} (ag(x) + bh(x)) p_X(x) \\ &= a \sum_{x \in S} g(x) p_X(x) + b \sum_{x \in S} h(x) p_X(x) \\ &= a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]. \end{aligned}$$

□

Let X be a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$. If $g, h: S \rightarrow \mathbb{R}$ satisfy $g(x) \leq h(x)$ for all $x \in S$ then

$$\mathbb{E}[g(X)] \leq \mathbb{E}[h(X)].$$

Proof. Observe that

$$\begin{aligned} \mathbb{E}[g(X)] &= \sum_{x \in S} g(x)p_X(x) \\ &\leq \sum_{x \in S} h(x)p_X(x) && (g(x) \leq h(x), p_X(x) > 0) \\ &= \mathbb{E}[h(X)]. \end{aligned}$$

□

9 Lecture 9

9.1 Special Mathematical Expectations

Definition. *Moments*

Let X be a discrete random variable taking values in a discrete set $S \subseteq \mathbb{R}$ and $b \in \mathbb{R}$. We define the r^{th} moment of X about b to be

$$\mathbb{E}[(X - b)^r] = \sum_{x \in S} (x - b)^r p_X(x).$$

When $b = 0$ we refer to this simply as the r^{th} moment of X .

Definition.

Let X be a discrete random variable. We define the *variance* of X to be

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

whenever it converges. We use the notation $\sigma_X^2 = \text{var}(X)$. The *standard deviation* of X is $\sigma_X = \sqrt{\text{var}(X)}$.

Theorem. If X is a discrete random variable and $a, b \in \mathbb{R}$ then:

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $\text{var}(aX + b) = a^2 \text{var}[X]$

Proof. Observe that

$$\begin{aligned} \mathbb{E}[aX + b] &= \sum_{x \in S} (ax + b)p_X(x) \\ &= a \sum_{x \in S} x \cdot p_X(x) + \sum_{x \in S} b \cdot p_X(x) \\ &= a\mathbb{E}[X] + b. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} \text{var}(aX + b) &= \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] \\ &= \mathbb{E}[(aX + b - a\mathbb{E}[X] - b)^2] \\ &= \mathbb{E}[(aX - a\mathbb{E}[X])^2] \\ &= a^2 \cdot \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= a^2 \text{var}(X). \end{aligned}$$

□

Theorem. If X is a discrete random variable then

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof. Let $\mu_X = \mathbb{E}[X]$, so then

$$\begin{aligned}\text{var}(X) &= \mathbb{E}[(X - \mu_X)^2] \\ &= \mathbb{E}[X^2 - 2\mu_X X + \mu_X^2] \\ &= \mathbb{E}[X^2] - 2\mu_X \mathbb{E}[X] + \mu_X^2 \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

□

Example. Let $m \geq 1$ and $X \sim \text{Uniform}(\{1, 2, \dots, m\})$. What is $\text{var}(X)$?
By using a Gaussian sum, we can see that $\mathbb{E}[X] = \frac{m+1}{2}$. To compute $\mathbb{E}[X^2]$, we have

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{x=1}^m x^2 \cdot \frac{1}{m} \\ &= \frac{1}{m} \cdot \frac{m \cdot (m+1) \cdot (2m+1)}{6} \\ &= \frac{(m+1)(2m+1)}{6}.\end{aligned}$$

Hence the variance is

$$\frac{1}{6}(m+1)(2m+1) - \left(\frac{m+1}{2}\right)^2 = \frac{m^2 - 1}{12}.$$

9.2 Moment Generating Functions

Definition. *Moment Generating Function (MGF)*

If X is a discrete random variable we define the *Moment Generating Function (MGF)* of X to be the function

$$M_X(t) = \mathbb{E}[e^{tX}],$$

whenever it exists.

Theorem. Let X be a discrete random variable with MGF $M_X(t)$ that is well-defined and smooth for $t \in (-\delta, \delta)$. Then

$$\frac{d^r}{dt^r} M_X|_{t=0} = \mathbb{E}[X^r].$$

Proof. Let S be the set of outputs of X . Then

$$\begin{aligned}M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{x \in S} e^{tx} p_X(x).\end{aligned}$$

Hence we have

$$\begin{aligned}\frac{\partial^r}{\partial t^r} M_X(t) &= \frac{\partial^r}{\partial t^r} \sum_{x \in S} e^{tx} p_X(x) \\ &= \sum_{x \in S} x^r e^{tx} p_X(x).\end{aligned}$$

Therefore

$$\frac{\partial^r}{\partial t^r} M_X(t)|_{t=0} = \sum_{x \in S} x^r p_X(x) = \mathbb{E}[X^r]$$

□

Theorem. Let X be a discrete random variable with MGF $M_X(t)$ that is well-defined and smooth for $t \in (-\delta, \delta)$.

