# Probability Theory

Raymond Bian

February 6, 2024

## Contents


**1 Basics of Probability** **1**
- 1.1 Basics of Set Theory . . . . . . . . 1
- 1.2 Back to Probability . . . . . . . . . 1
- 1.3 Counting . . . . . . . . . . . . . . . 2
- 1.4 Back to Probability Again . . . . . . 3
- 1.5 Measure Theory . . . . . . . . . . . 3
- 1.6 Properties of Event Spaces . . . . . 3
- 1.7 Examples . . . . . . . . . . . . . . . 4
- 1.8 Conditional Probability . . . . . . . 4
- 1.9 Bayes Theorem . . . . . . . . . . . 5

**2 Random Variables** **6**

**3 Multivariate Probability** **9**

**4 Probability Generating Functions** **12**


### Lecture 1: Intro to Probability

# 1 Basics of Probability

What data do you need to specify probability? You need the **set of all outcomes**, a list of everything that could possibly occur as a consequence, and the likelihood of each event.

> **Example.** For a roll of a dice, the set of all outcomes would be $\{1, 2, 3, 4, 5, 6\}$. The list could include things like "the result is 3", or "the result is $\geq 4$", and the likelihood would be $\frac{1}{6}$ for each of the results.

## 1.1 Basics of Set Theory

> **Definition 1.** A **set** is an unordered collection of elements. **Elements** are objects within sets.

> **Definition 2.** A set $A$ is a **subset** of a set $B$ if $a \in A \Rightarrow a \in B$

> **Definition 3.** The **union** of two sets $A$ and $B$ is the collection of elements that are in $A$ or $B$.

> **Definition 4.** The **intersection** of two sets $A$ and $B$ is the collection of elements that are in both $A$ and $B$.

> **Definition 5.** The **complement** of a set $A$ is everything not in $A$.

> **Definition 6.** A **finite set** is a set with finite number of elements.

> **Definition 7.** The **cartesian** product of two sets $A$ and $B$ denoted $A \times B$ is
> $$\{(a, b) \colon a \in A \wedge b \in B\}.$$
> Then, $|A \times B| = |A| \cdot |B|$.

## 1.2 Back to Probability

> **Definition 8.** A **sample space** is the set of al possible outcomes in an experiment.

> **Example.** The sample space $\Omega$ for a coin flip is $\{H, T\}$.

Note that **events** are just subsets of the sample space, and **elementary events** are just elements of the sample space.

> **Example.** For a dice roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, some events could be $\{1, 2\}$, $\{3, 6\}$, $\{3\}$. There are a total of $2^6$ events.

> **Definition 9.** If $\Omega$ is a finite set, a probability $P$ on $\Omega$ is a function: $P \colon 2^\Omega \to [0, 1]$ such that $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$.

> **Lemma 1.** If $A_1, \ldots, A_\alpha \subset \Omega$ are disjoint, $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$.

**Proposition 1.** Let $A = \{a_1, a_2, \ldots a_l\}$ such that $a_i$ are elementary events. Then,

$$\mathbb{P}(A) = \sum_{i=1}^{l} \mathbb{P}(\{a_i\}).$$

**Example.** For the dice roll, if $A = \{1, 3, 5\}$, then $\mathbb{P}(A) = 3 \cdot \frac{1}{6} = \frac{1}{2}$.

**Definition 10.** **Equiprobable outcomes**: Let's say we have the set $\Omega = \{\omega_1, \ldots, \omega_N\}$ and $\mathbb{P}(\omega_i) = \mathbb{P}(\omega_j)$ for all $i$ and $j$. Then, $\mathbb{P}(\omega) = \frac{1}{N}$ for all $\omega \in \Omega$ and $\mathbb{P}(A) = \frac{|A|}{N}$. In other words, when outcomes are probable,

$$\mathbb{P}(\text{event}) = \frac{\text{number of outcomes for that event}}{\text{number of possible outcomes}}.$$

## 1.3 Counting

Suppose 2 experiments are being performed. Let's say that experiment 1 has $m$ possible outcommes, and experiment 2 has $n$ possible outcomes. Then together, there are total of $n \cdot m$ total outcomes.

**Example.** Rolling a dice and then flipping a coin, how many possible outcomes are there?

**Explanation.** You have $6 \cdot 2 = 12$ outcomes.

**Example.** Let's say you have a college planning committee that consists of 3 freshman, 4 sophomores, 5 juniors, and 2 seniors. How many ways are there to select a subcommittee of 4 with one person from each grade?

**Explanation.** There are 4 events with 3, 4, 5, and 2 possible outcomes for each. Therefore, there are $3 \cdot 4 \cdot 5 \cdot 2 = 120$ total subcommittees.

**Example.** How many 7-place license plates are there if the first 3 are letters and the last 4 are numbers?

**Explanation.** There are $26^3 \cdot 10^4$ license plates.

**Definition 11.** A **permutation** is an ordering of elements in a set. The number of ways to order $n$ elements is given by $n!$.

**Example.** Alex has a bunny ranch with 10 bunnies. They are going to run an obstacle course and ranked 1-10 based on completion time. How many possible rankings are there (no ties)?

**Explanation.** There are 10! possible rankings.

**Example.** Assume 6 bunnies have straight ears and 4 have floppy ears. We rank the bunnies separately. How many possible rankings are there?

**Explanation.** There are $6! \cdot 4!$ possible outcomes.

**Definition 12.** A **combination** denotes the number of ways to choose $k$ elements from $n$ total elements (counting subsets).

**Example.** How many ways are there to pick a 2 person team from a set of 5 people?

**Explanation.** There are $C(5, 2) = \binom{5}{2} = \frac{5!}{2! \cdot 3!} = 10$ ways.

**Example.** How many committees consisiting of 2 women and 3 men can be formed from a group of 5 women and 7 men?

**Explanation.** We have $C(5, 2) \cdot C(7, 3)$ possible committees.

**Example.** What if two of the men do not want to serve on the committee together?

**Explanation.** The number of ways to choose the women stays the same. However, for the men we must subtract the number of committees that have both men. Therefore, we have $C(5, 2) \cdot (C(7, 3) - C(5, 1))$ possible committees.

**Example.** How many ways can we divide a 10 person class into 3 groups, sizes 3, 3, and 4?

**Explanation.** We just have 3 events, multiplying: $C(10, 3) \cdot C(7, 3) \cdot C(4, 4)$.

**Definition 13.** This is known as a **multinomial**, and is given by

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

It counts the number of ways to partition a set of size $n$ into sets of sizes $n_1, n_2 \ldots, n_r$.

