Thinking like a Frequentist

Student

January 2026

This page is intentionally left blank.

Preface

When I was learning Probability and Statistics (ProbStat for short), I didn't really understand why things were distributed *normally*. Were they *normal* because they were, or were *we just assuming "Everything is normal for simplicity"*? Sometimes, I felt we were overusing this term and in many situations, we were separating our theory from the actual data due to the intital assumption of *normal* distribution.
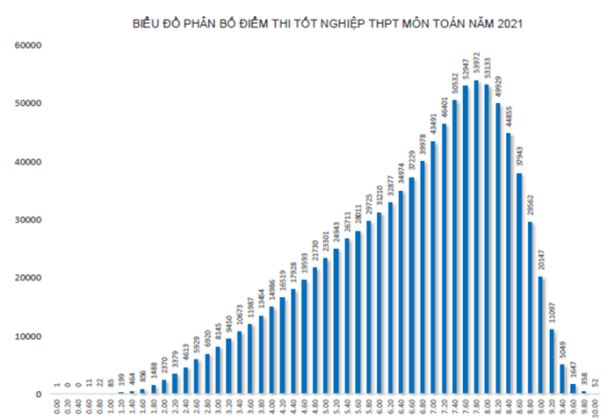

Figure 1: Distribution of Math scores in the 2021 National Entrance Exam

As you can see here, this graph **doesn't really resemble a bell-shaped curve**; and clearly, this data is not distributed normally. In fact, not much real-world statistical data fits a bell-shaped curve. But why are we still using it?

In my opinion (as a junior CE student), I think the main reason is due to **Central Limit Theorem** (CLT for short), every sample in the same population (regradless of how the data is distributed) has a **mean** that converges to a *normal distribution*, or **bell-shaped curve**. This topic will be explained in detail in **Chapter 8: Fundamentals of Statistics**.

This book approaches the fundamentals of probability and statistics in a very mathematically rigourous way to ensure maximum accuracy. I encourage readers to prove all of theorems, corollaries; and you should create your own examples for each theorem to gain a deeper understading of their origins.

—Student—

Acknowledgement

First of all, I want to thank you for reading this book. Although I am not a native English speaker, but I truly enjoy writing in English. I have tried my best to express my ideas in English, but minor grammatical or spelling errors are unavoidable. If you find any of them, I would be very happy to receive your feedback via my email address: `tinvu1309@gmail.com`

Secondly, I am extremly grateful and would like to express my thanks to the authors of the textbook "Probability and Statistics for Engineers and Scientists", 9th edition. This book truly saved my student life.

Finally, the idea of writing this book was inspired by Prof. Steve Brunton lectures on YouTube. You should check out his videos too!

—Student—

Contents

List of Figures

List of notations

Since much of my work is handwritten, so I've modified some commonly used notations for probability distribution funcions using curved script for convenience. You should notice that my conventions are not the international standards.

1. $\mathscr{I}$: Bernoulli distribution

2. $\mathscr{B}$: Binomial distribution

3. $\mathscr{B}^*$: Negative binomial distribution

4. $\mathscr{H}$: Hypergeometric distribution

5. $\mathscr{G}^*$: Geometric distribution

6. $\mathscr{P}$: Poisson distribution

7. $\mathscr{N}$: Normal distribution

8. $\mathscr{G}$: Gamma distribution

9. $\mathscr{E}$: Exponential distribution

10. $\mathscr{C}$: Chi-squared distributon

11. $\mathscr{T}$: t-distribution (or Student distribution)

Chapter 1

Introduction to Probability and Statistics

1.1 Tossing a coin

Imagine you have a coin, like this one. It's an ordinary coin that can be found everywhere.

Figure 1.1: Tossing a coin

Everyone knows that the chance of heads (H) appearing after each toss is 50%, no one even doubts that. So, the **probability** of an **event** that heads appearing is 50%. But because you're a very curious person, you don't easily accept the truth like others, so you will find a way to test it. The fact "The probability of heads appearing is 50%" will be tested, and it can be called a **hypothesis**.

