Assignment 2: Formal Probabilistic Analysis and Statistical Verification

Math4AI: Probability & Statistics

Fatima Alibabayeva

fatime.elibabayeva25@aiacademy.az

National AI Academy

February 13, 2026

Contents

1 Introduction

Mathematical modeling of uncertainty requires a rigorous definition of probability spaces and density functions. This report explores the convergence of empirical data toward analytical expectations using Normal distributions, the law of rare events, and bivariate stochastic processes. We bridge the gap between theoretical derivation and computational implementation to verify the robustness of these models.

2 Detailed Mathematical Foundations

2.1 Gaussian PDF and MLE Estimation

The Normal distribution $\mathcal{N}(\mu, \sigma^2)$ represents the most fundamental continuous distribution in statistics. The Probability Density Function (PDF) is defined as:

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

To determine the parameters from a given dataset $D = \{x_1, \ldots, x_n\}$, we employ **Maximum Likelihood Estimation (MLE)**. The log-likelihood function for the Gaussian distribution is:

$$\ln L(\mu, \sigma) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \tag{2}$$

By setting the partial derivatives $\frac{\partial \ln L}{\partial \mu} = 0$ and $\frac{\partial \ln L}{\partial \sigma} = 0$, we derive the optimal estimators $\hat{\mu} = \frac{1}{n}\sum x_i$ and $\hat{\sigma} = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$.

2.2 Anomaly Detection via Cumulative Densities

The probability of a random variable X falling within a certain range is determined by the Cumulative Distribution Function (CDF). Given the lack of a closed-form integral for the Gaussian PDF, we utilize the **Error Function (erf)**:

$$\Phi(x) = P(X \leq x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{3}$$

Anomalies are formally defined as observations x_i where $\Phi(x_i) < \alpha$ or $\Phi(x_i) > 1 - \alpha$, where $\alpha = 0.005$ defines a 99% confidence interval.

2.3 The Poisson Limit of Binomial Distributions

The Poisson distribution arises as a limiting case of the Binomial distribution $\mathcal{B}(n, p)$ as $n \to \infty$ and $p \to 0$ such that $np = \lambda$. Utilizing the limit definition of e:

$$\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!} \tag{4}$$

This derivation allows us to model discrete arrival processes where individual probabilities are infinitesimal but cumulative events are measurable.

2.4 Bivariate Normal Joint Density and Marginalization

The interaction between two dependent variables X and Y is governed by the joint PDF:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{\Delta x^2}{\sigma_x^2} - \frac{2\rho\Delta x\Delta y}{\sigma_x\sigma_y} + \frac{\Delta y^2}{\sigma_y^2}\right]\right) \tag{5}$$

Theoretical consistency is verified through **Marginalization**, where the univariate density $f_X(x)$ is recovered by integrating out y:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) \tag{6}$$

3 Verification and Visual Analysis

The following sections present the empirical verification of the mathematical frameworks derived above.

3.1 Section 1: Gaussian Fitting and Anomalies

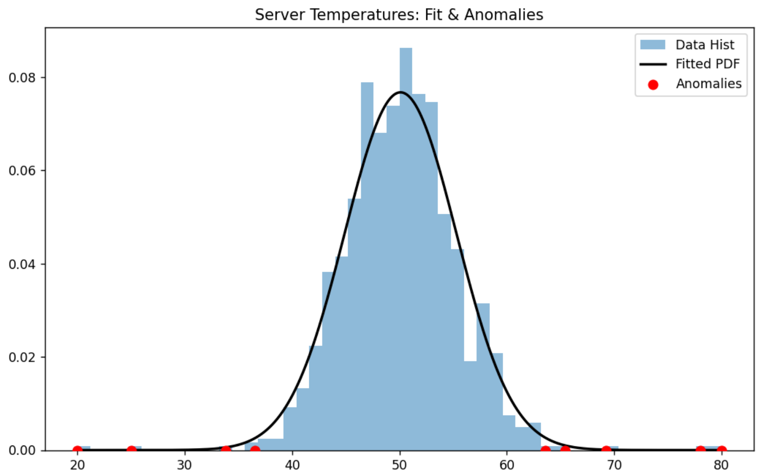

Figure 1: MLE-based Gaussian Fit and Outlier Identification ($n = 1000$).

