ETHIOPIAN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

eISSN: 2349-5715 pISSN: 2349-5707

Volume: 12, Issue 06, June-2025 SJIF 2019: 4.702 2020: 4.737 2021: 5.071 2022: 4.919 2023: 6.980 2024: 7,662

THE ROLE AND PRACTICAL APPLICATION OF THE TOTAL PROBABILITY AND BAYES' FORMULAS IN STATISTICAL ANALYSIS

Nurmatov Sardor Sidikovich

Institution: Kokand University

Abstract: This article explores two fundamental concepts in probability theory: the total probability formula and Bayes' formula. It discusses their theoretical foundations and examines their practical applications in various real-life fields, including medicine, artificial intelligence, risk analysis, and economic modeling.

Keywords: total probability formula, Bayes' formula, conditional probability, statistical analysis, artificial intelligence, probability theory, machine learning, medical diagnosis, risk analysis, probability updating

Introduction

In modern statistical modeling and data analysis, the role of probability theory is paramount. When information is uncertain, incomplete, or random in nature, precise mathematical tools are essential for decision-making, forming hypotheses, and updating existing knowledge. Two of the most important tools for such situations are the **total probability formula** and **Bayes' formula**.

The total probability formula expresses the probability of an event as the sum of the probabilities of all possible causes that can result in that event. This formula allows for analysis that considers all contributing factors. Bayes' formula, on the other hand, enables us to reassess the probability of possible causes after an outcome is observed, based on conditional probabilities. This makes Bayes' approach a dynamic model for updating beliefs in the light of new data.

Today, these formulas are widely used in interdisciplinary fields such as bioinformatics, epidemiology, economic forecasting, risk assessment, cryptography, and artificial intelligence. For instance, machine learning algorithms based on Bayes' methods effectively solve problems such as text classification, spam filtering, language detection, and medical diagnostics.

The importance of probabilistic models is growing in systems that support scientific decision-making. Traditional statistical methods are typically based on fixed probabilities, whereas the Bayesian approach updates probabilities based on new evidence, which is especially valuable in real-time systems.

Total Probability Formula

The total probability formula is used in probability theory to compute the overall probability of an event based on conditional probabilities. This is especially useful when the cause of the event is unknown, but all possible conditions that could lead to the event are known.

Assume a set of mutually exclusive events, $B_1, B_2, ..., B_n$ that together form a complete set ($B_i \quad B_j =$, for $i \quad j$) and together, completely covering the space ($U_{i=1}^{n} B_i = \Omega$) In this case, they form a complete system of events. If **A** is any event in the sample space, then the probability of event **A** can be calculated based on its relation to each event in this complete system as follows:

$$P(A)= \overset{n}{\underset{i=1}{}} P(B_i)\bullet P(A|B_i)$$

Where:

- $P(B_i) -$ is the prior probability of each cause,

eISSN: 2349-5715 pISSN: 2349-5707
Volume: 12, Issue 06, June-2025 SJIF 2019: 4.702 2020: 4.737 2021: 5.071 2022: 4.919 2023: 6.980 2024: 7,662

- $P(A \mid B_i) - B_i$ is the conditional probability of event A given that occurred.

This formula shows that if an event can occur in several different ways (through various causes), then the probability of the event is the sum of the products of the probability of each cause and the probability of the event occurring under that cause.

Example: Suppose there are two factories — Factory A and Factory B — both producing the same product. 60% of the products come from Factory A and 40% from Factory B. 3% of Factory A's products and 5% of Factory B's products are defective. What is the probability that a randomly selected product is defective?

$$P(Defective) = P(A)\bullet P(Defective \mid A) + P(Defective \mid B)$$

$$= 0.6\bullet 0.03 + 0.4\bullet 0.05 - 0.018 + 0.02 = 0.038$$

So, the probability that a randomly selected product is defective is 3.8%.

