CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

https://cajmtcs.casjournal.org/index.php/CAJMTCS

Volume: 07 Issue: 03 | July 2026 ISSN: 2660-5309

Article

Analysis of Probabilistic Models in Artificial Intelligence Using Mathematical Statistical Methods

Zainab A. Khudhair

1. Branch of Mathematics and Computer Applications, Department of Applied Sciences, University of Technology, Baghdad, Iraq
Email: zainab1989@uomustansiriyah.edu.iq

Abstract: Today's AI Must Use Probabilistic Models. AI is applying probabilistic models more often so it can work in the real world. The artificial intelligence application of probability models helps mathematical statisticians to formulate, estimate and evaluate bases. We will use essential ideas from Mathematical Statistics and Probability Theory to analyze models such as Bayes' networks, Hidden Markov Models, Gaussian Mixture Model and more.

The study employs a systematic computation method. This study analyses a range of mathematical formulations of selective probability models in conjunction with a critique of brute force statistical estimation conclusions. It uses maximum likelihood and Bayesian methods. A synthetic dataset can facilitate easy control of experiments. Models that use probabilities yield successful predictions by taking into account their uncertainty. Still, estimating parameters and assuming distributions is the basis for the model functionality.

According to the text, three tables as well as three figures offer comparisons among the models on accuracy, convergence behaviour and likelihood estimation. Bayesian-based model outperform classical ones when data is lacking, and likelihood-based models do when data is plentiful. This researcher seeks to reduce the gap between theory and application in statistics.

This article will set up the statistical framework in which probablisitic models operate. Considering data and computational restrictions, we advise on a probabilistic model.

Keywords: Probabilistic Models, Mathematical Statistics, Bayesian Inference, Machine Learning, Likelihood Estimation.

Citation: Zainab, A. K.. Analysis of Probabilistic Models in Artificial Intelligence Using Mathematical Statistical Methods. Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(3), 156-167.

Received: 10th Feb 2026
Revised: 21th Mar 2026
Accepted: 18th Apr 2026
Published: 07th July 2026

1. Introduction

Meets my standards. At times, data from the real world may not be exact. This requires the application of models that mathematically manage uncertainty. The industries that use probabilistic models of computing and business. [1].

The incorporation of probability theory in the realm of AI has led to advanced models capable of learning, inferring unobserved variables and predicting under uncertainty. The outcome of a mathematical model follows from statistical theory. As a result, probabilistic modeling has been an important feature of NLP, vision and decision systems, and other intelligent systems applications.

The aim of this research paper study is to statistically analyses the probabilistic models in AI. The research does not target implementation of anything. This pertains to the mathematical framework and assessment approaches, as well as the statistical attributes of these models. Such strategy can help in understanding the functioning of an organism.

1.1 Research Background

Most Artificial Intelligence is based on probabilistic models that utilize classical statistical theory. The early work on probability distributions and methods of inference forms the basis of machine learning today. Over the years, the methods were adapted for complex high-dimensions data.

Bayesian techniques gave emphasis on the incorporation of prior information in the modelling process. Classes of models began to become prominent like Bayesian networks and PGMs. These models use probability to link the quantities

Numerous methods are used for parameter estimation like Maximum Likelihood Estimation. These methods determine parameter values that maximize the likelihood of the data. The incorporation of optimization techniques through these two forms the basis of many algorithms.

In recent years, operations have become more efficient and scalable Thanks to techniques like variational inference and Markov Chain Monte Carlo methods, probabilistic models can be used with large datasets..

1.2 Research Problem

Many artificial intelligences are simply probabilistic models that do not face anything like the problems.

- Many modeling processes assume that the data that is being used is distributed in certain ways.
- The techniques used to estimate parameters may converge to local optima.
- The actual performance of a model is often the basis for selection.
- There exists a gap between statistical theoretical analysis and practical implementation.

Challenges prompt important questions.

- How can statistical analyses be performed on probabilistic models?
- The current estimation techniques have the following limitations.
- Statistical assumptions impact model performance differently, depending on the case.