## 1.4 Back to Probability Again

**Example.** Flip 10 fair coins. What is the likelihood of flipping 3 heads?

**Explanation.** Number of events of 3 heads is $C(10, 3)$. Total number of events is $2^{10}$. Therefore,

$$\mathbb{P}(10 \text{ heads}) = \frac{C(10, 3)}{2^{10}}.$$

In general, we have $\sum_{k=0}^{n} \mathbb{P}(k \text{ heads}) = 1$. In other words,

$$\frac{1}{2^{10}} \cdot \sum_{k=0}^{10} \binom{10}{k} = 1.$$

such that

$$\sum_{k=0}^{10} \binom{10}{k} = 2^{10}.$$

More generally,

**Definition 14.** The **binomial theorem** states that for all $x, y \in \mathbb{R}$, $n \geq 1$, $n \in \mathbb{N}$,

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

**Example.** Rolling 10 dice, what is the likelihood of exactly 2 outcomes each of 1,2,3,4, 1 outcome of 6, and 1 outcome of 5.

**Explanation.** There are total $6^{10}$ outcomes, and there are $\binom{10}{2,2,2,2,1,1}$ desired outcomes. Therefore, the probability of this event is $\frac{\binom{10}{2,2,2,2,1,1}}{6^{10}}$.

**Definition 15.** The **multinomial theorem** states that $(x_1 + \ldots + x_r)^n =$

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

## 1.5 Measure Theory

This is just a generalization of what we have seen before.

**Definition 16.** Let $\mathcal{F} \subset 2^{\Omega}$ be an "event space". A mapping $P : \mathcal{F} \to \mathbb{R}$ is a **probability measure** on $(\Omega, \mathcal{F})$ if

- $\mathbb{P}(A) \geq 0 \quad \forall A \in \mathcal{F}$
- $\mathbb{P}(\varnothing) = 0, \mathbb{P}(\Omega) = 1$
- If $A_1, A_2, \ldots$ are disjoint,

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

# Lecture 2: More Probabiliy

## 1.6 Properties of Event Spaces

**Definition 17.** A collection $\mathcal{F}$ of subsets of the sample space $\Omega$ is called an **event space** if

- $\mathcal{F}$ is non-empty.
- if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$.
- if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

**Theorem 1.** If $A \in \mathcal{F}$, then $\mathbb{P}(A) + \mathbb{P}(\Omega \setminus A) = 1$

**Proof.** Notice that $A$ and $\Omega \setminus A$ are disjoint. And, that $A \cup (\Omega \setminus A) = \Omega$. Then,

$$\mathbb{P}(A \cup (\Omega \setminus A)) = \mathbb{P}(\Omega) = 1.$$

∎

**Theorem 2.** If $A, B \in \mathcal{F}$ then $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$.

**Proof.** Note that $A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$. This is a union of disjoint sets, such that $\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$. Then, we have $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$, of which the RHS simplifies to $\mathbb{P}(A) + \mathbb{P}(B)$. ∎

**Theorem 3.** If $A, B \in \mathcal{F}$, and $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

**Proof.** We wish to show $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$. Then, $B = (B \setminus A) \cup (B \cap A) = (B \setminus A) \cup A$, such that $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A) \geq \mathbb{P}(A)$ because $\mathbb{P}(B \setminus A) \geq 0$. ∎

## 1.7 Examples

**Example.** What is the probability that one is dealt a full house?

**Explanation.** This is the number of ways one can get a full house, divided by the total number of poker hands (5 card). The total number of poker hands is $\binom{52}{5}$. The number of full houses is $\frac{52\cdot\binom{3}{2}\cdot 48\cdot 3}{2!3!}$. Another way we can count the number of full houses is $\binom{13}{1}\cdot\binom{4}{3}\cdot\binom{12}{1}\cdot\binom{4}{2}$. The result of the division is our answer.

**Example.** A box contains 3 marbles, 1 red 1 green and 1 blue. Consider an experiment that cnsists of us taking 1 marble, replacing it, and drawing another marble. What is the sample space?

**Explanation.**

$$\begin{aligned}\Omega = \{&(r,r),(r,b),(r,g),\\ &(b,r),(b,g),(b,b),\\ &(g,r),(g,g),(g,b)\}.\end{aligned}$$

**Example.** What about if we don't replace the first marble?

**Explanation.** Everything without $(r,r),(b,b),(g,g)$.

**Example.** What is the probability of being dealt a flush?

**Explanation.** This is just number of flushses divided by number of poker hands. The number of flushes is $\binom{4}{1}\cdot\binom{13}{5}$.

**Example.** What is the probability of being dealt a straight?

**Explanation.** We can do the probability of any straight, minus probability of straight flush. The number of straights is 10 number-wise. Therefore, the number of straights is $10\cdot(4^5-4)$. The probability can be then calculated.

**Example.** An urn contains $n$ balls. If $k$ balls are withdrawn one at a time, what is the probability that a special ball is chosen?

**Explanation.** $\mathbb{P}(\text{special}) = 1-\mathbb{P}(\text{special}^c)$. If the special ball is not chosen, it would be $\frac{(n-1)!}{(n-k-1)!}$. The total number of withdrawings is $\frac{n!}{k!}$. Then, the total probability is $1-\frac{n-k}{n}$.

**Example.** If $n$ people are present in a room, what is the prob that no two celebrate their birthday on the same date? How large must $n$ be such that this probability is $<\frac{1}{2}$.

**Explanation.** $\mathbb{P}(\text{no people with same birthday})$ is the number of no same birthday situations divided by the number of possibilities. Total possibilities is $365^n$. No same birthday situatons is $\mathbb{P}(365,n) = \frac{365!}{(365-n)!}$. For the second question, $n=23$.

## 1.8 Conditional Probability

**Definition 18.** If $A,B\in\mathcal{F}$ and $\mathbb{P}(B)>0$ then the **conditional probability** if $A$ given $B$ is denoted by $\mathbb{P}(A\mid B)$ and defined by

$$\mathbb{P}(A\mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$

**Theorem 4.** If $B\in\mathcal{F}$ and $\mathbb{P}(B)>0$ then $(\Omega,\mathcal{F},\mathbb{Q})$ is a pobability space where $\mathbb{Q}:\mathcal{F}\to\mathbb{R}$ is defined by $\mathbb{Q}(A)=\mathbb{P}(A\mid B)$

**Example.** Let's say a coin is flipped twice. What is the conditional probability that both flips land on heads, given that the first flip lands on heads?

