Discrete and Continuous Probability Distributions

Mathematics for Artificial Intelligence

1 Introduction

Probability distributions describe how probabilities are assigned to outcomes of random variables. They can be classified into:

- **Discrete distributions:** defined on countable outcomes.
- **Continuous distributions:** defined over intervals of real numbers.

This document covers four fundamental distributions:

$$\text{Bernoulli, Binomial, Poisson (Discrete)} \quad \text{and} \quad \text{Normal (Continuous)}$$

These serve as the foundation for machine learning, statistical modeling, and inference.

2 Bernoulli Distribution

2.1 Definition

A **Bernoulli random variable** takes value 1 with probability p (success) and 0 with probability $1 - p$ (failure):

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

2.2 Mean and Variance

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

2.3 Applications

- Coin flips (0 = tails, 1 = heads)
- Binary classification labels
- Yes/No outcomes in ML

3 Binomial Distribution

3.1 Definition

A **Binomial distribution** represents the number of successes in n independent Bernoulli trials with probability p of success.

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

3.2 Mean and Variance

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

3.3 Example

If a classifier has 80% accuracy ($p = 0.8$), the probability of correctly predicting exactly 7 out of 10 samples is:

$$P(X = 7) = \binom{10}{7} (0.8)^7 (0.2)^3.$$

3.4 Applications

- Counting correct predictions
- A/B testing results
- Number of defective items in a batch

4 Poisson Distribution

4.1 Definition

A **Poisson random variable** counts the number of events occurring in a fixed interval when:

- Events occur independently,
- The average rate λ is constant,
- Two events cannot occur at the same instant.

The PMF is:

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

4.2 Mean and Variance

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

4.3 When Poisson Approximates Binomial

If n is large and p small, with $\lambda = np$:

$$\mathrm{Binomial}(n, p) \approx \mathrm{Poisson}(\lambda).$$

4.4 Applications

- Number of customer arrivals per minute
- Rare event modeling (system failures)
- Number of messages received (email, network traffic)

5 Normal Distribution (Gaussian)

5.1 Definition

A continuous random variable X is said to follow a **Normal distribution** with mean μ and variance σ^2 if its probability density function (PDF) is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

We write:

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

5.2 Properties

- Symmetric around mean μ.
- Mean, median, and mode are equal.
- 68–95–99.7 rule:

$$P(|X-\mu| \le \sigma) \approx 0.68,$$
$$P(|X-\mu| \le 2\sigma) \approx 0.95,$$
$$P(|X-\mu| \le 3\sigma) \approx 0.997.$$

5.3 Standard Normal Distribution

If $\mu = 0$ and $\sigma^2 = 1$:

$$Z \sim \mathcal{N}(0, 1).$$

Conversion:

$$Z = \frac{X-\mu}{\sigma}.$$

5.4 Central Limit Theorem (CLT)

The sum (or average) of many independent random variables tends toward a normal distribution:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0, 1).$$

5.5 Applications

- Measurement errors
- Feature distributions in ML
- Gradient noise modeling in deep learning
- Initialization of neural network weights

6 Summary Table

Distribution	Type	Parameters	Mean / Variance
Bernoulli	Discrete	p	p, $p(1-p)$
Binomial	Discrete	(n, p)	np, $np(1-p)$
Poisson	Discrete	λ	λ, λ
Normal	Continuous	(μ, σ^2)	μ, σ^2