📘 Full Summary: Probability Distributions & Naïve Bayes

This document provides a detailed summary of the most commonly used probability distributions and Naïve Bayes models in data science.   
Each distribution includes its type, key formulas, usage in real-world scenarios, and interpretation in statistical modeling.

# 1. Probability Distributions

## Bernoulli Distribution

- Type: Discrete  
- Use Case: Models binary outcomes (success/failure)  
- Formula: P(X=1) = p, P(X=0) = 1-p  
- Mean = p, Variance = p(1 - p)  
- Example: Coin toss (Head = 1, Tail = 0)

## Binomial Distribution

- Type: Discrete  
- Use Case: Number of successes in a fixed number of independent Bernoulli trials  
- Formula: P(X = k) = C(n, k) \* p^k \* (1-p)^(n-k)  
- Mean = np, Variance = np(1 - p)  
- Example: Number of successful sales out of 10 calls

## Poisson Distribution

- Type: Discrete  
- Use Case: Counts of events in a fixed interval of time or space  
- Formula: P(X=k) = λ^k \* e^-λ / k!  
- Mean = Variance = λ  
- Example: Number of customer arrivals per hour

## Uniform Distribution

- Type: Continuous  
- Use Case: All outcomes equally likely over a given interval [a, b]  
- PDF: f(x) = 1 / (b - a)  
- Mean = (a + b)/2, Variance = (b - a)^2 / 12  
- Example: Rolling a fair die, random number generator

## Normal Distribution

- Type: Continuous  
- Use Case: Many natural phenomena (central limit theorem)  
- PDF: f(x) = (1 / sqrt(2πσ^2)) \* exp(- (x - μ)^2 / (2σ^2))  
- Mean = μ, Variance = σ²  
- Example: Heights, weights, standardized test scores

## Exponential Distribution

- Type: Continuous  
- Use Case: Time between events in a Poisson process  
- PDF: f(x) = λ \* exp(-λx) for x ≥ 0  
- Mean = 1/λ, Variance = 1/λ²  
- Example: Time between incoming customer calls

## Student’s t-Distribution

- Type: Continuous  
- Use Case: Estimating mean from small sample sizes  
- Heavy tails compared to normal distribution  
- Mean = 0, Variance = df / (df - 2) for df > 2  
- Example: Hypothesis testing when n < 30

## Chi-Square Distribution

- Type: Continuous  
- Use Case: Goodness-of-fit tests, variance estimation  
- Mean = k, Variance = 2k (k = degrees of freedom)  
- Example: Independence testing in contingency tables

## Gamma Distribution

- Type: Continuous  
- Use Case: Generalization of exponential distribution  
- PDF: f(x) = x^(k-1) \* exp(-x/θ) / (Γ(k) \* θ^k)  
- Mean = kθ, Variance = kθ²  
- Example: Modeling wait times, rainfall, insurance claims

## Beta Distribution

- Type: Continuous (0 to 1)  
- Use Case: Prior distributions in Bayesian analysis  
- Mean = α / (α + β), Variance = αβ / [(α+β)^2(α+β+1)]  
- Example: A/B testing, probability distributions

## Multinomial Distribution

- Type: Discrete  
- Use Case: Extension of Binomial for >2 outcomes  
- Mean = np\_i, Variance = np\_i(1 - p\_i)  
- Example: Predicting which product a customer chooses

## Multivariate Normal Distribution

- Type: Continuous  
- Use Case: Multiple correlated Gaussian variables  
- Parameters: Mean vector μ, Covariance matrix Σ  
- Example: Stock price modeling, multivariate regression

## Dirichlet Distribution

- Type: Continuous  
- Use Case: Probabilistic modeling of proportions  
- Mean = α\_i / Σα, used as a prior for multinomial  
- Example: LDA in topic modeling, proportions of topics in documents

# 2. Naïve Bayes Classifiers

Naïve Bayes classifiers apply Bayes' theorem with the “naïve” assumption of independence between every pair of features. Despite this strong assumption, they perform well in practice on many real-world problems, especially with high-dimensional data.

## Gaussian Naïve Bayes

- Suitable for: Continuous features  
- Assumes features follow a normal distribution  
- Use Case: Iris dataset, medical diagnosis  
- Formula: P(x\_i | y) = (1 / sqrt(2πσ²)) \* exp(- (x\_i - μ)² / (2σ²))

## Multinomial Naïve Bayes

- Suitable for: Discrete counts  
- Assumes data is generated from a multinomial distribution  
- Use Case: Text classification, NLP  
- Formula: P(x | y) = Π\_j P(x\_j | y) ^ x\_j

## Bernoulli Naïve Bayes

- Suitable for: Binary features (0/1)  
- Assumes features are binary and independent  
- Use Case: Word presence in document classification  
- Formula: P(x\_i | y) = p^x \* (1-p)^(1-x)