**Explanation.** $\frac{\mathbb{P}(\text{two heads}\cap\text{first heads})}{\mathbb{P}(\text{first heads})} = \frac{\mathbb{P}(\text{two heads})}{\mathbb{P}(\text{first heads})}$. This is just $\frac{1}{2}$.

**Example.** What if given at least one lands on heads?

**Explanation.** $\frac{\mathbb{P}(\text{two heads}\cap\text{at least one head})}{\mathbb{P}(\text{at least one head})} = \frac{\mathbb{P}(\text{two heads})}{\mathbb{P}(\text{at least one head})} = \frac{2}{3}$.

**Example.** In the card game bridge, the 52 cards are dealt equally. If North and South have a total of 8 spades among them, what is the probability that East has 3 of the 5 remaining spades?

**Explanation.** No rule: $\mathbb{P}(\text{E has 3 spades}) = \frac{\binom{5}{3}\cdot\binom{21}{10}}{\binom{26}{13}}$.

**Theorem 5.** Probability of intersection of three sets (insert from canvas).

**Definition 19.** We call two events $A, B$ **independent** if the occurence of one does not affect the other. Formally,

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \text{ and } \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

We can also check that $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

**Example.** Flip three fair coins. What is likelihood that all three come up heads?

**Explanation.** With the sample space approach: $\Omega = \{H, T\}^3$. Of 8 total elementary events, 1 has three heads, so the probability is $\frac{1}{8}$.

With independence: we know that each event is independent, and all events are $\frac{1}{2}$, so the probability is $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$.

**Definition 20.** Independence can be expanded to more than just two events (insert from canvas). However, note that events can be pairwise independent, but may not be all together independent.

**Lemma 2.**

$$\mathbb{P}(B \mid A) = \mathbb{P}(A \mid B)\frac{\mathbb{P}(B)}{\mathbb{P}(A)}.$$

**Proof.** The RHS is the same as $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \cdot \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B \mid A)$. ■

**Example.** There are $n$ balls that are sequentially chosen without replacement from $r$ red balls and $b$ blue balls. Given that $k$ of the $n$ balls are blue, what is the conditional probability that the first chosen is blue?

**Explanation.**

$$\begin{aligned} &\mathbb{P}(\text{first is blue} \mid k \text{ are blue}) \\ &= \mathbb{P}(k \text{ are blue} \mid \text{first is blue}) \\ &\cdot \frac{\mathbb{P}(\text{first is blue})}{\mathbb{P}(k \text{ are blue})} \cdots \end{aligned}$$

## Lecture 3: Bayes Theorem and Random Variables

Continuing on with conditional probabiltiy from last time,

**Example.** A total of $n$ balls are sequentially and randomly chosen without replacement from an urn containing $r$ red balls and $b$ blue balls ($n \leq r + b$). Given that $k$ of the $n$ balls are blue, what is the conditional probability that the first ball chosen is blue?

**Explanation.** We can use Lemma 2. Then, we have

$$\begin{aligned} \mathbb{P}(\text{first blue}) &= \frac{b}{r+b} \\ \mathbb{P}(\text{first } k \text{ are blue}) &= \frac{\binom{n}{k} P(b,k) P(r, n-k)}{P(r+b, n)}. \end{aligned}$$

$$\mathbb{P}(k-1 \text{ of remaining } n-1 \text{ slots are blue.}) =$$

$$\frac{\binom{n-1}{k-1} P(b-1, k-1) \cdot P(r, n-k)}{P(r+b-1, n-1)}.$$

$$\mathbb{P}(k-1 \text{ of rest } n-1 \text{ are blue}) \cdot \frac{\mathbb{P}(\text{first blue})}{\mathbb{P}(\text{first } k \text{ are blue})}.$$

will then be our answer.

### 1.9 Bayes Theorem

**Definition 21.** A **partition** of $\Omega$ is a collection $\{B_i : i \in I\}$ of disjoint events with union $\bigcup_i B_i = \Omega$.

**Theorem 6.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $\{B_1, B_2, \ldots\}$ is such a parition with $\mathbb{P}(B_i) > 0$, then

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) \quad \text{for } A \in \mathcal{F}.$$

**Example.** Flip a fair coin. If heads, roll a 6-sided fair die. If tails, roll two 4-sided dice and sum the total. What is the overall likelihood of an outcome of 3?

**Explanation.** Look at the event tree, and count the probabilities. The heads case is $\frac{1}{2} \cdot \frac{1}{6}$ and the tails case is $\frac{1}{2} \cdot \frac{1}{8}$. This is an informal Bayes Theorem.

**Theorem 7.** We can also rearrange Bayes' Theorem to yield

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

## 2 Random Variables

**Definition 22.** A **random variable** on $(\Omega, \mathbb{P})$, $|\Omega| < \infty$ is a function $X : \Omega \to \mathbb{R}$.

**Notation.** $\{X = a\} = \{\omega \in \Omega : X(\omega) = a\} = \ldots = X^{-1}(a)$.

**Example.** 3 balls are to be selected without replacement from an urn containing 20 balls numbered 1 to 20. What is the probability that at least one of the balls that are drawn has a number as large or larger than 17?

**Explanation.** $\Omega = \{1, 2, 3, \ldots, 20\}$. $|\Omega| = \binom{20}{3}$. Let our random variable $X : \Omega \to \mathbb{R}$, $X =$ largest of the three values. Let $E = \{X \geq 17\}$. Then, $\mathbb{P}(E) = 1 = \mathbb{P}(E^c) = \mathbb{P}(\text{all } < 17)$.

$$\mathbb{P}(\text{all } < 17) = \frac{|E^c|}{|\Omega|} = \frac{\binom{16}{3}}{\binom{20}{3}}$$

$$\mathbb{P}(E) = 1 - \mathbb{P}(E^c) = 1 - \frac{\binom{16}{3}}{\binom{20}{3}}.$$

**Definition 23.** $X$ is called **discrete** if $\exists$ a countable set $S \subset \mathbb{R}$ such that $\mathbb{P}(X \in S) = 1$.

**Definition 24.** The **probability mass function** $p(a) = \mathbb{P}(X = a)$ is positive for most a countable number of values of $a$.

