Contents

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

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 1.1

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,

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

□

Proposition 1.1

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

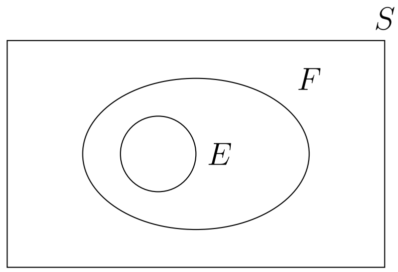
- $E \cap E^c = \varnothing$
- $E \cup E^c = S$

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

Proposition 1.2

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



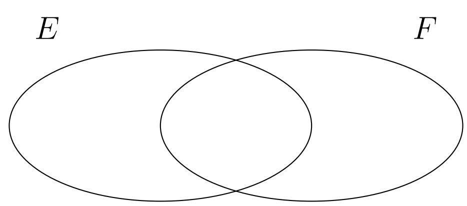
Note that $E \cap F = E$ and $E^c \cap F$ are exclusive events $(E \cap (E^c \cap F) = (E \cap E^c) \cap F = \varnothing \cap F = \varnothing)$, and $(E \cap F) \cup (E^c \cap F) = (E \cup E^c) \cap F = S \cap F = F$.

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 1.3

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

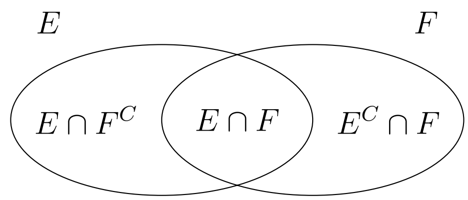
$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F).$$

Corollary 1.2: Union bound

$\mathbb{P}(E \cup F) \leq \mathbb{P}(E) + \mathbb{P}(F)$.

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

Proof. (Prop)



We have unions of exclusive events

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

By Corollary 1.1,

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

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

$\square$

Example 2. Play a game against Real Madrid.

- $\mathbb{P}(\text{Mbappé scores}) = 0.5$
- $\mathbb{P}(\text{Vinicius scores}) = 0.4$

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

Q. $\mathbb{P}(\text{Mbappé or Vinicius scores}) = ?$

Solution. Define events

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

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

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. $\mathbb{P}(\text{roll in divisible by 2, 3, or 5})$?

Solution. Let $E = \{\text{div. by 2}\}$, $F = \{\text{div. by 3}\}$, $G = \{\text{div. by 5}\}$.

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

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

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

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

Inclusion-Exclusion. What is $\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right)$?

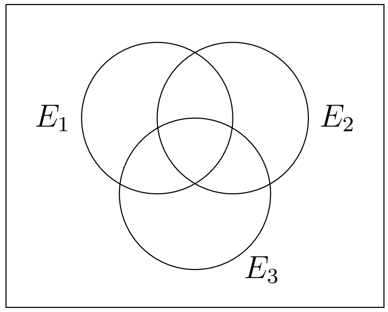
Use induction, we can get

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

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

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

Given every $J \subseteq [n]$, $\mathbb{P} \left(\displaystyle\bigcap_{j \subseteq J} E_j \right)$

$\displaystyle\bigcap_{j \subseteq J} E_j = \dot{\bigcup_{I : J \subseteq I}} E_I$

RHS:

$$\begin{aligned}
&\sum_{r=1}^{n} \sum_{\substack{J \subseteq [n] \\ |J| = r}} (-1)^{r+1} \mathbb{P} \left(\bigcap_{j \subseteq J} E_j \right) \\
&= \sum_{r=1}^{n} \sum_{\substack{J \subseteq [n] \\ |J| = r}} (-1)^{r+1} \mathbb{P} \left(\dot{\bigcup_{I : J \subseteq I}} E_I \right) \\
&= \sum_{r=1}^{n} (-1)^{r+1} \sum_{J \subseteq [n] |J| = r} \sum_{I : J \subseteq I} \mathbb{P}\left(E_I\right) && \text{(mutually exclusive)} \\
&= \sum_{\substack{I \subseteq [n] \\ I \neq \varnothing}} \left(\sum_{r=1}^{n} \sum_{\substack{J \subseteq [n] \\ |J| = r}} (-1)^{r+1} \right) \mathbb{P}(E_I)
\end{aligned}$$

