Introduction to Probability Theory

Lecturer: Shagnik Das
Scribe: Jenny

Abstract. These are notes from the course 'Introduction to Probability Theory,' given by Shagnik Das in the Spring Semester of 2025 at the National Taiwan University. Nobody bears any responsibility for mistakes in the script, but I would be grateful to be notified of any errors found.

This note lacks content of the lecture on 2/18, 2/20, 4/17, and 4/22.

Contents

CHAPTER 1

Axioms of Probability

Given a sample space S,

(1) For any event $E \subseteq S$, $0 \leq \mathbb{P}(E) \leq 1$.
(2) $\mathbb{P}(S) = 1$.
(3) For mutually exclusive events $E_1, E_2, \ldots$, $\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$.

Define $\varnothing = \{\}$ as the empty set.

Claim 1. $\mathbb{P}(\varnothing) = 0$.

Proof. Consider the sequence of events $E_1 = S$, $E_2 = \varnothing$ for all $i \geq 2$. These events are mutually exclusive. By Axiom 3,

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

$$\bigcup_{i=1}^{\infty} E_i = S \cup \varnothing \cup \varnothing \cup \cdots = S$$

$$\mathbb{P}(S) = \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \mathbb{P}(S) + \sum_{i=2}^{\infty} \mathbb{P}(\varnothing)$$

$$\Rightarrow \sum_{i=2}^{\infty} \mathbb{P}(\varnothing) \Rightarrow \mathbb{P}(\varnothing) = 0$$

□

Corollary 1. For any finite sequence of mutually exclusive events $E_1, E_2, \ldots, E_n$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathbb{P}(E_i).$$

Proof. Extend to an infinite sequence of exclusive events by adding the empty set $E_i = \varnothing$ for all $i \geq n+1$. Then $\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{\infty} E_i$.

By Axiom 3,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \\
&= \sum_{i=1}^{n} \mathbb{P}(E_i) + \sum_{i=n+1}^{\infty} \mathbb{P}(\varnothing) \\
\text{(since } \mathbb{P}(\varnothing) = 0\text{)} \qquad &= \sum_{i=1}^{n} \mathbb{P}(E_i)
\end{aligned}$$

□

Proposition 1. Given a probability space $(S, \mathbb{P})$, where S is the sample space and $\mathbb{P}$ is the probability function, we have

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E).$$

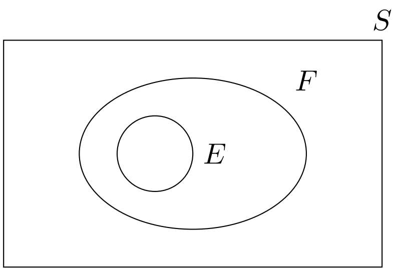
Proof. Note that

- $E \cap E^c = \varnothing$
- $E \cup E^c = S$

By Corollary, $1 = \mathbb{P}(S) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c)$. □

Proposition 2. Given a probability space $(S, \mathbb{P})$, and nested sets $E \subseteq F \subseteq S$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.

Proof. Venn diagrams



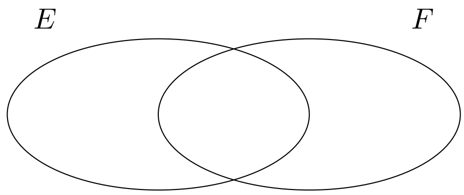
Note that $E \cap F = E$ and $E^c \cap F$ are exclusive events $(E \cap (E^c \cap F) = (E \cap E^c) \cap F = \varnothing \cap F = \varnothing)$, and $(E \cap F) \cup (E^c \cap F) = (E \cup E^c) \cap F = S \cap F = F$.

By Corollary, $\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F) = \mathbb{P}(E) + \mathbb{P}(E^c \cap F) \geq \mathbb{P}(E)$. □

Example 1. Rolling a fair six-sided dice.

$$\Rightarrow \mathbb{P}(\text{rolling a } 6) \leq \mathbb{P}(\text{rolling an even number})$$

For arbitrary events, we observe:



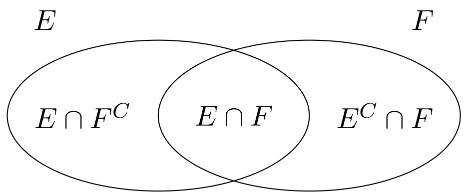
Proposition 3. In a probability space $(S, \mathbb{P})$, given any events $E, F \subseteq S$,

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F).$$

Corollary 2 (Union bound). $\mathbb{P}(E \cup F) \leq \mathbb{P}(E) + \mathbb{P}(F)$.

Proof. (Cor.) $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) \leq \mathbb{P}(E) + \mathbb{P}(F)$ □

Proof. (Prop.)



We have unions of exclusive events

- $E \cup F = (E \cap F^c) \dot\cup (E \cap F) \dot\cup (E^c \cap F)$
- $E = (E \cap F^c) \dot\cup (E \cap F)$, $F = (E \cap F) \dot\cup (E^c \cap F)$

By Corollary 1,

- $\mathbb{P}(E \cup F) = \mathbb{P}(E \cap F^c) + \mathbb{P}(E \cap F) + \mathbb{F}(E^c \cap F)$
- $\mathbb{P}(E) = \mathbb{P}(E \cap F^c) + \mathbb{P}(E \cap F)$
- $\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F)$

$$\begin{aligned}
\Rightarrow \mathbb{P}(E) + \mathbb{P}(F) &= \mathbb{P}(E \cap F^c) + \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F) \\
&= \mathbb{P}(E \cap F^c) + \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F) + \mathbb{P}(E \cap F) \\
&= \mathbb{P}(E \cup F) + \mathbb{P}(E \cap F)
\end{aligned}$$

□

Example 2. Play a game against Real Madrid.

- $\mathbb{P}$(Mbappé scores) $= 0.5$
- $\mathbb{P}$(Vinicius scores) $= 0.4$
- $\mathbb{P}$(Mbappé and Vinicius both scores) $= 0.2$

<u>Q.</u> $\mathbb{P}$(Mbappé or Vinicius scores) $=$?

Solution. Define events

- $E = \{$Mbappé scores$\}$
- $F = \{$Vinicius scores$\}$

$$
\begin{aligned}
&\mathbb{P}(E) = 0.5, \mathbb{P}(F) = 0.4, \mathbb{P}(E \cap F) = 0.2 \\
\overset{\text{Prop 3}}{\Rightarrow}\ &\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) = 0.7 \\
\overset{\text{Prop 1}}{\Rightarrow}\ &\mathbb{P}(E^c \cap F^c) = \mathbb{P}((E \cup F)^c) = 1 - \mathbb{P}(E \cap F) = 0.3
\end{aligned}
$$

Q. What can we say about $\mathbb{P}(E \cup F \cup G)$?

$$
\begin{aligned}
&\mathbb{P}(E \cup F \cup G) \\
&= \mathbb{P}((E \cup F) \cup G) \\
&= \mathbb{P}(E \cup F) + \mathbb{P}(G) - \mathbb{P}((E \cup F) \cap G) \\
&= \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) + \mathbb{P}(G) - \mathbb{P}((E \cup F) \cap G)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{P}((E \cup F) \cap G) &= \mathbb{P}((E \cap G) \cup (F \cap G)) \\
&= \mathbb{P}(E \cap G) + \mathbb{P}(F \cap G) - \mathbb{P}((E \cap G) \cap (F \cap G)) \\
&= \mathbb{P}(E \cap G) + \mathbb{P}(F \cap G) - \mathbb{P}(E \cap F \cap G)
\end{aligned}
$$

Therefore

$$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G).$$

Example 3. Roll a 60-sided dice. $\mathbb{P}$(roll in divisible by 2, 3, or 5)?
Solution. Let $E = \{\text{div. by } 2\}$, $F = \{\text{div. by } 3\}$, $G = \{\text{div. by } 5\}$.

$$\mathbb{P}(E) = \frac{\#\text{even numbers in } 1, 2, \ldots, 60}{60} = \frac{30}{60} = \frac{1}{2}.$$

$$\mathbb{P}(F) = \frac{1}{3}, \quad \mathbb{P}(G) = \frac{1}{5}.$$

$$
\begin{aligned}
\mathbb{P}(E \cap F) &= \mathbb{P}(\text{div by 2 of div by 3}) \\
&= \mathbb{P}(\text{div by 6}) = \frac{1}{6} \\
\mathbb{P}(E \cap G) &= \mathbb{P}(\text{div by 10}) = \frac{1}{10} \\
\mathbb{P}(F \cap G) &= \mathbb{P}(\text{div by 15}) = \frac{1}{15} \\
\mathbb{P}(E \cap F \cap G) &= \mathbb{P}(\text{div by 30}) = \frac{1}{30}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}(E \cup F \cup G) \\
&= \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G) \\
&= \frac{1}{2} + \frac{1}{3} + \frac{1}{5} - \frac{1}{6} - \frac{1}{10} - \frac{1}{15} + \frac{1}{30} = \frac{22}{30}
\end{aligned}
$$

CHAPTER 2

Inclusion-Exclusion

1. Inclusion-Exclusion Formula

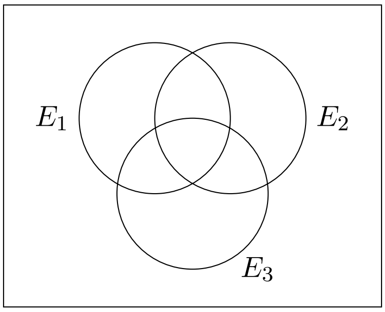
What is $\mathbb{P}\left(\bigcup_{i=1}^n E_i\right)$?
Use induction, we can get

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{n-1} E_i\right) \cup E_n\right) \\ &= \sum_{i=1}^n \mathbb{P}(E_i) - \sum_{1\le i_1<i_2\le n} \mathbb{P}(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \cdots\end{aligned}$$

Formally,

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{r=1}^n \sum_{1\le i_1<i_2<\cdots<i_r\le n} (-1)^{r+1} \mathbb{P}\left(\bigcap_{j=1}^r E_{i_j}\right).$$

Proof. (Inclusion-Exclusion Formula)



We can write all the events as mutually exclusive unions

$$E_I = \left(\bigcap_{i\in I} E_i\right) \cap \left(\bigcap_{i\notin I} E_i^C\right) \text{ for } I \subseteq [n]\,.$$

$$E_I = \{\text{outcomes where } E_i \text{ happens} \iff i \in I\}$$

For example, $\bigcup_{i=1}^n E_i = \dot\bigcup_{I:I\neq\varnothing} E_I$.

$$\Rightarrow \mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{I\neq\varnothing} \mathbb{P}(E_I) \quad (*)$$

Given every $J \subseteq [n]$, $\mathbb{P}\left(\bigcap_{j\subseteq J} E_j\right)$, $\bigcap_{j\subseteq J} E_j = \dot\bigcup_{I:J\subseteq I} E_I$.

RHS:

$$\sum_{r=1}^{n} \sum_{\substack{J\subseteq[n]\\|J|=r}} (-1)^{r+1}\mathbb{P}\left(\bigcap_{j\subseteq J} E_j\right)$$

$$= \sum_{r=1}^{n} \sum_{\substack{J\subseteq[n]\\|J|=r}} (-1)^{r+1}\mathbb{P}\left(\dot{\bigcup_{I:J\subseteq I}} E_I\right)$$

(mutually exclusive)
$$= \sum_{r=1}^{n}(-1)^{r+1} \sum_{J\subseteq[n]|J|=r} \sum_{I:J\subseteq I} \mathbb{P}\left(E_I\right)$$

$$= \sum_{\substack{I\subseteq[n]\\I\neq\varnothing}} \left(\sum_{r=1}^{n} \sum_{\substack{J\subseteq[n]\\|J|=r}} (-1)^{r+1}\right) \mathbb{P}(E_I)$$

Recall that the number of choices of J, $J \subseteq I$, $|J| = r$ is $\binom{|I|}{r}$.

$$\Rightarrow \sum_{r=1}^{n} \sum_{\substack{J\subseteq[n]\\|J|=r}} (-1)^{r+1} = \sum_{r=1}^{n} \binom{|I|}{r}(-1)^{r+1}$$

$$= \sum_{r=1}^{|I|} \binom{|I|}{r}(-1)^{r+1}$$

$$= \sum_{r=0}^{|I|} \binom{|I|}{r}(-1)^{r+1} - \binom{|I|}{0}(-1)^{0+1}$$

$$= -\sum_{r=0}^{|I|} \binom{|I|}{r}(-1)^{r} - (-1)$$

(binomial thm)
$$= -(-1+1)^{|I|} + 1 = 1$$

$$\therefore \sum_{r=1}^{n}(-1)^{r+1} \sum_{\substack{J\subseteq[n]\\|J|=r}} \sum_{I:J\subseteq I} \mathbb{P}\left(E_I\right)$$

$$= \sum_{\substack{I\subseteq[n]\\I\neq\varnothing}} 1 \cdot \mathbb{P}(E_I)$$

(*)
$$= \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right)$$

$\square$

Warm-up. Randomly shuffle a deck of cards. Turn them over, one-by-one, until the first Ace.
Q. What is the probability that the next card is

(a) Ace of spades?
(b) Two of clubs?

Attempt to answer:

(a) We remove A♠, shuffle remaining 51 cards, and place A♠ in a random position.
$\Rightarrow$ 51! ways to shuffle other cards
$\Rightarrow$ 52 positions available for A♠
For the event to occur, we must place the A♠ directly after the first ace.
$\Rightarrow \mathbb{P}(a) = \frac{1}{52}$
(b) Similarly, $\mathbb{P}(b) = \frac{1}{52}$.

Example 1. (Inclusion-Exclusion) There are a party with n people. They put their hats in a rack. When leaving, everybody takes a random hat from the rack.
Q. What is the probability that nobody gets their own hat?
Solution. $S = \{\text{bijection from hats to people}\}$, $|S| = n!$.
Let $E = \{\text{nobody gets their own hat}\}$.
To make things simpler, let $E_i = \{i\text{th person gets their own hat}\}$. Then

$$E = \bigcap_{i=1}^{n} E_i^C = \left(\bigcup_{i=1}^{n} E_i\right)^C$$
$$\Rightarrow \mathbb{P}(E) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right)$$

Therefore

$$\mathbb{P}(E_i) = \frac{1}{n},\ \mathbb{P}(E_i \cap E_j) = \frac{(n-2)!}{n!},\ \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) = \frac{(n-r)!}{n!}$$

Plug into Inclusion-Exclusion:

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{r=1}^{n} (-1)^{r+1} \sum_{1 \le i_1 < i_2 < \cdots < i_r \le n} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_r}) \\ &= \sum_{r=1}^{n} (-1)^{r+1} \sum_{1 \le i_1 < i_2 < \cdots < i_r \le n} \frac{(n-r)!}{n!} \\ &= \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \frac{(n-r)!}{n!} \\ &= \sum_{r=1}^{n} \frac{(-1)^{r+1}}{r!}\end{aligned}$$

$$\mathbb{P}(E) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \sum_{r=1}^{n} \frac{(-1)^{r+1}}{r!} = \sum_{r=0}^{n} \frac{(-1)^r}{r!}$$

As $n \to \infty$, $\mathbb{P}(E) \to \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} = e^{-1}$.

2. Bonferroni Inequalities

Inclusion-Exclusion:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_i \mathbb{P}(E_i) - \sum_{i_1<i_2} \mathbb{P}(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \cdots$$

Proposition 1. If t is odd, then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{r=1}^{t} (-1)^{r+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq n} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_r})$$

If t is even, then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \geq \sum_{r=1}^{t} (-1)^{r+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq n} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_r})$$

In particular, the case $t = 1$ is called the union bound:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(E_i).$$

Proof. Proof by induction on t.

$\bigcup_{i=1}^{n} E_i \to$ want to write as a union of mutually exclusive events

$$\bigcup_{i=1}^{n} E_i = E_1 \,\dot\cup\, (E_2 \cap E_1^C) \,\dot\cup\, (E_3 \cap E_1^C \cap E_2^C) \,\dot\cup \cdots \dot\cup\, (E_n \cap E_1^C \cap E_2^C \cap \cdots \cap E_{n-1})$$

$$\Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \mathbb{P}\left(\dot{\bigcup_{i=1}^{n}}\left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right)\right)$$

$$\Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathbb{P}\left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right) \tag{*}$$

Base case. ($t = 1$) For each i, $E_i \cap \left(\bigcap_{j<i} E_j^C\right) \subseteq E_i$.

$\overset{\text{Prop 2}}{\Rightarrow} \mathbb{P}\left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right) \leq \mathbb{P}(E_i)$ by (*).

Induction step.

$$E_i = \left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right) \dot\cup \left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)^C\right)$$

$$= \left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right) \dot\cup \left(E_i \cap \left(\bigcup_{j<i} E_j\right)\right)$$

$$\Rightarrow \mathbb{P}\left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right) = \mathbb{P}(E_i) - \mathbb{P}\left(E_i \cap \left(\bigcup_{j<i} E_j\right)\right)$$

$$\Rightarrow \mathbb{P}\left(E_i \cap \left(\bigcap_{j<i} E_j^C\right)\right) = \mathbb{P}(E_i) - \underbrace{\mathbb{P}\left(\bigcup_{j<i}(E_i \cap E_j)\right)}_{(\dagger)}$$

Apply the $(t-1)$-Bonferroni Inequality to $(\dagger)$.
For example: $(t=2)$ By the case of $t=1$,

$$\mathbb{P}\left(\bigcup_{j<i}(E_i \cap E_j)\right) \leq \sum_{j<i} \mathbb{P}(E_i \cap E_j)$$

plug $(*) \to (\dagger)$

$$\Rightarrow \mathbb{P}\left(E_i \cap \left(\bigcup_{j<i} E_j\right)\right) \geq \mathbb{P}(E_i) - \sum_{j<i} \mathbb{P}(E_i \cap E_j)$$

$$\overset{(*)}{\Rightarrow} \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \geq \sum_i \left(\mathbb{P}(E_i) - \sum_{j<i} \mathbb{P}(E_i \cap E_j)\right) = \sum_i \mathbb{P}(E_i) - \sum_{j<i} \mathbb{P}(E_i \cap E_j)$$

$\square$

CHAPTER 3

Continuity of Probability

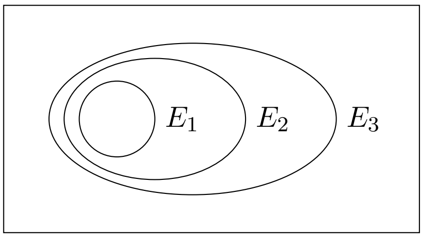
1. Increasing and Decreasing Sequences of Events

Definition 1. Let $E_1, E_2, E_3, \ldots$ be a sequence of sets. We say the sequence is increasing if $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$ and define $\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$.

The sequence is decreasing if $E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots$ and define $\lim_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$.

Proposition 1. If $E_1, E_2, E_3, \ldots$ is increasing or decreasing, then $\mathbb{P}\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} \mathbb{P}(E_n)$.

Proof. Suppose $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$. Then $\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$.



Let $F_n = E_n \setminus \left(\bigcup_{i=1}^{n-1} E_i\right)$.

Then $F_1, F_2, \ldots$ are mutually exclusive. Therefore $\bigcup_{i=1}^{n} F_i = E_n = \bigcup_{i=1}^{n} E_i$

$$
\begin{aligned}
\mathbb{P}\left(\lim_{n\to\infty} E_n\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \\
\text{(Axiom 3)} \qquad &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i\right) \\
\text{(def. of infinite sum)} \qquad &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(F_i) \\
\text{(Axiom 3)} \qquad &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} F_i\right) \\
&= \lim_{n\to\infty} \mathbb{P}(E_n)
\end{aligned}
$$

If $E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots$ is decreasing, then $E_1^C \subseteq E_2^C \subseteq E_3^C \subseteq \cdots$ is increasing and $(\lim_{n\to\infty} E_n)^C = \lim_{n\to\infty} E_n^C$.

$$\begin{aligned} \Rightarrow \mathbb{P}\left(\lim_{n\to\infty} E_n\right) &= 1 - \mathbb{P}\left(\left(\lim_{n\to\infty} E_n\right)^C\right) \\ &= 1 - \mathbb{P}\left(\lim_{n\to\infty} E_n^C\right) \\ \text{(above result)} \qquad &= 1 - \lim_{n\to\infty} \mathbb{P}(E_n^C) \\ &= 1 - \lim_{n\to\infty} (1 - \mathbb{P}(E_n)) \\ &= \lim_{n\to\infty} \mathbb{P}(E_n) \end{aligned}$$

□

Given any sequence of sets $E_1, E_2, E_3, \ldots$, we define

$$\limsup_{n\to\infty} E_n := \bigcap_{n=1}^{\infty} \left(\bigcup_{i=n}^{\infty} E_i \right) = \lim_{n\to\infty} \underbrace{\bigcup_{i=n}^{\infty} E_i}_{\text{decreasing sequence}} .$$

Remark 1. $\limsup_{n\to\infty} E_n := \bigcap_{n=1}^{\infty} (\bigcup_{i=n}^{\infty} E_i)$ is the event that infinitely many of events of the events E_n occur.

2. 1st Borel-Cantelli Lemma

Theorem 1 (1st Borel-Cantelli Lemma). If $E_1, E_2, E_3, \ldots$ is a sequence of events and $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$, then $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$.

Proof.

$$\begin{aligned} &\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) \\ \text{(continuity)} \qquad &= \mathbb{P}\left(\lim_{n\to\infty} \left(\bigcup_{i=n}^{\infty} E_i\right)\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^{\infty} E_i\right) \\ &\leq \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbb{P}(E_i) \to 0 \text{ since } \sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty \end{aligned}$$

□

Application. (1st Borel-Cantelli Lemma)
(1) Promotion in a restaurant: the nth customer rolls n dice. If all rolls are even, then they get free food for life!

Let $E_n = \{n\text{th customer gets free food for life}\}$. $S = \{1, 2, \ldots, 6\}^n$, $E_n = \{2, 4, 6\}^n$.

$$\mathbb{P}(E_n) = \frac{|\{2,4,6\}^n|}{|\{1,2,\ldots,6\}^n|} = \frac{3^n}{6^n} = 2^{-n}.$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \sum_{n=1}^{\infty} 2^{-n} = 1 < \infty$, the 1st Borel Cantelli Lemma states $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$. Therefore almost surely, only have to give finitely many customers free food!

(2) Roll a die infinitely many times. We are interested in the no. of even numbers.
Let $e_n = \frac{\#\{\text{even rolls in first } n \text{ rolls}\}}{n}$.
Fix $\varepsilon > 0$. Let $E_n = \left\{e_n \geq \frac{1}{2} + \varepsilon\right\}$.
$S = \{1,2,3,4,5,6\}^n$. Count E_n:
   (a) Choose how many even rolls r: $\left(\frac{1}{2}+\varepsilon\right) n \leq r \leq n$ (Apply the sum rule over choice of r).
   (b) Choose which rolls are even: $\binom{n}{r}$ choices.
   (c) Each roll has 3 choice $\{2,4,6\}$ if even, $\{1,3,5\}$ if odd. Product rule $\Rightarrow 3^n$ choice.

Putting it all togeher:

$$|E_n| = \sum_{r=\lceil * \rceil\left(\frac{1}{2}+\varepsilon\right)n}^{n} \binom{n}{r} 3^n$$

$$\mathbb{P}(E_n) = \frac{|E_n|}{|S_n|} = \frac{\sum_{r=\lceil * \rceil\left(\frac{1}{2}+\varepsilon\right)n}^{n} \binom{n}{r} 3^n}{6^n} = \frac{\sum_{r=\lceil * \rceil\left(\frac{1}{2}+\varepsilon\right)n}^{n} \binom{n}{r}}{2^n}$$

Approximation. If $\frac{1}{2} \leq \alpha \leq 1$,

$$\sum_{r=\lceil \alpha n \rceil}^{n} \binom{n}{r} \leq 2^{n\mathcal{H}(\alpha)}$$

where $\mathcal{H}$ is the binary entropy function, defined as $\mathcal{H}(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$.
$0 \leq \mathcal{H}(\alpha) \leq 1$ with $\mathcal{H}(\alpha) = 1$ iff $\alpha = \frac{1}{2}$.

$$\mathbb{P}(E_n) = \frac{\sum_{r=\lceil * \rceil\left(\frac{1}{2}+\varepsilon\right)n}^{n} \binom{n}{r}}{2^n} \leq \frac{2^{n\mathcal{H}\left(\frac{1}{2}+\varepsilon\right)}}{2^n} = 2^{-\delta n}$$

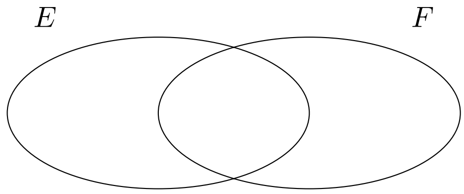
where $\mathcal{H}\left(\frac{1}{2}+\varepsilon\right) = (1-\delta)n$ for some $\delta = \delta(\varepsilon) > 0$.
$\Rightarrow \mathbb{P}(E_n) \leq 2^{-\delta n}$
$\Rightarrow \sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$
1st Borel Cantelli $\Rightarrow \mathbb{P}(\limsup_{n\to\infty} E_n) = 0$.
$\Rightarrow$ almost surely, there exists N such that for all $n \geq N$, E_n doesn't happen $e_n < \frac{1}{2} + \varepsilon$.
By symmetry, same is true for ratio of odd numbers.
$\Rightarrow$ exists N' such that for all $n \geq N'$, $e_n > \frac{1}{2} - \varepsilon$.
$\Rightarrow$ exists N'' such that for all $n \geq N''$, $\frac{1}{2} - \varepsilon < e_n < \frac{1}{2} + \varepsilon$.
Since $\varepsilon > 0$ is arbitrary, $\lim_{n\to\infty} e_n = \frac{1}{2}$.

CHAPTER 4

Conditional Probabilities

Example 1. Know that a die roll is prime. What is the probability that it is even?

$1 : 0 \quad 2 : \frac{1}{3} \quad 3 : \frac{1}{3} \quad 4 : 0 \quad 5 : \frac{1}{3} \quad 6 : 0 \quad \mathbb{P}(\text{even}) = \frac{1}{3}.$



Interested in probability of E.
$\rightarrow$ told that event F ocurrs
$\rightarrow$ for E to happen, $E \cap F$ must happen
Outcomes outside F now have zero probability $\Rightarrow$ to make total probability 1, we divide by $\mathbb{P}(F)$.

Definition 1. The conditional probability of E given F is

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

Observation.

- $E \cap F \subseteq F \Rightarrow 0 \leq \mathbb{P}(E \cap F) \leq \mathbb{P}(F) \Rightarrow 0 \leq \mathbb{P}(E|F) \leq 1$.
- If E, F are disjoint, then $\mathbb{P}(E|F) = 0$.
- $\mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F)$.