- $\frac{d}{dt} \ln M_X|_{t=0} = \mathbb{E}[X]$.
- $\frac{d^2}{dt^2} \ln M_X|_{t=0} = \text{var}(X)$.

10 Lecture 10

10.1 Binomial Random Variables

Definition. *Binomial Random Variable*

A *Bernoulli trial* is an experiment that has probability $p \in (0, 1)$ of success and probability $1 - p$ of failure.

- Suppose we run $n \geq 1$ independent, identical Bernoulli trials, and let X be the number of successes.
- Then we know that X is a discrete random variable taking values in the set $S = \{0, 1, \dots, n\}$.
- We say that X is a *Binomial random variable* with parameters n, p and write $X \sim \text{Binomial}(n, p)$.

Theorem. If $X \sim \text{Binomial}(n, p)$ then its PMF is

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{if } x \in \{0, 1, \dots, n\}.$$

Proof. Recall that $p_X(x) = \mathbb{P}[X = x]$. We know that there are $\binom{n}{x}$ ways to arrange exactly x “successes” in a total of n trials. Since each trial has a “success” rate of p , and a “failure” rate of $1 - p$, and they are independent, each arrangement has a $p^x (1-p)^{n-x}$ probability of occurring. Hence the total probability of any of the arrangements occurring (as they are mutually independent) is

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

□

Note. By the Binomial Theorem we have

$$\sum_{x=0}^n p_X(x) = (p + (1-p))^n = 1.$$

Theorem. If $X \sim \text{Binomial}(n, p)$, its MGF is

$$M_X(t) = (1 - p + pe^t)^n.$$

Proof. We compute

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tx}] \\ &= \sum_{x=0}^n e^{tx} p_X(x) \\ &= \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= (1 - p + pe^t)^n. \end{aligned}$$

□

Note. Recall that $(1 - p + pe^t)$ is the MFG of the Bernoulli random variable.

Theorem. If $X \sim \text{Binomial}(n, p)$ its mean is

$$\mathbb{E}[X] = np.$$

Proof. Recall that $M_X(t) = (pe^t + 1 - p)^n$ so $M'_X(t) = n(pe^t + 1 - p)^{n-1} \cdot pe^t$. Hence

$$\mathbb{E}[X] = M'_X(0) = n(p + 1 - p)^{n-1} \cdot p = np.$$

□

Note. Recall that p is the expected value of the Bernoulli random variable.

Theorem. If $X \sim \text{Binomial}(n, p)$ its variance is

$$\text{var}(X) = np(1 - p)$$

Proof. Recall that $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - (np)^2$. Furthermore, we just showed that

$$M'_X(t) = n(pe^t + 1 - p)^{n-1} \cdot pe^t,$$

so

$$M''_X(t) = n(n-1)(pe^t + 1 - p)^{n-2} \cdot (pe^t)^2 + n(pe^t + 1 - p)^{n-1} \cdot pe^t.$$

Hence

$$\begin{aligned} \mathbb{E}[X^2] &= M''_X(0) \\ &= n(n-1)(p + 1 - p)^{n-2} \cdot p^2 + n(p + 1 - p)^{n-1} \cdot p \\ &= n(n-1) \cdot p^2 + np. \end{aligned}$$

Therefore

$$\begin{aligned} \text{var}(X) &= n(n-1) \cdot p^2 + np - (np)^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 \\ &= np(1 - p). \end{aligned}$$

□

Note. Recall that $p(1 - p)$ is the variance of a regular Bernoulli random variable.

10.2 Geometric Random Variables

Definition. *Geometric Random Variable*

Suppose we repeatedly run independent, identical Bernoulli trials with probability $p \in (0, 1)$ of success.

- Let X be the trial on which we first achieve success.
- X is a discrete random variable taking values in the set $S = \{1, 2, 3, \dots\}$.
- We say that X is a *geometric random variable* with parameter p and write $X \sim \text{Geometric}(p)$.

Note. Sometimes geometric random variables are defined differently, but this is the definition that we will use for this class.