Recall that no. of choices of J, $J \subseteq I$, $|J| = r$ is $\binom{|I|}{r}$.

$$\begin{aligned}
\Rightarrow \sum_{r=1}^{n} \sum_{\substack{J \subseteq [n] \\ |J|=r}} (-1)^{r+1} &= \sum_{r=1}^{n} \binom{|I|}{r} (-1)^{r+1} \\
&= \sum_{r=1}^{|I|} \binom{|I|}{r} (-1)^{r+1} \\
&= \sum_{r=0}^{|I|} \binom{|I|}{r} (-1)^{r+1} - \binom{|I|}{0} (-1)^{0+1} \\
&= -\sum_{r=0}^{|I|} \binom{|I|}{r} (-1)^{r} - (-1) \\
&= -(-1+1)^{|I|} + 1 = 1 \qquad \text{(Binom Thm.)}
\end{aligned}$$

$$\begin{aligned}
\therefore \sum_{r=1}^{n} (-1)^{r+1} &\sum_{\substack{J \subseteq [n] \\ |J|=r}} \sum_{I : J \subseteq I} \mathbb{P}(E_I) \\
&= \sum_{\substack{I \subseteq [n] \\ I \neq \varnothing}} 1 \cdot \mathbb{P}(E_I) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n} E_i \right) \qquad (*)
\end{aligned}$$

$\square$

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 occus, we must place the A♠ directly after the first ace.

$\Rightarrow \mathbb{P}(a) = \dfrac{1}{52}$

(b) Similarly, $\mathbb{P}(b) = \dfrac{1}{52}$.

Example 4. (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. $S = \{\text{bijection from hats to people}\}$, $|S| = n!$.

$E = \{\text{nobody gets their own hat}\}$.

Simpler events: $E_i = \{i\text{th person gets their own hat}\}$

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

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

$\mathbb{P}(E_i) = \dfrac{1}{n}$, $\mathbb{P}(E_i \cap E_j) = \dfrac{(n-2)!}{n!}$,

$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) = \dfrac{(n-r)!}{n!}$

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 2.1

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.

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

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

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 $(*) \to (\dagger)$

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

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

$\square$

3 Continuity of Probability

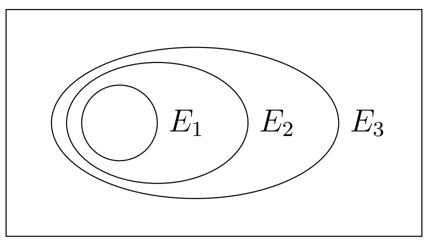
Definition 1. Let $E_1, E_2, E_3, \ldots$ be a sequence of sets. We say the sequence is *increasing* if $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$ and define $\lim\limits_{n\to\infty} E_n = \bigcup\limits_{n=1}^{\infty} E_n$.

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

Proposition 3.1

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

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



Let $F_n = E_n \setminus \left(\bigcup\limits_{i=1}^{n-1} E_i\right)$. Then $F_1, F_2, \ldots$ are mutually exclusive.

$\Rightarrow \bigcup\limits_{i=1}^{n} F_i = E_n = \bigcup\limits_{i=1}^{n} E_i$

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

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 $\left(\lim_{n\to\infty} E_n\right)^C = \lim_{n\to\infty} E_n^C$.

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

□

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

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

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

Theorem 3.1: 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 $\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0$.