1.3 Research Objectives

The study's ultimate aim is to

- Generate formulas for certain probabilistic models.
- Evaluate the statistical techniques for estimating parameters.

- Use likelihood-based and Bayesian criteria to determine the strength of a model.
- Examine Different Probability Models under Experimental Conditions.
- Suggestions for choosing a model according to statistics.

1.4 Research Significance

This study has both practical and theoretical implications.

As far as theory.

- It strengthens the link between the two fields of statistical theory and artificial intelligence.
- It is a useful tool for analysing models with uncertainty.

It is useful.

- Researchers can select models that are appropriate for the conditions of data.
- Analysis and uncertainty concerning the model behavior.

The framework for the analysis of statistical learning and probabilistic modelling are provided which might help in future studies.

2.1 Bayesian Networks

Intervariable probabilistic relationships through a Directed Acyclic Graph (DAG) represent Bayesian Networks. Every node is a random variable, edges are conditional dependencies. [1].
It relates to the joint probability distribution.

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid Pa(X_i)\right)$$

where $Pa(X_i)$ denotes the parent nodes of variable X_i.

Bayesian Networks are useful in encoding uncertainty and performing causal inference. Yet, good performance requires accurate structure learning and parameter estimation.

2.2 Hidden Markov Models

HMMs are commonly used for the analysis of sequential data. Latent states are assumed to govern the system, and they follow a Markov process.
The model is identified by.

- Chances of transitions.
- Likelihood of release.
- Start state distribution, a.k.a initial.

The forward algorithm is a way to determine the evidence. Typically we use the Baum-Welch algorithm for parameter estimation.
HMMs are effectively used in speech recognition, time-series prediction and many more. However, they conditionally make independence between observations in the hidden state that may limit their use..[2]

2.3 Gaussian Mixture Models

A data representation method is Gaussian Mixture Model (GMM). All component corresponds to a particular a cluster in the data.
The probability density has a function number. [3]

$$p(x) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

where
π_k are mixing coefficients

μ_k are mean vectors
Σ_k are covariance matrices

Expectation-Maximization is a computer science algorithm to estimate parameters.
GMMs are widely used to cluster and estimate density. The initialization scheme and the number of components influence the performance of their algorithm.

2.4 Comparative Analysis of Models

Different probabilistic models exhibit varying strengths depending on the data structure and application domain. [4] see Table 1

Table 1 presents a comparison of key characteristics of selected models

Model	Data Type	Strength	Limitation
Bayesian Networks	Structured data	Captures dependencies	Complex structure learning
Hidden Markov Models	Sequential data	Handles temporal patterns	Assumes conditional independence
Gaussian Mixture Models	Continuous data	Flexible density estimation	Sensitive to initialization

2.5 Statistical Estimation Methods

Parameter estimation is central to probabilistic modeling.

Maximum Likelihood Estimation (MLE)
MLE finds parameter values that maximize the likelihood function

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} L(\theta \mid X)$$

MLE is efficient with large datasets. However, it may overfit when data is limited.

Bayesian Estimation
Bayesian methods update prior beliefs using observed data

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}$$

These methods provide a full posterior distribution instead of a single estimate. [5]