Theorem. If $X \sim \text{Geometric}(p)$ then its PMF is

$$p_X(x) = (1-p)^{x-1}p \quad \text{if } x \in \{1, 2, 3, \dots\}.$$

Proof. The probability that $X = x$ is the probability that we have $x-1$ failures, followed by a success. Since the events are independent, this gives us the desired result. \square

Note. For the infinite series version of this, we have

$$\begin{aligned} \sum_{x=1}^{\infty} p_X(x) &= \sum_{x=1}^{\infty} (1-p)^{x-1}p \\ &= p \sum_{x=1}^{\infty} (1-p)^{x-1} \\ &= p \cdot \frac{1}{1-(1-p)} \\ &= 1. \end{aligned}$$

Theorem. If $X \sim \text{Geometric}(p)$ then its MGF is

$$M_X(t) = \frac{e^t p}{1 - (1-p)e^t} \quad \text{if } t < -\ln(1-p).$$

Proof. We have

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{x=1}^{\infty} e^{tx} p_X(x) \\ &= \sum_{x=1}^{\infty} e^{tx} (1-p)^{x-1} p \\ &= e^t p \sum_{x=1}^{\infty} (e^t (1-p))^{x-1} \\ &= e^t p \cdot \frac{1}{1 - (e^t (1-p))} \\ &= \frac{e^t p}{1 - (1-p)e^t}, \end{aligned}$$

as desired. \square

Note. The condition comes from the denominator, where

$$(1-p)e^t < 1.$$

Theorem. If $X \sim \text{Geometric}(p)$ then its mean is

$$\mathbb{E}[X] = \frac{1}{p}.$$

Proof. From the previous theorem, we showed that

$$M_X(t) = \frac{e^t p}{1 - e^t(1 - p)}.$$

Hence

$$\begin{aligned} \ln M_X(t) &= \ln \left(\frac{e^t p}{1 - e^t(1 - p)} \right) \\ &= \ln(e^t p) - \ln(1 - e^t(1 - p)) \\ &= t + \ln p - \ln(1 - e^t(1 - p)). \end{aligned}$$

Thus we have that

$$(\ln M_X)' = 1 - \frac{1}{1 - e^t(1 - p)} \cdot (p - 1)e^t = \frac{1}{1 - e^t(1 - p)}.$$

Therefore

$$\mathbb{E}[X] = (\ln M_X)'(0) = \frac{1}{p}.$$

□

Theorem. If $X \sim \text{Geometric}(p)$ then its variance is

$$\text{var}(X) = \frac{1 - p}{p^2}.$$

Proof. From the last problem we have that

$$(\ln M_X)' = \frac{1}{1 - e^t(1 - p)},$$

so

$$(\ln M_X)'' = -\frac{1}{(1 - e^t(1 - p))^2} \cdot (p - 1)e^t = \frac{e^t(1 - p)}{(1 - e^t(1 - p))^2}.$$

Therefore

$$\text{var}(X) = (\ln M_X)''(0) = \frac{1 - p}{p^2},$$

as desired.

□

11 Lecture 11

11.1 Negative Binomial Distribution

Definition. *Negative Binomial Distribution*

Suppose we repeatedly run independent, identical Bernoulli trials with probability $p \in (0, 1)$ of success.

- Let $r \geq 1$ and let X be the trial on which we first achieve the r^{th} success.
- X is a discrete random variable taking values in the set $S = \{r, r + 1, r + 2, \dots\}$.
- We say that X is a *negative binomial random variable* with parameters r, p and write $X \sim \text{Negative Binomial}(r, p)$.

Note. If we ask about a negative binomial with parameter $(1, p)$, we see that this should be the same as a Geometric random variable.

Theorem. If $X \sim \text{Negative Binomial}(r, p)$, then its PMF is

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \text{if } x \in \{r, r+1, \dots\}$$

Proof. If we want to get our r^{th} success on the x^{th} trial, then we must have gotten $r-1$ successes in the last $x-1$ trials, and a success on the last trial. Notice that the former is similar to a binomial distribution. Hence the PMF should be

$$\begin{aligned} p_X(x) &= \binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)} \cdot p \\ &= \binom{x-1}{r-1} p^r (1-p)^{x-r}. \end{aligned}$$

□

Theorem. If $r \geq 1$ is an integer and $0 < s < 1$ then

$$\left(\frac{1}{1-s} \right)^r = \sum_{x=r}^{\infty} \binom{x-1}{r-1} s^{x-r}.$$

Proof. Let $g(s) = (1-s)^{-r}$ be the left hand side of the above. We apply Taylor's theorem:

$$g(s) = \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \cdot \frac{d^\ell g}{ds^\ell}(0) s^\ell.$$

Observe that:

- If $\ell = 0$ then $\frac{d^\ell g}{ds^\ell}(0) = g(0) = 1$
- If $\ell = 1$ then $\frac{d^\ell g}{ds^\ell}(0) = g'(0) = r$
- If $\ell = 1$ then $\frac{d^\ell g}{ds^\ell}(0) = g'(0) = r(r-1)$