Example 2. (See Example 1.) There are a party with n people and n hats. What is the probability that nobody gets their own hat?
Solution. Before: calculated inclusion-exclusion

$$\mathbb{P}(0 \text{ people get own hats}) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \rightarrow e^{-1}$$

$$\mathbb{P}(n \text{ people get own hats}) = \frac{1}{n!}$$

Fix a set R of r people. Let $E_R = \{\text{people in } R \text{ get own hats and people not in } R \text{ don't}\}$.

$$\begin{aligned}\mathbb{P}(\text{exactly } r \text{ people get own hats}) &= \mathbb{P}\left(\dot{\bigcup_{R:|R|=r}} E_R\right)\\ &= \sum_{R:|R|=r} \mathbb{P}(E_R)\\ &= \binom{n}{r}\mathbb{P}(E_{\{1,\dots,r\}})\end{aligned}$$

$$E_R = \underbrace{\{r+1, r+2, \dots, n \text{ don't get own hats}\}}_{E} \cap \underbrace{\{1,2,\dots,r \text{ do get own hats}\}}_{F}$$

Use $\mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F)$.

$$\begin{aligned}\mathbb{P}(E|F) &= \mathbb{P}(\{\text{nobody gets own hate in a party of } n-r \text{ people}\})\\ &= \sum_{k=1}^{n-r} \frac{(-1)^k}{k!} \to e^{-1} \text{ if } n-r \to \infty\end{aligned}$$

Let $F_i = \{i\text{th person gets own hat}\}$. $F = F_1 \cap F_2 \cap \cdots \cap F_r$.

$$\begin{aligned}\mathbb{P}(F) &= \mathbb{P}((F_1 \cap F_2 \cap \cdots \cap F_{r-1}) \cap F_r)\\ &= \mathbb{P}(F_r | F_1 \cap F_2 \cap \cdots \cap F_{r-1})\mathbb{P}((F_1 \cap F_2 \cap \cdots \cap F_{r-2}) \cap F_{r-1})\\ &= \cdots = \mathbb{P}(F_r | F_1 \cap F_2 \cap \cdots \cap F_{r-1})\mathbb{P}(F_{r-1} | F_1 \cap F_2 \cap \cdots \cap F_{r-2}) \cdots \mathbb{P}(F_1)\end{aligned}$$

Observe that $\mathbb{P}(F_1) = \frac{1}{n}$, $\mathbb{P}(F_2|F_1) = \frac{1}{n-1}$,..., $\mathbb{P}(F_i | F_1 \cap F_2 \cap \cdots \cap F_{i-1}) = \frac{1}{n-i+1}$
$\Rightarrow \mathbb{P}(F) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \dots \cdot \frac{1}{n-r+1} = \frac{(n-r)!}{n!}$.

$$\mathbb{P}(\text{exactly } r \text{ people get own hats}) = \binom{n}{r}\mathbb{P}(E_{\{1,\dots,r\}}) \approx \binom{n}{r}\frac{1}{e} \cdot \frac{(n-r)!}{n!} = \frac{1}{r!e}$$

Suppose we can partition the sample space

$$S = F_1 \dot\cup F_2 \dot\cup \cdots \dot\cup F_n$$

Then for any event $E \subseteq S$,

$$E = E \cap S = E \cap \left(\dot{\bigcup_{i=1}^{n}} F_i\right) = \dot{\bigcup_{i=1}^{n}}(E \cap F_i)$$

$$\Rightarrow \mathbb{P}(E) \overset{\text{Axiom 3}}{=} \sum_{i=1}^{n} \mathbb{P}(E \cap F_i)$$

$$\Rightarrow \mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E|F_i)\mathbb{P}(F_i)$$

This is the Law of Total Probability.

Example 3. Go on holiday to Australia. Want to go to the beach. Maybe go swimming depending on the weather.

- if sunny: go swimming with probability 70%
- if not sunny: go swimming with probability 30%

The weather forecast says there is 80% chance of sunny. What is $\mathbb{P}(\text{swimming})$?
Solution.

$$\begin{aligned}&\mathbb{P}(\text{swimming})\\ &= \mathbb{P}(\text{swimming}|\text{sunny})\mathbb{P}(\text{sunny}) + \mathbb{P}(\text{swimming}|\text{not sunny})\mathbb{P}(\text{not sunny})\\ &= 0.7 \times 0.8 + 0.3 \times 0.2 = \boxed{0.62}\end{aligned}$$

Warm-up. Game show (Monty Hall)

- Three doors: behind one door is a car, behind the other two are goats.
- You choose one, then the host open another door that he knows has a goat.
- Offer you the option to switch doors. Should you?

Example 4. (See Example 3.) Let $\mathbb{P}(\text{sunny}) = 0.8$,

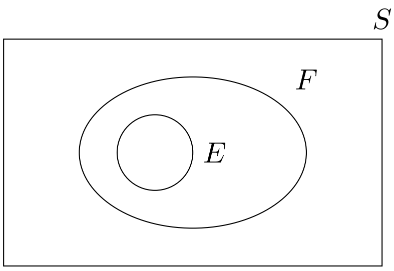
$$\mathbb{P}(\text{swim}|\text{sunny}) = 0.7, \quad \mathbb{P}(\text{swim}|\text{not sunny}) = 0.3,$$

$$\mathbb{P}(\text{bite}|\text{swim}) = 0.5, \quad \mathbb{P}(\text{bite}|\text{not swim}) = 0.01.$$

By law of total probability, $\mathbb{P}(\text{bite}) = 0.3138$.
<u>Q.</u> If I do get bitten by a shark, what is the probability it was sunny?
Solution.

$$\mathbb{P}(\text{sunny}|\text{bite}) = \frac{\mathbb{P}(\text{sunny} \cap \text{bite})}{\mathbb{P}(\text{bite})}$$

$$\mathbb{P}(\text{sunny} \cap \text{bite}) = \mathbb{P}(\text{bite} \cap \text{sunny}) = \mathbb{P}(\text{bite}|\text{sunny})\mathbb{P}(\text{sunny})$$



$$\begin{aligned}\mathbb{P}(\text{bite}|\text{sunny}) &= \mathbb{P}(\text{bite}|\text{swim, sunny})\mathbb{P}(\text{swim}|\text{sunny})\\ &\quad + \mathbb{P}(\text{bite}|\text{not swim, sunny})\mathbb{P}(\text{not swim}|\text{sunny})\\ &= \mathbb{P}(\text{bite}|\text{swim})\mathbb{P}(\text{swim}|\text{sunny}) + \mathbb{P}(\text{bite}|\text{not swim})\mathbb{P}(\text{not swim}|\text{sunny})\\ &= 0.5 \times 0.7 + 0.01 \times 0.3 = 0.353\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\text{sunny}|\text{bite}) &= \frac{\mathbb{P}(\text{sunny} \cap \text{bite})}{\mathbb{P}(\text{bite})}\\ &= \frac{0.353 \times 0.8}{0.3138} = \boxed{0.8999...}\end{aligned}$$

Theorem 1 (Bayes' Rule). If we have a partition $S = F_1 \cup F_2 \cup \cdots \cup F_n$ and an event $E \subseteq S$, then

$$\mathbb{P}(F_i|E) = \frac{\mathbb{P}(E|F_i)\mathbb{P}(F_i)}{\sum_{j=1}^{n} \mathbb{P}(E|F_j)\mathbb{P}(F_j)}.$$

Proof. By definition, $\mathbb{P}(F_i|E) = \frac{\mathbb{P}(F_i \cap E)}{\mathbb{P}(E)}$. By Law of total probability, $\mathbb{P}(E) = \sum_{j=1}^{n} \mathbb{P}(E|F_j)\mathbb{P}(F_j)$. This follows from $\mathbb{P}(F_i \cap E) = \mathbb{P}(E \cap F_i) = \mathbb{P}(E|F_i)\mathbb{P}(F_i)$. □

Example 5. 1% of the population has COVID. Rapid test for COVID has 95% accuracy, with 5% chance of "false positive" and 5% chance of "false negative".
Q. A random person tests positive. What is the probability they have COVID?
Solution. Let S be the population. Let

$$F_1 = \{\text{people with COVID}\}, \qquad \mathbb{P}(F_1) = 0.01$$
$$F_2 = \{\text{people without COVID}\}, \qquad \mathbb{P}(F_1) = 0.99$$

$$E = \{\text{test positive}\}, \qquad \mathbb{P}(E|F_1) = 0.95$$
$$\mathbb{P}(E|F_2) = 0.05$$

$$\text{(Bayes')} \qquad \mathbb{P}(F_1|E) = \frac{\mathbb{P}(E|F_1)\mathbb{P}(F_1)}{\mathbb{P}(E|F_1)\mathbb{P}(F_1) + \mathbb{P}(E|F_2)\mathbb{P}(F_2)} = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = \boxed{0.1610}$$

Example 6. DNA test:

- $\mathbb{P}(\text{positive}|\text{match}) = 1$
- $\mathbb{P}(\text{positive}|\text{not match}) = 0.0001$
- City of population 2500000
- Random person $\to$ DNA matches sample from the crime scene

What is $\mathbb{P}(\text{guilty})$?
Solution. Let $S = \{\text{all people in the city}\}$, $F_1 = \{\text{guilty}\}$, $F_2 = \{\text{not guilty}\}$.
$\mathbb{P}(F_1) = \frac{1}{2500000}$, $\mathbb{P}(F_2) = \frac{2499999}{2500000}$.
Let $E = \{\text{match on DNA test}\}$. $\mathbb{P}(E|F_1) = 1$, $\mathbb{P}(E|F_2) = 0.0001$.

$$\text{(Bayes')} \qquad \mathbb{P}(F_1|E) = \frac{\mathbb{P}(E|F_1)\mathbb{P}(F_1)}{\mathbb{P}(E|F_1)\mathbb{P}(F_1) + \mathbb{P}(E|F_2)\mathbb{P}(F_1)} = \frac{1 \times \frac{1}{2500000}}{1 \times \frac{1}{2500000} + \frac{1}{10000}\left(1 - \frac{1}{2500000}\right)} = \boxed{0.003984...}$$

CHAPTER 5

Independent Events

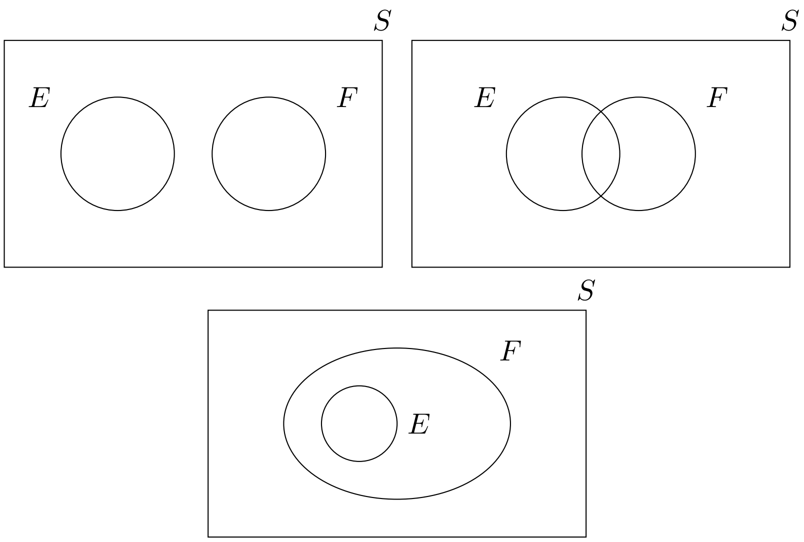
Definition 1. If $\mathbb{P}(E|F) = \mathbb{P}(E)$, then we say E and F are independent. Otherwise they are dependent.

Equivalently, E and F are independent iff

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F).$$

Corollary 1. Independence is symmetric in E, F.

Quiz. Which of the following pairs of events can be independent?



Example 1. Roll two dices.

$$E_1 = \{\text{first roll is a } 4\}\,,\ E_2 = \{\text{second roll is a } 3\}$$

$$F_1 = \{\text{sum is } 6\}\,,\ F_2 = \{\text{sum is } 7\}$$

Which pairs are independent?

Solution.

$$S = \{(1,1), \ldots (1,6), (2,1), \ldots, (2,6), \ldots, (6,1), \ldots, (6,6)\}$$

$$E_1 = \{(4,1), (4,2), \ldots, (4,6)\}\,, \quad \mathbb{P}(E_1) = \tfrac{6}{36} = \tfrac{1}{6}.$$

$$E_2 = \{(1,3), (2,3), \ldots, (6,3)\}\,, \quad \mathbb{P}(E_2) = \tfrac{6}{36} = \tfrac{1}{6}.$$

$$E_1 \cap E_2 = \{(4,3)\}\,, \mathbb{P}(E_1 \cap E_2) = \frac{1}{36} = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

$\Rightarrow E_1, E_2$ are independent.

$$F_1 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}, \quad \mathbb{P}(F_1) = \tfrac{5}{36}.$$
$$E_1 \cap F_1 = \{(4,2)\}, \mathbb{P}(E_1 \cap F_1) = \frac{1}{36} \neq \frac{5}{36} \cdot \frac{1}{6} = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

$\Rightarrow E_1, F_1$ are not independent.
F_1, F_2 not independent. They are disjoint.

$$F_2 = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}, \quad \mathbb{P}(F_1) = \tfrac{6}{36} = \tfrac{1}{6}.$$
$$E_i \cap F_2 = \{(4,3)\}, \mathbb{P}(E_i \cap F_2) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(E_i)\mathbb{P}(F_2).$$

$\Rightarrow E_1, E_2$ are both independent of F_2.

Claim 1. If E, F are independent, then E, F^C are independent.

Proof.
$$\begin{aligned}
\mathbb{P}(E \cap F^C) &= P(E) - \mathbb{P}(E \cap F) \\
&= \mathbb{P}(E) - \mathbb{P}(E)\mathbb{P}(F) && \text{(independence)} \\
&= \mathbb{P}(E)(1 - \mathbb{P}(F)) = \mathbb{P}(E)\mathbb{P}(F^C)
\end{aligned}$$

□

However, if

$$\begin{aligned}
&E_1, F \text{ are independent, and} \\
&E_2, F \text{ are independent,}
\end{aligned}$$

that doesn't mean

$$\begin{aligned}
&E_1 \cup E_2, F \text{ are independent, or} \\
&E_1 \cap E_2, F \text{ are independent.}
\end{aligned}$$

Definition 2. We say E_1, E_2, E_3 are (mutually) independent if:

- $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3)$
- $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$
- $\mathbb{P}(E_1 \cap E_3) = \mathbb{P}(E_1)\mathbb{P}(E_3)$
- $\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_2)\mathbb{P}(E_3)$

all hold.

There is a more general version:

Definition 3. Given a sequence of events $E_1, E_2, E_3, \ldots$, we say they are (mutually) independent if for any finite set I of indices,

$$\mathbb{P}\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \mathbb{P}(E_i)$$

Example 2. Inclusion-Exclusion for independent events.

Let $E_1, E_2, E_3, \ldots, E_n$ be independent.

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{\substack{I \subseteq [n] \\ I \neq \varnothing}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} E_i\right) \\
&= \sum_{\substack{I \subseteq [n] \\ I \neq \varnothing}} (-1)^{|I|+1} \prod_{i \in I} \mathbb{P}(E_i) \\
&= 1 - \prod_{i=1}^{n} (1 - \mathbb{P}(E_i))
\end{aligned}$$

Alternatively, use De Morgan to turn the union into an intersection:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) &= 1 - \mathbb{P}\left(\left(\bigcup_{i=1}^{n} E_i\right)^C\right) \\
&= 1 - \mathbb{P}\left(\bigcap_{i=1}^{n} E_i^C\right) \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}(E_i^C) = 1 - \prod_{i=1}^{n} (1 - \mathbb{P}(E_i))
\end{aligned}$$

Application. Suppose we have a test with a false negative rate of 1% and a false positive rate rate of 50%.

Suppose we can repeat the test independently.

If actually positive, $\mathbb{P}(\text{pos, pos}) = 0.99 \times 0.99 \geq 0.98$.

If actually negative, $\mathbb{P}(\text{pos, pos}) = 0.5 \times 0.5 = 0.25$.

Let $S = (0, 1]$, $z \in S$ be uniformly randomly chosen. That is, $\mathbb{P}(z \in (x, y]) = y - x$.

Let $E_1, E_2, \ldots$ be events in the probability space. Let $p_i = \mathbb{P}(E_i)$.

The 1st Borel-Cantelli Lemma states that if $\sum_{n=1}^{\infty} p_n < \infty$, then $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$.

Homework: if $\sum_{n=1}^{\infty}$, then it is possible that $\mathbb{P}(\limsup_{n\to\infty} E_n) = 1$.

Also possible that $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$. For example, $E_n = (0, \frac{1}{n}]$.

Theorem 1 (2nd Borel-Cantelli Lemma). If $E_1, E_2, \ldots$ are mutually independent events and $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$, then $\mathbb{P}(\limsup_{n\to\infty} E_n) = 1$.

Proof. Recall that $\limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \left(\bigcup_{i=n}^{\infty} E_n\right)$.

$$\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 1 \Rightarrow \mathbb{P}\left(\left(\limsup_{n\to\infty} E_n\right)^C\right) = 0$$

$$\left(\limsup_{n\to\infty} E_n\right)^C = \left(\bigcap_{n=1}^{\infty}\left(\bigcup_{i=n}^{\infty} E_n\right)\right)^C = \bigcup_{n=1}^{\infty}\left(\bigcup_{i=n}^{\infty} E_n\right)^C = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} E_n^C$$

$$\mathbb{P}\left(\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{i=n}^{\infty} E_n\right)\right)^C\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} E_n^C\right)$$

(continuity)
$$= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{\infty} E_n^C\right)$$

(independence, *)
$$= \lim_{n\to\infty} \prod_{i=n}^{\infty} \mathbb{P}(E_i^C)$$

$$= \lim_{n\to\infty} \prod_{i=n}^{\infty} (1 - \mathbb{P}(E_i))$$

(**)
$$= \lim_{n\to\infty} 0 = 0$$

(**) by convergence test for infinite product ($\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$) and (*) by

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} E_i^C\right) = \mathbb{P}\left(\lim_{N\to\infty}\bigcap_{i=n}^{N} E_i^C\right)$$

(continuity)
$$= \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{N} E_i^C\right)$$

(independence)
$$= \lim_{N\to\infty} \prod_{i=n}^{N} \mathbb{P}(E_i^C)$$

$$= \prod_{i=n}^{\infty} \mathbb{P}(E_i^C)$$

□

CHAPTER 6

Discrete Random Variables

1. Discrete Random Variable

Definition 1. Given a probability probability space $(S, \mathbb{P})$, a random variable is a function $X : S \to \mathbb{R}$. It is discrete if it only takes countably many value.

Observation. A discrete random variable defines a (simpler) probability space.

Let $x_1, x_2, x_3, \ldots$ be the values X can take. i.e. $X(S) = \{x_1, x_2, x_3, \ldots\}$. $\leftarrow$ new sample space $p(x_i) = \mathbb{P}(X(s) = x_i) = \mathbb{P}(\{s \in S \mid X(s) = x_i\})$.
Observation.

$$\begin{aligned} \sum_i p(x_i) &= \sum_i \mathbb{P}(X(s) = x) \\ &= \sum_i \mathbb{P}(X^{-1}(x_i)) && \text{(pairwise disjoint)} \\ &= \mathbb{P}(\cup_i X^{-1}(x_i)) \\ &= \mathbb{P}(S) = 1 \end{aligned}$$

Example 1. Multiple choice exam

- 5 questions, each question has 4 options, one is correct
- pick uniformly random answer on each question, independently

Q. What is the probability of getting non of them correct?
Solution. Let $X =$ the number of correct answers.

Calculate $\mathbb{P}(X = 0)$:
$\mathbb{P}(X = 0) = \mathbb{P}(F_1 \cap F_2 \cap \cdots \cap F_5)$, $F_i = \{\text{get } i\text{th question wrong}\}$. $\mathbb{P}(F_i) = \frac{3}{4}$.
By independence, $\mathbb{P}\left(\bigcap_{i=1}^5 F_i\right) = \prod_{i=1}^5 \mathbb{P}(F_i) = \left(\frac{3}{4}\right)^5$.

We can calculate

$$\mathbb{P}(X=0)=\left(\frac{3}{4}\right)^5$$
$$\mathbb{P}(X=1)=\binom{5}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^4$$
$$\mathbb{P}(X=2)=\binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3$$
$$\mathbb{P}(X=3)=\binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2$$
$$\mathbb{P}(X=4)=\binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)$$
$$\mathbb{P}(X=5)=\left(\frac{1}{4}\right)^5$$

Example 2. Promotion: n different types of prizes

In each attempt, we get a uniformly random prize, independent of previous attempt.

Q. How many attempts do we need to get all types of prizes?

Solution. Let $S=\{(s_1,s_2,s_3,\ldots)\,|\,1\le s_i\le n\}$, and $X((s_1,s_2,s_3,\ldots))=\min\{t\,|\,(s_1,s_2,s_3,\ldots)$ has all numbers from 1 to $n\}$.

If $t<n$, $\mathbb{P}(X=t)=0$.

$$\mathbb{P}(X=n)=\frac{n!}{n^n}\approx\frac{1}{(e+o(1))^n}$$

If $t>n$, $\mathbb{P}(X=t)=?$

$$\mathbb{P}(X>t)=\mathbb{P}\left(\bigcup_{i=1}^n E_i\right)\text{ where } E_i=\{i\text{th prize is ruisrily after } t \text{ attempts}\}$$

$$\mathbb{P}(E_i)=\left(\frac{n-1}{n}\right)^t\leftarrow\frac{n-1}{n}\text{ probability for each independent try}$$

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right)\overset{\text{inc-exc}}{=}\sum_{\varnothing\neq I\subseteq[n]}(-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I}E_i\right)$$

$$\mathbb{P}\left(\bigcap_{i\in I}E_i\right)=\left(\frac{n-|I|}{n}\right)^t\leftarrow n-|I|\text{ bid options for each attempt}$$

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right)=\sum_{I\neq\varnothing}(-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I}E_i\right)=\sum_{r=1}^n(-1)^{r+1}\binom{n}{r}\left(\frac{n-r}{n}\right)^t$$

Therefore

$$\mathbb{P}(X=t)=\mathbb{P}(X>t-1)-\mathbb{P}(X>t)=\sum_{r=1}^n(-1)^{r+1}\binom{n}{r}\left(\frac{n-r}{n}\right)^{t-1}\left(1-\frac{n-r}{n}\right)$$

2. Expectation

Definition 2. Given a probability space $(S, \mathbb{P})$ and a discrete random variable $X : S \to \mathbb{R}$ which takes values $x_1, x_2, \ldots$, the expectation of X is

$$\mathbb{E}[X] = \sum_i x_i p(x_i) = \sum_i x_i \mathbb{P}(X = x_i).$$

Example 3. (See Example 1.) Multiple choice exam

- 2 questions, each question has 4 options
- pick uniformly random answer on each question, independently

Q. What is the expected number of correct answers?
Solution. X takes values 0, 1, or 2.

$$p(0) = \left(\frac{3}{4}\right)^2 = \frac{9}{16},\ p(1) = \binom{2}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{6}{16},\ p(2) = \left(\frac{1}{4}\right)^2 = \frac{1}{16}$$

$$\mathbb{E}[X] = 0 \cdot \frac{9}{16} + 1 \cdot \frac{6}{16} + 2 \cdot \frac{1}{16} = \frac{8}{16} = \frac{1}{2}$$

Multiple choice, +1 point if answer correct and −1 point if answer is incorrect.
Let Y = score. What is the xpectation of Y?

X	Y	$p(Y)$
0	-2	$\frac{9}{16}$
1	0	$\frac{6}{16}$
2	2	$\frac{1}{16}$

$$Y = X - (2 - X) = 2X - 2$$

$$\mathbb{E}[Y] - \frac{9}{16} \cdot (-2) + \frac{6}{16} \cdot 0 + \frac{1}{16} \cdot 2 = -1 = 2 \cdot \frac{1}{2} - 2$$

Lemma 1 (Linearity of Expectation). Let $X_1, X_2, \ldots, X_n$ be random variables in a probability space $(S, \mathbb{P})$.

Let $Y = \sum_{i=1}^n \alpha_i X_i$ for some $\alpha_i \in \mathbb{R}$. Then $\mathbb{E}[Y] = \sum_{i=1}^n \alpha_i \mathbb{E}[X_i]$.

Proof. First, we prove the claim:

Claim 1. $\mathbb{E}[X] = \sum_{s \in S} X(s)\mathbb{P}(s)$.

Proof. (claim) By definition, if $X(S) = \{x_2, x_2, \dots\}$,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_i x_i p(x_i) \\
&= \sum_i x_i \mathbb{P}(\{s \in S \mid X(s) = x_i\}) \\
&= \sum_i x_i \mathbb{P}\left(\bigcup_{s \in X^{-1}(x_i)} \{s\} \right) \\
&= \sum_i x_i \sum_{s \in X^{-1}(x_i)} \mathbb{P}(s) \\
&= \sum_{s \in S} X(s)\mathbb{P}(s)
\end{aligned}$$

□

$$\begin{aligned}
\Rightarrow \mathbb{E}[Y] &= \sum_{x \in S} Y(s)\mathbb{P}(s) \\
&= \sum_{x \in S} \left(\sum_{i=1}^{n} \alpha_i X_i(s) \right) \mathbb{P}(s) \\
&= \sum_{x \in S} \sum_{i=1}^{n} \alpha_i X_i(s) \mathbb{P}(s) \\
&= \sum_{i=1}^{n} \alpha_i \sum_{x \in S} X_i(s)\mathbb{P}(s) \\
&= \sum_{i=1}^{n} \alpha_i \mathbb{E}[x_i]
\end{aligned}$$

□

Example 4. (See Example 1.) Multiple choice exam

- n questions, each question has k options
- pick uniformly random answer on each question, independently

Q. What is the expectation number of correct answers?
Solution. Let X = number of correct answers. Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th question is right } \left(\frac{1}{k}\right) \\ 0 & \text{otherwise } \left(\frac{k-1}{k}\right). \end{cases}$$

Then $X = \sum_{i=1}^{n} X_i$.

$$\overset{\text{LoE}}{\Rightarrow} \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{k} = \boxed{\frac{n}{k}}$$

Example 5. (See Example 1.) Multiple choice exam

- first 10 questions have 3 options
- last 5 questions have 5 options
- pick uniformly random answer on each question, independently

Q. What is

(a) the probability of getting exactly k correct?
(b) the expected number of correct answers?

Solution.

(a) Suppose we get l correct from the first 10, $0 \leq l \leq 10$.
$\Rightarrow k - l$ correct from last 5. Then the answer would be

$$\sum_{l=0}^{10} \binom{10}{l} \binom{5}{k-l} \left(\frac{1}{3}\right)^l \left(\frac{2}{3}\right)^{10-l} \left(\frac{1}{5}\right)^{k-l} \left(\frac{4}{5}\right)^{5-k+1}.$$

(Define $\binom{n}{r} = 0$ for $r > n$.)

