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.
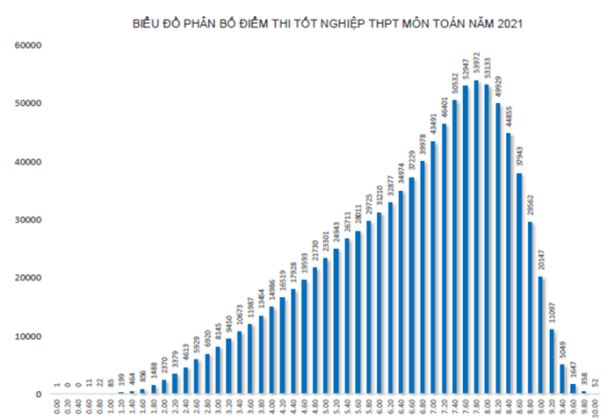

Figure 1: Distribution of Math scores in the 2021 National Entrance Exam

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. If today is the 10th and you haven't seen any rain since the beginning of the month, what are the probabilities of:

1. There will be 3 rainy days this month.

2. Tomorrow will be a rainy day.

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 to solve it!
The answer for the first question is:

$$P(X = 3) = \frac{e^{-\frac{5.21}{31}}\left(\frac{5.21}{31}\right)^3}{3!} \approx 21.89\%$$

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:
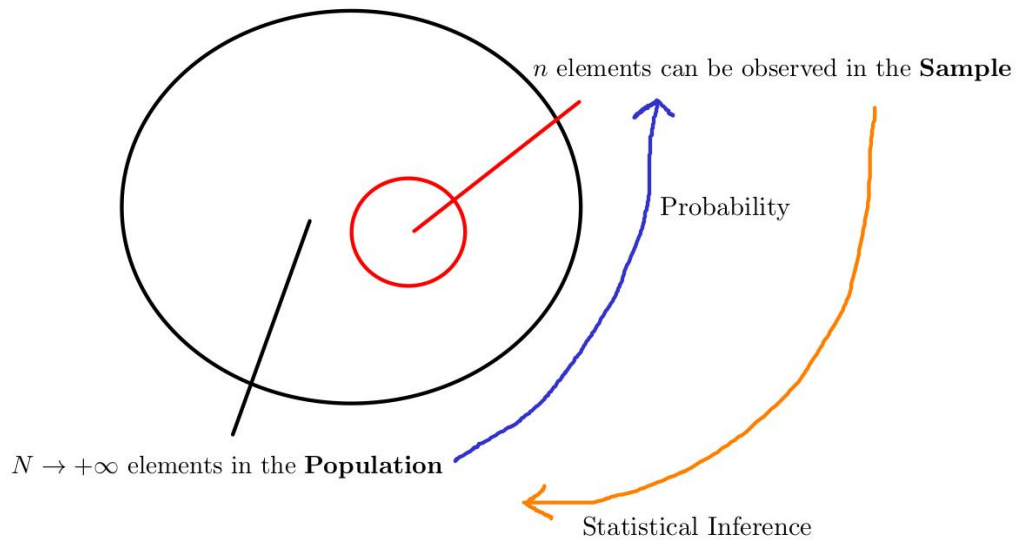
Figure 1.2: General model of simple ProbStat problems

You want to know some attributes of a very large **population** such as the probabilites, means,··· 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 appears:

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

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

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:

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

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 $\overline{E}$.*

Apply 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 law 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** instead.
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:*

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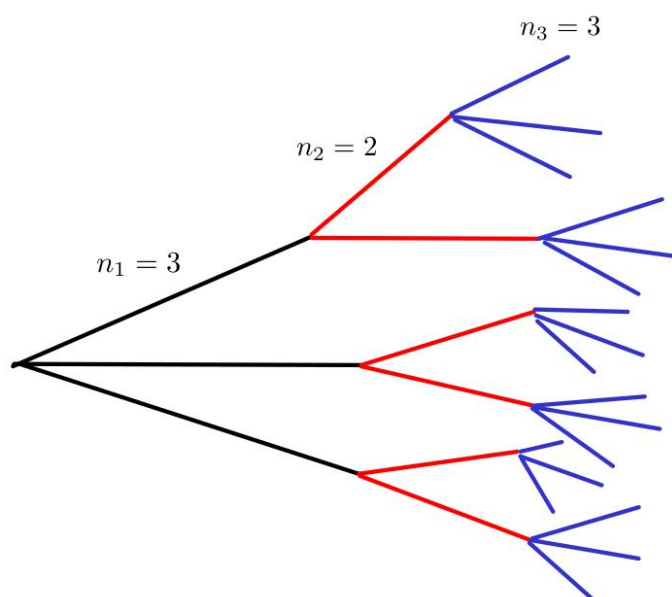
There are a total $n_1 n_2 n_3 = 18$ ways to perform this operation

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.

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 3 elements subsets from an original set with 10 elements is

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