In general,

$$\frac{d^\ell g}{ds^\ell}(0) = r(r+1) \cdots (r+\ell-1) = \frac{(r+\ell-1)!}{(r-1)!}.$$

Hence

$$g(s) = \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \cdot \frac{(r+\ell-1)!}{(r-1)!} \cdot s^{\ell}.$$

If we let $x = r + \ell$, we have $\ell = x - r$, so

$$\begin{aligned} g(s) &= \sum_{x=r}^{\infty} \frac{1}{(x-r)!} \cdot \frac{(x-1)!}{(r-1)!} \cdot s^{x-r} \\ &= \sum_{x=r}^{\infty} \binom{x-1}{r-1} s^{x-r}. \end{aligned}$$

□

Note. Recall that if $X \sim \text{Negative Binomial}(r, p)$ then

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}.$$

Hence

$$\begin{aligned} \sum_{x=r}^{\infty} p_X(x) &= \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= p^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} (1-p)^{x-r} \\ &= p^r \cdot (1 - (1-p))^{-r} \\ &= 1. \end{aligned}$$

Theorem. If $X \sim \text{Negative Binomial}(r, p)$ then its MGF is

$$M_X(t) = \left(\frac{e^t p}{1 - (1-p)e^t} \right)^r \quad \text{if } t < -\ln(1-p).$$

Proof. We compute that the MGF is

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{x=r}^{\infty} e^{tx} p_X(x) \\ &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= e^{tr} p^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} (e^t(1-p))^{x-r} \\ &= e^{tr} p^r \cdot \frac{1}{(1 - (e^t(1-p)))^r} \\ &= \left(\frac{e^t p}{1 - (1-p)e^t} \right)^r. \end{aligned}$$

□

Theorem. If $X \sim \text{Negative Binomial}(r, p)$ then

$$\mathbb{E}[X] = \frac{r}{p},$$

$$\text{var}(X) = \frac{r(1-p)}{p^2}.$$

Proof. Let

$$h(t) = \ln \left(\frac{pe^t}{1 - e^t(1-p)} \right).$$

Since X is a negative binomial random variable, its moment generating function is geometric. We know that for geometric random variables, $h'(0) = \frac{1}{p}$ and $h''(0) = \frac{1-p}{p^2}$. Hence if $X \sim \text{Negative Binomial}(r, p)$ then

$$\begin{aligned} \ln M_X(t) &= \ln \left(\frac{pe^t}{1 - e^t(1-p)} \right)^r \\ &= r \cdot h(t). \end{aligned}$$

Therefore

$$\mathbb{E}[X] = \frac{r}{p},$$

$$\text{var}(X) = \frac{r(1-p)}{p^2}.$$

□

11.2 Poisson Distribution

Definition. *Poisson Random Variable*

We make the following assumptions about arrivals in a given time interval:

- If the time intervals $(a_1, b_1], (a_2, b_2], \dots, (a_n, b_n]$ are disjoint then the number of arrivals in each time interval are independent.
- If $h = b - a$ is sufficiently small then the probability of exactly one arrival in the time interval $(a, b]$ is λh .
- If $h = b - a$ then the probability of having more than one arrival in the time interval $(a, b]$ converges to 0 as $h \rightarrow 0$.

An arrival process satisfying these assumptions is called an *approximate Poisson process*. If we take X to be the number of arrivals in a unit of time, then we call X a *Poisson random variable* and write $X \sim \text{Poisson}(\lambda)$.

Theorem. If $X \sim \text{Poisson}(\lambda)$ then it has PMF

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{if } x \in \{0, 1, 2, \dots\}.$$

Proof. Let X be the number of arrivals in a unit interval with width $\frac{1}{n}$. As n gets very large, we can think of the number of arrivals in each interval to be either 0 or 1 (Bernoulli trial with success rate

$\frac{\lambda}{n}$). Thus the overall setup looks like a Binomial distribution, so

$$X \approx \text{Binomial}\left(n, \frac{\lambda}{n}\right).$$

We now just need to take $n \rightarrow \infty$. Hence

$$\begin{aligned} p_X(x) &= \lim_{n \rightarrow \infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{x-n} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^x}{x!} \cdot \frac{n \cdot (n-1) \cdots (n-x+1)}{n \cdot n \cdots n} \cdot \left(1 - \frac{\lambda}{n}\right)^{x-n} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^{x-n} \\ &= \frac{\lambda^x}{x!} \cdot e^{-\lambda}. \end{aligned}$$

□