(b) Let X_i be the indicator random variable for the event that we got the i-th question right.

$$X_i = \begin{cases} 1 & \text{if } i\text{-th question correct} \\ 0 & \text{if not} \end{cases}$$

Then if $X =$ the number of correct answers, $X = \sum_{i=1}^{15} X_i$.
By linearity of expectation,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^{15} \mathbb{E}[X_i] = \sum_{i=1}^{15} \mathbb{P}(X_i = 1) \\ &= \sum_{i=1}^{10} \mathbb{P}(i\text{-th question correct}) + \sum_{i=11}^{10} \mathbb{P}(i\text{-th question correct}) \\ &= \sum_{i=1}^{10} \frac{1}{3} + \sum_{i=11}^{10} \frac{1}{5} = \boxed{\frac{13}{3}} \end{aligned}$$

Theorem 1 (Markov's Inequality). If X is a discrete random variable taking nonnegative values, then for any $t \in \mathbb{R}_{>0}$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Remark 1. (a) Nonnegativity is necessary. Consider

$$X = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

Then $\mathbb{E}[X] = 0$, but for $t \leq 1$, $\mathbb{P}(X \geq t) \geq \frac{1}{2} > 0$.

(b) Inequality is useless for $t \leq \mathbb{E}[X]$, but useful for saying a random variable is unlikely to be much bigger than its expectation.

Proof.

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_x xp(x) \\
&= \sum_{x:x<t} xp(x) + \sum_{x:x\geq t} xp(x) \\
(X \text{ is nonnegative}) \qquad &\geq \sum_{x:x<t} 0 + \sum_{x:x\geq t} tp(x) \\
&= t \sum_{x:x\geq t} p(x) \\
&= t \sum_{x:x\geq t} \mathbb{P}(\{X = x\}) \\
(\text{disjoint events}) \qquad &= t\mathbb{P}\left(\bigcup_{x:x\geq t} \{X = x\} \right) \\
&= t\mathbb{P}(X \geq t)
\end{aligned}
$$

□

From Markov's inequality, we can know that if $\mathbb{E}[X]$ is low, X is likely to be low.
Is the converse true? if $\mathbb{E}[X]$ is high, is X likely to be high?
This is in general not true. For example, let

$$
X = \begin{cases} 1000000 & \text{with probability } \frac{1}{1000} \\ 0 & \text{with probability } \frac{999}{1000}. \end{cases}
$$

Then $\mathbb{E}[X] = 1000000 \cdot \frac{1}{1000} + 0 \cdot \frac{999}{1000} = 1000$. But $\mathbb{P}(X > 0) = \frac{1}{1000}$.
Fun question. There are 3 investment option. Which one would you take?

$$
X_1 = 1 \text{ with probability } 1 \qquad \mathbb{E}[X_1] = 1
$$

$$
X_2 = \begin{cases} 1000 & \text{with probability } \frac{1}{1000} \\ 0 & \text{with probability } \frac{999}{1000} \end{cases} \qquad \mathbb{E}[X_2] = 1
$$

$$
X_3 = \begin{cases} \frac{2000}{999} & \text{with probability } \frac{999}{1000} \\ -1000 & \text{with probability } \frac{1}{1000} \end{cases} \qquad \mathbb{E}[X_3] = 1
$$

3. Variance

We want to know that how far from the expectation are we on average.

Definition 3. The variance of a random variable X with expectation μ is

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

Proposition 1. $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Proof.

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X-\mu)^2] \\
&= \sum_x (x-\mu)^2 p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}[X^2] - \mu^2
\end{aligned}$$

□

Example 6. Let X_1, X_2, X_3 be the investment strategies from before.

$$\begin{aligned}
\mathrm{Var}(X_1) &= \mathbb{E}[(X_1-1)^2] = 0 \\
\mathrm{Var}(X_2) &= \mathbb{E}[(X_2-1)^2] = 999^2 \cdot \frac{1}{1000} + (-1)^2 \cdot \frac{999}{1000} \\
&= \frac{999}{1000}(999+1) = 999 \\
&= \mathbb{E}[X_2^2] - \mathbb{E}[X_2] = \left(1000^2 \cdot \frac{1}{1000} + 0^2 \cdot \frac{999}{1000}\right) - 1^2 \\
&= 1000 - 1 = 999 \\
\mathrm{Var}(X_3) &= \mathbb{E}[(X_3-1)^2] = \mathbb{E}[X_3^2] - \mathbb{E}[X_3] \\
&= \left(\left(\frac{2000}{999}\right)^2 \cdot \frac{999}{1000} + (-1000)^2 \frac{1}{1000}\right) - 1 \\
&= \left(\frac{4000}{999} + 1000\right) - 1 = 1003\frac{4}{999}
\end{aligned}$$

Definition 4. The standard deviation of a random variable is the square root of its variance, often denoted by $\sigma(X)$.

Theorem 2 (Chebychev's Inequality). Let X be a random variable with expectation $E[X] = \mu$. Then for any $t > 0$,

$$\mathbb{P}(|X-\mu| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proof. Apply Markov's inequality to the nonnegative random variable $(X-\mu)^2$. Observe that

$$\{|X-\mu| \geq t\} = \left\{(X-\mu)^2 \geq t^2\right\}.$$

By Markov,

$$\mathbb{P}((X-\mu)^2 \geq t^2) \leq \frac{\mathbb{E}[(X-\mu)^2]}{t^2} = \frac{\operatorname{Var}(X)}{t^2}.$$

□

Corollary 1. The probability that X is at least k standard deviations away from its expectation is $\leq \frac{1}{k^2}$.

Remark 2. Let X be a random variable, $a, b \in \mathbb{R}$. Define $Y = aX + b$.

By linearity, $\mathbb{E}[Y] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

What about the variance?

$$\begin{aligned}
\operatorname{Var}(Y) &= \mathbb{E}[(Y - \mathbb{E}[Y])^2] \\
&= \mathbb{E}[(aX + b - (a\mathbb{E}[X] + b))^2] \\
&= \mathbb{E}[(a(X - E[X])^2)] \\
&= a^2\mathbb{E}[(X - \mathbb{E}[x])^2] = a^2 \operatorname{Var}(X)
\end{aligned}$$

CHAPTER 7

Discrete Distributions

1. Binomial Distribution

Setting:

- run n independent trial of a random experiment
- each trial is a success with probability p
- count the number of successes

Denoted by $\mathrm{Bin}(n,p)$.

Distribution: The possible values are $0,1,2,\dots,n$. The probability that we get k successes is

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

Observation.

$$\sum_k p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$$

Remark 1. When $n = 1$, we get a Bernoulli distribution, defined by

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

Denoted by $\mathrm{Ber}(p)$.

Therefore

$$\mathrm{Bin}(n,p) = \text{ sum of } n \text{ independent Bernoulli random variables.}$$

Statistics. Let $Y \sim \mathrm{Ber}(p)$ (Y be a $\mathrm{Ber}(p)$ random variable). Then

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

Let $X \sim \mathrm{Bin}(n,p)$. Then $X = \sum_{i=1}^{n} X_i$ where each $X_i \sim \mathrm{Ber}(p)$ independently.

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = \boxed{np}$$

To calculate the expectation of the binomial distribution manually, we use the binomial theorem.

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

$$\overset{\frac{\mathrm{d}}{\mathrm{d}x}}{\Rightarrow} \sum_{k=0}^{n} k \binom{n}{k} x^{k-1} y^{n-k} = n(x+y)^{n-1}$$

Multiply both side by x,

$$\sum_{k=0}^{n} k \binom{n}{k} x^{k} y^{n-k} = nx(x+y)^{n-1}.$$

Substitute $x = p$, $y = 1-p$, and we can get

$$\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^{k} (1-p)^{n-k} = np(p+(1-p))^{n-1} = \boxed{np}.$$

Now, to calculate the variance of the binomial distribution, we need to compute $\mathbb{E}[X^2]$. Observe

$$\sum_{k=0}^{n} k \binom{n}{k} kx^{k} y^{n-k} = nx(x+y)^{n-1}$$

$$\overset{\frac{\mathrm{d}}{\mathrm{d}x}}{\Rightarrow} \sum_{k=0}^{n} k^2 \binom{n}{k} kx^{k-1} y^{n-k} = n(x+y)^{n-1} + n(n-1)x(x+y)^{n-2}$$

Multiply both side by x,

$$\sum_{k=0}^{n} k^2 \binom{n}{k} kx^{k} y^{n-k} = nx(x+y)^{n-1} + n(n-1)x^2(x+y)^{n-2}$$

Substitute $x = p$, $y = 1-p$, and we can get

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^{k} (1-p)^{n-k} \\ &= np(p+(1-p))^{n-1} + n(n-1)p^2(p+(1-p))^{n-2} = \boxed{np + n(n-1)p^2}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= np + n(n-1)p^2 - n^2p^2 \\ &= np - np^2 = \boxed{np(1-p)}\end{aligned}$$

Also, We can calculate the variance of Bernoulli distribution:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

$$X^2 = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

$$\Rightarrow \mathbb{E}[X^2] = \mathbb{E}[X] = p$$

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= p - p^2 = \boxed{p(1-p)}\end{aligned}$$

Remark 2. We have the following observation:

(a) Let $X \sim \mathrm{Bin}(n,p)$. Then $\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$.
By Chebychev we can know that $\mathbb{P}(|X - np| \geq t) \leq \frac{np(1-p)}{t^2}$.
That is, even though there are $n+1$ values the distribution can take, the probability it is outside an interval of with $\Theta(\sqrt{n})$ around the expectation is very small.

(b) $\mathbb{E}[X^2] = \underbrace{\mathbb{E}[X(X-1)]}_{\sum_k k(k-1)p(k)} + \mathbb{E}[X]$.

2. Poisson Distribution

Setting:

- the number of earthquakes in Taiwan in a month
- on average, there are λ earthquakes in a month
- divide into n equal time intervals $\to$ expect $\frac{\lambda}{n}$ earthquakes in each interval

Assumption:

- At most one earthquakes per interval.
- Each interval is independent of the others.

The number of earthquakes $\sim \mathrm{Bin}(n, \frac{\lambda}{n})$.
Distribution:

$$\mathbb{P}(k \text{ earthquakes in a month}) \approx \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Take $n \to \infty$,

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \to \frac{\lambda^k}{k!}$$

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \to \frac{e^{-\lambda}}{1}$$

Therefore the Possion distribution with parameter $\lambda > 0$, $\mathrm{Poi}(\lambda)$ has distribution

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \text{for } k = 0, 1, 2, \ldots.$$

Fun fact. This is a distribution $p(k) \geq 0$ for all $k \geq 0$.

$$\begin{aligned}\sum_{k=0}^{\infty} p(k) &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}e^{\lambda} = 1\end{aligned}$$

Remark 3. $\text{Poi}(\lambda)$ is a good approximation for $\text{Bin}(n, \frac{\lambda}{n})$ when n is large.

That is to say, Poisson distribution is appropriate when we have many independent events, each with small probability.

For example,

- number of customers in a shop in an hour.
- number of people who will die in a day.
- radioactive decay.

Statistics. Let $X \sim \text{Poi}(\lambda)$. The expectation is

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} kp(k) \\
&= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\
&= \sum_{k=0}^{\infty} \frac{\lambda^{k+1} e^{-\lambda}}{k!} \\
&= \lambda \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \lambda \sum_{k=0}^{\infty} p(k) = \boxed{\lambda}
\end{aligned}$$

The variance is

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \lambda - \lambda^2
\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1)p(k)\\ &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!}\\ &= \sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!}\\ &= \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!}\\ &= \sum_{k=0}^{\infty} \frac{\lambda^{k+2} e^{-\lambda}}{k!} = \lambda^2\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X(X-1)] + \lambda - \lambda^2\\ &= \lambda^2 + \lambda - \lambda^2 = \boxed{\lambda}\end{aligned}$$

Like what we mentioned above, Poi $\approx \mathrm{Bin}(n, \frac{\lambda}{n})$, which has expectation $np = \lambda$ and variance $np(1-p) = n \cdot \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) \approx \lambda$.

The Poisson Paradigm. The Possion distribution is more widely applicable: if we have n events $E_1, E_2, E_3, \ldots, E_n$ such that

- $p_i = \mathbb{P}(E_i)$ is small for every i, and
- the events are "weakly independent": for $j \neq i$, $\mathbb{P}(E_i|E_j) \approx p_i$,

then if $\lambda = p_1 + p_2 + \cdots + p_n$, $\mathrm{Poi}(\lambda)$ is a good approximation to the number of events that occur.

Example 1. (See Example 1.) There are a party with n people and n hats. What is the probability that nobody gets their own hat?
Solution. Let $E_i = \{i\text{-th person gets own hat}\}$. Then $\mathbb{P}(E_i) = \frac{1}{n}$, $\mathbb{P}(E_i|E_j) = \frac{1}{n-1}$.

Therefore the Poisson paradigm applies. The number of correct hats $\approx$ Poi(1).

$$\mathbb{P}(\text{nobody gets own hat}) \approx \frac{1^0 e^{-1}}{0!} = \frac{1}{e}.$$

$$\mathbb{P}(\text{exactly } k \text{ gets own hat}) \approx \frac{1^k e^{-1}}{k!} = \frac{1}{k!e}.$$

Example 2. Toss a fair coin n times. Let L_n denote the length of longest sequence of consecutive heads.

$$\begin{aligned} E &= \{\text{there is a sequence of } k \text{ heads in a row}\} \\ &= \{L_n \geq k\} \\ &= \bigcup_{i=1}^{n-k+1} E_i, \text{ where } E_i = \{\text{tosses } i, i+1, \dots, i+k-1 \text{ are all heads}\} \end{aligned}$$

We have $\mathbb{P}(E_i) = \frac{1}{2^k}$. However, these events are far from independence:

$$\mathbb{P}(E_i | E_j) = \frac{1}{2^k} \text{ if } i - j \geq k,$$

but $\mathbb{P}(E_i | E_{i-1}) = \frac{1}{2}$. So the Poisson paradigm does not apply in this setting.
Fortunately, we can fix the problem by letting $E = \bigcup_{i=1}^{n-k+1} E_i'$, where

$$E_i' = \begin{cases} \text{tosses } i, i+1, \dots, i+k-1 \text{ are all heads AND } i+k \text{ is tail} & \text{if } 1 \leq i \leq n-k \\ \text{tosses } n-k+1, n-k+2, \dots, n \text{ are all heads} & \text{if } i = n-k+1. \end{cases}$$

Then

$$\mathbb{P}(E_i') \begin{cases} \frac{1}{2^{k+1}} & \text{if } 1 \leq i \leq n-k \text{ (fix outcome of } k+1 \text{ tosses)} \\ \frac{1}{2^k} & \text{if } i = n-k+1 \text{ (same as before)} \end{cases}$$

Hence we have

$$\mathbb{P}(E_i' | E_j') = \begin{cases} \mathbb{P}(E_i) & \text{if } i, j \text{ are far apart} \\ 0 & \text{if sequence overlap } \rightarrow \text{ close to } \mathbb{P}(E_i'). \end{cases}$$

Then Poisson paradigm applies. :)
The number of k heads followed by a tail at the end of tosses is

$$X_k \sim \text{Poi}\left(\frac{n-k}{2^{k+1}} + \frac{1}{2^k}\right) = \text{Poi}\left(\frac{n-k+2}{2^{k+1}}\right).$$

$$\{L_n \leq k\} = \{X_{k+1} = 0\}$$

By the Poisson paradigm,

$$\begin{aligned} \mathbb{P}(X_{k+1} = 0) &\approx \frac{\lambda_{k+1}^0 e^{-\lambda_{k+1}}}{0!} \\ &= e^{-\lambda_{k+1}}, \text{ where } \lambda_{k+1} = \frac{n-k+1}{2^{k+2}} \end{aligned}$$

$$\begin{aligned} \mathbb{P}(L_n \leq k) &\approx e^{-\frac{n-k+1}{2^{k+2}}} \\ &\approx e^{-\frac{n}{2^{k+2}}} \end{aligned}$$

Finally,

$$\begin{aligned}\mathbb{P}(L_n = k) &= \mathbb{P}(L_n \le k) - \mathbb{P}(L_n \le k-1)\\ &= e^{-\frac{n}{2^{k+2}}} - e^{-\frac{n}{2^{k+1}}}\\ &= e^{-\frac{n}{2^{k+2}}}\left(1 - e^{-\frac{n}{2^{k+2}}}\right)\end{aligned}$$

In order to have $\mathbb{P}(L_n = k) \not\to 0$, we need $e^{-\frac{n}{2^{k+2}}} \not\to 0$ and $e^{-\frac{n}{2^{k+2}}} \not\to 1$. Therefore we need $k \approx \log_2 n - 2$.

3. Geometric Distribution

Setting:

- Independent trials, successful with probability p.
- How many trials until our first success?

Denoted by $\text{Geom}(p)$.

Distribution: $\mathbb{P}(X = k) = \mathbb{P}(\overbrace{FFF\ldots F}^{\substack{\text{first } k-1\\ \text{trials failed}}}\underbrace{S}_{\substack{k\text{-th trial}\\ \text{success}}}) = (1-p)^{k-1}p$

Verify this is a valid distribution:

$$\begin{aligned}\sum_{k=1}^{\infty}\mathbb{P}(X = k) &= \sum_{k=1}^{\infty}(1-p)^{k-1}p\\ &= p\sum_{k=1}^{\infty}(1-p)^{k-1}\\ &= p\cdot\frac{1}{1-(1-p)} = \frac{p}{p} = 1\end{aligned}$$

Statistics. To calculate the expectation of the geometry distribution, we observe

$$\sum_{k=1}^{\infty} x^k = \frac{x}{1-x} \qquad \text{(geometric series)}$$

$$\overset{\frac{\mathrm{d}}{\mathrm{d}x}}{\Rightarrow} \sum_{k=1}^{n} kx^{k-1} = (1-x)^{-1} + x(1-x)^{-2} = \frac{1}{(1-x)^2}$$

Substitute $x = 1-p$, and we can get

$$\mathbb{E}[X] = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = \boxed{\tfrac{1}{p}}.$$

Example 3. A casino has a game where you gave a 50% chance of winning.

If you bet $\$x$, then if you win, you get $\$2x$.

If you lose, you get \$0.

Q1. What is your expected profit/loss?

Solution. Let X = profit. Then

$$X = \begin{cases} \$x & \text{if we win, } \mathbb{P} = \frac{1}{2} \\ -\$x & \text{if we lose, } \mathbb{P} = \frac{1}{2}. \end{cases}$$

We have $\mathbb{E}[X] = \frac{1}{2}\$x + \frac{1}{2}(-\$x) = \$0$.

Q2. You aren't happy with losing, so your strategy is to keep betting \$1 until you win. What is your expected profit/loss?

Solution.

$$\begin{aligned} X &= \$1 - (\text{number of losses}) \cdot \$1 \\ &= \$2 - \underbrace{(\text{number of trials})}_{\text{Geom}\left(\frac{1}{2}\right)} \cdot \$1 \end{aligned}$$

Let Y = number of trials until first win. Then $Y \sim \text{Geom}\left(\frac{1}{2}\right)$. Compute

$$\mathbb{E}[X] = \mathbb{E}[2 - Y] = 2 - \mathbb{E}[Y] = 2 - \frac{1}{\frac{1}{2}} = \boxed{0}.$$

Q3. You have a new strategy: every time we lose, we double our bet and go again. Repeat until we win.

number of games	profit	how much money we need
1	$+\$1$	$\$1$
2	$-\$1 + \$2 = +\$1$	$\$1 + \$2 = \$3$
3	$-\$1 - \$2 + \$4 = +\$1$	$\$1 + \$2 + \$4 = \$7$
$\vdots$	$\vdots$	$\vdots$
k	$-\$1 - \$2 - \cdots - \$2^{k-2} + \$2^{k-1} = +\$1$	$\$1 + \$2 + \$4 + \cdots + \$2^{k-1} = \$2^k - 1$

Note that no matter how many times you lose before you win, you win \$1 back.

Therefore $\mathbb{E}[X] = \$1$ since $\mathbb{P}(X = 1) = 1$.

However,

$$\begin{aligned} \mathbb{E}[\text{amount of money needed}] &= \sum_{k=1}^{\infty} (2^k - 1)\left(\frac{1}{2}\right)^k \\ &= \sum_{k=1}^{\infty} 1^k - \sum_{k=1}^{\infty}\left(\frac{1}{2}\right)^k \\ &= \infty - 1 \end{aligned}$$

Example 4. Coupon collecter (See Homework 3.2.).

There are n types of coupons. Every coupon we get is uniformly random, independent of previous coupons.

Q. How many coupon do we need to collect them all?

Solution. Let X_i be the number of coupons we need to get the i-th new coupon after we got the $(i-1)$-th. The answer we want is $X_1 + X_2 + \cdots + X_n$.

(first coupon is always new)
$$X_1 = 1$$
$$X_2 \sim \text{Geom}\left(\frac{n-1}{n}\right)$$
$$\to \text{ each coupon is independent}$$
$$\to \text{ probability of being new } = \frac{n-1}{n}$$
$$\to \text{ repeat until we get a new one}$$
$$X_i \sim \text{Geom}\left(\frac{n-i+1}{n}\right)$$

Therefore

(LoE)
$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] \\
&= \sum_{i=1}^n \frac{1}{\frac{n-i+1}{n}} = \sum_{i=1}^n \frac{n}{n-i+1} \\
&= \sum_{i=1}^n \frac{n}{i} = n\sum_{i=1}^n \frac{1}{i} \\
&= nH_n \approx n \log n
\end{aligned}$$

Calculate the variance of Geom(p):
$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \underbrace{\mathbb{E}[X]}_{\frac{1}{p}} - \underbrace{\mathbb{E}[X^2]}_{\frac{1}{p^2}}
\end{aligned}$$

To calculate $\mathbb{E}[X(X-1)]$, observe
$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$
$$\overset{\frac{\mathrm{d}}{\mathrm{d}x}}{\Rightarrow} \sum_{k=1}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

Multiply both side by x,
$$\sum_{k=1}^{\infty} k(k-1)x^{k-1} = \frac{2x}{(1-x)^3}.$$

Substitute $x = 1 - p$, and we can get

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1}p \\ &= p\frac{2(1-p)}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \boxed{\tfrac{1-p}{p^2}}.\end{aligned}$$

Example 5. Estimate $X =$ the number of dice rolls until the first 6.

Then $X \sim \mathrm{Geom}(\frac{1}{6})$.

$$\mathbb{E}[X] = \frac{1}{\frac{1}{6}} = 6$$

$$\mathrm{Var}(X) = \frac{1 - \frac{1}{6}}{\frac{1}{36}} = 30$$

4. Other Distributions

Negative Binomial Distribution.

- Repeat independent trials, each with success probability p, until r-th success.
- How many trials do we need?

Observation. When $r = 1$, this is just $\mathrm{Geom}(p)$.

In general, this is sum of r independent $\mathrm{Geom}(p)$ variables.

Distribution: $\mathbb{P}(X = n) = \binom{n-1}{r-1}p^r(1-p)^{n-r}$.

Hypergeometric Distribution.

- Bucket with N balls, m of which are good.
- We draw n balls from the bucket.
- How many are good?

Distribution: $\mathbb{P}(X = k) = \frac{\text{(choice of } k \text{ good balls)(choice of } N-k \text{ bad balls)}}{\text{(choice of } N \text{ balls)}} = \frac{\binom{m}{k}\binom{N-m}{m-k}}{\binom{N}{n}}$.

Statistics. We try to find the expectation of X.

Imagine we draw the balls one at a time. Let X_i be the indicator of the i-th ball being good. Then $X = \sum_{i=1}^{n} X_i$.

(LoE)
$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{n} \mathbb{E}[X_i] \\ &= \sum_{i=1}^{n} \mathbb{P}(X_i = 1) \\ &= \sum_{i=1}^{n} \mathbb{P}(i\text{-the ball is good})\end{aligned}$$

By careful obervation, we can find that any of the N balls is equally likely to be the i-th ball.

Therefore we can view the i-th ball as uniformly distributed.

Then $\mathbb{P}(i\text{-th ball is good}) = \frac{m}{N}$. Hence $\mathbb{E}[X] = \boxed{\frac{nm}{N}}$.

In conclusion,

Distribution	Definition	Expectation	Variance
$\mathrm{Bin}(n, p)$	number of successes in n trials, each is independent with success probability p	np	$np(1-p)$
$\lim_{n\to\infty} \mathrm{Bin}(n, \frac{\lambda}{n}) = \mathrm{Poi}(\lambda)$	number of rare independent events occurring in a fixed time frame	λ	λ
$\mathrm{NB}(1, p) = \mathrm{Geom}(p)$	number of trials needed, each is independent with success probability p, until first success	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathrm{NB}(r, p)$	number of trials needed, each is independent with success probability p, until r-th success	$\frac{r}{p}$	lack content
$\mathrm{Hypergeometric}(N, m, n)$	N outcomes, m of which are good, select n without replacement, number of good outcomes	$\frac{nm}{N}$	$\frac{nm(N-m)(N-n)}{N^2(N-1)}$

CHAPTER 8

Continuous Random Variables

1. Cumulative Distribution Function

Definition 1. Let X be a random variable. We define the cumulative distribution function $F_X : \mathbb{R} \to [0, 1]$ as

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

Observation. Given F_X, we have $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.

This can be obtained from the identity $\{X \leq b\} = \{X \leq a\} \mathbin{\dot\cup} \{a < X \leq b\}$ and thus $\mathbb{P}(X \leq b) = \mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b)$.

Some other properties:

- $F_X(x)$ is increasing in x.
- $\lim_{x\to\infty} F_X(x) = 1$. This is obtained from

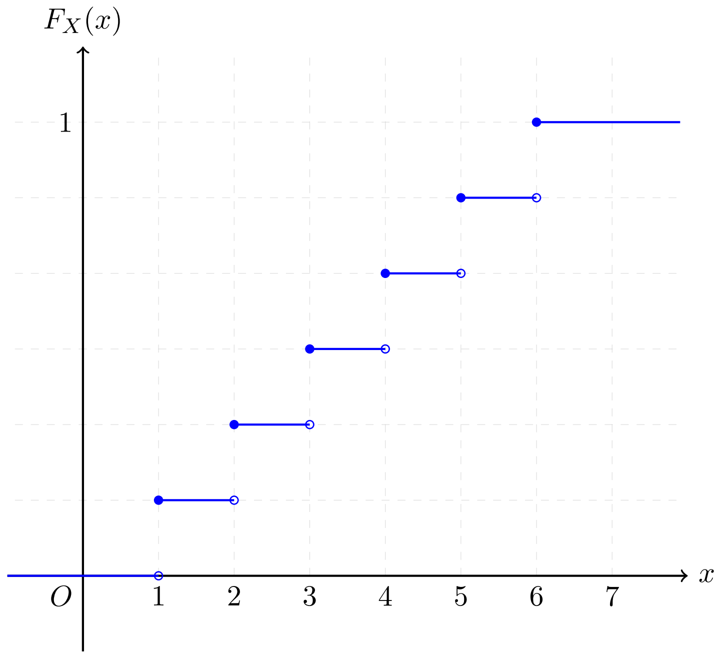
$$\lim_{x\to\infty} \mathbb{P}(\{X \leq x\}) \overset{\text{continuity}}{=} \mathbb{P}\left(\bigcup_{x\to\infty} \{X \leq x\}\right) = \mathbb{P}(X \in \mathbb{R}) = 1.$$

- $\lim_{x\to-\infty} F_X(x) = 0$.
- If $x_n \searrow x$, then $\lim_{n\to\infty} F_X(x_n) = F_X(x)$. (right continuity)
  This is obtained from $\bigcap_n \{X \leq x_n\} = \{X \leq x\}$.