**Example.** The pmf of a random variable $X$ is given by $p_X(i) = \frac{c\lambda^i}{i!}$, $i = 0, 1, 2 \ldots$ where $\lambda$ is some positive value. What is $\mathbb{P}(X = 0)$ and $\mathbb{P}(X > 2)$?

**Explanation.**

$$\sum p_X(i) = \sum_{i=0}^{\infty} \frac{c\lambda^i}{i!} = 1$$

$$\Rightarrow X = \frac{1}{\sum_{k=0}^{\infty} \frac{\lambda^i}{i!}} = e^{-\lambda}.$$

Then, $\mathbb{P}(X = 0) = p_X(0) = \frac{c\lambda^0}{0!} = c = e^{-\lambda}$.

Also, $\mathbb{P}(X > 2) = 1 - \mathbb{P}(X \leq 2)$.

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right).\end{aligned}$$

**Definition 25.** If $X$ is a discrete random variable, the **expectation** of $X$ is denoted by $\mathbb{E}(X)$ and is defined by

$$\mathbb{E}(X) = \sum_{x \in \operatorname{Im} X} x\mathbb{P}(X = x).$$

**Example.** We say that $I$ is an indicator variable for the event $A$ if

$$I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}.$$

Find $\mathbb{E}(I)$.

**Explanation.**

$$\begin{aligned}\mathbb{E}(I_A) &= 0 \cdot \mathbb{P}(I_A = 0) + 1 \cdot \mathbb{P}(I_A = 1) \\ &= \mathbb{P}(I_A = 1) \\ &= \mathbb{P}(\{\omega \in \Omega : I_A(\omega) = 1\}) \\ &= \mathbb{P}(\{\omega \in A\}) \\ &= \mathbb{P}(A).\end{aligned}$$

**Example.** A class of 120 students is driven in 3 buses to a performance, with 36, 40, and 44 students in the busees. Let $X$ denote the number of students on the bus of a randomly chosen student. Find $\mathbb{E}(X)$.

**Explanation.** Note that $\mathbb{P}(B_1) = \frac{36}{120}$, $\mathbb{P}(B_2) = \frac{40}{120}$ and $\mathbb{P}(B_3) = \frac{44}{120}$. Then,

$$\begin{aligned}\mathbb{E}(X) &= 36 \cdot \frac{36}{120} + 40 \cdot \frac{40}{120} + 44 \cdot \frac{44}{120} \\ &= \frac{36^2 + 40^2 + 44^2}{120}\end{aligned}$$

.

### Lecture 4: Random Variables and Expected Values

**Proposition 2.** If $X$ is a discrete random variables that takes on one of the values $x_i$, $I \geq 1$, with respective proabilities $p(x_i)$, then, for any real valued function $g$,

$$\mathbb{E}(g(X)) = \sum_i g(x_i)p(x_i).$$

In other words, $g(X)$ is also a random variable.

**Proof.** Done with a change of variables. ■

**Example.** Suppose $t$ units of a product are ordered, and $X$ = number of units sold is a random variable. Assume a net profit of $b$ per unit and a net loss of $l$ per unit left unsold. Compute *expected profit*.

**Explanation.** Our profit function is then $\gamma = bX - l(t - X)$. Then, $\mathbb{E}(\gamma) = \mathbb{E}(g(X))$ where $g(X) = (b + l)X - lt$. Then we have

$$\begin{aligned}\mathbb{E}(g(X)) &= \sum_{x \in \mathrm{Im} X} g(x) \cdot p_X(x) \\ &= (b + l) \underbrace{\sum_{x \in \mathrm{Im} X} x \cdot p_X(x)}_{\mathbb{E}(X)} \\ &\quad - lt \underbrace{\sum_{x \in \mathrm{Im} X} p_X(x)}_{1} \\ &= (b + l)\mathbb{E}(X) - lt.\end{aligned}$$

**Definition 26.** If $X$ is a random variable with mean $\mu$, then the **variance** of $X$, denoted by $\mathrm{Var}(X)$ is defined by

$$\mathrm{Var}(X) = \mathbb{E}((X - \mu)^2).$$

**Proposition 3.** $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

**Proof.** We have

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}((X - \mathbb{E}(X)^2)) \\ &= \mathbb{E}(X^2 - 2x \cdot \mathbb{E}(X) + (\mathbb{E}(X))^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X) + (\mathbb{E}(X))^2 \\ &\qquad (\mathbb{E}(c) = c \text{ for constant } c) \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2.\end{aligned}$$

■

**Proposition 4.** If $X$ is a discrete random variable with finitely many values, then $\mathrm{Var}(x) = 0 \Leftrightarrow X \equiv \mathbb{E}(X)$.

**Proof.** ($\Leftarrow$) Suppose $X = \mathbb{E}(X)$ Then,

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{i=1}^{n} x_i^2 p_X(x_i) \\ &= c^2 \cdot \sum_{i=1}^{n} p_X(x_i) \\ &= c^2.\end{aligned}$$

Plugging both sides back into $\mathrm{Var}(X)$, we have $\mathrm{Var}(X) = c^2 - c^2 = 0$. ($\Rightarrow$) Suppose $\mathrm{Var}(X) = 0$. Then,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = 0 \\ &= \underbrace{\sum_i (x_i - c)^2 \cdot p_X(i)}_{\text{every term} \geq 0} \\ &\Rightarrow (x_i = c \quad \forall i) \Rightarrow \mathbb{E}(X) = c.\end{aligned}$$

■

Note that $\mathrm{Var}(X)$ is very similar to standard deviation, and it measures the spread of how far apart data is from the mean.

**Definition 27.** Let $X : \Omega \to \mathbb{R}$ be a random variable. The **cumulative distributino function** (CDF) is defined as

$$F_X(a) = \mathbb{P}(X \leq a) = \mathbb{P}(X(\omega) \in (-\infty, a]).$$

**Definition 28.** We say $X \sim \mathrm{Bernoulli}(p)$ if

$$\mathbb{P}(X = 1) = p \quad \mathbb{P}(X = 0) = 1 - p \quad (p \in (0, 1)).$$

**Example.** It is known that screws produced will be defective with probability 0.1. The company sells screws in packages of 10 and gives a refund if more than 1 screw is defective. What proportion of packages must the company refund?

