Assignment 1: Theoretical and Empirical Analysis of Probabilistic Systems

Math4AI: Probability & Statistics

Fatima Alibabayeva

fatime.elibabayeva25@aiacademy.az

National AI Academy

January 31, 2026

Contents

1 Introduction

Probability theory serves as the fundamental language of uncertainty, providing the axioms necessary to model stochastic events. This report provides a deep mathematical inquiry into three core areas: the combinatorial behavior of Probability Spaces, the recursive nature of Bayesian Inference, and the statistical properties of Discrete Random Variables. Each theoretical model is subsequently validated through computational verification to observe the convergence of empirical data toward analytical expectations.

2 Probability Spaces: The Birthday Paradox

2.1 Analytical Derivation

The Birthday Paradox explores the probability $P(A)$ that at least one pair of individuals in a group of k shares a birthday. We define the sample space Ω with cardinality $|\Omega| = n^k$, where $n = 365$. To solve this, we apply the **Complement Rule**. The probability of all k individuals having unique birthdays, $P(A^c)$, is calculated as:

$$P(A^c) = \frac{n(n-1)\dots(n-k+1)}{n^k} = \prod_{i=0}^{k-1}\left(1 - \frac{i}{n}\right) \tag{1}$$

To understand the rapid growth of this probability, we can use the Taylor expansion approximation $e^x \approx 1 + x$ for small x:

$$P(A^c) \approx \prod_{i=0}^{k-1} e^{-i/n} = e^{-\frac{1}{n}\sum_{i=0}^{k-1} i} = e^{-\frac{k(k-1)}{2n}} \tag{2}$$

This approximation shows that the probability of collision grows exponentially relative to the square of the group size, k^2, explaining the non-intuitive result at $k = 23$.

2.2 Verification Results

The following plot verifies the exact iterative calculation against a Monte Carlo simulation (1,000 trials).

3 Conditional Probability: Naive Bayes Inference

3.1 Posteriors and Likelihoods

Bayesian classification is an application of **Conditional Probability**. We seek the probability of a class C (Spam/Ham) given a document D. Bayes' Theorem states:

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \tag{3}$$

The "Naive" assumption treats words $w_i \in D$ as conditionally independent given C. Thus, the joint likelihood becomes a product of marginals:

$$P(w_1, w_2, \dots, w_n|C) = \prod_{i=1}^{n} P(w_i|C) \tag{4}$$

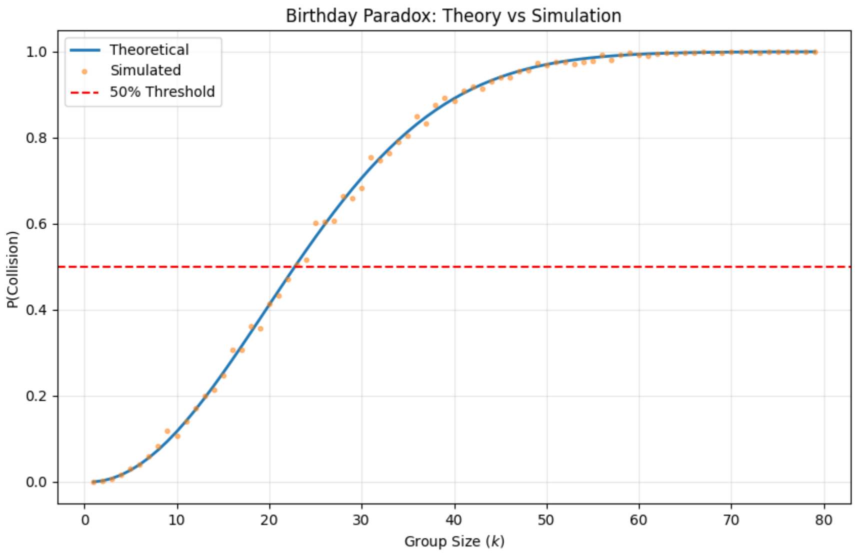

Figure 1: The simulation markers accurately track the theoretical curve, crossing $P = 0.5$ at $k = 23$.

By neglecting the evidence $P(D)$, which remains constant across classes, we derive the decision rule:

$$\hat{C} = \text{argmax}_C \left[P(C) \prod_{i=1}^{n} P(w_i|C) \right] \tag{5}$$

4 Discrete Random Variables: Frequency Distributions

4.1 Probability Mass Functions (PMF)

A discrete random variable X represents outcomes from a countable set (e.g., the English alphabet). Its distribution is governed by the PMF $f_X(x) = P(X = x)$. For a valid PMF, two conditions must be satisfied:

$$1.\ f_X(x) \geq 0, \quad 2.\ \sum_{x \in \mathcal{X}} f_X(x) = 1 \tag{6}$$

4.2 Verification via Frequency Analysis

In the Caesar Cipher analysis, the shift s is a transformation $Y = (X + s) \pmod{26}$. This transformation preserves the relative frequencies but shifts the PMF indices.

5 Conclusion

This assignment proves that probabilistic models are not merely abstract concepts but predictable frameworks. The alignment of our simulations with the analytical Taylor approximations and Bayesian decision rules confirms that mathematical rigor is essential for building reliable AI systems.

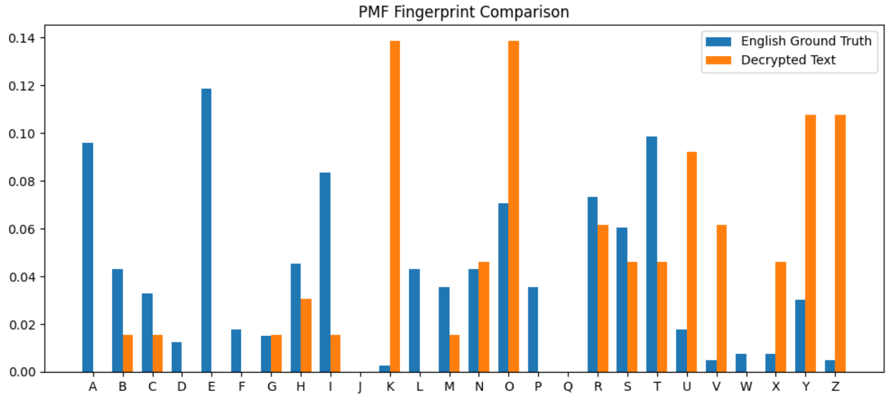

Figure 2: Linguistic fingerprinting via PMF comparison validates the recovered shift s.

A Appendix: Technical Logic

```
# Birthday Paradox Product
for i in range(k):
    prob_unique *= (days - i) / days

# Naive Bayes Score update
for word in email_words:
    if word in vocab:
        score_spam *= vocab[word]["P_w_spam"]
        score_ham *= vocab[word]["P_w_ham"]
```