Remark 1. If $x_n \nearrow x$, then $\bigcup_n \{X \leq x_n\} = \{X < x\}$, so

$$\lim_{n\to\infty} F_X(x_n) = F_X(x) - \mathbb{P}(X = x).$$

Example 1. Let X be the outcome of a roll of a die. Then the plot of its cdf F_X is shown below:


$F_X(x)$
1
x
O
1
2
3
4
5
6
7

CHAPTER 9

Continuous Random Variable

Many random situation have uncountably many possible outcomes. For example,

- How long over time will this lecture run?
- How many seconds will it take for the first student to fall asleep?

Definition 1. A random variable X is said to be (absolutely) continuous if there is a function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that the cummulative distribution function is given by

$$\mathbb{P}(X \leq x) = F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

f is called the probability density function (pdf).

Q. What does the pdf represent?

Observe that

$$\frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = f_X(x).$$

If f_X is continuous, then

$$\begin{aligned}\mathbb{P}(x - \varepsilon \leq X \leq x + \varepsilon) &= \mathbb{P}(x + \varepsilon) - \mathbb{P}(x - \varepsilon) \\ &= F_X(x + \varepsilon) - F_X(x - \varepsilon) \\ &= \int_{x-\varepsilon}^{x+\varepsilon} f_X(t)\,\mathrm{d}t\end{aligned}$$

More generally, for any event $E \subseteq \mathbb{R}$, $\mathbb{P}(E) = \int_E f_X(t)\,\mathrm{d}t$.

Since

$$\int_{x-\varepsilon}^{x+\varepsilon} f_X(t)\,\mathrm{d}t \overset{\substack{f_X \\ \text{continuous}}}{\approx} f_X(x)\,\mathrm{d}t = 2\varepsilon f_X(x),$$

therefore $f_X(x)$ approximately represents the likelihood of X being near x.

Example 1. Let

$$f(x) = \begin{cases} \frac{C}{x^3} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases}$$

for some constant C.

Q. What is C? What is $F(X)$?

Solution.

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{if } x < 1 \\ \int_1^x \frac{C}{t^3}\,\mathrm{d}t = \frac{-C}{2t^2}\Big|_1^x = \frac{C}{2} - \frac{C}{2x^2} & \text{if } x \geq 1 \end{cases}$$

Since the total probability is 1, we have

$$1 = \lim_{x \to \infty} F(x) = \lim_{x \to \infty} \frac{C}{2} - \frac{C}{2x^2} = \frac{C}{2}$$
$$\Rightarrow C = 2.$$

Therefore we have

$$F(x) = \begin{cases} 1 - \frac{1}{x^2} & \text{if } x \geq 1 \\ 0 & \text{if } x \leq 1. \end{cases}$$

1. Expectation

In the discrete setting, $\mathbb{E}[X] = \sum_i x_i \cdot \mathbb{P}(X = x_i)$.
For a continuous random variable, observe

$$\mathbb{P}(x - \varepsilon \leq X < x + \varepsilon) \approx 2\varepsilon f_X(x).$$

Therefore we define

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t f_X(t)\,\mathrm{d}t.$$

Example 2. Let X have pdf

$$f(x) = \begin{cases} 0 & \text{if } x \leq 1 \\ \frac{2}{x^3} & \text{if } x \geq 1. \end{cases}$$

Q. What is $\mathbb{E}[X]$?
Solution.

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} t f(t)\,\mathrm{d}t \\ &= \int_1^{\infty} t \frac{2}{t^3}\,\mathrm{d}t \\ &= \int_1^{\infty} \frac{2}{t^2}\,\mathrm{d}t \\ &= \left.\frac{-2}{t}\right|_1^{\infty} = \boxed{2} \end{aligned}$$

Example 3. The lecturer walks from their office to the lecture hall. The time of the walk is a random variable W with pdf f_W.

- If the lecturer arrives early, they incur a cost of c per minute.
- If the lecturer arrives late, then they incur a cost of k per minute.

Q1. If the lecturer leaves the office t before the lecture starts, what is the expected cost?
Q2. When should they leave to minimize the cost?
Solution.

The cost if the walk takes w minute is

$$g_t(w) := \begin{cases} c(t - w) & \text{if } w \leq t \\ k(w - t) & \text{if } w \geq t \end{cases}$$

The expectation cost is

$$\begin{aligned}
\mathbb{E}[g_t(w)] &= \int_{-\infty}^{\infty} g_t(w) f_W(w)\,\mathrm{d}w \\
&= \int_{0}^{\infty} g_t(w) f_W(w)\,\mathrm{d}w \\
&= \int_{0}^{t} g_t(w) f_W(w)\,\mathrm{d}w + \int_{t}^{\infty} g_t(w) f_W(w)\,\mathrm{d}w \\
&= \int_{0}^{t} c(t-w) f_W(w)\,\mathrm{d}w + \int_{t}^{\infty} k(w-t) f_W(w)\,\mathrm{d}w \\
&=: C(t)
\end{aligned}$$

To minimize the expected cost, differentiate with respect to t.

$$\begin{aligned}
\frac{\mathrm{d}C}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{0}^{t} c(t-w) f_W(w)\,\mathrm{d}w + \int_{t}^{\infty} k(w-t) f_W(w)\,\mathrm{d}w\right) \\
&= \cancel{c(t-w)f(w)|_{w=t}} + \int_{0}^{t} c f_W(w)\,\mathrm{d}w - \int_{t}^{\infty} k f_W(w)\,\mathrm{d}w - \cancel{k(w-t)f(w)|_{w=t}} \\
&= \int_{0}^{t} (c+k) f_W(w)\,\mathrm{d}w - \int_{0}^{\infty} k f_W(w)\,\mathrm{d}w \\
&= (c+k)F_W(t) - k
\end{aligned}$$

Setting the derivative equal to 0,

$$\begin{aligned}
\frac{\mathrm{d}C}{\mathrm{d}t} = 0 &\iff (c+k)F_W(t) - k \\
&\iff F_W(t) = \frac{k}{c+k}
\end{aligned}$$

Therefore the optimal t is $F_W^{-1}\left(\frac{k}{c+k}\right)$.

Observe that the linearity of expectation still works for continuous random variables (by the linearity of integral).

2. Variance

We define the variance as before

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (t-\mathbb{E}[X])^2 f(t)\,\mathrm{d}t.$$

Alternatively, $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Example 4. Let

$$f_X(x) = \begin{cases} \frac{2}{x^3} & \text{if } x \geq 1 \\ 0 & \text{if } x \leq 1 \end{cases}$$

We saw that $\mathbb{E}[X] = 2$. What is $\mathrm{Var}(X)$?

Compute

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} t^2 f(t)\,\mathrm{d}t \\ &= \int_1^{\infty} t^2 \cdot \frac{2}{x^3}\,\mathrm{d}t \\ &= \int_1^{\infty} \frac{2}{t}\,\mathrm{d}t \\ &= 2\ln t\Big|_1^{\infty} \\ &= \infty \qquad \text{(!!)}\end{aligned}$$

Example 5. Game show

Two envelopes are with $\$x$ and one with $\$y$, and $1 \le x < y$.

First choose an envelope and open it. Then decide whether to take it or take the other.

Q. What strategy can we use to maximize the chance of getting the more valuable envelope?

Solution. Attempts to give a lower bound:

50%. Choose a random envelope and keep it no matter what.

Can we do better from 50% chance?

Strategy: Choose a thresgold value $\$z$. Choose a random envelope.

If the amount is less than $\$z$, switch. Otherwise we keep it.

Envelope we choose \ Threshold	$z \le x < y$	$x < z \le y$	$x < y \le z$
$\$x$	$\$x$	$\$y$	$\$y$
$\$y$	$\$y$	$\$y$	$\$x$

$$\begin{aligned}\mathbb{P}(\text{get } \$y) &= \mathbb{P}(\text{get } \$y \mid z \le x < y)\mathbb{P}(z \le x < y) \\ &\quad + \mathbb{P}(\text{get } \$y \mid x < z \le y)\mathbb{P}(x < z \le y) \\ &\quad + \mathbb{P}(\text{get } \$y \mid x < y \le z)\mathbb{P}(x < y \le z) \\ &= \frac{1}{2}\mathbb{P}(z \le x < y) + \mathbb{P}(x < z \le y) + \frac{1}{2}\mathbb{P}(x < y \le z) \\ &= \frac{1}{2} + \frac{1}{2}\mathbb{P}(x < z \le y)\end{aligned}$$

Choose z according to a continuous random variable with $\mathbb{P}(z \in (x, y]) > 0$ for all $1 \le x < y$. For example, $f_X(x) = \begin{cases} \frac{2}{x^3} & \text{if } x \ge 1 \\ 0 & \text{if } x \le 1. \end{cases}$

CHAPTER 10

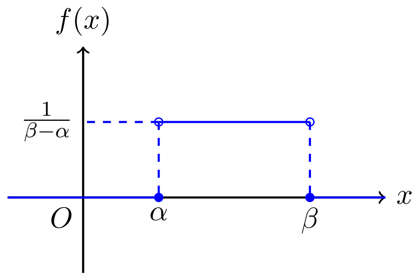
Continuous Distributions

1. Uniform Distribution

Setting: Want to choose a uniformly random number in α, β.

- run n independent trial of a random experiment
- each trial is a success with probability p
- count the number of successes

Denoted by Unif(α, β). The pdf is $f(x) = \frac{1}{\beta-\alpha}\chi_{(\alpha,\beta)}$.



Statistics. The cummulative distribution function is

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{if } x \leq \alpha \\ \int_\alpha^x \frac{1}{\beta-\alpha}\,\mathrm{d}t = \frac{x-\alpha}{\beta-\alpha} & \text{if } \alpha < x < \beta \\ 1 & \text{if } x \geq \beta \end{cases}$$

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} t f(t)\,\mathrm{d}t \\ &= \int_\alpha^\beta \frac{t}{\beta-\alpha}\,\mathrm{d}t \\ &= \frac{1}{\beta-\alpha}\,\frac{1}{2}t^2\Big|_\alpha^\beta \\ &= \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \boxed{\frac{\beta+\alpha}{2}} \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} t^2 f(t)\,\mathrm{d}t \\
&= \int_{\alpha}^{\beta} \frac{t^2}{\beta-\alpha}\,\mathrm{d}t \\
&= \frac{1}{\beta-\alpha}\,\frac{1}{3}t^3\bigg|_{\alpha}^{\beta} \\
&= \frac{\beta^3-\alpha^3}{2(\beta-\alpha)} = \frac{\beta^2+\alpha\beta+\alpha^2}{3}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2]-\mathbb{E}[X]^2 \\
&= \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \left(\frac{\beta+\alpha}{2}\right)^2 \\
&= \boxed{\frac{(\beta-\alpha)^2}{12}}
\end{aligned}
$$

Example 1. Bus arrives every 15 minutes in the hour (e.g. 7:00, 7:15, 7:30, 7:45,...). A passenger arrives at a uniformly random time between 7:00 and 7:30.

(a) What is $\mathbb{P}(\text{wait} > 5 \text{ minutes})$:
(b) What is $\mathbb{E}[\text{waiting time}]$?

Solution.

(a) Let X = arrival time and $X \sim \mathrm{Unif}(0,30)$.

$$
\begin{aligned}
\mathbb{P}(\text{wait} > 5 \text{ minutes}) &= \mathbb{P}(X \in (0,10) \cup (15,25)) \\
&= \mathbb{P}(X \in (0,10)) + \mathbb{P}(X \in (15,25)) \\
&= \int_0^{10} \frac{1}{30}\,\mathrm{d}t + \int_{15}^{25} \frac{1}{30}\,\mathrm{d}t \\
&= \frac{20}{30} = \boxed{\frac{2}{3}}
\end{aligned}
$$

(b) Let W = waiting time.

$$
W = \begin{cases} 15 - x & \text{if } 0 \le x \le 15 \\ 30 - x & \text{if } 15 < x \le 30 \end{cases}
$$

$$
\begin{aligned}
\mathbb{E}[W] &= \int_0^{30} W(t) f(t)\,\mathrm{d}t \\
&= \frac{1}{30}\int_0^{15} 15 - t\,\mathrm{d}t + \frac{1}{30}\int_{15}^{30} 15 - t\,\mathrm{d}t \\
&= \frac{2}{30}\int_0^{15} 15 - t\,\mathrm{d}t \\
&= \frac{1}{15}\left(15^2 - \frac{15^2}{2}\right) = \boxed{\frac{15}{2}}
\end{aligned}
$$

Generating Random Variables. We have a continuous random variable X. Let $F_X(x)$ be its cummulative distribution function.

Let $U \sim \text{Unif}(0,1)$ be uniform in $(0,1)$. Let $Y = F_X^{-1}(U)$.

Claim 1. Y has the same distribution as X.

Proof. $F_Y(x) = \mathbb{P}(Y \le x) = \mathbb{P}(F_X^{-1}(U) \le x) = \mathbb{P}(U \le F_X(x)) = F_U(F_X(x)) = F_X(x)$. □

2. Exponential Distribution

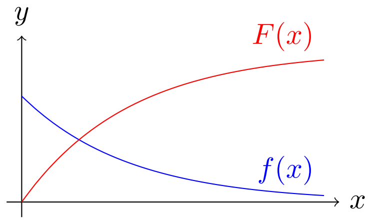
Let $X \sim \text{Exp}(\lambda)$ with $\lambda > 0$. The pdf is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Verify this is a valid distribution:

$$\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = \int_0^{\infty} \lambda e^{-\lambda t}\,\mathrm{d}t = -e^{-\lambda t}\Big|_0^{\infty} = 1.$$

The cdf is $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{if } x < 0 \\ \int_0^x \lambda e^{-\lambda x}\,\mathrm{d}t = -e^{-\lambda t}\big|_0^x = 1 - e^{-\lambda x} & \text{if } x \ge 0 \end{cases}$



Let $X \sim \text{Exp}(\lambda)$.

$$\mathbb{P}(X > s) = 1 - \mathbb{P}(X \le s) = 1 - F_X(s) = 1 - (1 - e^{-s\lambda}) = e^{-s\lambda}$$

$$\mathbb{P}(X > s + t | X > t) = \frac{\mathbb{P}(\{X > s + t\} \cap \{X > t\})}{\mathbb{P}(X > t)}$$

Since $\{X > s + t\} \subseteq \{X > t\}$,

$$\mathbb{P}(X > s + t | X > t) = \frac{\mathbb{P}(\{X > s + t\})}{\mathbb{P}(X > t)} = \frac{e^{-(s+t)\lambda}}{e^{-t\lambda}} = e^{-s\lambda} = \mathbb{P}(X > s).$$

This is an important property of the exponential distribution:

Definition 1. We say that a distribution is *memoryless* if $\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s)$.

For this reaon we often use exponentials to model the lifetime of appliances, radioactive decay, etc.

Statistics.

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} t f(t)\,\mathrm{d}t \\
&= \int_{0}^{\infty} t\lambda e^{-\lambda t}\,\mathrm{d}t \\
&= -te^{-\lambda t}\Big|_0^{\infty} - \int_0^{\infty} -e^{-\lambda t}\,\mathrm{d}t \\
&= -\frac{1}{\lambda}e^{-\lambda t}\Big|_0^{\infty} = \boxed{\tfrac{1}{\lambda}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} t^2 f(t)\,\mathrm{d}t \\
&= \int_{0}^{\infty} t\lambda e^{-\lambda t}\,\mathrm{d}t \\
&= -t^2e^{-\lambda t}\Big|_0^{\infty} - \int_0^{\infty} -2te^{-\lambda t}\,\mathrm{d}t \\
&= \frac{2}{\lambda}\int_0^{\infty} t\lambda e^{-\lambda t}\,\mathrm{d}t \\
&= \frac{2}{\lambda}\mathbb{E}[X] = \frac{2}{\lambda^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \boxed{\tfrac{1}{\lambda^2}}
\end{aligned}
$$

Example 2. The waiting time for a bus is exponentially distributed with expectation of $\frac{15}{2}$ minutes. What is the probability that we have to wait more than 5 minutes?
Solution. From $\mathbb{E}[X] = \frac{15}{2}$ we can get $\lambda = \frac{2}{15}$.

$$
\mathbb{P}(X > 5) = e^{-5\lambda} = \boxed{e^{-\frac{2}{5}}}.
$$

Example 3. Passengers can take one of the two buses.

- First bus's waiting time is exponential, with expectation $\frac{15}{2}$.
- Second bus's is exponential with expectation 15.

Q. What is the expected waiting time for a bus, if the two buses arrive independently?
Solution. Let (X, Y) = the waiting time for the (first, second) type of bus. Then $X \sim \mathrm{Exp}\left(\frac{2}{15}\right)$ and $Y \sim \mathrm{Exp}\left(\frac{1}{15}\right)$.

To calculate $\mathbb{E}[\min(X,Y)]$, we need the distribution of $\min(X,Y)$.

$$\begin{aligned}\mathbb{P}(\min(X,Y)\le t) &= 1-\mathbb{P}(\min(X,Y)>t)\\ &= 1-\mathbb{P}(\{X>t\}\cap\{Y>t\})\\ &= 1-\mathbb{P}(X>t)\mathbb{P}(Y>t)\\ &= 1-e^{-\frac{2}{15}t}e^{-\frac{1}{15}t} = 1-e^{-\frac{3}{15}t}\end{aligned}$$

Magically, $\min(X,Y)\sim \mathrm{Exp}\left(\frac{3}{15}\right)$. (!!!)
Therefore $\mathbb{E}[\min(X,Y)] = \frac{15}{3} = \boxed{5}$.
Generally, If $X_i\sim\mathrm{Exp}(\lambda_i)$, then $\min(X_i)_{1\le i\le n}\sim\mathrm{Exp}\left(\sum_{i=1}^n\lambda_i\right)$.

Theorem 1. The exponential distribution is the only memoryless distribution.

Proof. Let X be memoryless, and $g(X)=\mathbb{P}(X>x)$. Then

$$\begin{aligned}g(x+y) &= \mathbb{P}(X>x+y)\\ &= \mathbb{P}(X>x+y|X>y)\mathbb{P}(X>y)\\ &\overset{\text{memoryless}}{=} \mathbb{P}(X>x)\mathbb{P}(X>y) = g(x)g(y)\end{aligned}$$

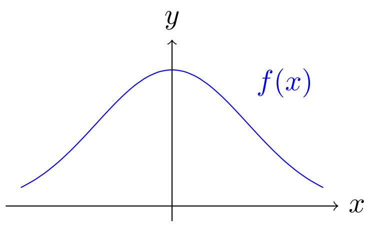
Let $g(1)=e^{-\lambda}$ for some $\lambda\ge 0$. Then we have $g(x)=e^{-\lambda x}$ for all $x\in\mathbb{Q}$.
Since g is right continuous, $g(x)=e^{-\lambda x}$.
Then we can obtain the pdf of X and see that $X\sim\mathrm{Exp}(\lambda)$. ☐

3. Normal Distribution

Also named Gaussian distribution.

The normal distribution with mean μ and standard deviation σ, denoted $N(\mu,\sigma^2)$, is the continuous random variable with probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in\mathbb{R}.$$



Verify this is a valid distribution: We need to show that $1=\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t$.
Let $y=\frac{t-\mu}{\sigma}$ and $\frac{\mathrm{d}y}{\mathrm{d}t}=\frac{1}{\sigma}$. Then the integral becomes

$$I=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-y^2}\,\mathrm{d}y$$

Square the integral to get $I^2=\frac{1}{2\pi}\iint_{y,z\in\mathbb{R}^2}e^{-(y^2+z^2)}\,\mathrm{d}y\,\mathrm{d}z$.
Then switch to polar coordinates to get $I^2=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}e^{-r^2}\cdot r\,\mathrm{d}r\,\mathrm{d}\theta=\int_0^{\infty}re^{-r^2}\,\mathrm{d}r=1$.

We call $N(0,1)$ the standard normal distribution.
Observation. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
We can obtain this by calculating the cdf:

$$\begin{aligned}
\mathbb{P}(Z \le z) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \le z\right) \\
&= \mathbb{P}(X \le \mu + z\sigma) \\
&= \int_{-\infty}^{\mu+z\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t \\
&= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{-y^2}{2}}\,\mathrm{d}y = \mathbb{P}(N(0,1) \le z)
\end{aligned}$$

So $Z \sim N(0,1)$.
We denote the cdf of the standard normal distribution as

$$\Phi(z) := \mathbb{P}(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Observe that $\Phi(z) = 1 - \Phi(-z)$ for $z > 0$.
Also if $X \sim N(\mu, \sigma^2)$, the cdf of X is $F_X(x) = \Phi\Bigg(\underbrace{\frac{x-\mu}{\sigma}}_{\text{"}z\text{-score"}}\Bigg)$.

Example 4. On a midterm exam, the grades are normally distributed with $\mu = 60$, $\sigma = 10$.
Students get a B if they score between 75 and 85. What proportion of students get a B?
Solution. Let $X \sim N(60, 10)$, so $\frac{X-60}{10} \sim N(0,1)$.

$$\begin{aligned}
\mathbb{P}(75 \le X \le 85) &= \mathbb{P}\left(1.5 \le \frac{X-60}{10} \le 2.5\right) \\
&= \mathbb{P}\left(\frac{X-60}{10} \le 2.5\right) - \mathbb{P}\left(\frac{X-60}{10} \le 1.5\right) \\
&= \Phi(2.5) - \Phi(1.5) \\
&\approx 0.9938 - 0.9332 = \boxed{0.0606}
\end{aligned}$$

Statistics. Let $Z \sim N(0,1)$. Compute the expectation

$$\mathbb{E}[Z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} te^{-\frac{t^2}{2}}\,\mathrm{d}t = -\frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\Big|_{-\infty}^{\infty} = 0.$$

If $X \sim N(\mu, \sigma^2)$, then $X = \sigma Z + \mu$, so $\mathbb{E}[X] = \sigma\mathbb{E}[Z] + \mu = \boxed{\mu}$.
Now compute the variance.

$$\begin{aligned}
\mathbb{E}[Z^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}}\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2\pi}} \left(t(-e^{-\frac{t^2}{2}})\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 1 \cdot -e^{-\frac{t^2}{2}}\,\mathrm{d}t \right) \\
&= \int_{-\infty}^{\infty} 1 \cdot -e^{-\frac{t^2}{2}}\,\mathrm{d}t = 1
\end{aligned}$$

Then $\mathrm{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 1 - 0 = 1$.

If $X \sim N(\mu, \sigma^2)$, then $\mathrm{Var}(X) = \mathrm{Var}(\sigma Z + \mu) = \sigma^2 \mathrm{Var}(Z) = \boxed{\sigma^2}$.

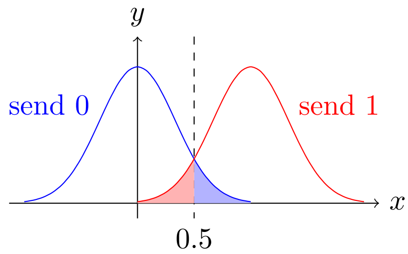
Example 5. Transmitting a binary string.

Send n bits, each bit is 0 or 1, across a cable.

Assume the cable introduces some noise $\sim N(0, 0.3^2)$.

That is, If we send 1, receiver gets $1 + X$ where $X \sim N(0, 0.3^2)$.

If we send 0, receiver gets $0 + X$ where $X \sim N(0, 0.3^2)$.



Decoding algorithm: If $x < 0.5$, decode as 0. If $x > 1.5$, decode as 1.

Q. What is the probability of misreading a bit (shaded area)?

Solution. Let X be the noise.

$$\begin{aligned}
\mathbb{P}(\text{error}) &= \mathbb{P}(X < -0.5) \\
&= \mathbb{P}\left(\frac{X-0}{0.3} \leq \frac{-0.5-0}{0.3}\right) \\
&= \Phi\left(-\frac{0.5}{0.3}\right) \approx \Phi(-1.67) \\
&= \mathbb{P}(Z \leq -1.67) = \mathbb{P}(Z \geq 1.67) \\
&= 1 - \mathbb{P}(Z \leq 1.67) = 1 - \Phi(1.67) \\
&= 1 - 0.9525 = \boxed{0.0475}
\end{aligned}$$

Normal Approximation to the Binomial. Recall that $\mathrm{Bin}(n, p)$ has mean np and variance $np(1-p)$.

If p is constant in $(0, 1)$ then as $n \to \infty$, $np, np(1-p) \to \infty$. Therefore we cannot use Poisson approximation.

But we do get an approximately normal distribution!

Theorem 2 (de Moivre-Laplace). Fix $p \in (0, 1)$. Then for any $a < b$, we have

$$\mathbb{P}\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) \to \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$

where $X \sim \mathrm{Bin}(n, p)$, as $n \to \infty$.

Proof. Later at 1. $\square$

Remark 1. (a) In practice, $np(1-p) \geq 10$ is a enough for a good approximation.

(b) Continuity correction: $\mathbb{P}(X = k) = \mathbb{P}\left(k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\right) \approx \Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$.

Example 6. Toss a fair coin 40 times.
Q. What is the probability of getting 20 heads?
Solution. Let $X =$ number of heads $\sim \text{Bin}(40, 0.5)$.

$$\mathbb{P}(X = 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{40} = \boxed{0.1254}$$

On the other hand, the normal approximation gives

$$\begin{aligned}
\mathbb{P}(X = 20) &= \mathbb{P}(19.5 \leq X \leq 20.5) \\
&= \mathbb{P}\left(\frac{19.5 - 20}{10} \leq \frac{X - 20}{10} \leq \frac{20.5 - 20}{10}\right) \\
&\approx \Phi\left(\frac{0.5}{\sqrt{10}}\right) - \Phi\left(\frac{-0.5}{\sqrt{10}}\right) \\
&= 2\Phi\left(\frac{0.5}{\sqrt{10}}\right) - 1 \\
&\approx 2 \cdot \Phi(0.16) - 1 \approx 0.1272
\end{aligned}$$

Example 7. A country has a population of N. The government wants to estimate the level of support for a new initiative.

Suppose p proportion of the population is for, $1 - p$ is against.

The government do a poll: Choose n people at random, ask if they are in favor.

Q. How large should n be to estimate p to within 1% with probability 95%?
Solution. Let $X =$ number of people in favor $\sim \text{Bin}(n, p)$.

$$\begin{aligned}
&\mathbb{P}(\text{good approximation}) \\
&= \mathbb{P}(pn - 0.01n \leq X \leq pn + 0.01n) \\
&= \mathbb{P}\left(\frac{-0.01n}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{0.01n}{\sqrt{np(1-p)}}\right) \\
\text{(since } p(1-p) \leq \tfrac{1}{4}\text{)} \quad &= \mathbb{P}\left(-0.02\sqrt{n} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq 0.02\sqrt{n}\right) \\
&\approx \Phi(0.02\sqrt{n}) - \Phi(-0.02\sqrt{n}) \\
&= 2\Phi(0.02\sqrt{n}) - 1
\end{aligned}$$

Therefore $\mathbb{P}(\text{good approximaiton}) \geq 0.95$ if $\Phi(0.02\sqrt{n}) \geq 0.975$.