**Explanation.** Let $X$ represent the number of defective screws. We wish to find $1 - \mathbb{P}(X \leq 1)$. This is just $1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1)$. Just apply the binomial formula to get your answer.

**Definition 29.** A random variable $X$ that takes on one of the values $0, 1, 2, \ldots$ is said to be a **Poisson** random variable with parameter $\lambda$ if for some $\lambda > 0$

$$p(i) = \mathbb{P}(X = i) = e^{-\lambda}\left(\frac{\lambda^i}{i!}\right).$$

## Lecture 5: More Distributions

Note that the Poisson distribution can be derived from the binomial distribution with $p = \frac{\lambda}{n}$.

**Example.** Let $X$ be a binomial random variable. Calculate $\mathbb{E}[X]$ and the variance.

**Explanation.** We have

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x \cdot q^{n-x} \\
&= \sum_{x=1}^{n} n \cdot \binom{n-1}{x-1} p^x q^{n-x} \\
&= np \cdot \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} \\
&= np \cdot (p+q)^{n-1} \\
&= np.
\end{aligned}$$

For the variance, we have

$$\operatorname{Var} X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

$$\begin{aligned}
\mathbb{E}(X^2) &= np \cdot \sum_{x=1}^{n} x \cdot \binom{n-1}{x-1} p^{x-1} q^{n-x} \\
&= np \cdot \mathbb{E}(Y+1) \quad (Y \sim \operatorname{Bin}(n-1, p)) \\
&= n \cdot p((n-1)p + 1).
\end{aligned}$$

such that

$$\operatorname{Var} X = np(1-p).$$

**Example.** Same thing, but with $X$ as poisson.

**Explanation.**

$$\begin{aligned}
\mathbb{E}X &= \sum_{x=0}^{\infty} x \cdot \left(\frac{e^{-\lambda}\lambda^x}{x!}\right) \\
&= e^{-\lambda} \cdot \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot e^{\lambda} \qquad \text{(Change of vars.)} \\
&= \lambda.
\end{aligned}$$

For the variance, we have

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} \frac{(y+1) \cdot e^{-\lambda}\lambda^y}{y!} \\
&= \lambda \left[ \sum_{y=0}^{\infty} y \cdot \frac{e^{-\lambda}\lambda^y}{y!} + \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} \right] \\
&= \lambda(\lambda+1).
\end{aligned}$$

such that

$$\operatorname{Var} X = \lambda.$$

**Example.** If $n$ people are present in the room, what is the probability that no two of them celebrate their birthday on the same day of the year? How large does $n$ be such that this probability is less than $\frac{1}{2}$?

**Explanation.** We compare $\binom{n}{2}$ times. Each probability for same birthday is $\frac{1}{365}$. Using Poisson,

$$\mathbb{P}(X = 0) = e^{-\lambda} = \exp\left(\frac{-n \cdot *(n-1)}{730}\right).$$

such that $n = 23$ is our threshhold.

**Definition 30.** A **geometric distribution** is the number of independent Bernoulli trials it takes for a single success. The pmf is

$$p_X(i) = (1-p)^{i-1} \cdot p.$$

**Definition 31.** If $X$ is a discrete random variable and $\mathbb{P}(B) > 0$, the **conditional expectation** of $X$ given $B$ is denoted by $\mathbb{E}(X \mid B)$ and defined by

$$\mathbb{E}(X \mid B) = \sum_{x \in \operatorname{Im} X} x \cdot \mathbb{P}(X = x \mid B).$$

**Definition 32.** If $X$ is a discrete random variable and $\{B_1, B_2, \ldots\}$ is a partition of the sample space such that $\mathbb{P}(B_i) > 0 \forall i$, then the **partition theorem** states that

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid B_i)\mathbb{P}(B_i).$$

## Lecture 6: Multivariate Probability

# 3 Multivariate Probability

Our objective is to treat random vectors $(X, Y_0 \in \mathbb{R}^2)$ together as

$$(X, Y) : \Omega^2 \to \mathbb{R}^2.$$

**Definition 33.** If $X$ and $Y$ are discrete random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, the **joint probability mass function** $P_{X,Y}(x, y)$ of $X$ and $Y$ is the function

$$p_{X,Y} : \mathbb{R}^2 \to [0, 1].$$

defined by

$$\begin{aligned} &p_{X,Y}(x, y) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x \text{ and } Y(\omega) = y\}). \end{aligned}$$

and abbreviated

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y).$$

**Note.** The sum of all options still remains one.

**Example.** Two cards are drawn at random from a dech of 52 cards. If $X$ denotes the number of aces drawn and $Y$ denotes the number of kings, display the join mass function of $X$, and $Y$ in tabular form.

**Explanation.** Note that $X = \{0, 1, 2\}, Y = \{0, 1, 2\}$. Then, we have

| | $X = 0$ | $X = 1$ | $X = 2$ |
|---|---|---|---|
| $Y = 0$ | $\frac{44}{52} \cdot \frac{43}{51}$ | $\frac{\binom{4}{1} \cdot \binom{44}{1}}{\binom{52}{2}}$ | $\frac{\binom{4}{2}}{\binom{52}{2}}$ |
| $Y = 1$ | $\frac{\binom{4}{1} \cdot \binom{44}{1}}{\binom{52}{2}}$ | $\frac{\binom{4}{1} \cdot \binom{4}{1}}{\binom{52}{2}}$ | 0 |
| $Y = 2$ | $\frac{\binom{4}{2}}{\binom{52}{2}}$ | 0 | 0 |

Note that we can expand this past 2 dimensions.

**Definition 34.** Suppose that each of $n$ experiments can result in any one of $r$ possible outcomes, with proabilities, $p_1, p_2, \ldots p_r$ which sum up to one. If we let $X_l$ denote the nubmer of the $n$ experiments that result in outcome number $i$, then the probability mass function is given by

$$\begin{aligned} &\mathbb{P}(X_1 = n_1, \ldots, X_r = n_r) \\ &= \binom{n}{n_1, n_2, \ldots, n_r} p_1^{n_1} \cdot p_2^{n_2} \cdot \ldots \cdot p_r^{n_r}. \end{aligned}$$

**Definition 35.** We have that

$$\begin{aligned} &\mathbb{E}(g(X, Y)) \\ &= \sum_{x \in \operatorname{Im} X} \sum_{y \in \operatorname{Im} Y} g(x, y) \mathbb{P}(X = x, Y = y). \end{aligned}$$

when this sum converges absolutely.

