

The Role of Statistics in Modern Science: Complementing Artificial Intelligence

Abstract

Statistics has been a cornerstone of scientific inquiry, providing essential tools for data analysis and inference. This paper examines the application of statistical methods in modern science, particularly in the context of biomedicine. It also explores the intersection between statistics and artificial intelligence (AI), highlighting the unique strengths of statistical approaches in the face of AI advancements. The paper concludes with a discussion on the future trajectory of statistics, emphasizing the importance of interdisciplinary collaboration.

Keywords Statistics, Biomedicine, Artificial Intelligence, Deep Learning, Interdisciplinary Collaboration

Introduction

Statistics has long been recognized for its pivotal role in scientific research, offering a rigorous framework for data collection, analysis, and interpretation. In the field of biomedicine, statistical methods are indispensable for drug development, disease diagnosis, and treatment evaluation. This paper discusses the application of statistics in biomedicine, its relationship with AI, and the future directions of statistical research.

History of Statistics

In the 19th century, statistics primarily consisted of sophisticated tricks but lacked a solid theoretical foundation [1].

In the early 20th century, Ronald A. Fisher revolutionized statistics with his work on experimental design, Maximum Likelihood Estimation (MLE), and analysis of variance (ANOVA). Fisher's development of ANOVA allowed for the testing of differences in means between two or more groups, which became a cornerstone in experimental statistics. He also introduced the concept of randomization in experimental design to minimize bias. Additionally, Fisher's work on MLE provided a method for estimating the parameters of a distribution based on observed data [1].

During the same period, Jerzy Neyman and Egon Pearson developed hypothesis testing and confidence intervals. Their collaboration led to the development of the Neyman-Pearson theory of hypothesis testing, which identified an alternative hypothesis and recognized two types of error. This approach focused on the power to detect alternative hypotheses and led to the identification of optimal test criteria in specific circumstances. Neyman also formulated his theory of statistical estimation via the method of confidence intervals [1].

In the mid-20th century, the development of computers in the 1950s and 1960s enabled more complex data analysis. This technological advancement allowed statisticians to handle larger datasets and perform more intricate statistical procedures, significantly expanding the scope and capabilities of statistical analysis [1].

Applications of Statistics in Biomedicine

Statistics in Drug Development

The process of drug approval involves stringent assessments of safety, toxicity, and efficacy, which are conducted through clinical trials. The design and analysis of these trials are heavily reliant on statistical methods such as hypothesis testing, regression analysis, and survival analysis. These methods provide the scientific basis for evaluating the effectiveness and safety of new drugs.

Statistics in Medical Research

Beyond drug development, statistics is widely applied in medical research. Survival analysis models, variable selection methods like Lasso, and pca help researchers extract valuable insights from complex data, aiding clinical decision-making.

Statistics in magic

The issue of seven shuffles being the fairest is somewhat related to the use of probability and conditional probability knowledge. This paper was written by Persi Diaconis. At the age of 14, he ran away from home to become a professional magician and traveled around the world. Later on, he developed an interest in mathematics, returned to school for further studies, and eventually earned his Ph.D. from Harvard University. He went on to become a professor of mathematics and statistics at Stanford University. His background in magic and his interest in probability and statistics led him to conduct numerous studies related to randomness [2].

The Relationship Between Statistics and Artificial Intelligence

Statistics and the Development of AI

The development of AI is deeply intertwined with statistical methods [3]. Many AI algorithms are essentially complex regression models. Besides, deep learning learns optimal function mappings through extensive data training. Statistical techniques such as regression analysis, model selection, and optimization form the theoretical foundation for AI advancements.

Differences Between Statistics and AI

While AI excels in data processing and pattern recognition, it often lacks interpretability. In fields like medicine, where sample sizes may be limited and data imbalanced, the black-box nature of AI models can be a significant drawback. Traditional statistical methods offer advantages in terms of interpretability, providing clearer causal relationships and model explanations [4]. Ethic issue should also been considered especially in the area of medicine.

As one zhihu blogger said in his comment of “the future trend of statistical research after 2020”: “ It must be acknowledged that in many real applications at present, statistics has been outperformed by deep learning. The theory is seemingly the only remaining advantage of statistics. The reason for this, in my opinion, lies in the era of big data. When data is abundant and the signal is strong, deep learning can simply perform brute force fitting and achieve remarkable results through sheer computational power. In contrast, in the early days, data in various fields was either limited in scale or highly noisy. The datasets available often failed to represent the true distribution. As a result, the statistics that developed in those times tended to focus on small-scale or high-noise datasets. While data fitting, including estimation and prediction, is crucial, people were also very concerned about the reliability of the estimates, that is, inference. This is a different path from the current AI. Does this mean that statistics is now useless? I don't think so. Firstly, many fields are still plagued by data scale or noise issues, such as genetic data and financial data. Moreover, in many scientific fields, people often care less about prediction accuracy and more about whether the method is interpretable and whether it can answer the questions that scientists are concerned with. This is precisely where statistics excels, such as in probabilistic modeling, inference, and testing. Deep learning finds it difficult to handle these tasks casually through a network. Instead, it requires a deep understanding of the dataset, skillful modeling, and some statistical analysis techniques. In many cases, these techniques have already been largely figured out by statisticians decades ago. I've read some applied papers before and can't help but marvel at the intelligence of the researchers. Their understanding of data, models, and statistics is truly impressive. For example, in genetics, there are well-known methods like LD Score regression and SuSiE, and in omics data analysis, there are methods like RCTD and GLUE. Reading their papers makes one exclaim: "Statistics is more like an art!" From a pragmatic perspective, if done well, the upper limit of such work could be a top-tier journal or Nature Communications (NC), although getting into the top-tier journals of the CNS group is still extremely challenging. In terms of impact, these journals are certainly more influential than those in the small circle of statistics.” [5]

Future Directions for Statistics

The Importance of Interdisciplinary Collaboration

As big data and AI technologies evolve, statistics must engage in closer collaboration with other disciplines. The intersection of biostatistics with bioinformatics, medicine, and computer science offers new approaches to complex biomedicine challenges. Interdisciplinary collaboration enhances research efficiency and fosters knowledge exchange and innovation across fields.

Education and Practice in Statistics

The education and practice of statistics must adapt to contemporary demands. University curricula should focus on developing students' interdisciplinary thinking and practical skills, encouraging participation in research projects and academic exchanges. The application of statistics should be grounded in real-world problems, emphasizing data quality and model interpretability to maximize its impact in scientific research and practical applications.

Conclusion

Statistics holds significant value in the biomedicine domain, particularly in drug development and medical research. Despite its foundational support to AI, statistics also retains its irreplaceable role in interpretability and model construction within the medical field that could not be replaced by AI. The future of statistics lies in closer collaboration with other disciplines, nurturing talents with interdisciplinary perspectives and practical skills to address the increasingly complex issues in biomedicine. We anticipate that statistics will continue to play a crucial role in biomedicine, contributing to advancements in human health.

References

- [1] Salsburg, D. (2002). The lady tasting tea: How statistics revolutionized science in the twentieth century. Macmillan.
- [2] Diaconis, P., & Skyrms, B. (2018). Ten great ideas about chance. Princeton University Press.
- [3] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
- [4] Fisher, R. A. (1925). Statistical Methods for Research Workers. Oliver and Boyd.
- [5]https://www.zhihu.com/question/624834153/answer/3243406072?utm_medium=social&utm_source=wechat_timeline