Proof.

$$
\begin{aligned}
&\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) \\
&= \mathbb{P}\left(\lim_{n\to\infty} \left(\bigcup_{i=n}^{\infty} E_n\right)\right) && \text{(continuity)}
\end{aligned}
$$

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

$$\leq \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbb{P}(E_i) \to 0 \text{ since } \sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$$

$\square$

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 $\displaystyle\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \sum_{n=1}^{\infty} 2^{-n} = 1 < \infty$, the 1st Borel Cantelli Lemma states $\mathbb{P}(\limsup_{n\to\infty} E_n) = 0$.
$\Rightarrow$ 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 $e_n = \dfrac{\#\{\text{even rolls in first } n \text{ rolls}\}}{n}$.

Fix $\varepsilon > 0$. Let $E_n = \left\{e_n \geq \dfrac{1}{2} + \varepsilon\right\}$.

$S = \{1, 2, 3, 4, 5, 6\}^n$. Count E_n:

(a) Choose how many even rolls r: $\left(\dfrac{1}{2} + \varepsilon\right) n \leq r \leq n$ (Apply the sum rule over choice of r).

(b) Choose which rolls are even: $\dbinom{n}{r}$ 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:

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

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

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 $\alpha = \frac{1}{2}$.

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

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 $n \geq N'$, $e_n > \frac{1}{2} - \varepsilon$.

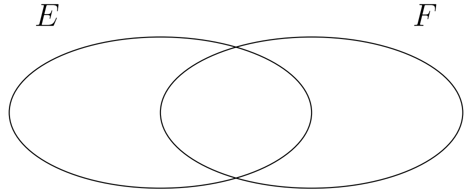
$\Rightarrow$ exists N'' such that for all $n \geq N''$, $\frac{1}{2} - \varepsilon < e_n < \frac{1}{2} + \varepsilon$.

Since $\varepsilon > 0$ is arbitrary, $\lim_{n\to\infty} e_n = \frac{1}{2}$.

4 Conditional Probabilities

Example 5. 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.

$\to$ told that event F ocurrs

$\to$ 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 2. 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 6. (See Example 4.) 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

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

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

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

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

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

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

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

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 7. 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%

Weather forecast: 80% chance of sunny. $\mathbb{P}(\text{swimming})$?

Solution.

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

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 8. (See Example 7.) $\mathbb{P}(\text{sunny}) = 0.8$

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

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

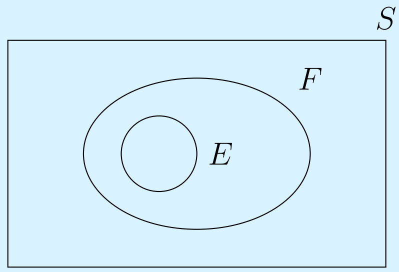
By law of total probability, $\mathbb{P}(\text{bite}) = 0.3138$.

Q. 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 4.1: Bayes' Rule

If we have a partition $S = F_1 \dot\cup F_2 \dot\cup \cdots \dot\cup F_n$ 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, $\mathbb{P}(F_i|E) = \dfrac{\mathbb{P}(F_i \cap E)}{\mathbb{P}(E)}$.

Law of total probability: $\mathbb{P}(E) = \displaystyle\sum_{j=1}^{n} \mathbb{P}(E|F_j)\mathbb{P}(F_j)$

$\mathbb{P}(F_i \cap E) = \mathbb{P}(E \cap F_i) = \mathbb{P}(E|F_i)\mathbb{P}(F_i)$ □

Example 9. 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

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

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

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

Example 10. 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

$\mathbb{P}(\text{guilty})$?

Solution. Let $S = \{\text{all people in the city}\}$, $F_1 = \{\text{guilty}\}$, $F_2 = \{\text{not guilty}\}$.

$\mathbb{P}(F_1) = \dfrac{1}{2500000}$, $\mathbb{P}(F_2) = \dfrac{2499999}{2500000}$.