**Corollary.**

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

**Proof.** Linearity :). ■

**Definition 36.** Two discrete random variables $X$ and $Y$ are **independent** if the pair of events $[X = x]$ and $[Y = y]$ are independent for all $x, y \in \mathbb{R}$. We write this as

$$\begin{gathered} \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ \forall x, y \in \mathbb{R}. \end{gathered}$$

**Corollary.** If $X$ is independent of itself, then $X$ is almost surely consant.

**Proposition 5.** $k$ random variables are independent of the product of all of them is the same as the multivariate probability of all.

**Definition 37.** The **indicator function** of an event $A$ is the function $\mathbb{I}_A$ defined by

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases}.$$

**Example.** Show that two events $A$ and $B$ are independent iff their indicator functions are independent random variables.

**Explanation.** Case work for the forward direction, and set construction for the second.

## Lecture 7: More Independence

**Example.** Suppose that $n + m$ independent trials with probability of success $p$ are performed. If $X$ is the number of successes of the first $n$, and $Y$ is the number of successes of the last $m$, then $X$ and $Y$ are independent.

**Explanation.** Look at $p_{x,y}(X=x,Y=y)$. We wish to show that this equals $p_X(x)\cdot p_Y(y)$. Let 1 be success, 0 be failure. Then, $\Omega = \{0,1\}^{n+m} = (a=\{0,1\}^n, b=\{0,1\}^m)$. Then

$$\begin{aligned}\mathbb{P}((a,b)) &= p^{x+y}\cdot(1-p)^{m+n-(x+y)} \\ &= p^x\cdot(1-p)^{n-x}\cdot p^y\cdot(1-p)^{m-y}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(x,y) &= \sum_{(a,b)\in\{X=x,Y=y\}} \mathbb{P}((a,b)) \\ &= \binom{n}{x}\binom{m}{y}\underbrace{p^x(1-p)^{n-x}p^y(1-p)^{n-y}}_{\mathbb{P}((a,b))} \\ &= \binom{n}{x}p^x(1-p)^{n-x}\binom{m}{y}p^y(1-p)^{m-y} \\ &= p_X(x)\cdot p_Y(y).\end{aligned}$$

**Theorem 8.** Discrete random variables $X,Y$ are independent iff

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y)).$$

**Proof.** We have that

$$\begin{aligned}\mathbb{E}(g(X)h(Y)) &= \sum_{x,y} g(x)h(y)\mathbb{P}(X=x,Y=y) \\ &= \sum_{x,y} g(x)h(y)\mathbb{P}(X=x)\mathbb{P}(Y=y) \\ &= \mathbb{E}(g(X)) + \mathbb{E}(h(Y)).\end{aligned}$$

In the other direction, assume that

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))..$$

We must show that

$$p_{X,Y}(x,y) = p_X(a)p_Y(b).$$

for all real numbers $a,b$. Define the indicator functions

$$g(x) = \begin{cases}1, & \text{if } x=a \\ 0 & \text{if } x\neq a\end{cases} \quad h(y) = \begin{cases}1, & \text{if } y=b \\ 0 & \text{if } y\neq b\end{cases}.$$

Then, we have that

$$\begin{aligned}\mathbb{E}(g(X)h(Y)) &= \sum_{x,y} g(x)h(y)\mathbb{P}(X=x,Y=y) \\ &= \mathbb{P}(X=a,Y=b).\end{aligned}$$

We also have that

$$\mathbb{E}(g(X))\mathbb{E}(h(Y)) = \mathbb{P}(X=a)\mathbb{P}(Y=b).$$

Putting these two together, we have that $X,Y$ are independent, as desired. ∎

**Example.** Suppose that $X$ has distribution given by $\mathbb{P}(X=-1)=\mathbb{P}(X=0)=\mathbb{P}(X=1)=\frac{1}{3}$ and $Y$ is given by

$$Y = \begin{cases}0, & \text{if } X=0 \\ 1 & \text{if } X\neq 0\end{cases}.$$

**Explanation.** We have that $\mathbb{E}X\mathbb{E}Y = \mathbb{E}XY$ if (not only if) $X,Y$ independent. However, $X$ and $Y$ are dependent here.

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[X\cdot|X|] \\ &= \frac{1}{3}\cdot -1\cdot|-1| + \frac{1}{3}\cdot 0\cdot|0| + \frac{1}{3}\cdot 1\cdot|1| \\ &= 0.\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}X &= \sum_x x\cdot\mathbb{P}(X=x) \\ &= -1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0.\end{aligned}$$

and

$$\mathbb{E}Y = \mathbb{E}(|X|) = \frac{2}{3}.$$

This means that

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y.$$

which shows that this property is not bidirectional.

**Theorem 9.** (Convolution Formula) Set $Z = X+Y$, $X,Y$ independent. Then for all $z\in\mathbb{R}$,

$$\begin{aligned}\mathbb{P}(Z=z) &= \sum_x \mathbb{P}(X=x, Y=z-x) \\ &= \sum_x \mathbb{P}(X=x)\mathbb{P}(Y=z-x).\end{aligned}$$

**Example.** If $X$ and $Y$ are independent discrete random variables, $X$ having the Poisson distribution with parameter $\lambda$ and $Y$ has Poisson disrubtion with parameter $\mu$, show that $X+Y$ has poisson distrubtion with parameter $\lambda+\mu$.

**Explanation.** Let $Z = X+Y$. Remember that

$$\mathbb{P}(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{and} \quad \mathbb{P}(Y=y) = \frac{e^{-\mu}\mu^y}{y!}.$$

Then,

$$\begin{aligned}
\mathbb{P}(Z=z) &= \sum_{x=0}^{\infty} \mathbb{P}(X=x)\mathbb{P}(Y=z-x) \\
&= \sum_{x=0}^{z} \mathbb{P}(X=x)\mathbb{P}(Y=z-x) \\
&= \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\mu}\mu^{z-x}}{(z-x)!} \\
&= \sum_{x=0}^{z} e^{-\lambda-\mu} \frac{1}{z!} \frac{z!}{x!(z-x)!} \cdot \lambda^x \mu^{z-x} \\
&= e^{-\lambda-\mu} \frac{1}{z!} \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x} \\
&= e^{-\lambda-\mu} \frac{1}{z!} (\lambda+\mu)^z.
\end{aligned}$$

which is precisely the Poisson distribution with $\lambda + \mu$.