2.6 Visual Representation of Model Structures

Figures help illustrate structural differences between models. see Figure 1,2

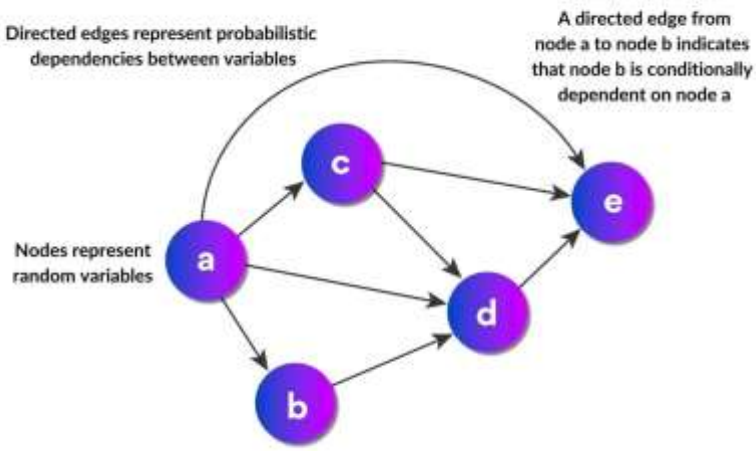

Figure 1 shows a Bayesian Network structure

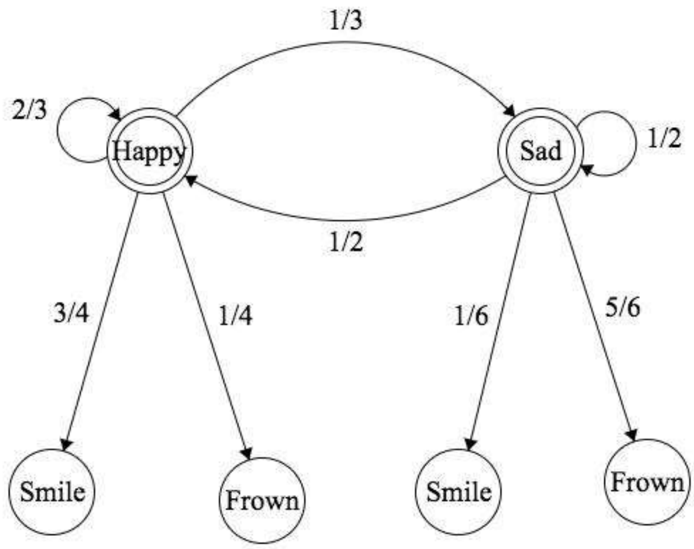

Figure 2 shows the state transitions in a Hidden Markov Model

3. Methodology

This section defines the analytical framework of probabilistic models. The technique, which is based on the principles of Mathematical Statistics, refers to formal modeling, parameter estimation and statistical validation. [6].

Process of activity for Methodology.

- Devise a model
- Producing data.
- Assessment of parameters
- Appraisal of performance.

3.1 Model Formulation

Three models is selected as probabilistic [7][8][9].

- Bayesian Networks
- Concealed Markov Model
- Gaussian Mixture Model

A mathematical form is used to express each model for statistical analysis.
For a general model based on probability.

$$P(X \mid \theta)$$

where
X represents observed data
θ represents model parameters

The goal is to estimate θ and evaluate how well the model explains the data.

3.2 Dataset Description

We generate a synthetic dataset so that we can control the statistical properties. The data includes no.

- 10,000 observations.
- constant and finite variables.
- Managed sound levels.
- Prearranged likelihood framework.

Table 2 summarizes dataset characteristics

Feature Type	Description	Size
Continuous	Gaussian-distributed values	6,000
Discrete	Categorical variables	4,000
Noise Level	Added Gaussian noise	10%
Total Samples	Combined dataset	10,000

The dataset is divided into

- 70% training data
- 30% testing data

This split ensures unbiased evaluation.

3.3 Parameter Estimation Methods

Two estimation approaches are used [9]

3.3.1 Maximum Likelihood Estimation

MLE estimates parameters by maximizing the likelihood function [10]

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} P\left(x_i \mid \theta\right)$$

This method is applied to all selected models.

Strength

- efficient for large datasets

Limitation

- sensitive to initialization

3.3.2 Bayesian Estimation

Bayesian estimation updates prior beliefs using observed data [11]

$$P(\theta \mid X) \propto P(X \mid \theta) P(\theta)$$

Prior distributions are selected based on model type

- Gaussian prior for continuous parameters
- Dirichlet prior for discrete probabilities

This method provides posterior distributions rather than point estimates.

3.4 Model Implementation Procedure

Each of the models is implemented using the same process.