The simplest way to test your hypothesis is tossing a coin many times. You might toss a coin $N = 100$ times and see headss appear $n = 40$ times, so your actual probability is:

$$p = \frac{n}{N} = 0.4$$

But your previous assumtion (or hypothesis) states that the ideal probability is:

$$\hat{p} = 0.5$$

Is there anything wrong here? Dissatisfied, you continue your experiment. This time, you toss a coin $N = 500$ times and see headss appear $n = 270$ times, so the new actual probability is:

$$p = \frac{n}{N} = 0.54$$

This time the result is closer with intial hypothesis $\hat{p}$, but your hand now must be very tired after tossing a coin 600 times. So do you think you have **enough** evidence to conclude that if $N \to +\infty$, then $p \to \hat{p}$? If so, your assumption is correct because *can't be refuted*; and if not, our common sense might be wrong because the evidence *strongly refutes it.*

Another strategy to toss a coin is fixing and dividing N. Instead of tossing $N = 600$ times and only getting the final p value, you can divide large number of coin tosses $N = 600$ to smaller tosses $N_1 = N_2 = \cdots = N_6 = 100$ (but don't too small, at least each $N_i > 30$), and get 6 values of p_i. Suppose that you would get:

p_1	p_2	p_3	p_4	p_5	p_6
0.43	0.46	0.56	0.53	0.49	0.51

You might see that the value $\hat{p} = 0.5$ too idealistic to occur in our experiment, and p constantly changes. Therefore instead of determining the **exact value** of p, we **predict** that the p value lies inside a **closed interval** with a **certainty** of $(1-\alpha)100\%$. For example, we can confidently conclude that the value of p lies within the range (0.48,0.51) with 95% accuracy.

1.2   Weather forecasting

You don't need to know anything about geography to predict what the weather will be tomorrow. Everything you need is just knowledge about Poisson, exponential distribution and the **average number** of rainy days per month where you are living.
For example, in December there are **average** 5 rainy days. What are the probabilities of:

1. There will be 3 rainy days this month.

2. Tomorrow will be a rainy day, if today is the 10th and you haven't seen any rain since the beginning of the month.

This problem will be covered in detail in **Chapter 5: Discrete Random Variables** and **Chapter 6: Continuous Random Variables**, but if you do have experience with random variables, you can try solving it!
The answer for the first question is:

$$P(X = 3) = \frac{e^{-5}5^3}{3!} \approx 14.03\%$$

The answer for the second question is:

$$P(X < 11|X > 10) = 1 - P(X > 11|X > 10) = 1 - e^{\frac{-5}{31}} \approx 14.89\%$$

Since 14.89% is quite low, so tomorrow you don't have to bring an umbrella.

1.3   Relationship between Probability and Statistics

The formal definitions of **probability** and **statistics** will be discussed later in **Chapter 2: Fundamentals of Probability** and **Chapter 8: Fundamentals of Statistics**, but now let's focus on the general model below:
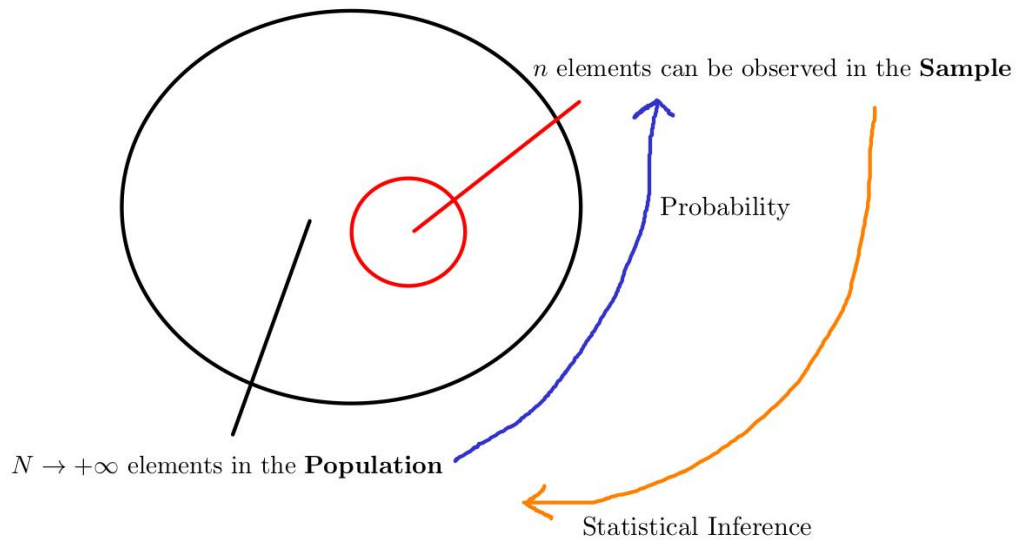
Figure 1.2: General model of simple ProbStat problems