Let $E = \{\text{match on DNA test}\}$. $\mathbb{P}(E|F_1) = 1$, $\mathbb{P}(E|F_2) = 0.0001$.

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

5 Independent Events

Definition 3. If $\mathbb{P}(E|F) = \mathbb{P}(E)$, then we say E and F are *independent*. Otherwise they are *dependent*.

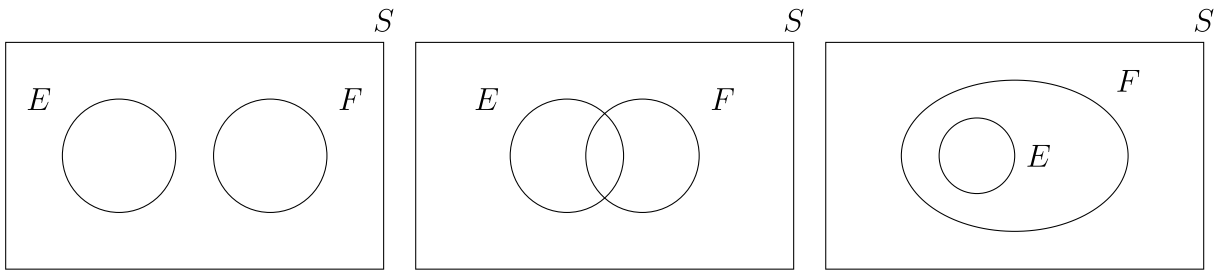
Equivalently, E and F are independent iff

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

Corollary 5.1

Independence is symmetric in E, F.

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



Example 11. $E_1 = \{\text{first roll is a 4}\}$, $E_2 = \{\text{second roll is a 3}\}$

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

Which pairs are independent?

Solution.

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

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

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

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

$\Rightarrow E_1, E_2$ are independent.

$$F_1 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}, \quad \mathbb{P}(F_1) = \frac{5}{36}.$$

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

$\Rightarrow E_1, F_1$ are not independent.

F_1, F_2 not independent. They are disjoint.

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

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

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

Claim. 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

$$E_1, F \text{ are independent, and}$$
$$E_2, F \text{ are independent,}$$

that doesn't mean

$$E_1 \cup E_2, F \text{ are independent, or}$$
$$E_1 \cap E_2, F \text{ are independent.}$$

Definition 4. 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 5. 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 12. 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 $\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0$.

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

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

Theorem 5.1: 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 $\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 1$.

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

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

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

$$\begin{aligned}
\mathbb{P}\left(\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{i=n}^{\infty} E_n\right)\right)^C\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} E_n^C\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{\infty} E_n^C\right) && \text{(continuity)} \\
&= \lim_{n\to\infty} \prod_{i=n}^{\infty} \mathbb{P}(E_i^C) && \text{(independence, *)} \\
&= \lim_{n\to\infty} \prod_{i=n}^{\infty} (1 - \mathbb{P}(E_i)) = \lim_{n\to\infty} 0 = 0
\end{aligned}$$

by convergence test for infinite product ($\lim_{n\to\infty} \prod_{i=n}^{\infty} \mathbb{P}(E_i) = \infty$)

$$\begin{aligned}
(*) \quad & \mathbb{P}\left(\bigcap_{i=1}^{\infty} E_i^C\right) - \mathbb{P}\left(\lim_{N\to\infty} \bigcap_{i=n}^{N} E_i^C\right) \\
&= \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{N} E_i^C\right) && \text{(continuity)} \\
&= \lim_{N\to\infty} \prod_{i=n}^{N} \mathbb{P}(E_i^C) && \text{(independence)} \\
&= \prod_{i=n}^{\infty} \mathbb{P}(E_i^C)
\end{aligned}$$