**Theorem 10.** Let $A_1, A_2 \ldots A_n$ be events. Then, we have that

$$\sum_{i=0}^{n} \mathbb{I}_{A_i}(\omega) = \text{number of events that } \omega \text{ occurs.}$$

**Example.** The $2n$ seats around a circular table are numbered clockwise. Queens sit in odd numbered seats and Kings in even numbesr. Let $N$ be the number of queens sitting next to their king. Find the mean and variance of $N$.

**Explanation.** Let $A_i$ be the event that the $i$-th pair sit together. Then,

$$N = \sum_{i=1}^{n} \mathbb{I}_{A_i}.$$

Note that $\mathbb{P}(A_i) = \frac{2}{n}$. Think of a fixed king permutation, then there are two spots out of $n$ spots for the queen to sit. Next,

$$\begin{aligned}
\mathbb{E}N &= \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{I}_{A_i}\right) \\
&= \sum_{i=1}^{n} \mathbb{E}\mathbb{I}_{A_i} \\
&= \sum_{i=1}^{n} \mathbb{P}(A_i) \\
&= \sum_{i=1}^{n} \frac{2}{n} = 2.
\end{aligned}$$

The variance calculation is more involved. Remember that

$$\mathrm{Var}(N) = \mathbb{E}N^2 - (\mathbb{E}N)^2.$$

Then, we have

$$\begin{aligned}
\mathbb{E}(N^2) &= \mathbb{E}\left(\left[\sum_{i=1}^{n} \mathbb{I}_{A_i}\right]^2\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{I}_{A_i}\mathbb{I}_{A_j}\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{I}_{A_i}\mathbb{I}_{A_i} + 2\sum_{1\le i<j\le n} \mathbb{I}_{A_i}\mathbb{I}_{A_j}\right) \\
&= \sum_{i=1}^{n} \mathbb{E}(\mathbb{I}_{A_i}\mathbb{I}_{A_i}) + 2\sum_{1\le i<j\le n} \mathbb{E}(\mathbb{I}_{A_i}\mathbb{I}_{A_j}) \\
&= \sum_{i=1}^{n} \mathbb{E}(\mathbb{I}_{A_i}) + 2\sum_{1\le i<j\le n} \mathbb{E}(\mathbb{I}_{A_i}\mathbb{I}_{A_j}) \\
&= \sum_{i=1}^{n} \underbrace{\mathbb{P}(A_i)}_{\frac{2}{n}} + 2\sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j).
\end{aligned}$$

From here, we need to calculate $\mathbb{P}(A_i \cap A_j)$.

## Lecture 8: Generating Functions

Continuing on with the Variance calculation, we have that

**Explanation.** Note that $\mathbb{P}(A_i \cap A_j)$ is given by (WLOG)

$$\begin{aligned}
\mathbb{P}(A_i \cap A_j) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \\
&= \frac{2}{n} \cdot \left[\frac{1}{n-1}\frac{1}{n-1} + \frac{n-2}{n-1}\frac{2}{n-1}\right] \\
&= \frac{2}{n} \cdot \frac{2n-3}{(n-1)^2}.
\end{aligned}$$

Continuing on, we have that

$$\mathbb{E}(N^2) = n \cdot \frac{2}{n} + \frac{4}{n} \cdot \frac{2n-3}{(n-1)^2}\binom{n}{2}.$$

**Example.** My squad of bunnies have been training all summer. Each bunny is ready for the bunny mission with probability $p$. If I have $n$ bunnies in the squad and need $k$ for the mission, find $\mathbb{E}$ of the number of $k$-large teams that I can form.

**Explanation.** Let $A_i$ be the event that bunny $i$

is ready. Then,

$$X = \sum_{i=1}^{n} \mathbb{I}_{A_i}.$$

counts the number of bunnies ready. We want to compute for $X \geq k$ how many $k$-large teams is possible. This is just $\binom{X}{k}$. Note that

$$\binom{X}{k} = \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{I}_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_n}}.$$

This means that

$$\begin{aligned}
\mathbb{E}\binom{X}{k} &= \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{E}\left(\mathbb{I}_{A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_n}}\right) \\
&= \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n}) \\
&= \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_n}) \\
&\qquad \text{(Independent)} \\
&= \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} p^k \\
&= \binom{n}{k} p^k.
\end{aligned}$$

**Example.** A grove of 52 trees is araranged in a circular fashion. If 15 chipmunks live in these trees, show that there is a group of 7 consecutive trees that together house at least 3 chipmunks.

**Explanation.** We find

$$\mathbb{E}(\text{num chipmunks that lie in 7 con. trees}) > 2.$$

In other words, if on average there are 3, then for one group there must be 3. Let $X$ be the number of chipmunks that lie in a random tree and 6 neighbors clockwise. Let

$$X_i = \begin{cases} 1, & \text{if chipmunk } i \text{ lives in nbhd} \\ 0 & \text{otherwise.} \end{cases}.$$

We know that

$$X = \sum_{i=1}^{15} X_i \quad \text{and} \quad \mathbb{E}X = \sum_{i=1}^{15} \mathbb{E}X_i.$$

Then, we have that

$$\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \frac{7}{52}.$$

. Therefore,

$$\mathbb{E}X = 7 \cdot \frac{15}{52} = \frac{105}{52} > 2.$$

## 4 Probability Generating Functions

**Definition 38.** Consider the sequence $u_0, u_1, u_2 \ldots$ of real numbers. We can write down the **generating function** of this sequence as

$$U(s) = u_0 + u_1 s + u_2 s^2 + \ldots.$$

**Example.** The sequence given by

$$u_n = \begin{cases} \binom{N}{n}, & \text{if } n = 0, 1, 2, \ldots, N \\ 0 & \text{otherwise} \end{cases}.$$

has generating function

$$U(s) = \sum_{n=0}^{N} \binom{N}{n} s^n = (1+s)^N.$$

**Example.** If $u_0, u_1, \ldots$ has generating function $U(s)$ and $v_0, v_1, \ldots$ has generating function $V(s)$, find $V(s)$ in terms of $U(s)$ when (a) $v_n = 2u_n$ and (b) $v_n = u_1 + 1$, and (c) $v_n = nu_n$.