The total probability formula is widely used in real-world statistical analysis, risk assessment, medical diagnosis, signal processing, and many other fields. Correct understanding and application of this formula are particularly important in complex systems with multiple potential causes.

Bayes' Formula

Bayes' formula is one of the most powerful tools in probability theory. It allows us to revise the probabilities of causes based on observed outcomes using conditional probabilities. In other words, Bayes' formula helps convert prior beliefs into updated (posterior) probabilities after new data is observed.

Bayes' formula is written as:

$$P(B_i \mid A) = \frac{P(B_i)\,P(A \mid B_i)}{{}^{n}_{j=1} P(B_j)\,P(A \mid B_i)}$$

Where:

- $B_1, B_2, \ldots B_n -$ is a complete set of events,
- $P(B_i)$- is the prior probability,
- $P(A|B_i)$- conditional probability (likelihood),
- $P(A_i|B)$- is the posterior probability of B_i given that A occurred.

Bayes' formula is mainly used to calculate reverse probabilities. For instance, if we observe a result (e.g., symptoms in a patient), we use Bayes' formula to determine the likelihood of different possible causes (e.g., specific diseases).

Example:

Suppose a disease affects 1% of the population:

- P(K)=0.01 — disease present
- P(¬K)=0.99 — disease absent

A medical test detects the disease correctly in 99% of cases but has a 5% false positive rate:

- P(+｜K)=0.99
- P(+｜¬K)=0.05

What is the probability that a person has the disease if their test is positive?

$$P(K \mid +) = \frac{P(K)\,P(+ \mid K)}{P(K)\,P(+ \mid K) + P(\quad K)\,P(+ \mid \quad K)}$$

$$= \frac{0.01 \quad 0.99}{0.01 \quad 0.99 + 0.99 \quad 0.05} = \frac{0.0099}{0.0099 + 0.0495} \approx 0.1667$$

That is, even if the test result is positive, the probability that the person actually has the disease is only 16.67%.

eISSN: 2349-5715 pISSN: 2349-5707
Volume: 12, Issue 06, June-2025 SJIF 2019: 4.702 2020: 4.737 2021: 5.071 2022: 4.919 2023: 6.980 2024: 7,662

Applications of Bayes' formula:

Medicine: in diagnosis Forensics: identifying suspects in a crime, Artificial intelligence: text classification (Naive Bayes classifier), spam detection,
Data science: updating probabilities based on data,
Robotics: analyzing uncertain information obtained from sensors.

Theoretical significance: Bayes' formula represents probability as a dynamic model of knowledge. As new information becomes available, it recalculates probabilities and adjusts the system's level of knowledge accordingly. This makes it the foundation of modern statistical approaches, particularly Bayesian inference, Bayesian networks, Markov chains, and other advanced methods.

Conclusion

The total probability and Bayes' formulas are fundamental components of probability theory. They help simplify complex statistical problems, draw conclusions based on data, and generate updated predictions. In conditions of uncertainty, these formulas are indispensable tools for decision-making.

While the total probability formula analyzes the probability of an event based on its possible causes, Bayes' formula allows us to revise these probabilities based on observed results. These approaches have proven effective across a wide range of applications — from medical diagnostics to artificial intelligence.

The formulas discussed in this article embody both theoretical depth and practical versatility in statistical modeling. Mastery of these concepts is essential for addressing the challenges of today's information-driven society, making them an integral part of modern statistical thinking.

References:

1. Feller, W. (1968). An Introduction to Probability Theory and Its Applications. Wiley.
2. Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London.
3. DeGroot, M. H., & Schervish, M. J. (2012). Probability and Statistics (4th ed.). Pearson.
4. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). CRC Press.
5. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
6. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
7. Jaynes, E. T. (2003). Probability Theory: The Logic of Science. Cambridge University Press.
8. Ross, S. (2014). Introduction to Probability Models (11th ed.). Academic Press.
9. Kass, R. E., & Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules. Journal of the American Statistical Association, 91(435), 1343–1370.
10. Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.