You want to know some attributes of a very large **population** such as the probabilites, means,$\cdots$ of the quantities. But there's no way you can observe the entire population (in many cases, $N \to +\infty$) , so you have to take some **samples** (n elements) and use **probability rules** to process them. After that, you can try applying **statistical inference rules** to draw conclusions about the original population. That's how it works!

Referring to the previous example, you want to check if the probability of heads appearing is $\hat{p} = 50\%$, so you toss a coin multiple times. But since you can only toss a coin for 600 times consecutively, so you conclude based on your final result $p = 0.54$ (or closed interval with a certainty) that might be if $N \to +\infty$ then $p \to \hat{p} = 0.5$. You took the **samples** ($N = 600$ or $N_i = 100$), used **probability rule** to calculate the actual p value, and then applied **statistical inference rules** (hypothesis or closed interval) to draw conclusions respectively.

Chapter 2

Fundamentals of Probability

2.1 Sample Space and Events

2.1.1 Sample Space

If you toss a coin, you will see that there are 2 possible outcomes: heads and tails, denoted by capital letters H and T; or if you roll a die, you will see that there are 6 possible outcomes, from 1 to 6. Tossing a coin, or rolling a die are typical examples of **experiments**.

Definition 2.1.1. *An **experiment** is the process that generates a set of outcomes (or data).*

All of possible outcomes are collected into a single set.

$$S_1 = \{H, T\}$$

$$S_2 = \{1, 2, 3, 4, 5, 6\}$$

Definition 2.1.2. *The **sample space** is the set of all possible outcomes of an **experiment**.*

The sample space is usually represented by the symbol S or Ω. You can easily see that S is not always a countable or finite set. For instance, S_3 is the set of all random numbers you can choose within the range $(0, 1)$:

$$S_3 = \{x \mid 0 < x < 1\}$$

S_4 is the set of all number of coin tosses until first heads appear:

$$S_4 = \{1, 2, 3, \cdots\}$$

It's possible that you toss a coin forever, and heads never appear.

Definition 2.1.3. *A **sample point** is a single outcome of the sample space S.*

S_1, S_2 have 2 and 6 sample points respectively, while S_3, S_4 have infinite sample points.

2.1.2 Events

For any given experiment, we are often interested in the occurence of certain **events** rather than a specific element (or **sample point**) in the sample space. For example, in the die roll experiment, you may want to know when the outcome is an even number. This will happen if the result is an element of the subset E_2 of the sample space S_2:

$$E_2 = \{2, 4, 6\}$$

10

Definition 2.1.4. *An **event** is the subset of a sample space.*

Events are always denoted by capital letters like $A, B, C, \cdots$. Similarly, an event is not always a countable or finite subset.

The **complement** of an event E_2 with respect to S_2 is the set of all *odd outcomes*, and can be represented as follows:
$$\overline{E_2} = \{1, 3, 5\}$$

There are many ways to denote the **complement** of an event: $\overline{E}, E', E^c$, but in this book I choose overline notation for convenience and make it easy to relate with Boolean algebra.