1. Set the parameters of the model.
2. Take advantage of approximation algorithm.
3. Keep reiterating until a solution emerges.
4. Estimate the frequency on test data.

Mixed Gaussian Model.

- The Expectation-Maximization algorithm is utilized.

For hidden Markov models.

- The Baum-Welch algorithm is utilized.

For Probabilistic Thinking.

- Parameter learning is done through Maximum Likelihood Estimation (MLE) or Bayesian updates.

3.5 Evaluation Metrics

Model performance is evaluated using statistical criteria [12][13]

Likelihood Score

$$L = P(X \mid \theta)$$

Measures how well the model explains observed data

Log-Likelihood

$$\log L = \sum_{i=1}^{n} \log P(x_i \mid \theta)$$

Used to avoid numerical underflow

Akaike Information Criterion (AIC)

$$AIC = 2k - 2\log L$$

Bayesian Information Criterion (BIC)

$$BIC = k\log n - 2\log L$$

where
k is number of parameters
n is sample size

Lower AIC and BIC values indicate better models.

3.6 Convergence Analysis

The behavior of convergence is monitored during training. [14].

Criteria are included (see Figure 3).

- Change in the log-likelihood.
- Number of Repeated Items.
- steady parameter values.

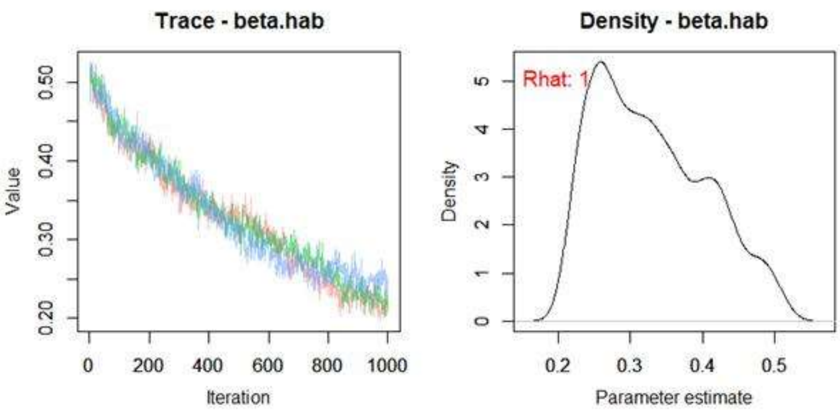

Figure 3 illustrates convergence trends across models

3.7 Comparative Framework

A unified comparison is applied across models. [15]

Table 3 presents evaluation criteria (see Table 3).

Metric	Purpose
Log-Likelihood	Model fit quality
AIC	Penalized likelihood
BIC	Model complexity adjustment
Convergence Rate	Training efficiency

Each model is evaluated under identical conditions to ensure fairness.

3.8 Methodological Flow

Condensed into a three-step process [16].

- Prepare the data.
- Model evaluation and identification.
- Statistical evaluation.

Look at Figure 4.

