The Atom of Probability: A Monograph on Bernoulli Trials and the Bernoulli Distribution

The Bernoulli distribution represents the atom of probability. It is the mathematical formalization of the simplest possible experiment: an action with exactly two mutually exclusive and exhaustive outcomes

The binary setup isn't just a mathematical tool, it mirrors how we naturally see the world. We often think in pairs: true or false, success or failure, presence or absence, 0 or 1. The Bernoulli distribution simply gives a precise way to describe these two-outcome situations.

The idea starts with something as basic as flipping a coin. Even though a single yes/no outcome looks simple, it actually forms the foundation for many powerful tools. When you combine many such outcomes, you get the Binomial distribution, which counts how many times you succeed. When you track how long it takes to succeed, you get the Geometric and Negative Binomial distributions. Today, this basic yes/no logic is used everywhere—from predicting credit risk to training machine-learning models and even to defining the basic unit of information in computers.

Historical Context: Jacob Bernoulli and the *Ars Conjectandi*

The distribution bears the name of Jacob Bernoulli, a titan of the Bernoulli family of Swiss mathematicians who dominated the field in the 17th and 18th centuries. His seminal work, *Ars Conjectandi* (The Art of Conjecturing), published posthumously in 1713, is a foundational text in the history of probability.

Bernoulli's key idea was to define a 'trial'. Doing the same process again and again under the same conditions. He noticed that even though one trial, like a coin flip, is impossible to predict, many repeated trials show a stable pattern. This insight became the Law of Large Numbers. It was a huge breakthrough because it showed that probability isn't just guesswork or gambling, it can be a reliable scientific method for understanding real-world patterns.

Defining the Bernoulli Trial

To treat binary events mathematically, we must first define the experimental framework. A random experiment is classified as a **Bernoulli trial** if and only if it satisfies specific, rigid conditions.

A Bernoulli experiment must have only two possible outcomes. We usually call them 'Success'***(S)*** and 'Failure'***(F)***, but these words don't carry any moral meaning. They're just labels. In a medical test, 'Success' might mean detecting a serious disease, while in engineering it might mean a part breaking. 'Success' simply marks the outcome we are focusing on, nothing more.

The chance of getting the outcome we call 'Success' must stay the same every time and is written as ***p***. That automatically means the chance of 'Failure' is ***1 – p (or q).*** This value ***p*** tells us how the process behaves. If ***p*** keeps changing from one trial to the next, then it's no longer a basic Bernoulli process.

When we repeat Bernoulli trials, each one must be independent. What happens in one trial shouldn't affect the next, each flip or test starts fresh. Even though we're mainly talking about a single trial here, this independence is what allows Bernoulli trials to build into larger distributions like the Binomial.

2. The Mathematical Object: Definitions and Notation

2.1 The Random Variable Mapping

To perform statistical analysis, we must map the physical outcomes of the Bernoulli trial to numerical values. We define the Bernoulli random variable X as:

$$X : \{\text{Success}, \text{Failure}\} \to \{0, 1\}$$

The mapping is standard and universally accepted:

- $X(\text{Success}) = 1$
- $X(\text{Failure}) = 0$

This binary encoding is crucial because it enables the property of **countability via summation.** If we perform n trials and sum the resulting Bernoulli variables ($\sum X_i$), the result is exactly the count of successes. This algebraic utility explains why we do not map outcomes to $\{-1, 1\}$ or $\{1, 2\}$ in standard introductory contexts, although such mappings exist in physics (e.g., Ising models of spin).

3. Characterizing Uncertainty: Moments and Properties

To truly understand the Bernoulli distribution, we need to work out its moments (like the mean and variance). Doing this is a standard example of how to handle discrete random variables.

3.1 The First Moment: Expectation (Mean)

Expected Value of a Bernoulli Random Variable

For a Bernoulli variable $X \in \{0, 1\}$ with success probability p:

$$\mu = E[X] = p$$

Derivation (Simple):

$$\begin{aligned} E[X] &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= 0 \cdot (1 - p) + 1 \cdot p \\ &= p \end{aligned}$$

Insight (Simplified):

The expected value doesn't mean the variable will ever *equal* that number.
A Bernoulli trial only outputs **0 or 1**, never anything in between.

For example, if $p = 0.3$, the expected value is 0.3, but every single trial still gives only **0** or **1**. The value 0.3 appears only as the **long-run average** over many trials.

I. Mean (μ) of a Bernoulli Distribution

The mean (expected value) of a Bernoulli distribution tells us the long-run average outcome of many repeated Bernoulli trials. Since each trial gives either **0** (failure) or **1** (success), the mean simply equals the **probability of success**

3.2 The Second Central Moment: Variance

3.2 The Second Central Moment: Variance

Variance measures how spread out the outcomes are from the mean.

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

Step 1: Compute $E[X^2]$

For a Bernoulli variable, powers don't change the value because:

- $0^n = 0$
- $1^n = 1$

So $X^n = X$, and therefore:

$$E[X^2] = 0^2(1-p) + 1^2(p) = p$$

In fact, all raw moments satisfy:

$$E[X^n] = p$$

Step 2: Compute the Variance

$$\mathrm{Var}(X) = p - p^2 = p(1-p)$$

Letting $q = 1 - p$:

$$\mathrm{Var}(X) = pq$$

The standard deviation is:

$$\sigma = \sqrt{pq}$$

Behavior of the Variance

- **Minimum:**
  When $p = 0$ or $p = 1$, variance is **0** → the outcome is certain.
- **Maximum:**
  Differentiate $p(1-p)$:

  $$\frac{d}{dp}(p - p^2) = 1 - 2p = 0 \Rightarrow p = 0.5$$

  So the variance is highest at $p = 0.5$.

Interpretation:
Uncertainty is maximized when success and failure are equally likely—like a fair coin toss.

Hypothesis testing

Bernoulli distributions play a key role in hypothesis testing whenever we deal with proportions or success rates. Typical uses include:

- **A/B Testing:** Comparing success rates of two versions of a product or strategy.
- **Quality Control:** Checking whether the defect rate is higher than an acceptable level.
- **Medical Trials:** Testing if a new treatment has a higher success rate than a placebo.

In these situations, the **null hypothesis** usually assumes a specific value of the success probability ppp. The **alternative hypothesis** suggests that the true ppp is different, and we use observed data to decide whether to reject the null.

Difference between Bernoulli Distribution and Binomial Distribution

The Bernoulli Distribution and the Binomial Distribution are both used to model random experiments with binary outcomes, but they differ in how they handle multiple trials or repetitions of these experiments.

Basis	Bernoulli Distribution	Binomial Distribution
Number of Trials	Single trial	Multiple trials
Possible Outcomes	2 outcomes (1 for success, 0 for failure)	Multiple outcomes (e.g., success or failure)
Parameter	Probability of success is p	Probability of success in each trial is p and the number of trials is n
Random Variable	X can only be 0 or 1	X can be any non-negative integer (0, 1, 2, 3, ...)
Purpose	Describes single trial events with success/failure.	Models the number of successes in multiple trials.
Example	Coin toss (Heads/Tails), Pass/Fail, Yes/No, etc.	Counting the number of successful free throws in a series of attempts, number of defective items in a batch, etc.