This can be satisfied if $0.02\sqrt{n} \geq 1.96$ and thus $\boxed{n \geq 9604}$.

CHAPTER 11

Function of Random Variables

Let X be a continuous random variable with pdf $f_X(x)$.

Let $Y = g(X)$ be a function $g : \mathbb{R} \to \mathbb{R}$.

Q. What can we say about the distribution of Y?

For example, let $Y = g(X) = X^2$. Then

$$\begin{aligned}\mathbb{P}(Y \leq y) &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(t)\,\mathrm{d}t\end{aligned}$$

Theorem 1. Let $Y = g(X)$ where g is monotone increasing. Then if X has pdf f_X, the pdf of Y is

$$f_Y(y) = \begin{cases} 0 & \text{if } y \notin g(\mathbb{R}) \\ f_X(g^{-1}(y)) \cdot \frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)) & \text{otherwise .} \end{cases}$$

where the sum is over all x such that $g(x) = y$.

Proof.

$$\begin{aligned}F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(X) \leq y) \\ &= \begin{cases} 0 \text{ or } 1 & \text{if } y \notin g(\mathbb{R}) \\ \mathbb{P}(X \leq g^{-1}(y)) & \text{otherwise} \end{cases}\end{aligned}$$

Taking the derivative with respect to y;

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \begin{cases} 0 & \text{if } y \notin g(\mathbb{R}) \\ f_X(g^{-1}(y)) \cdot \frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)) & \text{otherwise .} \end{cases}$$

Note that $\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)) = \frac{1}{g'(g^{-1}(y))}$. $\square$

1. Measurable Sets

Q. Can we define a uniform probability distribution on $(0,1)$? i.e. $\mathbb{P} : 2^{(0,1)} \to [0,1]$.

Desirable properties:

(1) $\mathbb{P}((0,1)) = 1$.

(2) If $E_1, E_2, E_3, \ldots \in 2^{(0,1)}$ are pairwise disjoint, then $\mathbb{P}\left(\bigcup_n E_n\right) = \sum_n \mathbb{P}(E_n)$.
(3) Also, consider representing numbers in $(0,1)$ in binary.

$$x = \frac{x_1}{2^1} + \frac{x_2}{2^2} + \frac{x_3}{2^3} + \ldots$$

where $x_i \in \{0,1\}$.
Let $A_i(x) = (x_1, x_2, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots)$, where $x = (x_1, x_2, \ldots)$.
For any $E \subseteq (0,1)$ and any $i \in \mathbb{N}$, We want $\mathbb{P}(A_i(E)) = \mathbb{P}(E)$.

Theorem 2 (Vitali, 1905). Such a $\mathbb{P}$ does not exist.

Proof. Define an equivalence relation on $\{0,1\}^{\mathbb{N}}$: $\vec{x} \sim \vec{y}$ if $|\{i : x_i \neq y_i\}| < \infty$.
Let $[\vec{x}]$ denote the equivalence class of $\vec{x}$.

Claim 1. $[\vec{x}]$ is countable for all $\vec{x}$.

This can be obtained by

$$[\vec{x}] = \bigcup_{\substack{I \subseteq \mathbb{N} \\ |I| < \infty}} A_I(\vec{x}).$$

where $A_I(\vec{x}) = A_{i_1}(A_{i_2}(\cdots A_{i_r}(\vec{x})))$ and $I = \{i_1, i_2, \ldots, i_r\}$.
Such I is countable since the choices are $\binom{\mathbb{N}}{0} \cup \binom{\mathbb{N}}{1} \cup \binom{\mathbb{N}}{2} \cup \cdots$, countable terms.
For each equivalence class, choose a representative, and let E be the set of representatives.
This naturally gives that $(0,1) = \dot{\bigcup}_{|I|<\infty} A_I(E)$.
Therefore $1 \overset{(1)}{=} \mathbb{P}((0,1)) = \mathbb{P}\left(\bigcup_{|I|<\infty} A_I(E)\right) \overset{(2)}{=} \sum_{|I|<\infty} \mathbb{P}(A_I(E)) \overset{(3)}{=} \sum_{|I|<\infty} \mathbb{P}(E) = \infty$. $\square$

[Insert the note on 4/17, 4/22 here.]

CHAPTER 12

Independent Random Variables

Definition 1. We say two random variables X and Y are independent if $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ for all A, B.

Remark 1. This is equivalent to

$$F(a,b) = F_X(a)F_Y(b) \iff \begin{array}{ll} p(x,y) = p_X(x)p_Y(y) & \text{(discrete)} \\ f(x,y) = f_X(x)f_Y(y) & \text{(continuous).} \end{array}$$

Example 1. Let (X,Y) be uniformly distributed in the unit square $[0,1]^2$.
Q. Are X and Y independent?
Solution. We have

$$f(x,y) = 1_{[0,1]^2}(x,y) = 1_{[0,1]}(x) \cdot 1_{[0,1]}(y)$$

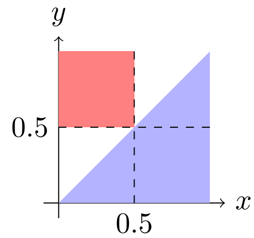
Thus they are independent.

Example 2. Let (X,Y) be uniformly distributed in the triangle with vertices $(0,0)$, $(1,0)$, $(1,1)$.
Q. Are X and Y independent?
Solution. We have

$$\mathbb{P}(X \in (0, \tfrac{1}{2}), Y \in (\tfrac{1}{2}, 1)) = 0$$

but

$$\mathbb{P}(X \in (0, \tfrac{1}{2}))\mathbb{P}(Y \in (\tfrac{1}{2}, 1)) = \frac{1}{4} \cdot \frac{1}{4} \neq 0$$



Thus they are not independent.

1. Sums of Independent Random Variables

If X and Y are independent, with densities f_X and f_Y respectively, what is the distribution of their sum $X + Y$?

$$\begin{aligned}
F_{X+Y}(a) &= \mathbb{P}(X+Y \le a) \\
&= \iint_{x+y\le a} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
\text{(independence)} \qquad &= \iint_{x+y\le a} f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{a-y} f_X(x)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f_Y(y)F_X(a-y)\,\mathrm{d}y
\end{aligned}$$

Therefore

$$\begin{aligned}
f_{X+Y}(a) &= \frac{\mathrm{d}}{\mathrm{d}a}F_{X+Y}(a) \\
&= \frac{\mathrm{d}}{\mathrm{d}a}\int_{-\infty}^{\infty} f_Y(y)F_X(a-y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f_Y(y)\left(\frac{\mathrm{d}}{\mathrm{d}a}F_X(a-y)\right)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f_X(x)f_Y(a-x)\,\mathrm{d}x
\end{aligned}$$

which is the convolution of f_X and f_Y. This gives us the following result:

Proposition 1. If X and Y are independent, then

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_Y(y)f_X(a-y)\,\mathrm{d}y = \int_{-\infty}^{\infty} f_X(x)f_Y(a-x)\,\mathrm{d}x.$$

2. Sums of Uniform Random Variables

Let X, Y be i.i.d (independent identically distributed) Unif$(0,1)$ random variables.
Q. What is the distribution of $X+Y$?
Solution. We have

$$f_X(x) = f_Y(x) = \begin{cases} 1 & \text{if } x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

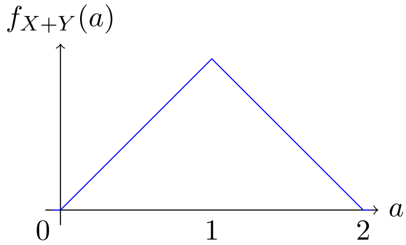
$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_Y(y)f_X(a-y)\,\mathrm{d}y = \int_0^1 f_X(a-y)\,\mathrm{d}x.$$

Since $0 \le X, Y \le 1$, we have $f_{X+Y}(a) = 0$ if $a < 0$ or $a > 2$.
If $0 \le a \le 1$, then

$$f_{X+Y}(a) = \int_0^1 f_X(a-y)\,\mathrm{d}y = \int_0^a 1\,\mathrm{d}y = a.$$

If $1 \le a \le 2$, then

$$f_{X+Y}(a) = \int_0^1 f_X(a-y)\,\mathrm{d}y = \int_{a-1}^1 1\,\mathrm{d}y = 1-(a-1) = 2-a.$$



This is called "the triangular distribution".
Q. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d Unif$(0,1)$. What can we say about $\sum_{i=1}^n X_i$?

Proposition 2. For $n = 1, 2, \ldots$,

$$f_{\sum_{i=1}^n X_i}(x) = \frac{x^{n-1}}{(n-1)!} \quad \text{if } 0 \le x \le 1.$$

Proof. We have

$$\begin{aligned}
f_{\sum_{i=1}^{n+1} X_i}(x) &= f_{X_{n+1}+\sum_{i=1}^n X_i}(x) \\
&= \int_{-\infty}^{\infty} f_{X_{n+1}}(y) f_{\sum_{i=1}^n X_i}(x-y)\,\mathrm{d}y \\
&= \int_0^1 f_{\sum_{i=1}^n X_i}(x-y)\,\mathrm{d}y \\
\text{(induction hypothesis)} \qquad &= \int_0^x \frac{(x-y)^{n-1}}{(n-1)!}\,\mathrm{d}y \\
&= \left.\frac{-(x-y)^n}{n!}\right|_0^x = \frac{x^n}{n!}
\end{aligned}$$

By induction, we can get the result. □

Example 3. A person has 10 cakes in their fridge. Each day their hunger is independent, urging them to eat x cakes, where $x \sim$ Unif$(0,1)$.
Q. What is the expected number of days until the first cake is finished?
Solution. Let N = number of days to finish the first cake.
Let S_n = the amount of cake eaten after n days.
Then $S_n = \sum_{i=1}^n X_i$ where $X_i \sim$ Unif$(0,1)$ i.i.d.

By definition, $N = \min\{n : S_n \geq 1\}$.

$$\begin{aligned}
\mathbb{P}(N > n) &= \mathbb{P}(S_n < 1) \\
&= \mathbb{P}(S_n \leq 1) \\
&= \int_0^1 f_{S_n}(x)\,\mathrm{d}x \\
&= \int_0^1 \frac{x^{n-1}}{(n-1)!}\,\mathrm{d}x \\
&= \left.\frac{x^n}{n!}\right|_0^1 = \frac{1}{n!}
\end{aligned}$$

Therefore

$$\mathbb{P}(N = n) = \mathbb{P}(N > n-1) - \mathbb{P}(N > n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}.$$

Thus

$$\mathbb{E}[N] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(N = k) = \sum_{k=1}^{\infty} \frac{k(k-1)}{k!} = \sum_{k=2}^{\infty} \frac{1}{(k-2)!} = \sum_{n=0}^{\infty} \frac{1}{n!} = \boxed{e}.$$

2.1. Gamma Distribution. Recall that

$$\Gamma(\alpha, \lambda) : f_Y(y) = \begin{cases} \frac{\lambda e^{-\lambda y}(\lambda y)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 3. If X, Y are independent, $X \sim \Gamma(\alpha, \lambda)$, $Y \sim \Gamma(\beta, \lambda)$, then $X + Y \sim \Gamma(\alpha + \beta, \lambda)$.

Corollary 1. The sum of n i.i.d $\mathrm{Exp}(\lambda)$ random variables is $\Gamma(n, \lambda)$.

Proof.

$$\begin{aligned}
f_{X+Y}(a) &= \int_{-\infty}^{\infty} f_Y(y) f_X(a-y)\,\mathrm{d}y \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^a \left(\lambda e^{-\lambda y}(\lambda y)^{\beta-1}\right)\left(\lambda e^{-\lambda(a-y)}(\lambda(a-y))^{\alpha-1}\right)\,\mathrm{d}y \\
&= \frac{\lambda^{\alpha+\beta} e^{-\lambda a}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^a y^{\beta-1}(a-y)^{\alpha-1}\,\mathrm{d}y \qquad (*)
\end{aligned}$$

Let $x = \frac{y}{a}$. Then

$$\begin{aligned}
(*) &= \frac{\lambda^{\alpha+\beta} e^{-\lambda a}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (xa)^{\beta-1} a^{\alpha-1} (1-x)^{\alpha-1} \cdot a\,\mathrm{d}x \\
&= \frac{\lambda^{\alpha+\beta} e^{-\lambda a}}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\left(\int_0^1 x^{\beta-1}(1-x)^{\alpha-1}\,\mathrm{d}x\right)}_{C'_{\alpha,\beta}} a^{\alpha+\beta-1}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow f_{X+Y}(a) &= C'_{\alpha,\beta} \cdot e^{-\lambda a} a^{\alpha+\beta-1} \\
&= C_{\alpha,\beta} \cdot e^{-\lambda a} (\lambda a)^{\alpha+\beta-1}
\end{aligned}$$

where $C_{\alpha,\beta} = \frac{\lambda}{\Gamma(\alpha+\beta)}$ since this pdf has the form of a Gamma distribution with rate λ and order $\alpha+\beta$.

Therefore $X + Y \sim \Gamma(\alpha+\beta, \lambda)$. $\square$

Application. Let $Z_i \sim N(0,1)$ i.i.d.
Q. What is the distribution of $\sum_{i=1}^n Z_i^2$?
Solution. The distribution of Z_i^2 is given by

$$\begin{aligned} F_{Z_i^2}(a) &= \mathbb{P}(Z_i^2 \leq a) \\ &= \mathbb{P}(-\sqrt{a} \leq Z_i \leq \sqrt{a}) \\ &= \int_{-\sqrt{a}}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x \end{aligned}$$

$$\begin{aligned} \Rightarrow f_{Z_i^2}(a) &= \frac{\mathrm{d}}{\mathrm{d}a} F_{Z_i^2}(a) \\ &= \frac{\mathrm{d}}{\mathrm{d}a}\left(\int_{-\sqrt{a}}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x\right) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\bigg|_{x=\sqrt{a}} \cdot \frac{\mathrm{d}}{\mathrm{d}a}(\sqrt{a}) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\bigg|_{x=-\sqrt{a}} \cdot \frac{\mathrm{d}}{\mathrm{d}a}(-\sqrt{a}) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{a}{2}} \cdot \frac{1}{2\sqrt{a}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{a}{2}} \cdot \frac{-1}{2\sqrt{a}} \\ &= \frac{1}{\sqrt{2\pi a}} e^{-\frac{a}{2}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{a}{2}} \left(\frac{a}{2}\right)^{\frac{1}{2}-1} \end{aligned}$$

Therefore $Z_i^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ (and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$).

Hence $\sum_{i=1}^n Z_i^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ has density function $\frac{\frac{1}{2} \cdot e^{\frac{1}{2}a} \cdot \left(\frac{1}{2}a\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)}$ where

$$\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right) = \begin{cases} \left(\frac{n}{2}-1\right)! & \text{if } n \text{ is even} \\ \sqrt{\pi}\left(\frac{n}{2}-\frac{1}{2}\right)! & \text{if } n \text{ is odd} \end{cases}$$

Remark 2. The distribution of $\sum_{i=1}^n Z_i^2$ is called the chi-squared distribution with n degrees of freedom, denoted by χ_n^2.

2.2. Normal Distribution.

Proposition 4. Let $X_1, X_2, \ldots, X_n$ be independent $N(\mu_i, \sigma_i^2)$ random variables.

Then

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i\right).$$

Proof. We complete the proof in the following steps:

(1) Show that $n = 2$ case implies the general case.
(2) $n = 2$: Show that $\mu_1 = \mu_2 = 0$ and $\sigma_1 = 1$, $\sigma_2 = \sigma$ implies the general case.
(3) $n = 2$: $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = \sigma$.

(1) Proof by induction on n.
Base case. The case $n = 2$ is $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
Induction step. We have

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n$$

where $\sum_{i=1}^{n-1} X_i \sim N\left(\sum_{i=1}^{n-1} \mu_i, \sum_{i=1}^{n-1} \sigma_i^2\right)$ and $X_n \sim N(\mu_n, \sigma_n^2)$.
By induction hypothesis we have the desired result.

(2) Let $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$.
Then

$$\begin{aligned} X_1 + X_2 &= (X_1 - \mu_1) + (X_2 - \mu_2) + (\mu_1 + \mu_2) \\ &= \sigma_1 \left(\underbrace{\frac{X_1 - \mu_1}{\sigma_1}}_{\sim N(0,1)} + \underbrace{\frac{X_2 - \mu_2}{\sigma_1}}_{\sim N\left(0, \frac{\sigma_2^2}{\sigma_1^2}\right)} \right) + (\mu_1 + \mu_2) \end{aligned}$$

By (3), we have

$$\begin{aligned} &\frac{X_1 - \mu_1}{\sigma_1} + \frac{X_2 - \mu_2}{\sigma_1} \sim N\left(0, 1 + \frac{\sigma_2^2}{\sigma_1^2}\right) \\ \Rightarrow\ &\sigma_1 \left(\frac{X_1 - \mu_1}{\sigma_1} + \frac{X_2 - \mu_2}{\sigma_1}\right) \sim N(0, \sigma_1^2 + \sigma_2^2) \\ \Rightarrow\ &X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \end{aligned}$$

(3) Let $X_1 \sim N(0, 1)$, $X_2 \sim N(0, \sigma^2)$ independently. Then

$$\begin{aligned} f_{X+Y}(a) &= \int_{-\infty}^{\infty} f_X(a - y) f_Y(y)\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(a-y)^2}{2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-y^2}{2\sigma^2}}\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{\frac{-a^2}{2}} \cdot e^{-\frac{1}{2}\left(y^2 + \frac{y^2}{\sigma^2} - 2ay\right)}\, \mathrm{d}y \end{aligned}$$

Compute

$$\begin{aligned} y^2 + \frac{y^2}{\sigma^2} - 2ay &= y^2 \cdot \frac{\sigma^2 + 1}{\sigma^2} - 2ay \\ &= \frac{\sigma^2 + 1}{\sigma^2} \left(\left(y - \frac{\sigma^2 a}{\sigma^2 + 1}\right)^2 - \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2 a^2 \right) \\ &= \frac{\sigma^2 + 1}{\sigma^2} \left(y - \frac{\sigma^2 a}{\sigma^2 + 1}\right)^2 - \frac{\sigma^2}{\sigma^2 + 1} a^2 \end{aligned}$$

Therefore

$$\begin{aligned}
f_{X+Y}(a) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{\frac{-a^2}{2}} \cdot e^{-\frac{1}{2}\left(\frac{\sigma^2+1}{\sigma^2}\left(y-\frac{\sigma^2 a}{\sigma^2+1}\right)^2 - \frac{\sigma^2}{\sigma^2+1}a^2\right)} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{\frac{-a^2}{2}+\frac{\sigma^2}{\sigma^2+1}\cdot\frac{a^2}{2}} \cdot e^{-\frac{1}{2}\cdot\frac{\sigma^2+1}{\sigma^2}\left(y-\frac{\sigma^2 a}{\sigma^2+1}\right)^2} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} e^{-\frac{1}{2}\cdot\frac{1}{\sigma^2+1}\cdot a^2} \cdot e^{-\frac{1}{2}\cdot\frac{\sigma^2+1}{\sigma^2}\left(y-\frac{\sigma^2 a}{\sigma^2+1}\right)^2} \mathrm{d}y \\
&= \frac{1}{2\pi\sigma} e^{-\frac{a^2}{2(\sigma^2+1)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot\frac{\sigma^2+1}{\sigma^2}\left(y-\frac{\sigma^2 a}{\sigma^2+1}\right)^2} \mathrm{d}y
\end{aligned}$$

Let $z = y - \frac{\sigma^2 a}{\sigma^2+1}$ and $\mathrm{d}y = \mathrm{d}z$. So

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot\frac{\sigma^2+1}{\sigma^2}\left(y-\frac{\sigma^2 a}{\sigma^2+1}\right)^2} \mathrm{d}y = \int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot\frac{\sigma^2+1}{\sigma^2}z^2} \mathrm{d}z := C$$

Then $f_{X+Y}(a) = C' e^{-\frac{a^2}{2(\sigma^2+1)}}$ for some constant C'.
Hence $X+Y$ is a normal random variable with mean 0 and variance $1+\sigma^2$. $\square$

Application. There is a stock with price $S(n)$ after n weeks.
We introduce the definition:

Definition 2. X is log-normally distributed with parameters (μ, σ) if $\log(X) \sim N(\mu, \sigma^2)$.

The random variables $\frac{S(n)}{S(n-1)}$ are independent, log-normed with $(\mu, \sigma) = (0.0165, 0.0730)$.

(a) What is the probability that $S(n)$ increases every week for a year?
(b) What is the probability that $S(n)$ increases for at least 30 weeks in a year?
(c) What is the probability that the value of the stock increase by the end of the year?

Solution.

(a) We have

$$\mathbb{P}(S(n) \geq S(n-1)) = \mathbb{P}\left(\frac{S(n)}{S(n-1)} \geq 1\right) = \mathbb{P}\left(\log\left(\frac{S(n)}{S(n-1)}\right) > 0\right)$$

where $\log\left(\frac{S(n)}{S(n-1)}\right) \sim N(0.0165, 0.0730^2)$.
Therefore the probability is equal to $\mathbb{P}\left(Z \geq \frac{0-0.0165}{0.0730}\right) = \Phi\left(\frac{0.0165}{0.0730}\right) \approx 0.5894$.
Then we can compute

$$\begin{aligned}
&\mathbb{P}(S(n) \geq S(n-1) \text{ for } n = 1, 2, \ldots, 52) \\
&= \prod_{n=1}^{52} \mathbb{P}(S(n) \geq S(n-1)) \\
&\approx 0.5894^{52} \approx \boxed{1.15 \times 10^{-12}}
\end{aligned}$$

(b) The goal is $\mathbb{P}$(stock price goes up for ≥ 30 weeks in the year).
Let X = the number of weeks the stock price goes up in the year. Then $X \sim \text{Bin}(52, 0.5894)$. We have

$$\mathbb{P}(X \geq 30) = \sum_{k=30}^{52} \binom{52}{k} 0.5894^k \times 0.4106^{52-k}.$$

We can use a normal approximation

$$X \simeq Y \sim N(52 \times 0.5894, 52 \times 0.5894 \times 0.4106).$$

Then we can compute

$$\mathbb{P}(X \geq 30) \approx \mathbb{P}(Y \geq 29.5) = \mathbb{P}\left(Z \geq \frac{29.5 - \mu}{\sigma}\right) \approx \boxed{0.627}.$$

(c)

$$\begin{aligned}
\mathbb{P}(S(52) \geq S(0)) &= \mathbb{P}\left(\frac{S(52)}{S(0)} \geq 1\right) \\
&= \mathbb{P}\left(\frac{S(52)}{S(51)} \times \frac{S(51)}{S(50)} \times \cdots \times \frac{S(1)}{S(0)} \geq 1\right) \\
&= \mathbb{P}\left(\log\left(\frac{S(52)}{S(51)} \times \frac{S(51)}{S(50)} \times \cdots \times \frac{S(1)}{S(0)}\right) \geq 0\right) \\
&= \mathbb{P}\left(\sum_{n=1}^{52} \log\left(\frac{S(n)}{S(n-1)}\right) \geq 0\right) \approx \boxed{0.9484}
\end{aligned}$$

since $\log\left(\frac{S(n)}{S(n-1)}\right) \sim N(0.0165, 0.0730^2)$ and by independence, $\sum_{n=1}^{52} \log\left(\frac{S(n)}{S(n-1)}\right) \sim N(52 \times 0.0165, 52^2 \times 0.0730^2)$.

3. Sums of Discrete Random Variables

Let X and Y be independent discrete random variables.

To calculate $\mathbb{P}(X + Y = a)$, we partition the sample space based on value of X, and use total probability.

$$\begin{aligned}
&\mathbb{P}(X + Y = a) \\
&= \sum_{x \in \text{Range}(X)} \mathbb{P}(X + Y = a | X = x)\mathbb{P}(X = x) \\
&= \sum_x \mathbb{P}(Y = a - x | X = x)\mathbb{P}(X = x) \\
\text{(independence)} \quad &= \sum_x \mathbb{P}(Y = a - x)\mathbb{P}(X = x) \\
&= \sum_x p_Y(a - x) p_X(x)
\end{aligned}$$

Example 4. Let $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ independently. What is $\mathbb{P}(X+Y=k)$?
Solution. Compute

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_x \mathbb{P}(Y=k-x)\mathbb{P}(X=x) \\
&= \sum_{x=0}^{k} \mathbb{P}(Y=k-x)\mathbb{P}(X=x) \\
&= \sum_{x=0}^{k} \frac{e^{-\lambda_2}\lambda_2^{k-x}}{(k-x)!} \cdot \frac{e^{-\lambda_1}\lambda_1^x}{x!} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{x=0}^{k} \frac{\lambda_1^x \lambda_2^{k-x}}{x!(k-x)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{x=0}^{k} \frac{k!}{x!(k-x)!} \lambda_1^x \lambda_2^{k-x} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{x=0}^{k} \binom{k}{x} \lambda_1^x \lambda_2^{k-x} \\
\text{(binomial thm)} \qquad &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} (\lambda_1+\lambda_2)^k
\end{aligned}$$

Therefore $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$.

CHAPTER 13

Conditional Distributions

Setting: We have two (not necessarily independent) random variables. If we know the value of one, what can we say about the other?

1. Discrete Conditional Distribution

First we introduce the notation:

Definition 1. Given discrete random variables X and Y, the conditional probability mass function of X given $Y = y$ is

$$p_{X|Y}(x|y) := \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p(x, y)}{p_Y(y)}.$$

Observation. If X and Y are independent, then $p_{X|Y}(x|y) = p_X(x)$.

This is followed by $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} \overset{\text{indep.}}{=} \frac{p_X(x)p_Y(y)}{p_Y(y)}$.

Example 1. Let $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ independently.

What is $\mathbb{P}(X = k|X + Y = n)$?

Solution. We have

$$\mathbb{P}(X = k|X + Y = n) = \frac{\mathbb{P}(X = k, X + Y = n)}{\mathbb{P}(X + Y) = n} = \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y) = n}.$$

Since X and Y are independent, we have the following facts:

(1) $\mathbb{P}(X = k, Y = n - k) = \mathbb{P}(X = k)\mathbb{P}(Y = n - k) = \frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}$.

(2) $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$, so $\mathbb{P}(X + Y = n) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}$

Therefore

$$\begin{aligned}
&\mathbb{P}(X = k|X + Y = n) \\
&= \frac{\frac{\cancel{e^{-\lambda_1}}\lambda_1^k}{k!} \cdot \frac{\cancel{e^{-\lambda_2}}\lambda_2^{n-k}}{(n-k)!}}{\frac{\cancel{e^{-(\lambda_1+\lambda_2)}}(\lambda_1+\lambda_2)^n}{n!}} \\
&= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}
\end{aligned}$$

Hence $X|(X + Y = n) \sim \text{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

Example 2. Let $X \sim \text{Bin}(n, p)$ and $\vec{o} =$ order of outcomes $\in \{S, F\}^n$ consists of sucesses and failures.