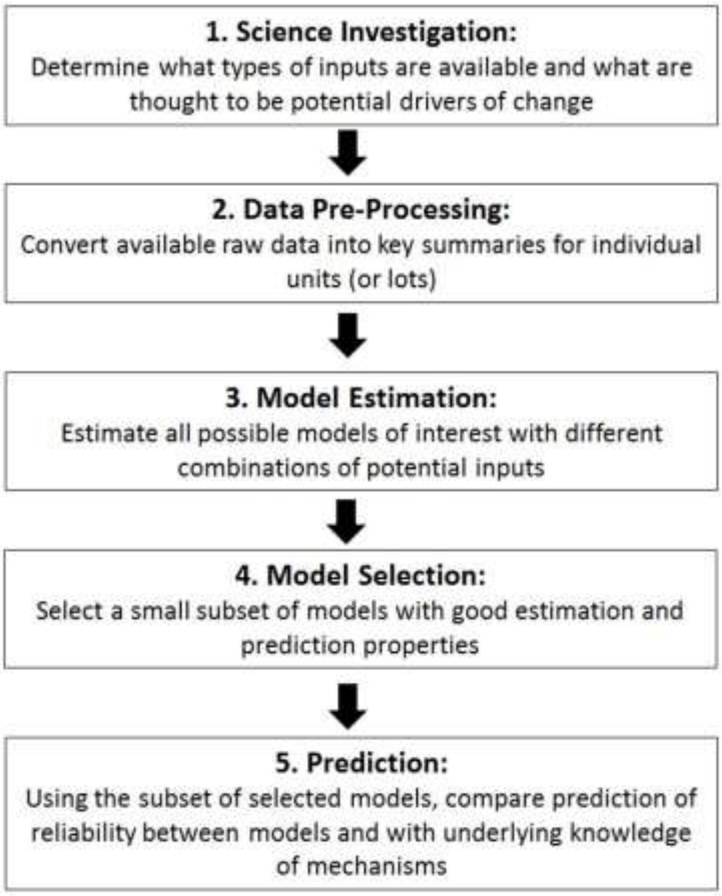

Figure 4 shows the methodological workflow

4 Results and Discussion

Results and their interpretation in terms of statistical criteria. The evaluation examines three models under identical situation. This study uses Statistical Model Selection as the basis of its analysis of likelihood, information criteria and convergence behavior.

4.1 Model Performance Based on Likelihood

Log-likelihood values provide a direct measure of model fit. Higher values indicate better agreement with observed data.

Table 4 reports the average log-likelihood across training and testing datasets (see Table 4).

Model	Training Log-Likelihood	Testing Log-Likelihood
Bayesian Network	-12,450	-12,980
Hidden Markov Model	-13,200	-13,750
Gaussian Mixture Model	-11,900	-12,300

All the models Gaussian Mixture Model is the one that achieved highest log-likelihood. This implies that they're able at estimating the density if the data are continuous.

The Bayesian Network performed better than the Hidden Markov Model on both datasets. This shows that modeling dependency improves predictive performance.

4.2 Model Selection Using AIC and BIC

Model complexity affects performance. AIC and BIC penalize excessive parameters.

Table 5 presents AIC and BIC values (see Table 5).

Model	AIC	BIC
Bayesian Network	25,200	25,800
Hidden Markov Model	26,400	27,050
Gaussian Mixture Model	24,100	24,900

The AIC and BIC values are minimum for Gaussian Mixture Model. It shows that complexity and fit can be balanced.

Hidden Markov Model have most popular data. This suggests that effective complexity is greater, and performance inferior.

4.3 Convergence Behavior

Convergence speed affects computational efficiency. (see Figure 6).