□

6 Discrete Random Variables

6.1 Discrete Random Variable

Definition 6. 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 13. 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 $\mathbb{P}(X = 0)$:
$\mathbb{P}(X = 0) = \mathbb{P}(F_1 \cap F_2 \cap \cdots \cap F_5)$, $F_i = \{\text{get } i\text{th question wrong}\}$. $\mathbb{P}(F_i) = \dfrac{3}{4}$.
independence$\Rightarrow \mathbb{P}\left(\bigcap_{i=1}^{5} F_i\right) = \prod_{i=1}^{5} \mathbb{P}(F_i) = \left(\dfrac{3}{4}\right)^5$.
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 14. Promotion: n different types of prizes

each attempt $\to$ get a uniformly random prize, independent of previous attempt.

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

Solution. Let $S = \{(s_1, s_2, s_3, \dots) \mid 1 \le s_i \le n\}$, and

$X((s_1, s_2, s_3, \dots)) = \min\{t \mid (s_1, s_2, s_3, \dots) \text{ has all numbers from 1 to } n\}$.

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

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

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

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

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

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

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

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

Therefore

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

6.2 Expectation

Definition 7. 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 15. (See Example 13.) 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 6.1: 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. **Claim.** $\mathbb{E}[X] = \sum_{s \in S} X(s)\mathbb{P}(s)$.

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

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

□

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

$$= \sum_{i=1}^{n} \alpha_i \sum_{x \in S} X_i(s)\mathbb{P}(s)$$
$$= \sum_{i=1}^{n} \alpha_i \mathbb{E}[x_i]$$

□

Example 16. (See Example 13.) 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

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

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 17. (See Example 13.) 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, $0 \le l \le 10$.

$\Rightarrow k - l$ 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 6.1: Markov's Inequality

If X is a discrete random variable taking nonnegative values, then for any $t \in \mathbb{R}_{>0}$,

$$\mathbb{P}(X \ge t) \le \frac{\mathbb{E}[X]}{t}.$$

Remark.

(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.

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

$\square$

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?

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

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

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

6.3 Variance

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

Definition 8. The *variance* of a random variable X with expectation μ is

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

Proposition 6.1

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

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

Theorem 6.2: 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,

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

□

Corollary 6.1

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

Remark. 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?

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

7 Discrete Distributions

7.1 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 $\operatorname{Bin}(n, p)$.

Distribution: The possible values are $0, 1, 2, \ldots, n$. 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. 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 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 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.

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

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

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. 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 $\mathbb{P}(|X - np| \geq t) \leq \dfrac{np(1-p)}{t^2}$.
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) $\mathbb{E}[X^2] = \underbrace{\mathbb{E}[X(X-1)]}_{\sum_k k(k-1p(k))} + \mathbb{E}[X]$.

7.2 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 $\dfrac{\lambda}{n}$ earthquakes in each interval

Assumption:

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

The number of earthquakes $\sim$ Poi$(n, \frac{\lambda}{n})$.

Distribution:

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

Take $n \to \infty$,

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

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

Therefore the Possion distribution with parameter $\lambda > 0$, Poi(λ) 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. Poi(λ) is a good approximation for 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

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

$$= \sum_{k=0}^{\infty} \frac{\lambda^{k+2} e^{-\lambda}}{k!} = \lambda^2$$

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 $\simeq$ Bin$(n, \frac{\lambda}{n})$, which has expectation $np = \lambda$ and variance $np(1-p) = n \cdot \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) \simeq \lambda$.

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 $j \neq i$, $\mathbb{P}(E_i | E_j) \simeq p_i$,

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

Example 19. (See Example 4.) 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 $\simeq$ Poi(1).

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

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

Example 20. Toss a fair coin n times. Let L_n denote the length of longest sequence of consecutive heads.

$$E = \{\text{there is a sequence of } k \text{ heads in a row}\}$$

$$= \{L_n \geq k\}$$

$$= \bigcup_{i=1}^{n-k+1} E_i, \text{ where } E_i = \{\text{tosses } i, i+1, \ldots, i+k-1 \text{ are all heads}\}$$

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

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

but $\mathbb{P}(E_i|E_{i-1}) = \dfrac{1}{2}$. 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

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

Then

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

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. 😊

$\Rightarrow$ 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,

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

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

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 $k \simeq \log_2 n - 2$.

