# Key Distributions

Antonio Aguirre

**University of California, Santa Cruz**
Winter 2025

## Introduction

This document provides a structured overview of six key probability distributions used in various applications across statistics, engineering, and data science. These distributions are mathematical models that describe and predict the behavior of random quantities.

Understanding probability distributions is essential because they serve as a bridge between **theory and real-world applications**. Each distribution models a specific type of **random behavior**. For example:

- The **Binomial distribution** models the number of successes in repeated independent trials.
- The **Normal distribution** describes continuous data that clusters around an average, such as human heights.
- The **Poisson distribution** models the number of random events occurring in a fixed interval, such as call arrivals at a help center.

Key Takeaways

- Probability distributions are **models**, not just formulas—they help capture and describe real-world uncertainty.
- Each distribution has specific **assumptions** that must be considered before using them.
- Distributions serve as **building blocks** for statistical inference, machine learning, and data analysis.

# 1. Bernoulli Distribution

## Definition

The Bernoulli distribution models the outcome of a single trial with two possible outcomes: success (1) or failure (0).

$$P(X = x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}.$$

## Parameters

- $p$: Probability of success ($0 \leq p \leq 1$).

## Support

$$x \in \{0, 1\}$$

## Assumptions

- A single trial with only two possible outcomes.
- The probability of success ($p$) remains constant.

**Mean and Variance**

$$\mathbb{E}[X] = p, \quad \mathrm{Var}(X) = p(1-p).$$

## Application

The Bernoulli distribution is commonly used in quality control to model whether a product passes (1) or fails (0) an inspection.

## Random Quantity

The Bernoulli distribution models the outcome of a **single** binary trial.

# 2. Binomial Distribution

## Definition

The Binomial distribution models the number of successes in $n$ independent trials, each with success probability $p$.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}.$$

## Parameters

- $n$: Number of independent trials.
- $p$: Probability of success in each trial.

## Support

$$k \in \{0, 1, \ldots, n\}$$

## Assumptions

- Trials are independent.
- The probability of success $p$ is constant across trials.

**Mean and Variance**

$$\mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1-p).$$

## Application

In surveys, the Binomial distribution is used to model the number of respondents who answer positively out of $n$ total participants.

## Random Quantity

The Binomial distribution models the **number of successes** in $n$ independent trials.

# 3. Poisson Distribution

## Definition

The Poisson distribution models the number of events occurring in a fixed interval, assuming events happen independently and at a constant rate $\lambda$.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \dots\}.$$

## Parameters

- $\lambda$: Average number of events per fixed interval ($\lambda > 0$).

## Support

$$k \in \{0, 1, 2, \dots\}$$

## Assumptions

- Events occur independently of one another.
- The rate $\lambda$ is constant over time.

**Mean and Variance**

$$\mathbb{E}[X] = \lambda, \quad \mathrm{Var}(X) = \lambda.$$

## Application

The Poisson distribution is widely used in call center analytics to model the number of incoming calls per hour.

## Random Quantity

The Poisson distribution models the **number of events** occurring in a fixed interval.

# Exercises

## Problem 1: Teaching Success

Antonio, a dedicated TA, gives a weekly statistics quiz to his class of 10 students. Historically, Antonio has observed that each student has a 70% chance of passing the quiz. Assume that the outcomes for each student are independent.

(a) What is the probability that exactly 7 students pass the quiz this week?

(b) What is the probability that at least 8 students pass the quiz?

(c) What is the expected number of students who will pass the quiz, and what is the variance?

## Problem 2: Emails Before the Final

The day before the final exam, Antonio receives an overwhelming number of emails from students asking for full explanations of specific topics they haven't studied. Historically, Antonio receives an average of 5 such emails per hour the day before the final. Assume these emails arrive according to a Poisson process.

(a) What is the probability that Antonio receives exactly 3 emails in a one-hour period?

(b) What is the probability that Antonio receives more than 7 emails in a one-hour period?

(c) If Antonio checks his email over a 2-hour period, what is the expected number of emails, and what is the standard deviation?

## Problem 3: Grading Speeds

Despite his underpayment as a TA, Antonio grades exams at an average speed of 12 minutes per exam with a standard deviation of 3 minutes. Note: his pay rate has nothing to do with the problem.

(a) What is the probability that Antonio grades an exam in less than 10 minutes?

(b) What is the probability that Antonio takes between 11 and 14 minutes to grade an exam?

(c) What is the probability that Antonio takes between 6 and 18 minutes to grade an exam?

(d) What is the probability that Antonio grades an exam in more than 21 minutes?