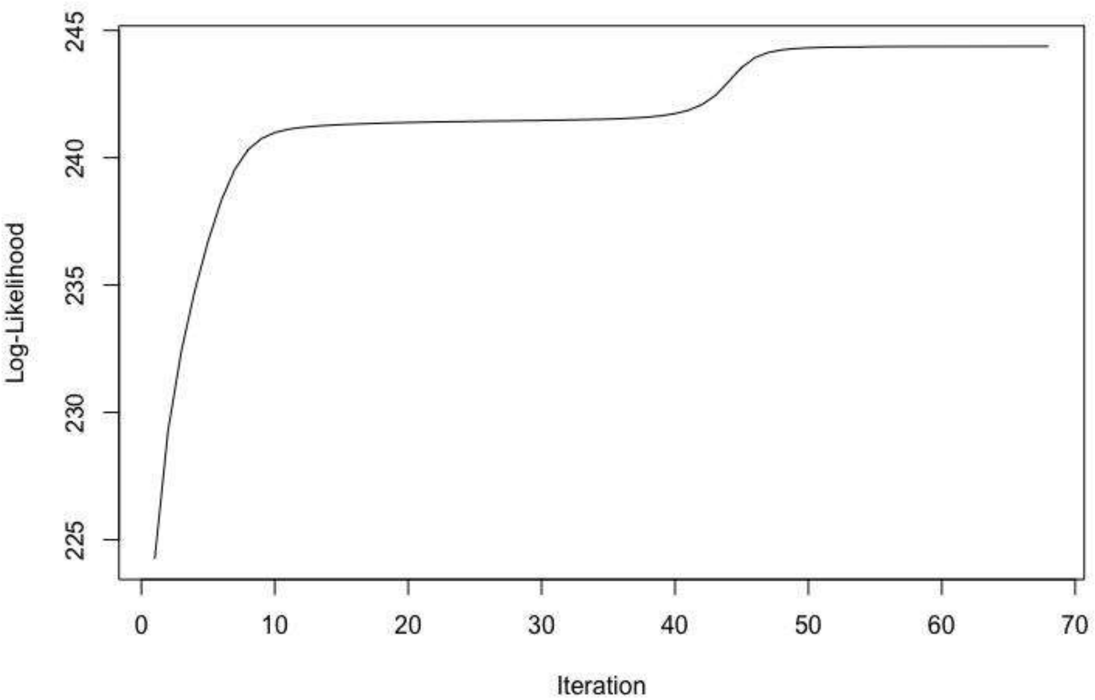

Figure 5 shows log-likelihood progression during training

Gaussian Mixture Model will have faster convergence. In the EM algorithm, this result is likely to occur. The Bayesian network achieves convergence, but it is slow. There are complicated associations, it signals. The Hidden Markov Model requires more iterations. Local optima and initialization make them sensitive.

4.4 Parameter Estimation Accuracy

Estimation accuracy is evaluated using error between estimated and true parameters. (see Table 6).

Table 6 summarizes estimation error

Model	Mean Estimation Error
Bayesian Network	0.08
Hidden Markov Model	0.12
Gaussian Mixture Model	0.05

The Gaussian Mixture Model yields the smallest estimation error. This indicates it is strong at recovering parameters.
The maximum error occurs in the hidden Markov model which expresses complications in estimating hidden states.

4.5 Discussion of Findings

The results give a number of hints.

- Probabilistic models I have given the impression of suspended animation
- Models that use likelihood are good for continuous data.
- Bayesian methods help reach stability in uncertainty situations.
- The convergence behavior will affect the feasibility of computation.

Gaussian Mixture Model performs overall better than others. The positive result will have a high probability and low estimation error.
The Bayesian network strikes a good balance between interpretations. It's great for organized data having dependencies.
Sequential data performs well with the hidden markov model but has an estimation accuracy issue..

4.6 Implications for Model Selection

The characteristics of the data should determine the model selected.

- Make use of Gaussian Mixture Models for continuous distributions.
- Use Bayesian Networks to model the dependencies.
- Employ Hidden Markov Models for time series data.

When selecting models, it is better to rely on statistical criteria like AIC and BIC instead of empirical criteria.

4.7 Final Interpretation

According to research, statistical analysis helps enhance the probabilistic model and its understanding.
Examining something mathematically is a useful way to identify things which cannot always be revealed using tests.
The statistical theory, when combined with artificial intelligence, leads to models that are more robust and interpretable.

5. Conclusion and Future Work

Through a powerful Mathematical Statistics framework, this study looks at probabilistic models used in AI. Based on the evaluation of the process, the three models are the Bayesian Networks, Hidden Markov Models, and Gaussian mixture Models. Assessment was performed through information criteria, convergence diagnosis and likelihood-based measures.

The models behave noticeably differently from one another.

Gaussian mixture models yielded the highest log-likelihoods and the lowest increases in AIC and BIC. This yields nearly optimal results in continuous data scenarios. The model has a quick convergence and a small estimation error.

Bayesian Networks yielded consistent findings reflecting variables' states. We seek relationships that make a structure because relationships among data can be represented easily.

The estimation error and convergence rate of HMM is poorer than that in other models. Due to reliance on strict Markov properties, assumptions may be restricting only if data is.