Definition 2.1.5. *The **complement** of an event E with respect to S is the subset of all elements of S that are not in E, and can be denoted by the symbol $\overline{E}$.*

Applying set theory, we can perform many set operations like joint, disjoint, union,$\cdots$

Definition 2.1.6. *The **intersection** of two events A and B, denoted by the symbol $A \cap B$ is the event containing all elements that are common to A and B.*

For example, if A and B are the subset of S_3 and defined as:
$$A = \{x \mid 0 < x < 0.7\}$$
$$B = \{x \mid 0.2 < x < 0.9\}$$
$$\Rightarrow A \cap B = \{x \mid 0.2 < x < 0.7\}$$

Definition 2.1.7. *Two events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \varnothing$, that is A and B have nothing in common.*

For example, E_2 and $\overline{E_2}$ are mutually exclusive.

Definition 2.1.8. *The **union** of the two events A and B, denoted by the symbol $A \cup B$ is the event containing all of the elements that belong to A or B or both.*

For example, $A \cup B = \{x \mid 0 < x < 0.9\}$, and $(A \cup B) \subset S_3$. Note that $E_2 \cup \overline{E_2} = S_2$ is an useful result and can be generalized to corollary below:

Corollary 2.1.0.1. *If A is an event with respect to sample space S, then $A \cup \overline{A} = S$*

De Morgan's laws can also be applied to set theory.

Corollary 2.1.0.2. *If A and B are events with respect to sample space S, then*
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

These results above are really useful in many cases, but you don't have prove them since they are pretty simple. Proving them rigourously is not the focus of Probability and Statistics book.

2.2   Counting Sample Points

In this section, we develop some *counting techniques* to count the number of points in the sample space S and its event subset E without actually listing each elements. These techniques play an important role in solving some simple probability problems. Notice that these techniques can only be applied when your sample space is finite and countable.

Obviously, you can't count the number of elements of S_3 and S_4, since they are infinite and uncountable sets.

2.2.1 Rule of Product

You toss a pair of coins. How many sample points are there in the sample space? First you can try to list all of the possible outcomes:

$$S = \{HH, HT, TH, TT\}$$

There are total 4 sample points in S. But listing all of the elements might be not so clever idea, so you use **rule of product**.

The first coin can land heads or tails, so can the second coin. We multiply the number of possible outcomes of the two coins:

$$n_1 n_2 = 2.2 = 4 \text{ (possible outcomes)}$$

Definition 2.2.1. *If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so fourth, the the sequence of k operations can be performed in:*

$$\prod_{i=1}^{k} n_i \text{ (ways)}$$
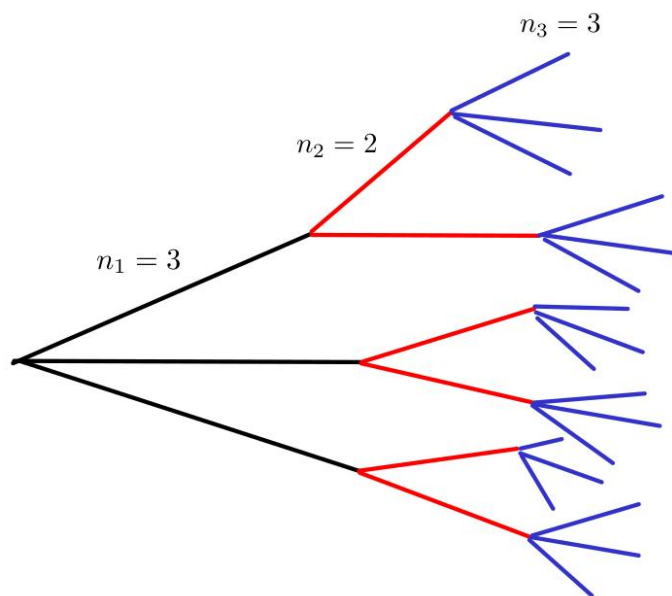
There are a total $n_1 n_2 n_3 = 18$ ways to perform this operation

Figure 2.1: Tree diagram for rule of product

2.2.2 Permutations

Definition 2.2.2. *A **permutation** is an arrangement of **all** or **part** of a set of objects.*

For instance, the number of permutations of n distinct objects can be counted by using rule of product:

$$n(n-1)(n-2)\cdots 3.2.1$$

We introduce a new notation for such a number.