**Explanation.** We have (a) $2U(S)$, (b) $U(S) + \frac{1}{1-s}$, and (c) $S \cdot U'(S)$.

### Lecture 9

**Theorem 11.** We have that $G_X(s) = \mathbb{E}(s^X)$.

**Proof.**

$$\begin{aligned}
\mathbb{E}(s^X) &= \sum_{i=0}^{\infty} \mathbb{P}(X = i) \cdot S^i \\
&= G_X(s).
\end{aligned}$$

■

**Example.** What is the PGF for $X \equiv 0$?

**Explanation.** We have that $p_i = \mathbb{P}(X = i)$, such that $p_0 = 1$, $p_i = 0$ for all $i > 0$. Then,

$$\begin{aligned}
\mathbb{E}(s^X) &= p_0 s^0 + \ldots + p_i s^i + \ldots = 1 \\
&= G_X(s).
\end{aligned}$$

**Example.** What is the PGF for $X \sim$ Bernoulli$(p)$?

**Explanation.** Remember that $p$ is the probability for success, and $1-p$ is the probability for failure (Binomial). We have that $p_0 = 1-p$ and $p_1 = p$. Then,

$$\begin{aligned}\mathbb{E}(s^X) = G_X(s) &= p_0 s^0 + p_1 s^1 \\ &= (1-p) + p \cdot s.\end{aligned}$$

**Theorem 12.** Given $X$ with geometric distribution with parameter $p$, we have that

$$\mathbb{P}(X = k) = pq^{k-1}.$$

where $p + q = 1$, and $X$ has probability generating function

$$\frac{p}{q} \cdot \frac{1}{1-qs}.$$

**Definition 39.** Let $k \geq 1$. The $k$th **moment** of the random variable $X$ is the quantity $\mathbb{E}(X^k)$.

**Theorem 13.** The $r$th derivative of $G_X(s)$ for $s = 1$ is $\mathbb{E}(X[X-1] \ldots [X-r+1])$ for $r = 1, 2, \ldots$. In other words, with $s = 1, r = 1$ we can get $G'_X(1) = \mathbb{E}(X)$.

**Example.** Use the method of generating functions to show that a random variable with Poisson distribution with parameter $\lambda$ has both mean and variance equal to $\lambda$.

**Explanation.** Note that $G''(1) = \mathbb{E}(X[X-1]) = \mathbb{E}X^2 - \mathbb{E}X$. Then,

$$\begin{aligned}\mathbb{E}X^2 &= \mathbb{E}[X(X-1) + X] \\ &= \mathbb{E}(X(X-1)) + \mathbb{E}X \\ &= G''_X(1) + G'_X(1).\end{aligned}$$

Which means that

$$\begin{aligned}\operatorname{Var} X &= \mathbb{E}(X^2) - (\mathbb{E}X)^2 \\ &= G''_X(1) + G'_X(1) - (G'_X(1))^2.\end{aligned}$$

Recall that $G_X(s) = e^{\lambda(s-1)}$. Note that $G'_X(1) = \lambda$, which is the expectation (mean). Also, $G''_X(1) = \lambda^2$ which means that

$$\operatorname{Var} X = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

**Theorem 14.** Suppose we have $X, Y$ such that

$$G_X(s) = G_Y(s).$$

This means that

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k) \quad \forall k.$$

**Theorem 15.** If $X$ and $Y$ are independent random variables, then $X + Y$ has generating function

$$G_{X+Y}(s) = G_X(s) G_Y(s).$$

**Proof.**

$$\begin{aligned}G_{X+Y}(s) &= \mathbb{E}(s^{X+Y}) \\ &= \mathbb{E}(s^X s^Y) \\ &= \mathbb{E}(s^X)\mathbb{E}(s^Y) \qquad \text{(Independence)} \\ &= G_X(s) G_Y(s).\end{aligned}$$

■

**Theorem 16.** (Random sum formula) Let $N$ and $X_1, X_2 \ldots$ be independent random variables taking values in $\mathbb{Z}_{>0}$. If $X_i$ are identically distributed with common PGF $G_X$, then

$$S = X_1 + X_2 + \ldots + X_N.$$

has PGF

$$G_S(s) = G_N(G_X(s)).$$

**Proof.** Note that $G_S(t) = \mathbb{E}t^S$. Recall that for partitions $E_i$ of $\Omega$,

$$\mathbb{E}X = \sum_{i=1}^{\infty} \mathbb{E}(X \mid E_i)\mathbb{P}(E_i).$$

Applying that

$$G_S(t) = \mathbb{E}t^S = \mathbb{E}(t^{X_1 + X_2 + \ldots + X_N}).$$

we have that

$$
\begin{aligned}
G_S(t) &= \sum_{n=0}^{\infty} \mathbb{E}(t^{X_1+\ldots+X_N} \mid N = k)\mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} \mathbb{E}(t^{X_1+\ldots+X_n})\mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} \mathbb{E}(t^{X_1})\ldots\mathbb{E}(t^{X_n})\mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} (\mathbb{E}(t^{X_1}))^n\mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} G_{X_1}(t)^n\mathbb{P}(N = n) \\
&= G_N(G_{X_1}(t)).
\end{aligned}
$$

■

**Example.** The hutch in the garden contains 20 pregnant rabbits. The hutch is insecure and each rabbit has a $\frac{1}{2}$ chance of escaping overnight. The next morning, each remaining rabbit gives birth to a litter, with each mother having a random number of offspring with Poisson distribution with parameter 3.

**Explanation.** Let $S$ be the number of baby bunnies. We wish to compute $G_S(t)$ and $\mathbb{E}S$. Let $X_i$ be the number of rabbits in the $i$th litter. Let $N$ be the number of rabbits in the hutch the next morning. Note that $N$ is binomial with $p = \frac{1}{2}$ and 20 trials.

Then, $S = X_1 + \ldots + X_N$. Then,

$$
\begin{aligned}
G_S(t) &= G_N(G_X(t)) \\
&= G_N(e^{3(t-1)}).
\end{aligned}
$$

Also,

$$
G_N(t) = \left(\frac{1}{2} + \frac{1}{2}t\right)^{20}.
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}(S) &= G_S'(1) \\
&= G_X'(1)G_N'(G_X(1)) \\
&= 3e^{3(1-1)} \cdot \frac{1}{2} \cdot 20 \left(\frac{1}{2} + \frac{1}{2}e^{3(1-1)}\right)^{19} \\
& \qquad \text{(at } t = 1\text{)} \\
&= 3 \cdot \frac{1}{2} \cdot 20 \cdot 1 = 30.
\end{aligned}
$$