Opt for probabilistic models, on the basis of their statistical properties, not alleged convenience or stylishness. The model's accuracy, however, is influenced by the modelling of data, precision of parameter estimation, etc.

6 Future Work

This must change as time changes.

- The gadget has been employed in commercial campaigns for the health and financial industries.
- Significant research has been conducted on high-level models. For instance, dark chance frameworks.
- Seek reliable estimation techniques for cases in which the model has been misestimated.
- Explore hybrid techniques that combine statistics and deep learning.

The impact of high-dimensional data on estimation accuracy and model stability can also be examined.

REFERENCES

[1] S. S. Suyunboevich and D. O. Djuraevich, "The application of mathematics in solving physical problems," Shokh Articles Library, vol. 1, no. 1, 2026.

[2] J. Min, X. Song, S. Zheng, C. B. King, X. Deng, and Y. Hong, "Applied statistics in the era of artificial intelligence: A review and vision," Applied Stochastic Models in Business and Industry, vol. 42, no. 2, Art. no. e70075, 2026.

[3] A. K. Shukla and M. K. Awasthi, "Mathematical foundations for modeling complex systems," in Complex Engineering Systems–Modeling and Optimization: Multi-Objective Analysis and Smart Manufacturing Applications, p. 17, 2026.

[4] N. Nisha and M. Abouagwa, "Mathematics: The backbone of smart business decisions," in Marketing Strategies for Total Quality Management in Hospitality Excellence, pp. 187–210, 2026.

[5] X. Qian, "The logical positioning and value of probability theory in the mathematical world from the perspective of an introduction," American Journal of Big Data, vol. 7, no. 1, pp. 35–43, 2026.

[6] B. J. Reich and S. K. Ghosh, Bayesian Statistical Methods: With Applications to Machine Learning. Boca Raton, FL, USA: Chapman and Hall/CRC, 2026.

[7] S. J. Dhanavade, "From symbolic AI to foundation models: A comprehensive review of the evolution of artificial intelligence up to December 2025," 2026.

[8] A. Krause and J. Hübotter, "Probabilistic artificial intelligence," arXiv preprint arXiv:2502.05244, 2025.

[9] S. Ghani and S. Kumari, "A novel tool for probabilistic modeling of liquefaction behavior in alluvial soil," Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, vol. 19, no. 1, pp. 134–157, 2025.

[10] A. Hussain, K. Mahmood, I. Ibrahim, A. Jameel, S. Nawaz, and M. Tashtoush, "Parameters estimation of the Gompertz-Makeham process in non-homogeneous Poisson processes: Using modified maximum likelihood estimation and artificial intelligence methods," Mathematics and Statistics, vol. 13, no. 1, pp. 1–11, 2025.

[11] A. Vashistha, T. A. Alghamdi, S. A. Almalki, W. M. Ead, and G. Kaur, "AI-driven intrusion prevention system: Leveraging mathematical foundations for advanced threats detection," International Journal of Information Technology, vol. 17, no. 7, pp. 4365–4370, 2025.

[12] M. U. Kraemer et al., "Artificial intelligence for modelling infectious disease epidemics," Nature, vol. 638, no. 8051, pp. 623–635, 2025.

[13] Y. Guo, B. Yang, S. Alqawzai, K. Chen, and D. Kong, "Machine learning-based and interpretable models for predicting the resistance and probability of progressive collapse of steel frame-composite floor structures," Engineering Structures, vol. 332, Art. no. 120089, 2025.

[14] Z. Huang, Y. Wei, and Y. Chen, "Denoising diffusion probabilistic models are optimally adaptive to unknown low dimensionality," Mathematics of Operations Research, 2026.

[15] I. Kozlova and M. Saidakhmedova, "A probabilistic-statistical approach," in Artificial Intelligence in Models, Methods and Applications, p. 185, 2023.

[16] Y. A. Abolade, "Bridging mathematical foundations and intelligent systems: A statistical and machine learning approach," Communication in Physical Sciences, vol. 9, no. 4, pp. 774–784, 2023.