Definition 2.2.3. *For any non-negative integer n, n!, called "n factorial", is defined as:*

$$n! = n(n-1)(n-2)\cdots 3.2.1$$

with special case $0! = 1$

Theorem 2.2.1. *The number of **permutations** of n distinct objects is n!*

In general, the number of permutations of n distinct objects taken r at a time can also be counted by using rule of product:

$$n(n-1)(n-2)\cdots(n-r+2)(n-r+1)$$

This product can be represented by the new symbol nPr.

Theorem 2.2.2. *The number of **permutations** of n distinct objects taken r at a time is:*

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

But how about our n objects are not distinct? Assume that n_1 are of one kind, n_2 are of a second kind,$\cdots$ and n_k of a kth kind.

Theorem 2.2.3. *The number of **permutations** of n things of which n_1 are of one kind, n_2 of a second kind, and so forth is:*

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

with $\sum_{i=1}^{k} n_i = n$.

For example, from the digits 1 to 9, we can form 9! numbers made up of 9 distinct digits, or we can form $9P5$ five-digit numbers such that all the digits are different. If we allow repetition, five-digits numbers are now formed by 2 digits 1, 2 digits 2 and 1 digit 3, then the number of permutation that satisfy is:

$$\frac{5!}{2!2!1!} = 30$$

2.2.3 Combinations

Consider another problem, now you have a set of n elements. How many ways you can partition it into k cells with n_1 elements in the first cell, n_2 elements in the second cell, and so forth? Coincidentally, the equation in **Theorem 2.2.3** appears again.

Theorem 2.2.4. *The number of ways of partitioning a set of n objects into k cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is:*

$$\binom{n}{n_1, n_2, \cdots, n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$$

with $\sum_{i=1}^{k} n_i = n$.

In many problems, we are interested in the number of ways of selecting k objects from n without regrad to order. These selections are called **combinations**. It's not too hard to realize that a **combination** is just a partition with 2 cells, one cell containing the k objects and the other containing the $(n-k)$ objects.

$$\binom{n}{k, n-k} \text{ is often shortened to } \binom{n}{k}$$

Definition 2.2.4. *A **combination** is a selection of items from a set that has distinct elements, such that **the order of selection does not matter**.*

Theorem 2.2.5. *The number of combinations of n distinct objects taken k at a time is:*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

For example, the number of subsets of 3 elements that can be obtained from an original set of 10 elements is:

$$\binom{10}{3} = 120$$

For the rest of this book, we mainly focus on **combinations**, and derive many probability distribution functions based on them. Like set theory, we don't go deeply into **counting techniques** since they are not the main concern of ProbStat.

2.3 Probability of an Event

Everyone knows the basic idea of probability. If I toss a **fair** coin once, and I want to know the probability of getting heads; how can I determine it? Strictly, I have to define S is the sample space of this experiment and A is the event "Getting heads after one toss".

$$S = \{H, T\}$$

$$A = \{H\}$$

So, the probability of event A occuring is:

$$P(A) = \frac{\text{number of sample points inside } A}{\text{number of sample points inside } S} = \frac{1}{2}$$

Very easy and intuitive. But how about tossing 4 **fair** coins once, and determining the probability of getting 2 heads? Now sample space S and its event set can be represented as:

$$S = \{HHHH, HHHT, HHTH, \cdots\}$$

$$A = \{TTHH, THTH, \cdots\}$$

Now we change our strategy using **counting techniques**:

$$P(A) = \frac{\text{number of sample points inside } A}{\text{number of sample points inside } S} = \frac{\binom{4}{2}}{2^4} = \frac{3}{8} = 0.375$$

In fact, the chance of getting heads is slightly greater than tails (you know, because coins are asymmetrical). Assume that probability of heads appearing is 60%, and tails appearing is 40%. Since the role of **sample points** inside set S and subset A are not **equal** anymore, so we can't use **counting techniques** blindly.

$$P(A) = \binom{4}{2} 0.6^2 0.4^2 = 0.3456$$

Coin tossing is a typical example of **Bernoulli trial**, and the experiment tossing unfair coins is **Bernoulli process**. The probability $P(A)$ can be calculated by using **Binomial distribution** formula. You don't have to worry about these terms, we will cover them in **Chapter 5: Discrete Random Variables** very carefully.