7.3 Geometric Distribution

Setting:

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

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

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

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

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

Example 21. 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.

<u>Q1.</u> What is your expected profit/loss?

Solution. Let X = profit. Then

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

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

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

Solution.

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

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

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

<u>Q3.</u> 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 22. 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 $X_1+X_2+\cdots+X_n$.

$$X_1=1 \qquad \text{(first coupon is always new)}$$

$$X_2\sim\text{Geom}\left(\frac{n-1}{n}\right)$$

$\rightarrow$ each coupon is independent

$\rightarrow$ probability of being new $=\dfrac{n-1}{n}$

$\rightarrow$ repeat until we get a new one

$$X_i\sim\text{Geom}\left(\frac{n-i+1}{n}\right)$$

Therefore

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

Calculate the variance of $\mathrm{Geom}(p)$:

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

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

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

Example 23. Estimate $X =$ the number of dice rolls until the first 6.

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

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

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

7.4 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 $\text{Geom}(p)$.

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

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

Hypergeometric Distribution.

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

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

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 food.

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

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] \qquad \text{(LoE)}$$

$$= \sum_{i=1}^{n} \mathbb{P}(X_i = 1)$$
$$= \sum_{i=1}^{n} \mathbb{P}(i\text{-the ball is good})$$

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

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	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
$\mathrm{NB}(r,p)$	number of trials needed, each is independent with success probability p, until r-th success	$\dfrac{r(1-p)}{p}$	$\dfrac{r(1-p)}{p^2}$
$\mathrm{Hypergeometric}(N,m,n)$	N outcomes, m of which are good, select n without replacement, number of good outcomes	$\dfrac{nm}{N}$	$\dfrac{nK(N-K)(N-n)}{N^2(N-1)}$

8 Continuous Random Variables

8.1 Cumulative Distribution Function

Definition 10. 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 $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ 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

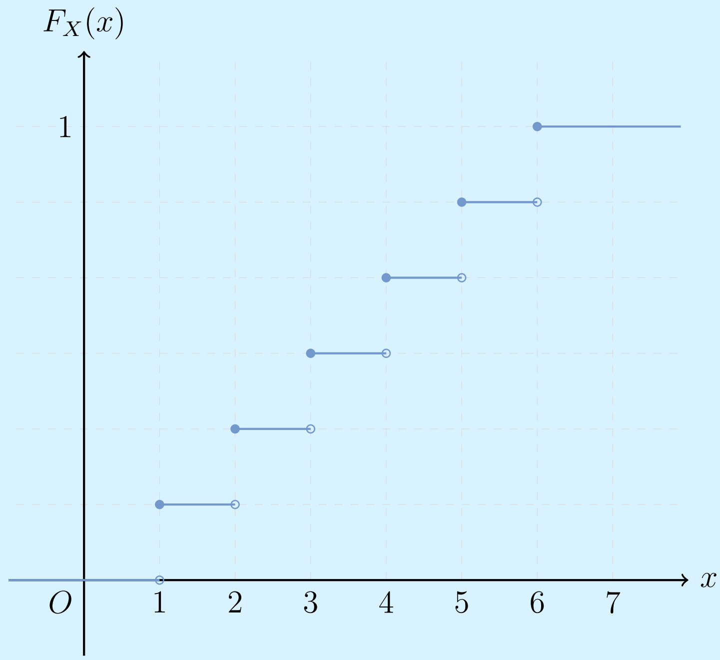
$$\lim_{x\to\infty} \mathbb{P}(\{X \leq x\}) \overset{\text{continuity}}{=} \mathbb{P}\left(\bigcup_{x\to\infty} \{X \leq x\}\right) = \mathbb{P}(X \in \mathbb{R}) = 1.$$