Claim 1. For any vector $\vec{v}$ with exactly k S's, we have $p_{\vec{o}|X}(\vec{v}|k) = \frac{1}{\binom{n}{k}}$.
i.e. the order of outcomes is uniformly random once we condition on the number of successes.

Proof. By definition,

$$\begin{aligned} p_{\vec{o}|X}(\vec{v}|k) &= \frac{\mathbb{P}(\vec{o} = \vec{v}, X = k)}{\mathbb{P}(X = k)} \\ (\{\vec{o} = \vec{v}\} \subseteq \{X = k\}) \qquad &= \frac{\mathbb{P}(\vec{o} = \vec{v})}{\mathbb{P}(X = k)} \\ &= \frac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{1}{\binom{n}{k}} \end{aligned}$$

□

Remark 1. If X, Y are independent, then $\underbrace{p_{X|Y}(x|y)}_{\text{conditonal}} = \underbrace{p_X(x)}_{\text{marginal}}$.

This result is from the generalization of the above proof

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} \overset{\text{indep.}}{=} \frac{p_X(x) p_Y(y)}{p_Y(y)} = p_X(x).$$

2. Continuous Conditional Distribution

Definition 2. Given continuous random variables X and Y, the conditional probability density function of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Then, given a set A, $\mathbb{P}(X \in A | Y = y) = \int_A f_{X|Y}(x|y)\, \mathrm{d}x$.
Note that Even though $\mathbb{P}(Y = y) = 0$, this allows us to condition in the value of Y.

Justification: Consider

$$\begin{aligned} &\mathbb{P}(x \le X \le x + \mathrm{d}x | y \le Y \le y + \mathrm{d}y) \text{ as } \mathrm{d}x,\, \mathrm{d}y \to 0 \\ &= \frac{\mathbb{P}(x \le X \le x + \mathrm{d}x | y \le Y \le y + \mathrm{d}y)}{\mathbb{P}(y \le Y \le y + \mathrm{d}y)} \end{aligned}$$

where

$$\begin{aligned} \mathbb{P}(y \le Y \le y + \mathrm{d}y) &= \int_y^{y+\mathrm{d}y} f_Y(t)\, \mathrm{d}t \\ &\approx f_Y(y)\, \mathrm{d}y \end{aligned}$$

$$\begin{aligned} \mathbb{P}(x \le X \le x + \mathrm{d}x | y \le Y \le y + \mathrm{d}y) &= \int_x^{x+\mathrm{d}x} \int_y^{y+\mathrm{d}y} f(u,v)\, \mathrm{d}u\, \mathrm{d}v \\ &\approx f(x,y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

Therefore

$$f_{X|Y}(x|y)\,\mathrm{d}x \approx \mathbb{P}(x \le X \le x + \mathrm{d}x | y \le Y \le y + \mathrm{d}y) \approx \frac{f(x,y)\,\mathrm{d}x\,\mathrm{d}y}{f_Y(y)\,\mathrm{d}y}$$

$$\Rightarrow f_{X|Y}(x|y) \approx \frac{f(x,y)}{f_Y(y)}$$

Example 3. Let

$$f_{X,Y}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}}e^{-y}}{y} & \text{if } x,y>0 \\ 0 & \text{otherwise.} \end{cases}$$

Q. What is $\mathbb{P}(X > 1|Y = y)$?
Solution.

$$\mathbb{P}(X > 1|Y = y) = \int_1^\infty f_{X|Y}(x|y)\,\mathrm{d}x$$

where $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x \\ &= \int_0^\infty \frac{e^{-\frac{x}{y}}e^{-y}}{y}\,\mathrm{d}x \\ &= \frac{e^{-y}}{y}\int_0^\infty e^{-\frac{x}{y}}\,\mathrm{d}x \\ &= \frac{e^{-y}}{y} \cdot -ye^{-\frac{x}{y}}\Big|_{x=0}^{\infty} \\ &= e^{-y} \end{aligned}$$

Therefore

$$f_{X|Y}(x,y) = \frac{\frac{e^{-\frac{x}{y}}e^{-y}}{y}}{e^{-y}} = \frac{1}{y}e^{-\frac{x}{y}}.$$

Then

$$\mathbb{P}(X > 1|Y = y) = \int_1^\infty \frac{1}{y}e^{-\frac{x}{y}}\,\mathrm{d}x = -e^{-\frac{x}{y}}\Big|_{x=1}^{\infty} = \boxed{e^{-\frac{1}{y}}}.$$

Example 4. We have a possibly weighted coin. It comes up heads with probability, where $p \sim \mathrm{Unif}(0,1)$.

We toss the coin n times, and find that we get k heads.

What is the distribution of p given this outcome?

Solution. Let $X =$ the number of heads. By Bayes' theorem,

$$
\begin{aligned}
f_{p|X}(p'|k) &= \frac{\mathbb{P}(\{X = k\} \cap \{p = p'\})}{\mathbb{P}(X = k)} \\
&= \frac{\mathbb{P}(X = k) f_p(p')}{\mathbb{P}(X = k)} \\
&= \frac{\binom{n}{k} p'^k (1 - p')^{n-k}}{\mathbb{P}(X = k)}
\end{aligned}
$$

where $\mathbb{P}(X = k)$ is constant, independent of p'. Therefore

$$f_{p|X}(p'|k) = cp'^k (1 - p')^{n-k}$$

where c is such that $\int_0^1 cp'^k (1 - p')^{n-k} \, \mathrm{d}p' = 1$.

CHAPTER 14

Joint Distributions of Functions of Random Variables

Proposition 1. Let X_1 and X_2 be continuous random variables with joint pdf f_{X_1,X_2}.
We define new random variables $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ where

(a) $(X_1, X_2) \mapsto (Y_1, Y_2)$ is a bijection, and
(b) continuous partial derivatives of g_1, g_2 satisfies

$$J(x_1, x_2) = \left| \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} \right| \neq 0.$$

Then Y_1, Y_2 are jointly continuous random variables with pdf

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2) \cdot |J(x_1, x_2)|^{-1}$$

where $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$.

Proof. (Sketch) We have

$$\begin{aligned} F_{Y_1,Y_2}(y_1, y_2) &= \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2) \\ &= \iint_{\substack{g_1(x_1,x_2) \leq y_1 \\ g_2(x_1,x_2) \leq y_2}} f_{X_1,X_2}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \end{aligned}$$

Then differentiate with respect to y_1 and y_2. □

Example 1. Given X_1, X_2 random variables, let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.
Determine f_{Y_1,Y_2} in terms of f_{X_1,X_2}.

Solution. Observe that

(a) $X_1 = \frac{1}{2}(Y_1 + Y_2)$ and $X_2 = \frac{1}{2}(Y_1 - Y_2)$, so we do have a bijection, and
(b)

$$J(x_1, x_2) = \left| \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right| = 2 \neq 0.$$

By Proposition 1,

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2) \cdot |J(x_1, x_2)|^{-1} \\ &= \frac{1}{2} f_{X_1,X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2} \right) \end{aligned}$$

Example 2. Continue from the previous example.
Let $X_1, X_2 \sim N(0, 1)$ independently. Then $Y_1 \sim N(0, 2)$ and $Y_2 \sim N(0, 2)$.

To calculate the joint density,

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{2\pi}e^{-\frac{x_1^2}{2}-\frac{x_2^2}{2}},$$

and thus

$$\begin{aligned}
f_{Y_1,Y_2}(y_1,y_2) &= \frac{1}{2}f_{X_1,X_2}\left(\frac{y_1+y_2}{2},\frac{y_1-y_2}{2}\right) \\
&= \frac{1}{4\pi}e^{-\frac{\left(\frac{y_1+y_2}{2}\right)^2}{2}-\frac{\left(\frac{y_1-y_2}{2}\right)^2}{2}} \\
&= \frac{1}{4\pi}e^{-\frac{y_1^2+y_2^2}{4}} \\
&= \left(\frac{1}{\sqrt{2\cdot 2\pi}}e^{-\frac{y_1^2}{2\cdot 2}}\right)\left(\frac{1}{\sqrt{2\cdot 2\pi}}e^{-\frac{y_2^2}{2\cdot 2}}\right)
\end{aligned}$$

That means, in fact, Y_1 and Y_2 are independent.

Actually, if X_1, X_2 are independent and $Y_1 = X_1 + X_2, Y_2 = X_1 - X_2$ are independent, then X_1, X_2 are normal.

More generally, if we have $X_1, X_2, \ldots, X_n$ and we define $Y_i = g_i(X_1, X_2, \ldots, X_n)$ for $1 \le i \le n$ such that $(x_1, x_2, \ldots, x_n) \to (y_1, y_2, \ldots, y_n)$ is a bijection and

$$J(x_1,x_2,\ldots,x_n) = \left|\begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{pmatrix}\right| \neq 0,$$

then

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1,y_2,\ldots,y_n) = f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)\cdot|J(x_1,x_2,\ldots,x_n)|^{-1}$$

where $(x_1, x_2, \ldots, x_n) \mapsto (y_1, y_2, \ldots, y_n)$.

1. Expectation

Proposition 2. Let X be a non-negative continuous random variable. Then

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \ge x)\,\mathrm{d}x.$$

Proof.

$$\begin{aligned}
\int_0^\infty \mathbb{P}(X \ge t)\,\mathrm{d}t &= \int_0^\infty \int_t^\infty f(x)\,\mathrm{d}x\,\mathrm{d}t \\
&= \int_0^\infty \int_0^x f(x)\,\mathrm{d}t\,\mathrm{d}x \\
&= \int_0^\infty xf(x)\,\mathrm{d}x = \mathbb{E}[X]
\end{aligned}$$

$\square$

Proposition 3. If X and Y are random variables
(discrete) with joint probability mass function p, then for any $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}[g(x, y)] = \sum_{x,y} g(x, y)p(x, y).$$

(continuous) with joint probability density function f, then for any $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}[g(x, y)] = \iint_{\mathbb{R}^2} g(x, y)f(x, y) \,\mathrm{d}x \,\mathrm{d}y.$$

Proof. (continuous) If $g \geq 0$, we have

$$\begin{aligned}
\text{(Proposition 2)} \qquad \mathbb{E}[g(X, Y)] &= \int_0^\infty \mathbb{P}(g(X, Y) \geq t) \,\mathrm{d}t \\
&= \int_0^\infty \iint_{(x,y):g(x,y)\geq t} f(x, y) \,\mathrm{d}x \,\mathrm{d}y \,\mathrm{d}t \\
&= \iint_{\mathbb{R}^2} \int_0^{g(x,y)} f(x, y) \,\mathrm{d}t \,\mathrm{d}x \,\mathrm{d}y \\
&= \iint_{\mathbb{R}^2} g(x, y)f(x, y) \,\mathrm{d}x \,\mathrm{d}y
\end{aligned}$$

General case in homework. □

Example 3. On a road of length L, an accident occurs at position X, uniformly distributed along the road.

An ambulance is at position Y, also independently uniformly distributed along the road.

Q. What is the expected distance the ambulance has to travel to arrive at the accident?
Solution. Let $X \sim \mathrm{Unif}(0, L)$ and $Y \sim \mathrm{Unif}(0, L)$ independently. The distance is $g(X, Y) = |X - Y|$. Since X and Y are independent and $f_X = f_Y = \frac{1}{L}1_{[0,L]}$, its expectation is

$$\mathbb{E}[|X - Y|] = \int_0^L \int_0^L |x - y|f(x, y) \,\mathrm{d}x \,\mathrm{d}y = \frac{1}{L^2} \int_0^L \int_0^L |x - y| \,\mathrm{d}x \,\mathrm{d}y.$$

Compute

$$\begin{aligned}
\int_0^L |x - y| \,\mathrm{d}y &= \int_0^x x - y \,\mathrm{d}y + \int_x^L y - x \,\mathrm{d}y \\
&= \left(xy - \frac{1}{2}y^2\right)\Big|_{y=0}^{x} + \left(\frac{1}{2}y^2 - xy\right)\Big|_{y=x}^{L} \\
&= \frac{1}{2}x^2 + \left(\frac{1}{2}L^2 - xL - \left(-\frac{1}{2}x^2\right)\right) \\
&= \frac{1}{2}L^2 + x^2 - xL
\end{aligned}$$

$$\begin{aligned}\int_0^L \frac{1}{2}L^2 + x^2 - xL\,\mathrm{d}x &= \left(\frac{1}{2}L^2x + \frac{1}{3}x^3 - \frac{1}{2}x^2L\right)_{x=0}^{L} \\ &= \frac{1}{2}L^3 + \frac{1}{3}L^3 - \frac{1}{2}L^3 = \frac{1}{3}L^3\end{aligned}$$

Therefore

$$\mathbb{E}[|X-Y|] = \frac{1}{L^2}\int_0^L\int_0^L |x-y|\,\mathrm{d}x\,\mathrm{d}y = \boxed{\frac{L}{3}}$$

Example 4. (linearity) Let $g(X,Y) = aX + bY$. Then

$$\begin{aligned}\mathbb{E}[g(X,Y)] &= \mathbb{E}[aX+bY] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)\,\mathrm{d}x\,\mathrm{d}y + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= a\int_{-\infty}^{\infty}x\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y\,\mathrm{d}x + b\int_{-\infty}^{\infty}y\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ \text{(by definition)} \quad &= a\int_{-\infty}^{\infty}xf_X(x)\,\mathrm{d}x + b\int_{-\infty}^{\infty}yf_Y(y)\,\mathrm{d}y \\ &= a\mathbb{E}[X] + b\mathbb{E}[Y]\end{aligned}$$

Application. (Monotonicity) If $X \geq Y$, then $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

Proof. If $X \geq Y$, then $X - Y \geq 0$. Then $\mathbb{E}[X] - \mathbb{E}[Y] = \mathbb{E}[X-Y] \geq 0$. □

Application. (Union Bound, Boole's Inequality) Let $A_1, A_2, \ldots, A_n$ are events, then

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i).$$

Proof. For each i, let I_i be the indicators

$$I_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise.}\end{cases}$$

Then $\mathbb{E}[I_i] = \mathbb{P}(I_i = 1) = \mathbb{P}(A_i)$. Let $X = \sum_{i=1}^n I_i$ be the number of events that occur. Let $Y = I_{\{\bigcup_{i=1}^n A_i\}} = \begin{cases} 1 & \text{if } X \geq 1 \\ 0 & \text{if } X = 0\end{cases}$ be the indicator of the union of events. Then $Y \leq X$. By the above application, we have

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{E}[Y] \leq \mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n I_i\right] = \sum_{i=1}^n \mathbb{E}\left[I_i\right] = \sum_{i=1}^n \mathbb{P}(A_i).$$

□

Application. A coin is tossed n times. Each toss is heads with probability p and tails with probability $1-p$ independently.

How many different runs (= sequence of successive heads/tails) do we expect?

$$\underbrace{HHHHHHHHHH}_{\text{1 run}}$$

$$\underbrace{HTTHTHHHTT}_{\text{6 runs}}$$

Let X_i be the indicator of the event that a new run starts from position i.

Then $X =$ number of runs $= \sum_{i=1}^n X_i$. Therefore

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mathbb{P}(\text{a new run starts from position } i) =: p_i.$$

We have $p_1 = 1$ and

$$\begin{aligned} p_i &= \mathbb{P}(\text{outcome } i \neq \text{outcome } i-1) \\ &= \mathbb{P}(\text{outcome } i = T, \text{ outcome } i-1 = H) + \mathbb{P}(\text{outcome } i = H, \text{ outcome } i-1 = T) \\ &= (1-p)p + p(1-p) = 2p(1-p) \end{aligned}$$

Therefore $\mathbb{E}[X] = 1 + 2p(1-p)(n-1)$.

Application. (Analysis of Quicksort)

Sorting problem: Given a permutation of $1, 2, \ldots, n$, we want an algorithm that outputs a sorted list in comparison-based model.

Quicksort algorithm: Pick a pivot x uniformly at random from the list, and then compare everthing else to x.

Then repeat the process on the left and right sublists.

Q. On average, how many comparisons does Quicksort need?

Solution. For every $1 \leq i \leq j \leq n$, let $X_{i,j}$ be the indicator of the event that i and j are compared.

Then $X = \sum_{1 \leq i < j \leq n} X_{i,j}$ is the total number of comparisons. Therefore

$$\mathbb{E}[X] = \sum_{1 \leq i < j \leq n} \mathbb{E}[X_{i,j}] = \sum_{1 \leq i < j \leq n} \mathbb{P}(i, j \text{ are compared}).$$

Consider the interval $i, i+1, \ldots, j$.

- If the pivot x is outside this interval, then the interval stays together for the next round.
- If the pivot x is inside the interval,
  - if $x = i$ or $x = j$, then i and j are compared.
  - if $x \neq i, j$, then i and j are not compared.

Therefore

$$\begin{aligned}
\mathbb{P}(i, j \text{ are compared}) &= \mathbb{P}\left(\substack{\text{when pivot } x \text{ satisfies } i \leq x \leq j, \\ \text{we have } x \in \{i,j\}}\right) \\
&= \mathbb{P}(x = i \text{ or } j | x \sim \mathrm{Unif}(\{i, i+1, \ldots, n\})) \\
&= \frac{2}{j - i + 1}.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{E}[X] = \sum_{1 \leq i < j \leq n} \mathbb{E}[X_{i,j}] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j - i + 1} \\
&= \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} \frac{2}{j+1} \\
&\approx 2 \sum_{i=1}^{n-1} \ln(n - i + 1) = 2 \sum_{i=2}^{n} \ln i \\
&\approx 2 \int_2^n \ln x \, \mathrm{d}x \\
&\approx 2n \ln n
\end{aligned}$$

CHAPTER 15

Moments of Numbers of Events

Setting:

- Let $A_1, A_2, \dots, A_n$ be events in a probability space.
- Let $I_1, I_2, \dots, I_n$ be the corresponding indicators random variables.

Then $X = \sum_{i=1}^{n} I_i$ is the number of events that occur, and $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(A_i)$.
Now consider $Y = \sum_{1 \le i < j \le n} I_i I_j$. We have

$$I_i I_j = \begin{cases} 1 & \text{if } A_i \cap A_j \text{ occur} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore Y is the number of pairs of events that occur $= \binom{X}{2}$.
By linearity of expectation,

$$\begin{aligned} \mathbb{E}\left[\binom{X}{2}\right] &= \mathbb{E}\left[\sum_{1 \le i < j \le n} I_i I_j\right] \\ &= \sum_{1 \le i < j \le n} \mathbb{E}[I_i I_j] \\ &= \sum_{1 \le i < j \le n} \mathbb{P}(A_i \cap A_j) \end{aligned}$$

More generally, $\binom{X}{k} = \sum_{i_1 < i_2 < \cdots < i_k} I_{i_1} I_{i_2} \cdots I_{i_k}$ and

$$\mathbb{E}\left[\binom{X}{k}\right] = \sum_{i_1 < i_2 < \cdots < i_k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

Example 1. Let $X \sim \mathrm{Bin}(n, p)$. Let $A_i = \{\text{the } i\text{-th trial is a success}\}$.
Then $X = \sum_{i=1}^{n} I_i$ where I_i is the indicator random variable for A_i.

$$\begin{aligned} \mathbb{E}\left[\binom{X}{2}\right] &= \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\ \text{(independence)} \qquad &= \sum_{i<j} \mathbb{P}(A_i)\mathbb{P}(A_j) \\ &= \sum_{i<j} p^2 = \binom{n}{2} p^2 \end{aligned}$$

This is equal to the result of the calculation

$$\begin{aligned}\mathbb{E}\left[\binom{X}{2}\right] &= \mathbb{E}\left[\frac{X(X-1)}{2}\right] \\ &= \frac{1}{2}\mathbb{E}[X^2 - X] \\ &= \frac{1}{2}\left(\mathbb{E}[X^2] - \mathbb{E}[X]\right)\end{aligned}$$

Therefore

$$n(n-1)p^2 = \mathbb{E}[X^2] - \mathbb{E}[X] = \mathbb{E}[X^2] - np$$
$$\mathbb{E}[X^2] = n(n-1)_n p$$

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np - np^2 = np(1-p)\end{aligned}$$

Example 2. Hat matching problem.

Let X = number of people getting own hat and $A_i = \{i\text{-th person gets own hat}\}$, $I_i = I_{A_i}$.
Then $X = \sum_{i=1}^n I_i$ and $\mathbb{P}(A_i) = \frac{1}{n}$.
By linearity, $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^n \frac{1}{n} = 1$.

$$\begin{aligned}\mathbb{E}\left[\binom{X}{2}\right] &= \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\ &= \sum_{i<j} \mathbb{P}(A_i)\mathbb{P}(A_j|A_i) \\ &= \sum_{i<j} \frac{1}{n} \cdot \frac{1}{n-1} = \binom{n}{2}\frac{1}{n(n-1)}\end{aligned}$$

Therefore $\mathbb{E}[X(X-1)] = 1$ and thus $\mathrm{Var}(X) = 1$.
More generally,

$$\begin{aligned}&\mathbb{E}[X(X-1)\cdots(X-k+1)] \\ &= k!\mathbb{E}\left[\binom{X}{k}\right] \\ &= k! \sum_{i_1<i_2<\cdots<i_k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \\ &= k! \sum_{i_1<i_2<\cdots<i_k} \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}|A_{i_1}) \cdots \mathbb{P}(A_{i_k}|A_{i_1} \cap \cdots \cap A_{i_{k-1}}) \\ &= k!\binom{n}{k}\frac{1}{n(n-1)\cdots(n-k+1)} = 1\end{aligned}$$

Example 3. Coupon collecter.

Every time we get a uniformly random type of coupon from 1 to n.

Let X = number of coupons we collect until we have a full set. Then

$$\mathbb{E}[X] = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \approx n \log n.$$

This means on average, we expect to have $\log n$ copies of each coupon.

Let Y = number of types of coupon that we only have a SINGLE copy of when we finish our first set.

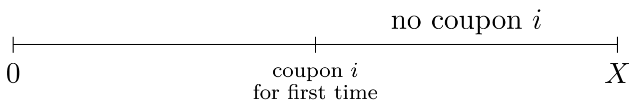
We already know that $Y \geq 1$. What is $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$?

Solution. Let $A_i = \{$the i-th type of coupon we get appears uniquely$\}$.

Let $I_i = I_{A_i}$ be the indicator random variable for A_i and $Y = \sum_{i=1}^n I_i$.

Note that

$$\begin{aligned} A_i &= \left\{\begin{array}{c}\text{in the coupons after getting coupon } i \text{ for the first time,}\\ \text{we get coupons } i, i+1, \ldots, n \text{ before getting } i\end{array}\right\} \\ &= \left\{\begin{array}{c}\text{in the set of coupons } \{i, i+1, \ldots, n\},\\ \text{coupon } i \text{ comes last}\end{array}\right\}\end{aligned}$$



Therefore $\mathbb{P}(A_i) = \frac{1}{n-i+1}$ by symmetry. Then

$$\mathbb{E}[Y] = \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^n \frac{1}{n-i+1} = \sum_{i=1}^n \frac{1}{i} \approx \log n.$$

Then in order to compute $\mathrm{Var}(Y)$, we look at the second moment.

Let $S_{i,j} = \left\{\begin{array}{c}\text{among the coupons } \{i,i+1,\ldots,n\} \text{ after seeing coupon } i \text{ for the first time,}\\ i \text{ is not in the first } j-i\end{array}\right\}$, and $T_{i,j} = \left\{\begin{array}{c}\text{among the coupons } \{i,j,j+1,\ldots,n\} \text{ after seeing coupon } j \text{ for the first time,}\\ i, j \text{ are the last two to appear}\end{array}\right\}$.

By uniformity, coupon i is equally likely to be in any position. Therefore $\mathbb{P}(S_{i,j}) = 1 - \frac{j-i}{n-i+1}$. Then

$$\mathbb{P}(S_{i,j} \cap T_{i,j}) = \mathbb{P}(S_{i,j})\mathbb{P}(T_{i,j} | S_{i,j}) = \mathbb{P}(S_{i,j})\mathbb{P}(T_{i,j})$$

since the "tail" coupons are independent of previous ones.

By uniformity, we have $\mathbb{P}(T_{i,j}) = \frac{2}{(n-j+2)(n-j+1)}$. Therefore

$$\begin{aligned}
\mathbb{E}\left[\binom{Y}{2}\right] &= \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\
&= \sum_{i<j} \mathbb{P}(S_{i,j})\mathbb{P}(T_{i,j}) \\
&= \sum_{i<j} \left(1 - \frac{j-i}{n-i+1}\right) \cdot \frac{2}{(n-j+2)(n-j+1)} \\
&= \sum_{i<j} \frac{2}{(n-i+1)(n-j+2)}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\
&= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 - 2\mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

CHAPTER 16

Covariance, Variance of Sums, and Correlation

1. Covariance

Proposition 1. If X and Y are independent, then for any function g, h,

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

Proof. (Continuous)

$$\begin{aligned}\mathbb{E}[g(X)h(Y)] &= \iint_{\mathbb{R}^2} g(X)h(Y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{\mathbb{R}^2} g(X)h(Y)f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\mathbb{R}} g(X)f_X(x)\,\mathrm{d}x \int_{\mathbb{R}} h(Y)f_Y(y)\,\mathrm{d}y \\ &= \mathbb{E}[g(X)]\mathbb{E}[h(Y)]\end{aligned}$$

□

Definition 1. Let X and Y be random variables. We define the covariance of X and Y to be

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])].$$

Observe

$$\begin{aligned}&\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X\mathbb{E}[Y]] - \mathbb{E}[Y\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

This implies that if X, Y are independent, then $\mathrm{Cov}(X,Y) = 0$.
WARNING: The converse is NOT true. Counterexample: Let

$$X = \begin{cases} -1 & \text{with probability } \frac{1}{3} \\ 0 & \text{with probability } \frac{1}{3}, \\ 1 & \text{with probability } \frac{1}{3} \end{cases} \qquad Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{if } X = 0. \end{cases}$$

Then $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$ but $p_{X,Y}(1,1) = 0 \neq \frac{1}{3}\cdot\frac{1}{3} = p_X(1)p_Y(1)$.

Proposition 2. The covariance has the following properties:

(i) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$
(ii) $\mathrm{Var}(X)=\mathrm{Cov}(X,X)$
(iii) $\mathrm{Cov}(aX,Y)=a\,\mathrm{Cov}(X,Y)$
(iv) $\mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right)=\sum_{i=1}^n\sum_{j=1}^m \mathrm{Cov}(X_i,Y_j)$

Proof. (i)-(iii) Followed by definition.