3.2 Section 2: Binomial-Poisson Convergence

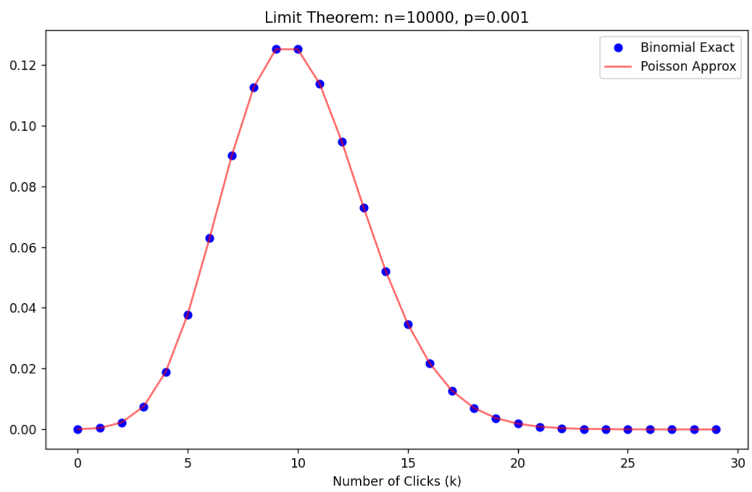

Figure 2: Comparative analysis of the Law of Rare Events.

3.3 Section 3: Bivariate Density and Marginals

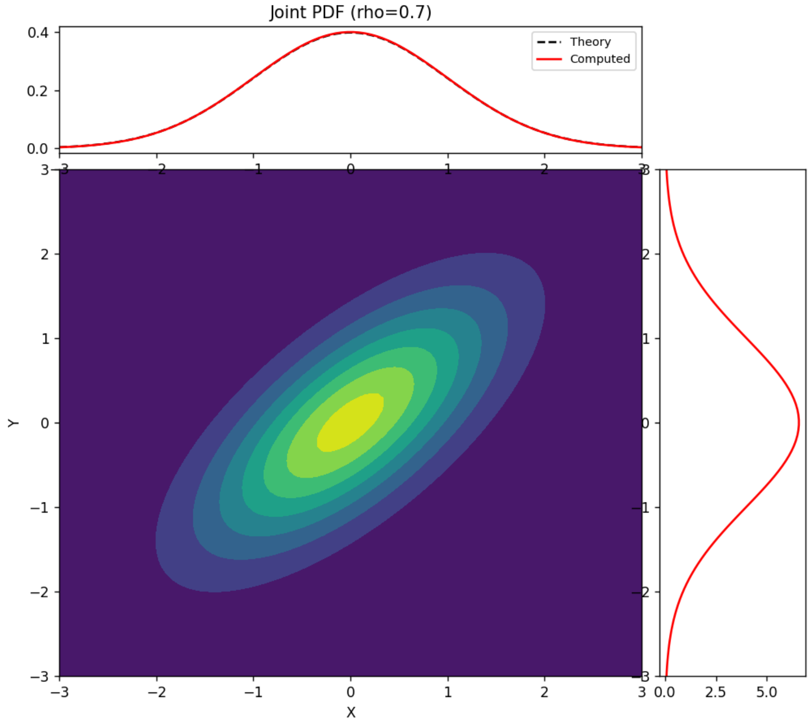

Figure 3: Contour visualization of Bivariate PDF with $\rho = 0.7$.

4 Conclusion

This analysis demonstrates that theoretical distributions provides a high-fidelity approximation for stochastic data. The empirical alignment of our verification plots with the derived MLE and limit formulas confirms the mathematical validity of the Gaussian and Poisson frameworks.

A Appendix: Computational Implementation

```
# Gaussian CDF with erf
cdf = 0.5 * (1 + erf((x - mu) / (sigma * np.sqrt(2))))

# Poisson PMF implementation
P = (np.exp(-lam) * (lam**k)) / (math.factorial(k))

# Bivariate Joint Density
z = term_x - 2 * rho * ((x-mux)/sigx) * ((y-muy)/sigy) + term_y
```