- $\lim_{x\to-\infty} F_X(x) = 0$.
- 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. 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 24. Let X be the outcome of a roll of a die. Then the plot of its cdf F_X is shown below:



8.2 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 11. 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

$$\mathbb{P}(x - \varepsilon \leq X \leq x + \varepsilon) = \mathbb{P}(x + \varepsilon) - \mathbb{P}(x - \varepsilon)$$

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

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

Since

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

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

Example 25. Let

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

for some constant C.

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

Solution.

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

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

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

8.3 Expectation

In the discrete setting, $\mathbb{E}[X] = \displaystyle\sum_i x_i \cdot \mathbb{P}(X = x_i)$.

For a continuous random variable, observe

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

Therefore we define

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

Example 26. Let X have pdf

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

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 27. 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).

8.4 Variance

We define the variance as before

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

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

Example 28. Let

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

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

Compute

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

Example 29. Game show

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

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

<u>Q.</u> 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	$z \leq x < y$	$x < z \leq y$	$x < y \leq z$
\$$x$	\$$x$	\$$y$	\$$y$
\$$y$	\$$y$	\$$y$	\$$x$

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

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

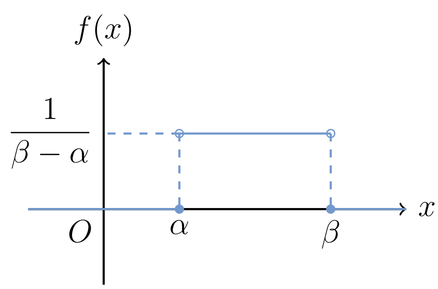
9 Continuous Distributions

9.1 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 $\text{Unif}(\alpha, \beta)$. The pdf is $f(x) = \dfrac{1}{\beta - \alpha}\chi_{(\alpha,\beta)}$.



Statistics. The cummulative distribution function is

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

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

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

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

Example 30. 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 $X \sim \mathrm{Unif}(0, 30)$.

$$\mathbb{P}(\text{wait} > 5 \text{ minutes}) = \mathbb{P}(X \in (0, 10) \cup (15, 25))$$

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

(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. Y has the same distribution as X.

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

9.2 Exponential Distribution

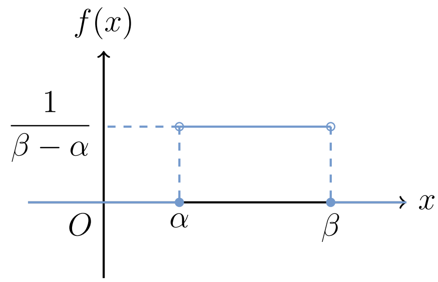
Let $X \sim \text{Exp}(\lambda)$ with $\lambda > 0$. The pdf is given by

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

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



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

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

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

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 12.** 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.

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

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

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

Example 31. 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 32. 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 $\min(X, Y)$.

$$\begin{aligned}
\mathbb{P}(\min(X,Y) \le t) &= 1 - \mathbb{P}(\min(X,Y) > t) \\
&= 1 - \mathbb{P}(\{X > t\} \cap \{Y > t\})
\end{aligned}$$

$$= 1 - \mathbb{P}(X > t)\mathbb{P}(Y > t)$$
$$= 1 - e^{-\frac{2}{15}t}e^{-\frac{1}{15}t} = 1 - e^{-\frac{3}{15}t}$$

Magically, $\min(X, Y) \sim \text{Exp}\left(\frac{3}{15}\right)$!

Therefore $\mathbb{E}[\min(X, Y)] = \frac{15}{3} = \boxed{5}$.

Generally, If $X_i \sim \text{Exp}(\lambda_i)$, then $\min(X_i)_{1 \leq i \leq n} \sim \text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right)$.