(iv) Let $\mu_i=\mathbb{E}[X_i]$, $\nu_j=\mathbb{E}[Y_j]$. Then by linearity of expectation,

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right]=\sum_{i=1}^n \mathbb{E}[X_i]=\sum_{i=1}^n \mu_i$$
$$\mathbb{E}\left[\sum_{j=1}^m Y_j\right]=\sum_{j=1}^m \nu_j$$

$$\begin{aligned}
\Rightarrow \mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i-\mathbb{E}\left[\sum_{i=1}^n X_i\right]\right)\left(\sum_{j=1}^m Y_j-\mathbb{E}\left[\sum_{j=1}^m Y_j\right]\right)\right]\\
&= \mathbb{E}\left[\left(\sum_{i=1}^n X_i-\sum_{i=1}^n \mu_i\right)\left(\sum_{j=1}^m Y_j-\sum_{j=1}^m \nu_j\right)\right]\\
&= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i-\mu_i)\right)\left(\sum_{j=1}^m (Y_j-\nu_j)\right)\right]\\
&= \mathbb{E}\left[\sum_{i=1}^n\sum_{j=1}^m (X_i-\mu_i)(Y_j-\nu_j)\right]\\
\text{(LoE)} \qquad &= \sum_{i=1}^n\sum_{j=1}^m \mathbb{E}\left[(X_i-\mu_i)(Y_j-\nu_j)\right]\\
&= \sum_{i=1}^n\sum_{j=1}^m \mathrm{Cov}(X_i,Y_j)
\end{aligned}$$

□

Corollary 1. Let $X=\sum_{i=1}^n X_i$ be a sum of random variables. Then

$$\mathrm{Var}(X)=\mathrm{Cov}(X,X)\overset{\text{(b)}}{=}\mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right)\overset{\text{(d)}}{=}\sum_{i=1}^n\sum_{j=1}^n \mathrm{Cov}(X_i,X_j).$$

Therefore, by (a) and (b),

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{Var}(X_i)+2\sum_{i<j}\mathrm{Cov}(X_i,X_j).$$

Corollary 2. If $X_1, X_2, \dots, X_n$ are pairwise independent, then

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i).$$

We show that the independence is necessary: Let $X_2 = -X_1$, where X_1 is a non-constant random variable.

Then $\operatorname{Var}(X_2) = (-1)^2 \operatorname{Var}(X_1) = \operatorname{Var}(X_1) \neq 0$, but

$$0 = \operatorname{Var}(X_1 + X_2) \neq \operatorname{Var}(X_1) + \operatorname{Var}(X_2).$$

Example 1. Let $X \sim \operatorname{Bin}(n, p)$. Then $X = \sum_{i=1}^n X_i$, where $X_i \sim \operatorname{Ber}(p)$ are i.i.d Bernoulli random variables.

Since X_i are independent,

$$\operatorname{Var}(X) = \sum_{i=1}^n \operatorname{Var}(X_i) = n \operatorname{Var}(X_1).$$

$$\begin{aligned}
\mathbb{E}[X_1] &= 1\mathbb{P}(X_1 = 1) + 0\mathbb{P}(X_1 = 0) = p \\
\mathbb{E}[X_1^2] &= \mathbb{E}[X_1] = p \\
\Rightarrow \operatorname{Var}(X_1) &= \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = p(1-p) \\
\Rightarrow \operatorname{Var}(X) &= np(1-p)
\end{aligned}$$

Example 2. Let $X_1, X_2, \dots, X_n$ be i.i.d random variables with mean μ and variance σ^2.

Define the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then

$$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] \overset{\text{LoE}}{=} \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^n \mu = \boxed{\mu}.$$

$$\operatorname{Var}(\overline{X}) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^n X_i\right) \overset{\text{indep.}}{=} \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

Compared to using just X_1 to extimate

$$\mathbb{E}[X_1] = \mu, \quad \text{but } \operatorname{Var}(X_1) = \sigma^2.$$

Therefore using the average $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ reduces the variance, hence a more accurate estimate.

Definition 2. We define the deviation of X_i as $X_i - \overline{X}$. Note that $\sum_i (X_i - \overline{X}) \equiv 0$.

Let $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ (square to remove cancellation). What is $\mathbb{E}[S^2]$?

$$\mathbb{E}[(n-1)S^2] = \mathbb{E}\left[\sum_{i=1}^n (X_i - \overline{X})^2\right]$$

where

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X})^2\\ &= \sum_{i=1}^{n}\left((X_i-\mu)^2+(\overline{X}-\mu)^2-2(X_i-\mu)(\overline{X}-\mu)\right)\\ &= \sum_{i=1}^{n}(X_i-\mu)^2+n(\overline{X}-\mu)^2-2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}[(n-1)S^2] &= \mathbb{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]\\ &= \sum_{i=1}^{n}\mathbb{E}[(X_i-\mu)^2]-n\mathbb{E}[(\overline{X}-\mu)^2] \qquad \text{(LoE)}\\ &= \sum_{i=1}^{n}\operatorname{Var}(X_i)-n\operatorname{Var}(\overline{X})\\ &= n\sigma^2-n\frac{\sigma^2}{n}=(n-1)\sigma^2\\ &\Rightarrow \mathbb{E}[S^2]=\sigma^2\end{aligned}$$

In summary, given samples $X_1, X_2, \ldots, X_n$, $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ is an unbiased estimator of the mean μ, and $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ is an unbiased estimator of the variance σ^2.

Definition 3. Let X, Y be random variable with $\operatorname{Var}(X), \operatorname{Var}(Y)>0$. Then we define the correlation of X and Y to be

$$\rho(X,Y)=\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Claim 1. For any X, Y, $-1\le\rho(X,Y)\le 1$.

Proof. Let $\sigma_X=\sqrt{\operatorname{Var}(X)}$, $\sigma_Y=\sqrt{\operatorname{Var}(Y)}$. Consider $\operatorname{Var}\left(\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}\right)$:

$$\begin{aligned}0\le\operatorname{Var}\left(\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}\right) &= \operatorname{Var}\left(\frac{X}{\sigma_X}\right)+\operatorname{Var}\left(\frac{Y}{\sigma_Y}\right)+2\operatorname{Cov}\left(\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right)\\ &= \frac{\operatorname{Var}(X)}{\sigma_X^2}+\frac{\operatorname{Var}(Y)}{\sigma_Y^2}+2\frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}\\ &= 2(1+\rho(X,Y))\end{aligned}$$

Therefore $\rho(X,Y)\ge -1$.

Similarly,

$$0 \le \operatorname{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \operatorname{Var}\left(\frac{X}{\sigma_X}\right) + \operatorname{Var}\left(\frac{Y}{\sigma_Y}\right) - 2\operatorname{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$
$$= 2(1 - \rho(X,Y))$$

Therefore $\rho(X,Y) \le 1$. $\square$

- Note that

$$\begin{aligned}\rho(X,Y) = 1 &\Rightarrow \operatorname{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 0 \\ &\Rightarrow \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \text{ is constant} \\ &\Rightarrow Y = aX + b, a > 0\end{aligned}$$

- $\rho(X,Y) = 0 \Leftarrow X, Y$ are independent. (The reverse is not true.)
  If $\rho(X,Y) = 0$, we say X, Y are unrelated.
- $\rho(X,Y) = -1 \Rightarrow Y = aX + b, a < 0$

Example 3. Let A, B be events in a probability space.
Let I_A, I_B be their indicator random variables. Then

$$\operatorname{Cov}(I_A, I_B) = \mathbb{E}[I_A I_B] - \mathbb{E}[I_A]\mathbb{E}[I_B] = \mathbb{E}[I_{A\cap B}] - \mathbb{E}[I_A]\mathbb{E}[I_B] = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B).$$

We can understand this covariance as

$$\begin{aligned}\operatorname{Cov}(I_A, I_B) &= \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(B)\left(\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} - \mathbb{P}(A)\right) \\ &= \mathbb{P}(B)\left(\mathbb{P}(A|B) - \mathbb{P}(A)\right)\end{aligned}$$

2. Conditional Expectation

(Discrete case) Given X, Y, we can look at the distribution of X given the value of Y

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}.$$

This defines a (conditional) distribution.
Therefore we can discuss what its expectation is.

$$\mathbb{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y).$$

Example 4. Let $X, Y \sim \text{Bin}(n, p)$ be independent. What is $\mathbb{E}[X|X+Y=m]$?
Solution. Compute

$$\begin{aligned} p_{X|X+Y}(k|m) &= \frac{\mathbb{P}(X=k, X+Y=m)}{\mathbb{P}(X+Y=m)} \\ &= \frac{\mathbb{P}(X=k, Y=m-k)}{\mathbb{P}(X+Y=m)} \\ &= \frac{\mathbb{P}(X=k)\mathbb{P}(Y=m-k)}{\mathbb{P}(X+Y=m)} \end{aligned}$$

where

$$\begin{aligned} \mathbb{P}(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ \mathbb{P}(Y=m-k) &= \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)} \\ \mathbb{P}(X+Y=m) &= \binom{2n}{m} p^m (1-p)^{2n-m} \end{aligned}$$

since $X+Y \sim \text{Bin}(2n, p)$. Continue the computation:

$$\begin{aligned} p_{X|X+Y}(k|m) &= \frac{\mathbb{P}(X=k)\mathbb{P}(Y=m-k)}{\mathbb{P}(X+Y=m)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)}}{\binom{2n}{m} p^m (1-p)^{2n-m}} \\ &= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}, \quad 0 \leq k \leq \min\{n, m\} \end{aligned}$$

Therefore $X|X+Y=m$ has the hypergeometric distribution with parameters $2n, n, m$. Using what we know about the distribution, $\mathbb{E}[X|X+Y=m] = \frac{m}{2}$.

(Continuous case) Given the value of Y, we can define the conditional distribution of X through the conditional density function

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

The conditional expectation is then

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) \, \mathrm{d}x.$$

Example 5. Let

$$f_{X,Y}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}} e^{-y}}{y} & \text{if } x, y > 0 \\ 0 & \text{otherwise }. \end{cases}$$

What is $\mathbb{E}[X|Y=y]$?

Solution. First calculate the conditional density:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x \\ &= \int_0^{\infty} \frac{e^{-\frac{x}{y}}e^{-y}}{y}\,\mathrm{d}x \\ &= \frac{e^{-y}}{y}\int_0^{\infty} e^{-\frac{x}{y}}\,\mathrm{d}x \\ &= \frac{e^{-y}}{y}\cdot -ye^{-\frac{x}{y}}\Big|_{x=0}^{\infty} = e^{-y} \end{aligned}$$

Therefore

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{1}{y}e^{-\frac{x}{y}}e^{-y}}{e^{-y}} = \frac{1}{y}e^{-\frac{x}{y}} \quad \text{when } x,y>0.$$

We find that $X|Y=y \sim \mathrm{Exp}(\frac{1}{y})$, and thus $\mathbb{E}[X|Y=y] = \boxed{y}$.

Now we have

$$\mathbb{E}[X|Y=y] = \begin{cases} \sum_x x p_{X|Y}(x|y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

Note that $\mathbb{E}[X|Y=y]$ is a function of y and $\mathbb{E}[X|Y]$ is a random variable.

Proposition 3. For any two random variables X, Y, $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$.

Proof. (both X and Y are discrete) Observe

$$\begin{aligned}
\mathbb{E}[X|Y=y] &= \sum_x x p_{X|Y}(x|y) = \sum_x x \cdot \frac{p(x,y)}{p_Y(y)} \\
\mathbb{E}[g(Y)] &= \sum_y g(y) p_Y(y) \qquad (\text{apply } g(y) = \mathbb{E}[X|Y=y]) \\
\Rightarrow \mathbb{E}[\mathbb{E}[X|Y]] &= \sum_y \mathbb{E}[X|Y=y] p_Y(y) \\
&= \sum_y \left(\sum_x x \cdot \frac{p_{X,Y}(x,y)}{p_Y(y)} \right) p_Y(y) \\
&= \sum_y \sum_x x p(x,y) \\
&= \sum_x \sum_y p(x,y) \\
&= \sum_x x p(x) = \mathbb{E}[X]
\end{aligned}$$

□

Example 6. A stop has a random number of customers each day, whose distribution is Poi(50). Each customer spends a random amount of money, uniform on \$[0, 1000], independent of all other customers. What is the expected daily income?

Solution. Let N = number of customers $\sim$ Poi(50).

Let X_i = amount of money spent by customer i, so $X_i \sim \text{Unif}[0, 1000]$.

Let X be the total income $X = \sum_{i=1}^{N} X_i$. Then

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|N]] \\
\mathbb{E}[X|N=n] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i] = 500n \qquad \text{(LoE)}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \mathbb{E}[X] &= \mathbb{E}[500N] = 500\mathbb{E}[N] \\
&= 500 \cdot 50 = \boxed{25000}
\end{aligned}$$

Example 7. n students take an exam. Their grades are independent and random. Each student gets

$$\begin{aligned} &\text{A with probability } p_1,\\ &\text{B with probability } p_2,\\ &\text{C with probability } p_3,\\ &\text{Fail with probability } p_4. \end{aligned}$$

Let N_i be the number of students who get the i-th grade respectively.
Q. What is $\mathbb{E}[N_i|N_j > 0]$?
Solution. Define a new random variable

$$I = \begin{cases} 1 & \text{if } N_j > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $\mathbb{E}[N_i|N_j > 0] = \mathbb{E}[N_i|I = 1]$.
By our proposition,

$$\begin{aligned} \mathbb{E}[N_i] &= \mathbb{E}[\mathbb{E}[N_i|I]] \\ &= \mathbb{E}[N_i|I = 1]\mathbb{P}(I = 1) + \mathbb{E}[N_i|I = 0]\mathbb{P}(I = 0) \end{aligned}$$

Therefore

$$\mathbb{E}[N_i|I = 1] = \frac{\mathbb{E}[N_i] - \mathbb{E}[N_i|I = 0]\mathbb{P}(I = 0)}{\mathbb{P}(I = 1)}.$$

Note that $N_i \sim \text{Bin}(n, p_i)$. Therefore $\mathbb{E}[N_i] = np_i$.
Then note that $N_i|I = 0 \sim \text{Bin}\left(n, \frac{p_i}{1-p_j}\right)$. Therefore $\mathbb{E}[N_i|I = 0] = \frac{np_i}{1-p_j}$.
Also, $\mathbb{P}(I = 0) = (1 - p_j)^n$ and thus $\mathbb{P}(I = 1) = 1 - (1 - p_j)^n$.
Therefore we can compute $\mathbb{E}[N_i|I = 1]$ by plugging in the values:

$$\mathbb{E}[N_i|I = 1] = \frac{np_i - \frac{np_i}{1-p_j} \cdot (1 - p_j)^n}{1 - (1 - p_j)^n}.$$

Example 8. (Variance of a geometric random variable) There is a sequence of independent trials, each is successful with probability p.

Let N be the number of trials until the first success. Then $N \sim \text{Geom}(p)$.
Let Y be the indicator random variable for the event that the first trial is successful.
Our goal is to compute $\mathbb{E}[N^2]$.

$$\begin{aligned} \mathbb{E}[N^2] &= \mathbb{E}[\mathbb{E}[N^2|Y]] \\ &= \mathbb{E}[N^2|Y = 0]\mathbb{P}(Y = 0) + \mathbb{E}[N^2|Y = 1]\mathbb{P}(Y = 1) \\ &= \mathbb{E}[N^2|Y = 1]p + \mathbb{E}[N^2|Y = 0](1 - p) \end{aligned}$$

$$\mathbb{E}[N^2|Y = 1] = 1 \quad \text{since } Y = 1 \Rightarrow N = 1$$

$$\mathbb{E}[N^2|Y = 0] = \mathbb{E}[(N + 1)^2] \quad \substack{\text{since the number of addititonal trials} \\ \text{has the same geometric distribution}}$$

Therefore

$$\mathbb{E}[N^2] = \underbrace{\mathbb{E}[(N+1)^2]}_{\mathbb{E}[N^2]+2\mathbb{E}[N]+1}(1-p) + 1 \cdot p$$

$$p\mathbb{E}[N^2] = \frac{2(1-p)}{p} + 1$$

$$\mathbb{E}[N^2] = \frac{2(1-p)}{p^2} + \frac{1}{p}$$

$$\mathrm{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Example 9. There are r tech companies. The i-th company has n_i billion dollars, $n_i \in \mathbb{N}$.

Two companies will go to court for copyright infringement. There is a probability of $\frac{1}{2}$ for each company to win, and the loser pays the winner \$1 billion.

Repeat until all but one company is bankrupt, and the remaining company has n billion dollars, where $n = \sum_{i=1}^{r} n_i$.

Q. What is the expected value of court cases?

Solution. We start with the case $r = 2$.

There are n billion dollars in total. Let X_i = the number of rounds if the first company starts with i billion, $1 \leq i \leq n-1$.

Let Y be the outcome of the first case

$$Y = \begin{cases} 1 & \text{if the first company wins} \\ 2 & \text{if the second company wins.} \end{cases}$$

Let A_i be the number of additional rounds after the first if the first company starts with i billion. Therefore $X_i = 1 + A_i$.

Define $m_i = \mathbb{E}[X_i]$. Then $m_i = \mathbb{E}[X_i] = \mathbb{E}[1 + A_i] = 1 + \mathbb{E}[A_i]$.

$$\mathbb{E}[A_i | Y = 1] \sim X_{i+1}$$
$$\mathbb{E}[A_i | Y = 2] \sim X_{i-1}$$

Therefore $\mathbb{E}[A_i] = \mathbb{E}[\mathbb{E}[A_i|Y]] = \frac{1}{2}\mathbb{E}[X_{i+1}] + \frac{1}{2}\mathbb{E}[X_{i-1}]$. That is, $m_i = 1 + \frac{1}{2}m_{i+1} + \frac{1}{2}m_{i-1}$. Solve the recurrence relation and we can get that $m = i(n-i)$.

General case. Let $X^{(i)}$ be the number of court cases involving the i-th company. Then $X = \frac{1}{2}\sum_{i=1}^{r} X^{(i)}$, and thus $\mathbb{E}[X] = \frac{1}{2}\sum_{i=1}^{r} \mathbb{E}[X^{(i)}]$.

For $X^{(i)}$, its distribution is the same as the two-company process with initial amounts n_i and $n - n_i$. (Why?)

Therefore $\mathbb{E}[X^{(i)}] = n_i(n - n_i)$, and

$$\mathbb{E}[X] = \frac{1}{2}\sum_{i=1}^{r} n_i(n - n_i) = \boxed{\frac{1}{2}\left(n^2 - \sum_{i=1}^{r} n^2\right)}.$$

Example 10. A professor examines n students. She wants to give the best student an A+. But she has to give the grades immediately.

Suppose the students come in a random order. How can the professor maximize the probability of giving the best student A+?

Solution. We attempt to find some strategies:

Strategy 1: Give the first student A+. This strategy has a probability of $\frac{1}{n}$ to success.

Strategy k: Look at the first k students and don't give them an A+. Then when she sees a student who is better than everyone before, give them an A+.

Let $E = \{\text{give the best student the A+}\}$ and $X =$ position of the best student. Then $\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E|X = i)\mathbb{P}(X = i)$.

If $i \le k$, then $\mathbb{P}(E|X = i) = 0$.

If $i > k$, for $E|X = i$ to happen, the best of the first $i - 1$ students has to be in the first k. Thus $\mathbb{P}(E|X = i) = \frac{k}{i-1}$. Therefore

$$\begin{aligned}
\mathbb{P}(E) &= \sum_{i=1}^{n} \mathbb{P}(E|X = i)\mathbb{P}(X = i) \\
&= \sum_{i=k+1}^{n} \mathbb{P}(E|X = i)\mathbb{P}(X = i) \\
&= \sum_{i=k+1}^{n} \frac{k}{i-1} \cdot \frac{1}{n} \\
&= \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1} \\
&= \frac{k}{n} \left(\sum_{i=1}^{n-1} \frac{1}{i} - \sum_{i=1}^{k} \frac{1}{i} \right) \\
&\approx \frac{k}{n} \ln \frac{n}{k}
\end{aligned}$$

Optimizing for the best choice of k, we get $k \approx \frac{n}{e}$ and $\mathbb{P}(E) = \frac{1}{e}$.

3. Conditional Variance

Definition 4. The conditional variance of X given Y is

$$\mathrm{Var}(X|y) = \mathbb{E}[(X - \mathbb{E}[X|Y = y])^2|Y = y] = \mathbb{E}[X^2|Y = y] - \mathbb{E}[X|Y = y]^2.$$

We have the amazing result:

Proposition 4.

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y]).$$

Proof. Homework :) □

CHAPTER 17

Moment Generating Functions

Definition 1. Given a random variable X, its moment generating function is defined by

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

Proposition 1. For any $n \in \mathbb{N}$, $\mathbb{E}[X^n] = M_X^{(n)}(0)\ \big(= \frac{\mathrm{d}^n}{\mathrm{d}t^n} M(0)\big)$.

Proof.

$$\begin{aligned} M_X &= \mathbb{E}[e^{tX}] \\ M^{(n)}(t) &= \frac{\mathrm{d}^n}{\mathrm{d}t^n}\mathbb{E}[e^{tX}] \overset{(*)}{=} \mathbb{E}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n} e^{tX}\right] = \mathbb{E}[X^n e^{tX}] \\ \Rightarrow M^{(n)}(0) &= \mathbb{E}[X^n e^{0X}] = \mathbb{E}[X^n] \end{aligned}$$

□

Remark 1. In (*) we exchanged the order of differentiation and expectation. This is generally valid for all distributions we care about.

Fact 1. The distribution of X is determined by its moment generating function.

Example 1. Let $X \sim \mathrm{Bin}(n, p)$. Then

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{k=0}^{n} e^{tk}\binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1-p)^{n-k} \\ \text{(binomial thm)} \qquad &= (e^t p + (1-p))^n \end{aligned}$$

$$\mathbb{E}[X] = M'(0) = n(e^t p + (1-p))^{n-1} p e^t \big|_{t=0} = np$$

$$\begin{aligned} \mathbb{E}[X^2] &= M''(0) \\ &= \big(n(n-1)(e^t p + (1-p))^{n-2}(pe^t)^2 + n(e^t p + (1-p))^{n-1} pe^t\big)\big|_{t=0} \\ &= n(n-1)p^2 + np \end{aligned}$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = n(n-1)p(1-p)$$

Example 2. Let $X \sim \text{Poi}(\lambda)$. Then

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{n=0}^{\infty} e^{tn} \frac{e^{-\lambda}\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda + \lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

$$\mathbb{E}[X] = M'(0) = \lambda e^t e^{\lambda(e^t-1)}\big|_{t=0} = \lambda$$

$$\begin{aligned} \mathbb{E}[X^2] &= M''(0) \\ &= \left(\lambda e^t \cdot e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)}\right)\Big|_{t=0} \\ &= \lambda + \lambda^2 \end{aligned}$$

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

Example 3. Let $X \sim \text{Exp}(\lambda)$. Then

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} \, \mathrm{d}x \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x} \, \mathrm{d}x \\ &= \lambda \cdot \frac{e^{(t-\lambda)x}}{t - \lambda}\Bigg|_{t=0}^{\infty} \\ \text{(if } t < \lambda) \qquad &= \frac{\lambda}{\lambda - t} \end{aligned}$$

$$\mathbb{E}[X] = M'(0) = \frac{\lambda}{(\lambda - t)^2}\Bigg|_{t=0} = \frac{1}{\lambda}$$

$$\begin{aligned} \mathbb{E}[X^2] &= M''(0) \\ &= \frac{2\lambda}{(\lambda - t)^3}\Bigg|_{t=0} = \frac{2}{\lambda^2} \end{aligned}$$

$$\Rightarrow \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}$$

Example 4. Let $X \sim N(\mu, \sigma^2)$.

Case: $\mu = 0$, $\sigma^2 = 1$. $Z \sim N(0,1)$. Then

$$\begin{aligned} M_Z(t) &= \mathbb{E}[e^{tZ}] \\ &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, \mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2tz)}\, \mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} e^{\frac{t^2}{2}}\, \mathrm{d}z \\ &= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}}}_{\text{pdf of } N(t,1)}\, \mathrm{d}z = e^{\frac{t^2}{2}} \end{aligned}$$

Case: $X \sim (\mu, \sigma^2)$.

We already know that if $Z \sim N(0,1)$, then $\mathbb{E}[e^{tZ}] = e^{\frac{t^2}{2}}$. Then

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \mathbb{E}[e^{t(\sigma Z + \mu)}] \\ &= e^{t\mu}\mathbb{E}[e^{(t\sigma)Z}] \\ &= e^{t\mu} M_Z(t\sigma) \\ &= e^{t\mu} e^{\frac{(t\sigma)^2}{2}} = e^{t\mu + \frac{t^2\sigma^2}{2}} \end{aligned}$$

$$\mathbb{E}[X] = M'(0) = \left(\mu + t\sigma^2\right) e^{t\mu + \frac{t^2\sigma^2}{2}} \Big|_{t=0} = \mu$$

$$\begin{aligned} \mathbb{E}[X^2] &= M''(0) \\ &= \left(\sigma^2 e^{t\mu + \frac{t^2\sigma^2}{2}} + (\mu + t\sigma^2)^2 e^{t\mu + \frac{t^2\sigma^2}{2}}\right)\Bigg|_{t=0} \\ &= \mu^2 + \sigma^2 \end{aligned}$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2$$

Claim 1. If X and Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Proof.

$$\begin{aligned} M_{X+Y}(t) &= \mathbb{E}[e^{t(X+Y)}] \\ &= \mathbb{E}[e^{tX}e^{tY}] \\ \text{(independence)} \qquad &= \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}] \\ &= M_X(t)M_Y(t) \end{aligned}$$

□

Example 5. Let $X \sim \text{Poi}(\lambda_1)$, $Y \sim \text{Poi}(\lambda_2)$ independently. Then

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

which is the moment generating function of $\text{Poi}(\lambda_1+\lambda_2)$.

Therefore $X+Y \sim \text{Poi}(\lambda_1+\lambda_2)$.

Example 6. Let N be a random variable taking non-negative integer values. Let $X_1, X_2, X_3, \ldots$ be i.i.d random variables.

Let $Y = \sum_{i=1}^N X_i$. What is the distribution of Y?

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] \\ &= \mathbb{E}\left[e^{t\sum_{i=1}^N X_i}\right] \end{aligned}$$

To evaluate this, we condition on the value of N:

$$\mathbb{E}[e^{tY}] = \mathbb{E}\left[\mathbb{E}\left[e^{tY}|N\right]\right].$$

Compute

$$\begin{aligned} \mathbb{E}\left[e^{tY}|N=n\right] &= \mathbb{E}\left[e^{t\sum_{i=1}^n X_i}\right] \\ \text{(independence)} \qquad &= \prod_{i=1}^n \mathbb{E}\left[e^{tX_i}\right]^n \\ &= M_X(t)^n \end{aligned}$$

where $X_i \sim X$.

Therefore

$$\begin{aligned} M_Y(t) = \mathbb{E}[e^{tY}] &= \mathbb{E}\left[M_X(t)^N\right] \\ \Rightarrow \mathbb{E}[Y] = M_Y'(0) &= \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}M_X(t)^N\bigg|_{t=0}\right] \\ &= \mathbb{E}[N \cdot \underbrace{M_X(0)^{N-1}}_{-1} \cdot \underbrace{M_X'(0)}_{\mathbb{E}[X]}] \\ &= \mathbb{E}[N\mathbb{E}[X]] = \mathbb{E}[N]\mathbb{E}[X] \end{aligned}$$

CHAPTER 18

Limit Theorems

Let X be a random variable, and let $X_1, X_2, \dots$ be i.i.d random variables with the same distribution as X.

