Probability Theory

Raymond Bian

January 11, 2024

Contents

Lecture 1: Intro to Probability

1 Basics of Probability

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

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

1.1 Basics of Set Theory

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

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

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

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

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

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

Definition 7. The **cartesian** product of two sets A and B denoted $A \times B$ is

$$\{(a, b) \colon a \in A \wedge b \in B\}.$$

Then, $|A \times B| = |A| \cdot |B|$.

1.2 Back to Probability

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

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

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

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

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

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

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

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

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

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

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

1.3 Counting

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

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

Proof. You have $6 \cdot 2 = 12$ outcomes.

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

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

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

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

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

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

Proof. There are $10!$ possible rankings.

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

Proof. There are $6! \cdot 4!$ possible outcomes.

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

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

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

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

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

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

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

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

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

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

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

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

1.4 Back to Probability Again

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

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

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

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

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

such that

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

More generally,

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

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

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

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

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

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

1.5 Measure Theory

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

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

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

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

Lecture 2: More Probabiliy

1.6 Properties of Event Spaces

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

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

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

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

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

□

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

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

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

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

1.7 Examples

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

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

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

Proof.

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

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

Proof. Everything without $(r,r),(b,b),(g,g)$.

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

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

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

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

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

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

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

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

1.8 Conditional Probability

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lemma 2.

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

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

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

Proof.

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