Consider another problem, what's the probability of getting 0.7 when randomly choosing a number (assume that the role of every numbers are equal) inside the interval $(0,1)$?

$$S_3 = \{x \mid 0 < x < 1\}$$

$$A = \{0.7\}$$

$$P(A) = \frac{\text{number of sample points inside } A}{\text{number of sample points inside } S} = \frac{1}{+\infty} = 0 \ (?????)$$

Since $P(A) = 0$, so may we conclude 0.7 will never be chosen? Absolutely incorrect, even in common thinking. From previous examples, now we see the **limitation** of our "common definition" of probability in real life. Mathematically, our definition is just a very special case of the formal one.

Definition 2.3.1. *The **probability** of an event A is the sum of the weights (or probabilities) of all sample points in A. Therefore:*

$$0 \le P(A) \le 1, \quad P(\varnothing) = 0, \quad P(S) = 1$$

*If A and B are **mutually exclusive (or disjoint)**, then $P(A \cup B) = P(A) + P(B)$*

This definition can also be called Kolmogorov's axioms. An intuitive way to understand it is sketching a sample space with some events inside.
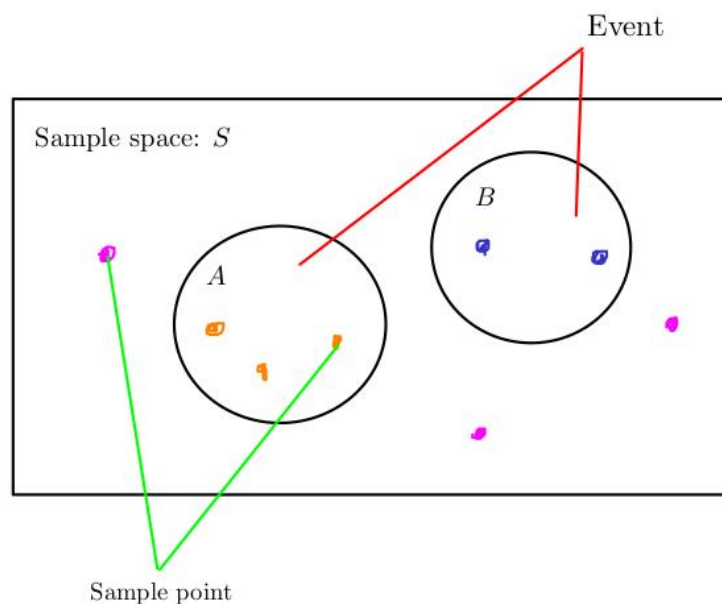

Figure 2.2: Visualize definition of probability

Literally, probability is just a number that describes how likely an event can occur. I often relate it with "weight" quantity. As you can see in the sample space, I can assign arbitrarily the "weights" (or probabilites) of sample points as follows:

- Each blue point is 0.1

- Each orange point is 0.2

- Each pink point is 0.066

Using **Definition 2.3.1**, now we can obtain these resuls:

$$P(A) = 3.0.2 = 0.6$$

$$P(B) = 2.0.1 = 0.2$$

$$P(A \cup B) = P(A) + P(B) = 0.6 + 0.2 = 0.8$$

$$P(\overline{A \cup B}) = 3.0.066 \approx 0.2$$

$$P(S) = 2.0.1 + 3.0.2 + 3.0.066 \approx 1$$

You should verify that our "special definition" of probabilty above satisfies the axioms of probability, and can be formalized by a theorem.