Define $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. What can we say about S_n?

There are two types of results:

- "Law of Large Numbers": We might see that S_n is "close" to $\mathbb{E}[X]$.
- "Central Limit Theorem": S_n is "close" to a normal distribution.

Theorem 1 (Weak Law of Large Numbers). Let $X_1, X_2, \dots$ be i.i.d random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$.

Let $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then for any $\varepsilon > 0$,

$$\mathbb{P}\left(|S_n - \mu| > \varepsilon\right) \to 0 \text{ as } n \to \infty.$$

Proof. It is easy to see that $\mathbb{E}[S_n] = \mu$ by LoE.

Also,

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left(X_i\right) = \frac{\sigma^2}{n}
\end{aligned}$$

Then by Chebychev's inequality,

$$\begin{aligned}
\mathbb{P}\left(|S_n - \mu| > \varepsilon\right) &= \mathbb{P}\left(|S_n - \mathbb{E}[S_n]| > \varepsilon\right) \\
&\le \frac{\mathrm{Var}(S_n)}{\varepsilon^2} \\
&= \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ as } n \to \infty
\end{aligned}$$

This completes the proof. □

Remark 1. In fact, we do not need the variance to be finite. (more complicated proof)

Theorem 2 (Strong Law of Large Numbers). Let $X_1, X_2, \dots$ be i.i.d random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{E}[X_i^4] = K < \infty$.

Then if $S_n = \frac{1}{n}\sum_{i=1}^n X_i$, we have

$$\mathbb{P}(S_n \to \mu \text{ as } n \to \infty) = 1.$$

Proof. Case: $\mu = 0$.

Define $Y_n = \sum_{i=1}^n X_i$. Then

$$Y_n^4 = \left(\sum_{i=1}^n X_i\right)^4 = \sum_{1\le i,j,k,l\le n} X_iX_jX_kX_l$$

$$\mathbb{E}[Y_n^4] = \sum_{i,j,k,l} \mathbb{E}[X_iX_jX_kX_l].$$

Observation. By independence, if one term $\mathbb{E}[X_iX_jX_kX_l]$ has one index of i,j,k,l different from others, then the term $= 0$.

Therefore the only surviving terms are $\mathbb{E}[X_i^4]$ and $\mathbb{E}[X_i^2X_j^2]$. Compute

$$\begin{aligned}\mathbb{E}[Y_n^2] &= \sum_{i=1}^n \mathbb{E}[X_i^4] + \sum_{i<j} 6\mathbb{E}[X_i^2X_j^2]\\ &= n\mathbb{E}[X_1^4] + 6\binom{n}{2}\mathbb{E}[X_1^2]^2\\ &\le n\mathbb{E}[X_1^4] + 6\binom{n}{2}\mathbb{E}[X_1^4]\\ &\le 4n^2K\end{aligned}$$

Let E_n be the event

$$\begin{aligned}\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \ge \frac{1}{\log n}\right\} &= \left\{\left|\frac{1}{n}Y_n\right| \ge \frac{1}{\log n}\right\}\\ &= \left\{|Y_n| \ge \frac{n}{\log n}\right\}\\ &= \left\{Y_n^4 \ge \frac{n^4}{\log^4 n}\right\}.\end{aligned}$$

By Markov's inequality,

$$\begin{aligned}\mathbb{P}(E_n) &= \mathbb{P}\left(Y_n^4 \ge \frac{n^4}{\log^4 n}\right)\\ &\le \frac{\mathbb{E}[Y_n^4]}{\frac{n^4}{\log^4 n}}\\ &\le \frac{4Kn^2}{\frac{n^4}{\log^4 n}} = \frac{4K\log^4 n}{n^2}.\end{aligned}$$

Since $\sum_{n=1}^\infty \mathbb{P}(E_n) \le \sum_{n=1}^\infty \frac{4K\log^4 n}{n^2} < \infty$, by Borel-Cantalli, $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$. i.e., only finitely many of the events E_n occur.

Therefore, with probability 1, there is some $N \in \mathbb{N}$ such that for all $n \geq N$, E_n does not hold. That is, $\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \leq \frac{1}{\log N}$ for all $n \geq N$, and thus $\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \to 0 = \mu$.

Case: general case $\mu \neq 0$.

Let $X_i' = X_i - \mu$. Then $X_1', X_2', \ldots$ are i.i.d with mean 0 and $\mathbb{E}[(X_i')^4] < \infty$.

By the above case, we have $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i' \to 0\right) = 1$, and thus $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu\right) = 1$. $\square$

Remark 2. We only need pairwise independence in WLLN and 4-wise independence in SLLN.

Theorem 3 (Central Limit Theorem). Let $X_1, X_2, \ldots$ be i.i.d random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Then $\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ (pointwisely) tends to the standard normal distribution as $n \to \infty$.

That is, for any $a \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \leq a\right) \to \Phi(a) \text{ as } n \to \infty.$$

For our proof, we will show that

(a) if the moment generating function $M_{X_i}(t)$ exists, so does that of $\frac{X_i}{\sqrt{n}}$,
(b) use the fact that if we have a sequence of random variables Z_i, with mgfs $M_{Z_i}(t)$ and cdfs $F_{Z_i}(z)$, and a random variable Z with mdf $M_Z(t)$ and cdf $F_Z(z)$, if $M_{Z_i}(t) \to M_Z(t)$ for all t, then $F_{Z_i}(t) \to F_Z(t)$ where $F_Z(t)$ is continuous.

Proof. Case: $\mu = 0$, $\sigma^2 = 1$.

Let M be the moment generating function for X_i. Then $M(t) = \mathbb{E}[e^{tX}]$.

Then the moment generating function of $\frac{X_i}{\sqrt{n}}$ is $\mathbb{E}\left[e^{\frac{tX_i}{\sqrt{n}}}\right] = M\left(\frac{t}{\sqrt{n}}\right)$.

Therefore the moment generating function of $\frac{X_1+X_2+\cdots+X_n}{\sqrt{n}}$ is

$$\mathbb{E}\left[e^{t\cdot\frac{X_1+X_2+\cdots+X_n}{\sqrt{n}}}\right] = \mathbb{E}\left[e^{t\frac{X_1}{\sqrt{n}}}e^{t\frac{X_2}{\sqrt{n}}}\cdots e^{t\frac{X_n}{\sqrt{n}}}\right] = M\left(\frac{t}{\sqrt{n}}\right)^n$$

Our goal is to show that

$$M\left(\frac{t}{\sqrt{n}}\right)^n \to M_Z(t)e^{\frac{t^2}{2}} \text{ for all } t,$$

where $M_Z(t) = e^{\frac{t^2}{2}}$ is the moment generating funtion of $Z \sim N(0,1)$.

Taking natural logarithm, it suffices to show that $n \log M\left(\frac{t}{\sqrt{n}}\right) \to \frac{t^2}{2}$.

Define $L(t) = \log M(t)$. We want to show that $nL\left(\frac{t}{\sqrt{n}}\right) = \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} \to \frac{t^2}{2}$.

Note that

$$\begin{aligned}
L(0) &= \log M(0) = \log M(1) = 0 \\
L'(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \log M(t) = \frac{M'(t)}{M(t)} \\
&\Rightarrow L'(0) = \frac{\mu}{1} = 0 \\
L''(0) &= \frac{M''(t)M(t) - M'(t)^2}{M(t)^2} \\
&\Rightarrow L''(0) = \frac{(\sigma^2 + \mu^2)1 - \mu^2}{1} = 1
\end{aligned}$$

Therefore

$$\begin{aligned}
&\text{(L'Hopital)} & \lim_{n\to\infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} &= \lim_{n\to\infty} \frac{\frac{\mathrm{d}}{\mathrm{d}n} L\left(\frac{t}{\sqrt{n}}\right)}{\frac{\mathrm{d}}{\mathrm{d}n} n^{-1}} \\
&& &= \lim_{n\to\infty} \frac{L'\left(\frac{t}{\sqrt{n}}\right) \cdot \frac{\frac{1}{2}t}{n^{\frac{3}{2}}}}{-n^{-2}} \\
&& &= \lim_{n\to\infty} \frac{tL'\left(\frac{t}{\sqrt{n}}\right)}{2n^{-\frac{1}{2}}} \\
&\text{(L'Hopital)} & &= \lim_{n\to\infty} \frac{\frac{\mathrm{d}}{\mathrm{d}n} tL'\left(\frac{t}{\sqrt{n}}\right)}{\frac{\mathrm{d}}{\mathrm{d}n} 2n^{-\frac{1}{2}}} \\
&& &= \lim_{n\to\infty} \frac{L''\left(\frac{t}{\sqrt{n}}\right) \cdot \frac{-\frac{1}{2}t^2}{n^{\frac{3}{2}}}}{2 \cdot -\frac{1}{2} n^{-\frac{3}{2}}} \\
&& &= \lim_{n\to\infty} \frac{t^2 L''\left(\frac{t}{\sqrt{n}}\right)}{2} = \frac{t^2}{2}
\end{aligned}$$

Hence $M\left(\frac{t}{\sqrt{n}}\right) \to e^{\frac{t^2}{2}}$. Thus, by our face, if $\mathbb{E}[X_i] = 0$ and $\operatorname{Var}(X_i) = 1$, for any a,

$$\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \leq a\right) \to \Phi(a) \text{ as } n \to \infty.$$

Case: general case.

Let $Y_i = \frac{X_i - \mu}{\sigma}$. Then $\mathbb{E}[Y_i] = 0$ and $\operatorname{Var}(Y_i) = 1$.

Therefore $\mathbb{P}\left(\frac{Y_1+Y_2+\cdots+Y_n}{\sqrt{n}} \leq a\right) \to \Phi(a)$ as $n \to \infty$ and thus $\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \Phi(a)$ as $n \to \infty$. □

Remark 3. (1) We showed that for any $a \in \mathbb{R}$, $\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \Phi(a)$ as $n \to \infty$. One can show the convergence is uniform in a.

(2) The CLT applies more generally. Let $X_1, X_2, \ldots$ be independent random variables such that

1. there exists $m < \infty$ such that $\mathbb{P}(|X_i| \leq m) = 1$ for all i,
2. $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) = \infty$.

Then
$$\frac{(X_1 - \mathbb{E}[X_1]) + (X_2 - \mathbb{E}[X_2]) + \cdots + (X_n - \mathbb{E}[X_n])}{\sqrt{\sum_{i=1}^{n} \mathrm{Var}(X_i)}}$$
tends to $N(0,1)$ in distribution.

(3) The convergence is typically very fast.

Example 1. Normal approximation to the binomial.

Let $X \sim \mathrm{Bin}(n,p)$ be a sum of n i.i.d Bernoulli variables with $\mathbb{E}[X_i] = p$ and $\mathrm{Var}(X_i) = p(1-p)$.

By the CLT, $\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n-np}{\sqrt{np(1-p)}} \leq a\right) \to \Phi(a)$. That is, X is approximated by $N(np, np(1-p))$.

Example 2. A fair die is rolled 10 times. What is the probability that the sum lies in $[30,40]$?
Solution. Let X_i be the outcome of the i-th roll. Then X_i's are i.i.d random variables with $\mathbb{E}[X_i] = \frac{7}{2}$ and $\mathrm{Var}(X_i) = \frac{35}{12}$.

By the CLT, $\frac{X_1+X_2+\cdots+X_{10}-35}{\sqrt{\frac{350}{12}}}$ tends to $N(0,1)$.

Therefore
$$\begin{aligned}
\mathbb{P}(30 \leq X_1 + \cdots + X_{10} \leq 40) &= \mathbb{P}(29.5 \leq X_1 + \cdots + X_{10} \leq 40.5) \\
&= \mathbb{P}\left(\frac{29.5-35}{\sqrt{\frac{350}{12}}} \leq \frac{X_1 + \cdots + X_{10} - 35}{\sqrt{\frac{350}{12}}} \leq \frac{40.5-35}{\sqrt{\frac{350}{12}}}\right) \\
&\approx \Phi\left(\frac{5.5}{\sqrt{\frac{350}{12}}}\right) - \Phi\left(\frac{-5.5}{\sqrt{\frac{350}{12}}}\right) \\
&= 2\Phi\left(\frac{5.5}{\sqrt{\frac{350}{12}}}\right) - 1 \approx \boxed{0.692}
\end{aligned}$$

CHAPTER 19

The Probabilistic Method in Extremal Combinatorics

What is Comvinatorics? It is the study of “discrete structures”. For example, graphs, set systems, number theory, etc.

What is extremal combinatorics? It asks the question: How big/small can structures be if it has a certain property?

For example, how many edges can an n-vertex graph have if you can draw it without crossing edges?

Let $f(n)$ be $\max\{|E(G)| : G \text{ is an } n\text{-vertex planar graph}\}$. In fact, $f(n) = 3n - 6$.

1. Sum-free Sets

In a game, the foal is to choose a set of integers from $\{1, 2, \ldots, n\}$ as possible without a solution to $x + y = z$.

Q. How large can a sum-free subset of $\{1, 2, \ldots, n\}$ be?

For example, $A = \{1, 3, 5, 7, \ldots\}$ or $A = \left\{\left\lfloor \frac{n}{2} \right\rfloor + 1, \left\lfloor \frac{n}{2} \right\rfloor + 2, \ldots, n\right\}$, $|A| \approx \frac{n}{2}$. This gives a lower bound. We can find an upper bound:

Claim 1. If A is a sum-free subset of $\{1, 2, \ldots, n\}$, then $|A| \leq \frac{n+1}{2}$.

Proof. Let $m = \max A$. Consider the disjoint pairs $\{1, m-1\}$, $\{2, m-2\}$, …, $\left\{\left\lfloor \frac{m}{2} \right\rfloor, \left\lceil \frac{m}{2} \right\rceil\right\}$.

Since A is sum-free, we have at most one element from each pair.

Therefore $A \setminus \{m\} \leq$ number of pairs $= \left\lfloor \frac{m}{2} \right\rfloor$, and thus

$$|A| \leq \left\lfloor \frac{m}{2} \right\rfloor + 1 = \left\lceil \frac{m+1}{2} \right\rceil \leq \left\lceil \frac{n+1}{2} \right\rceil.$$

□

Now we play a REAL game. Given a set $S \subseteq \mathbb{N}$, $|S| = n$. The goal is to find the largest sum-free subset of S.

Definition 1. Let $\sigma(S) = \max\{|A| : A \subseteq S \text{ is sum-free}\}$.

For example, $\sigma(\{1, 2, \ldots, n\}) \approx \frac{n}{2}$, and $\sigma(\{1, 3, 9, 27, \ldots, 3^{n-1}\}) = n$.

In a two-player game, the first player chooses a set S and the second player finds a large sum-free subset A. We want to find a strategy for the first player.

Definition 2. Let $\sigma(n) = \min\{\sigma(S) : S \subseteq \mathbb{N}, |S| = n\}$.

From the above discussion, we have an upper $\sigma(n) \leq \frac{n}{2}$.

Theorem 1 (Erdős,1965). $\sigma(n) \geq \frac{1}{3}(n+1)$.

The probabilistic method. Erdős's brilliant insight: To provide a lower (upper) bound for a max (min) problem, you do not need to construct an example, you only need to prove an example exists.

Consider a random construction. If $\mathbb{P}(\{\text{construction has good properties}\}) > 0$ then such a construction exists.

Proof. Idea: Given a set $S \subseteq \mathbb{N}$, $|S| = n$,

- consider a random subset $A \subseteq S$,
- show that with positive probability,
  - $|A| \geq \frac{1}{3}(n+1)$,
  - A is sum-free.

This will imply $\sigma(n) \geq \frac{1}{3}(n+1)$.

Attempt: Choose every element $s \in S$ to be in A with probability p. Then $|A| \sim \mathrm{Bin}(n,p)$, and thus $\mathbb{E}[|A|] = np$.

Is A sum-free? Let $X = \#\{\text{sums } x+y=z \text{ in } A\}$ and $X = \sum X_{x,y}$ where

$$X_{x,y} = \begin{cases} 1 & \text{if } x, y \in A \text{ and } x+y \in A \\ 0 & \text{otherwise.} \end{cases}$$

If $x \neq y$, then $\mathbb{P}(X_{x,y} = 1) \leq p^3$. If $x = y$, then $\mathbb{P}(X_{x,y} = 1) \leq p^2$. We can compute

$$\mathbb{E}[X] = \sum \mathbb{E}[X_{x,y}] = \sum \mathbb{P}(X_{x,y} = 1) = \sum_{x \neq y} \mathbb{P}(X_{x,y} = 1) + \sum_{x} \mathbb{P}(X_{x,x} = 1) \leq n^2p^3 + np^2.$$

Observation. If $\mathbb{E}[X] \leq 1$, then $\mathbb{P}(X = 0) > 0$.

We want $n^2p^3 + np^2 < 1$, so we need $p < n^{-\frac{2}{3}}$. Then $\mathbb{E}[|A|] \leq np \leq n^{\frac{1}{3}}$, which is too small. We need a different model for a random subset.

Observation. If $q = 3k+2$ is a prime, then $\{k+1, k+2, \ldots, 2k+1\}$ is sum-free in $\mathbb{Z}_q$.

Corollary 1. The projection $\pi : S \to \mathbb{Z}_q$ satisfies that $\pi^{-1}(\{k+1, k+2, \ldots, 2k+1\})$ is a sum-free subset in S.

Let $A = \{s \in S : xs \bmod q \in \{k+1, k+2, \ldots, 2k+1\}\}$, where $x \in \mathbb{Z}_q \setminus \{0\}$ is chosen uniformly at random.

Observation.

$$xs \bmod q = \begin{cases} 0 & \text{if } s \equiv 0 \bmod q \\ \text{uniformly in } \{1, 2, \ldots, q-1\} & \text{otherwise.} \end{cases}$$

Given $s \in S$, $s \not\equiv 0 \bmod q$,

$$\mathbb{P}(xs \bmod q \in \{k+1, k+2, \ldots, 2k+1\}) = \frac{|\{k+1, \ldots, 2k+1\}|}{|\{1, \ldots, 3k+1\}|} > \frac{1}{3}.$$

$$
\begin{aligned}
\mathbb{E}[|A|] &= \mathbb{E}\left[\sum_{s\in S} X_s\right] \qquad \text{where } X_s = 1 \text{ if } s \in A \\
&= \sum_{s\in S} \mathbb{E}\left[X_s\right] = \sum_{s\in S} \mathbb{P}(s \in A) \\
&= \sum_{s\in S} \mathbb{P}(xs \bmod q \in \{k+1,\dots,2k+1\}) \\
&= \sum_{s\in S, s\not\equiv 0 \bmod q} \mathbb{P}(xs \bmod q \in \{k+1,\dots,2k+1\}) \\
&> \frac{1}{3}\left|\{s \in S : s \not\equiv 0 \bmod q\}\right|
\end{aligned}
$$

Choose $q > \max S$. Then $\mathbb{E}[|A|] > \frac{1}{3}|S|$. In particular, there is a choice of x that gives $A \subseteq S$, $|A| > \frac{1}{3}n$. Thus $|A| \geq \frac{1}{3}(n+1)$ and this A is guaranteed to be sum-free. ☐

History:

$$\sigma(n) \geq \frac{1}{3}(n+1) \qquad \text{[Erdős,1965]}$$

$$\sigma(n) \geq \frac{1}{3}(n+2) \qquad \text{[Bourgain,1997]}$$

$$\sigma(n) \leq \frac{1}{2}n$$

$$\vdots$$

$$\sigma(n) \leq \left(\frac{1}{3} + o(1)\right) n \qquad \text{[Eberhard, Green, Manners, 2014]}$$

2. Ramsey Theory

<u>Problem.</u> Given k, what is the largest $R(k)$ such that we can red/blue- color the edges of a complete graph on $R(k)$ vertices without a subset of k vertices where edges are all the same?

History: $R(3) = 5$, $R(4) = 17$, $42 \leq R(5) \leq 45$.

$$R(k) < \infty. \quad \text{Ramsey, 1930}$$

$$c\sqrt{2}^k \overset{\substack{\text{Erdős}\\(1947)}}{\leq} R(k) \overset{\substack{\text{Erdős, Szelens}\\(1935)}}{\leq} 4^k$$

$$2c\sqrt{2}^k \overset{\substack{\text{Spencer}\\(197?)}}{\leq} R(k) \overset{\substack{\text{Campos, Coffin,}\\\text{Moins, Schasrabudhe}\\(2022)}}{\leq} (4-\varepsilon)^k = 3.7\dots^k$$

Theorem 2. There exists $c > 0$ such that for all $k \in \mathbb{N}$, $R(k) \geq c\sqrt{2}^k$.

Proof. Idea: Take N vertices, and color each edge randomly, independently.

For every set S of k vertices, let E_S be the event that all edges in S are the same color.

$$\mathbb{P}(E_S) = 2 \cdot \left(\frac{1}{2}\right)^{\binom{k}{2}}$$

$$\mathbb{P}(\exists \text{ monocromatic } k\text{-set}) = \mathbb{P}\left(\bigcup_S E_S\right) \leq \sum_S \mathbb{P}(E_S)$$

$$\mathbb{P}(\text{coloring is bad}) \leq \binom{N}{k} \cdot 2 \cdot \left(\frac{1}{2}\right)^{\binom{k}{2}} < 2N^k \left(\frac{1}{2}\right)^{\frac{1}{2}k(k-1)}$$

$$2 \cdot \left(\frac{N}{2^{\frac{k-1}{2}}}\right)^k \to 0 \quad \text{as } k \to \infty \text{ if } N < 2^{\frac{k-1}{2}}$$

Therefore $\mathbb{P}(\text{coloring is good}) \to 1$ and thus good coloring exists. $\square$

There are some major open problems:

(a) Determine the correct base of the exponent.
(b) Find a good coloring on 1.00000001^k vertices without using randomness.

Property B. Situation: You are packing for a holiday. To make sure there are no luggage mishaps, you decide to pack two suitcases of clothes.

Different categories of clothing:

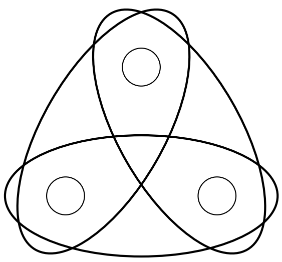
- Tops
- Trousers
- Shorts
- Skirts
- Underwear
- Formal
- Casual
- Beachwear
- Red
- Blue
- Green

Our goal is to assign clothes to suitcases in such a way that every category is represented in both suitcases.

We have one problem: we might have a category with only one item.

Assume each category has at least k items.

More problems: For example,



Definition 3. Given $k \geq 2$, let $m(k) = \min \left\{ \begin{array}{c} \text{no. of sets of size at least } k, \\ \text{such that in any two-coloring of the elements,} \\ \text{one of the sets is monochromatic} \end{array} \right\}$.

For small k:

$$m(1) = 1$$
$$m(2) = 3$$
$$m(3) = 7$$

Proposition 1. $m(k) \leq \binom{2k-1}{k} \approx \frac{c2^{2k}}{\sqrt{k}}$.

Proof. Take $2k-1$ elements, and all subsets of size k.

By the pigeonhole principle, in any 2-coloring, there will be k elements of the same color.

Since all k-sets are in our colleciton, there is a monochromatic set. □

Theorem 3 (Erdős, 1963). For all $k \geq 1$, $m(k) \geq 2^{k-1}$.

Proof. Let f be a collection of $m < 2^{k-1}$ sets of size k. Color the elements uniformly, independently.

For every set $F \in f$ in our collection, let E_F be the (bad) events that it is monochromatic.

$$\mathbb{P}(E_F) = 2 \cdot \left(\frac{1}{2}\right)^k$$

$$\begin{aligned} \mathbb{P}(\text{coloring is bad}) &= \mathbb{P}\left(\bigcup_{F \in f} E_F\right) \\ \text{(union bound)} \qquad &\leq \bigcup_{F \in f} \mathbb{P}(E_F) \\ (m < 2^{k-1}) \qquad &= m2^{1-k} < 1 \end{aligned}$$

Therefore $\mathbb{P}(\text{coloring is good}) > 0$, and thus good coloring exists. □

Theorem 4 (Erdős, 1964). There exists $C > 0$ such that for all $k \geq 2$, $m(k) \leq Ck^2 2^k$.

Proof. Consider a good set of n elements to be determined color.

Idea: Choose m sets at random.

Choose each set S_i from the $\binom{n}{k}$ possibilities uniformly, independently.

The goal is that every subset of $\frac{n}{2}$ elements contains one of our tandom sets, so in every coloring, the more popular color contains a set.

For every set X of $\frac{n}{2}$ elements, define the (bad) event

$$E = \{\text{none of our } m \text{ random sets } S_i \subseteq X\}.$$

$$\mathbb{P}(S_i \subseteq X) = \frac{\binom{\frac{n}{2}}{k}}{\binom{n}{k}} \approx \frac{\frac{\left(\frac{n}{2}\right)^k}{k!}}{\frac{n^k}{k!}} = \frac{1}{2^k}.$$

$$\begin{aligned}\mathbb{P}(E_X) &= \mathbb{P}(\forall i, S_i \not\subseteq X)\\ &= \prod_{i=1}^{m} \mathbb{P}(S_i \subseteq X)\\ &= \left(1 - \frac{\binom{\frac{n}{2}}{k}}{\binom{n}{k}}\right)^m\\ &\approx \left(1 - \frac{1}{2^k}\right)^m \approx e^{-\frac{m}{2^k}}\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{P}\left(\substack{\text{there is a 2-coloring of the elements}\\ \text{with no } S_i \text{ monochromatic}}\right) &\leq \mathbb{P}\left(\bigcup_X E_X\right)\\ &\leq \sum_X \mathbb{P}(E_X) = \binom{n}{\frac{n}{2}}\left(1 - \frac{\binom{\frac{n}{2}}{k}}{\binom{n}{k}}\right)^m\\ &\leq 2^n e^{-\frac{m}{2^k}} < 1 \quad \text{if } m > 2^k n \ln 2\end{aligned}$$

Thus if $m > 2^k n \ln 2$, $\mathbb{P}(\text{exists 2-coloring}) > 0$. Hence there exists a collection of m sets with no 2-coloring, and $m(k) \leq m$.

If we do the colorlatices coefectly, we need $n \geq Ck^2$. □

Now we have $2^{k-1} \leq m(k) \leq Ck^2 2^k$. The lower bound was improved:

Theorem 5 (Beck). $m(k) \geq ck^{\frac{1}{3}} 2^k$.

Theorem 6 (Radhakrishnan, Srinivasan (2000)). $m(k) \geq ck^{\frac{1}{2}} 2^k$.

In 1992, an algorithm is developed:

Theorem 7 (Pluhár (1992)). $m(k) \geq ck^{\frac{1}{4}} 2^k$ via a random greedy algorithm.

Theorem 8 (Cherkashin, Kozak (2015)). $m(k) \geq ck^{\frac{1}{2}} 2^k$ (same algorithm).