Theorem 2.3.1. *If an experiment can result in any one of N different **equally** likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is:*

$$P(A) = \frac{n}{N}$$

Logically, you can view probability as a mapping from a set to a closed interval $[0, 1]$, then you can freely define the mapping P by yourself as long as it satisfies Kolmogorov's axioms.

$$P : \text{Set} \to [0, 1]$$

$$A \xrightarrow{P} P(A)$$

By applying set theory, we can derive several extremely useful theorems and corollaries:

Theorem 2.3.2. *If A and B are two events, then:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Theorem 2.3.3. *If A is an event of sample space S, then:*

$$P(S) = P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$$

Corollary 2.3.3.1. *If A is an event of sample space S, then:*

$$P(\overline{A}) = 1 - P(A)$$

Again, De Morgan's laws can also be applied:

Corollary 2.3.3.2. *If A and B are events of sample space S, then:*

$$P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$$
$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$

2.4   Conditional Probability

2.4.1   Conditional Probability

Imagine you now have a perfectly fair coin. Obviously if you define events A and B are heads and tails appearing after one toss, respectively, you will conclude:

$$P(A) = P(B) = 0.5$$

But I might ask you "If event A **did** occur, might event B would have any chance to occur?" . Since there's no way heads and tails can simultaneously appear, so your answer must be "No!". Now we can form an equation to represent the probability of a "special event" (or formally, conditional event) that "If A occured, then B would occur.":

$$P(B|A) = 0$$

Now we shift our attention to a more complex problem. I give you a die, and you roll it. Events E_1 and E_2 can be defined as: "Landing a number that greater than 3" and "Landing a number that divisible by 2":

$$P(E_1) = P(E_2) = \frac{1}{3}$$

Sample points of 2 events can be listed:

$$E_1 = \{4, 5, 6\}$$

$$E_2 = \{2, 4, 6\}$$

If a number greater than 3 landed, what's the probability that it could be divisible by 2? You can count the number of sample points inside E_1 and draw a result:

$$P(E_2|E_1) = \frac{2}{3}$$

Conversely, if a number divisible by 2 landed, what's the probability that it could be greater than 3?

$$P(E_1|E_2) = \frac{2}{3}$$

As you can see here, if one event occurs before another event, probability will be completely **changed**. So we call the pairs of events A and B, E_1 and E_2 **dependent events** because they depend on each other. The "special" probabilities $P(B|A), P(E_2|E_1), P(E_1|E_2)$ are called **conditional probability**.

Definition 2.4.1. *The **conditional probability** of B, given A, denoted by $P(B|A)$, is defined:*

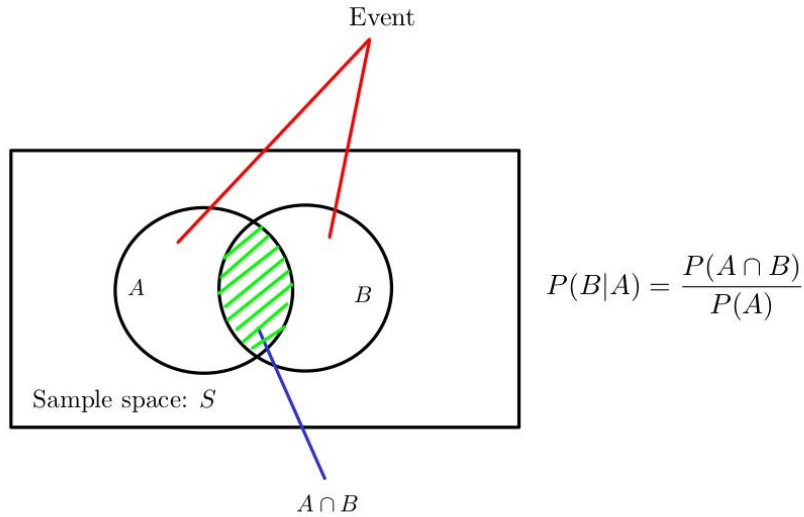$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Figure 2.3: Visualize how conditional probability is calculated

Now back to our previous problems, these conditional probabilities can easily be determined using **Definition 2.4.1** above:

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{2/6}{1/2} = \frac{2}{3}$$

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{2/6}{1/2} = \frac{2}{3}$$

2.4.2 Independent Events

Intuitively, we can see that if events A and B don't influence each other, then:

$$P(B|A) = P(B)$$

$$P(B|A) = P(B) \Leftrightarrow \frac{P(B \cap A)}{P(A)} = P(B) \Leftrightarrow P(A \cap B) = P(B)(A) \text{ (if } P(A) > 0)$$

Definition 2.4.2. *Two events A and B are independent if and only if:*

$$P(B|A) = P(B)$$

assuming $P(A) > 0$

Theorem 2.4.1. *Two events A and B are independent if and only if*

$$P(A \cap B) = P(A)P(B)$$

There are many examples of independent events, like if we define 2 events E_1: "Getting heads on the first toss." and E_2: "Getting heads on the second toss.". Intuitively you can see that E_1 and E_2 are unrelated, so they are **independent events**. You can also verify this fact:

$$P(E_1 \cap E_2) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} = P(E_1)P(E_2)$$

It's very important to note that determining the independence of events if completely **unrelated** to their mutual exclusion. Events are considered independent if and only if they satisfy **Theorem 2.4.1**.

2.5 Total Probability and Bayes's rule

2.5.1 Total Probability

In many situations, we don't know directly the information of $P(A)$ in **Definition 2.4.1**(or denomirator part); so in this section, we will develop a simple formula to handle this problem.
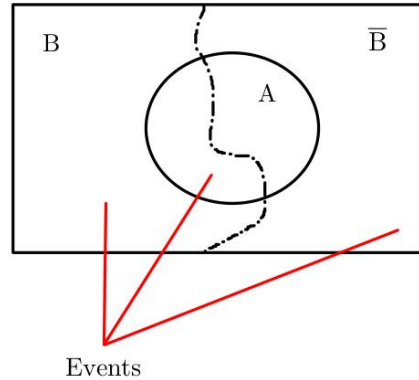

Figure 2.4: Total probability

$$A = A \cap S = A \cap (B \cup \overline{B}) = (A \cap B) \cup (A \cap \overline{B})$$

Because $A \cap B$ and $A \cap \overline{B}$ are independent events, so:

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

Theorem 2.5.1. *If A and B are two events of sample space S, then:*

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

Total probability is an useful formula, especially when you're conducting surveys in practice. For example, in my university, there are 2 types of students: those who "studied the whole semester" and those who "studied only one night before the exam". After the final exam, I asked everyone in my class about their scores and totaled the results. Event A: "Got $A+$" and event $\overline{A}$: "Didn't get $A+$"; event B: "Studied the whole semester" and event $\overline{B}$: "Studied only one night".

$$P(A|B) = 0.8, \ P(A|\overline{B}) = 0.3, \ P(B) = 0.4$$

Using total probability formula, I obtained the probability of A:

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) = 0.8.0.4 + 0.3.(1 - 0.4) = 0.5$$

Wow, that was an impressive ratio. Half of a class received perfect score. Was this course too easy?

2.5.2 Bayes's rule

If I studied hard, I would get $A+$. But how about me, who wasn't keen on studying boring courses like Computer Architecture but *still survived after final test and even got "A+"*? I could questioned myself "Was I too lucky?". To answer myself, I had to calculate $P(\overline{B}|A)$, if the result is not so high, perhaps I was lucky. Bayes's rule was what I needed.

Theorem 2.5.2. *If A and B are two events of sample space S, then:*

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})}$$

Bayes's rule seems very simple. Indeed an average high school student has no difficulty finding it, but its idea is brilliant. Before Bayes, we typically only reasoned about problems in terms of cause first, then effect. But Bayes's rule opens up a completely new way of thinking for us: knowing the effect beforehand, then understanding how cause influences it.

Back to my previous problem, I applied Bayes's rule:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})} = \frac{0.8.0.4}{0.5} = 0.64 \Rightarrow P(\overline{B}|A) = 1 - P(B|A) = 0.36$$

Since 0.36 was not a high number, indeed I was incredibly lucky and the course was not as easy as I thought.