A review of recent researches for student t-distribution

Yuhuan Ma

Nanjing Foreign Language School, Nanjing, Jiangsu, China, 210008

15996267211@163.com

Abstract. The t-distribution, often known as the student t-distribution, can be referred to simply as the t-distribution. The development of the aforementioned concept was initially documented by William Gosse in the year 1908, during his tenure at the renowned Guinness Distillery located in Dublin. Due to the author's inability to publish under his own name, the piece employed the pseudonym "Student". The t-test and its associated theories were further advanced by Ronald Fisher, who referred to this particular distribution as the student distribution. In the field of statistics, the t-distribution is employed to make estimations regarding the mean value of a population that conforms to a normal distribution with an unknown variance. This work employs a literature review methodology to examine the historical development of research on the t-distribution. Additionally, it investigates recent advancements in t-distribution research and its use in many domains. This work aims to offer readers a comprehensive understanding of the various implications associated with the t-distribution, thereby shedding light on potential avenues for future research in this area.

Keywords: T-Distribution, Statistics, Mathematics.

1. Introduction

The t-distribution, also known as the student t-distribution, is a statistical distribution that can be commonly referred to as the t-distribution. The derivation of the aforementioned concept was initially published by William Gosse in the year 1908, during his tenure at the Guinness Distillery located in Dublin. Due to the author's inability to write under his actual identity, the paper opted to employ the pseudonym "Student." The development of the t-test and its associated ideas can be attributed to the contributions of Ronald Fisher, who referred to this particular distribution as the student distribution.

In recent times, extensive study has yielded numerous novel insights into the t-distribution, rendering it applicable across various dimensions. This work employs a literature review methodology to examine the historical development of research on the t-distribution. It also investigates the contemporary advancements in t-distribution research and its utilization in many domains. This work aims to offer readers a comprehensive understanding of the various findings associated with the t-distribution, so shedding light on potential avenues for future research in this area.

2. Overview of the research into t-distribution

The T-distribution was introduced by William Sedgwick, who formulated it based on empirical investigations conducted in the context of beer brewing. Simultaneously, Bessel employed this particular distribution in his publication titled "The Theory of Probability." Subsequently, numerous renowned

DOI: 10.54254/2753-8818/9/20240753

statisticians globally, including Spearman, Pearson, Nieman, Fisher, and others, have employed the T-distribution in their scholarly investigations. In the year 1927, Sneijder formally designated the statistical distribution as the "T-distribution".

2.1. *A new fault diagnosis method*

In recent times, the advancement in statistical research has led to the identification of additional uses for the t-distribution. In the year 2023, a proposal was put forth by Chinese academics ZHU Xihai-lun and YI Can-can. The proposal outlined a method for mechanical fault identification that is grounded in t-distribution random neighbor embedding. The primary objective is to establish historical monitoring indicators for mechanical failures, thereby creating an extensive repository of unique features. The t-SNE dimension reduction algorithm is employed to extract the primary characteristics of the matrix representing mechanical failure signals. The clustering algorithm is utilized to identify the clustering center for each sample period. The eccentric distance is then computed for each sampling instance. Accurate fault prediction has been achieved using the accumulation of the normalized eccentricity distance matrix. The findings indicate that the method developed in this study is capable of effectively discerning various failure modes with a high level of accuracy. Therefore, it is advantageous to guarantee the optimal functioning and reliability of the equipment [1].

2.2. *Application of t-distribution in computer science*

In the field of computer science, BI Xiu-Chun and YANG Hao-Feng utilized the t-distribution within the framework of Bayesian Deep Learning (BDL) Model. A BDL framework was developed by employing the t distribution and the cyclic stochastic gradient Hamiltonian Monte Carlo sampling procedure. In order to assess the credibility and practicality of the framework, this research establishes corresponding Bayesian Deep Learning (BDL) models utilizing three distinct types of neural networks: artificial neural network (ANN), convolutional neural network (CNN), and recurrent neural network (RNN). These models are subsequently employed to forecast the behavior of 15 global stock indices. The empirical findings indicate that the proposed framework demonstrates applicability across various neural network architectures, namely Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN). Furthermore, the predictive performance of all indices within the framework is found to be excellent. Moreover, when considering prediction accuracy and applicability, the Bayesian Deep Learning (BDL) models based on the t distribution exhibit notable advantages over those based on the normal distribution. Additionally, the Mean Absolute Error (MAE) achieved under a specific uncertainty threshold surpasses the original MAE. This observation suggests that the uncertainty measure defined in this study is effective and holds significant importance in the context of uncertainty modeling. Considering the several benefits offered by the BDL framework, such as enhanced forecasting accuracy, ease of scalability, and the ability to evaluate forecasting uncertainty, it is evident that this framework holds significant potential for application in finance and other domains characterized by complex data attributes [2].

2.3. *Application of t-distribution in mathematics*

In the field of mathematics, researchers Li Yongyi, Zhang Jianzhu, and Lian Wei utilized the t-distribution for the purpose of optimization. The proposed approach involves incorporating an adaptive t-distribution variation operator into the slime mold algorithm. The number of iterations of the algorithm is utilized as the degree of freedom parameter of the t distribution. This modification aims to enhance the diversity of the slime mold search and prevent the algorithm from converging to local optima. The study conducted experiments on 12 benchmark functions and the multi-threshold segmentation point engineering problem in image analysis. The findings indicate that the slime mold optimization technique, when based on t-distribution, has superior optimization accuracy, convergence speed, and robustness compared to the conventional slime mold algorithm. The user's text does not provide any information to rewrite in an academic manner. Furthermore, within the same field, researchers Huang Zhen and Wu Jun utilized the student t-distribution within the Bayesian Unscented Kalman Filter (UKF) technique, and subsequently in the context of unmanned ship alignment. This study introduces a novel algorithm, namely the adaptive robust variational Bayes Unscented Kalman Filter (St-VB UKF). The algorithm incorporates the student's t-distribution to model the measured noise, thereby capturing the thick-tail characteristics of the noise. Additionally, it employs the variational Bayes method to estimate the statistical properties of the measured noise in real-time, allowing for adaptive estimation [3]. The simulation findings demonstrate that the proposed approach outperforms both the classic Unscented Kalman Filter (UKF) technique and the Sage-Husa adaptive UKF algorithm in terms of alignment accuracy and speed, particularly in situations where short-term and instantaneous significant interference is present. Furthermore, in cases where there exist outliers in the equivalency and the presence of thick-tail features, the method being provided demonstrates the ability to successfully mitigate the impact of these extreme values. As a result, the accuracy of alignment remains unaffected, hence enhancing the overall resilience of the algorithm [4].

2.4. Application of t-distribution in the planning of aerial vehicles

In June 2022, a group of researchers, namely Li Nan, Xue Jiankai, and Shu Huisheng, conducted a study on the sparrow search algorithm utilizing t-distribution. They then used this algorithm in the domain of aerial vehicle planning. In this study, we have developed a model for UAV track planning that incorporates both the three-dimensional space model and the track cost model. By transforming the track planning problem into a multi-dimensional function optimization problem, we have successfully optimized and solved it. The findings from the simulation results demonstrate that the enhanced algorithm exhibits superior performance in UAV path planning compared to both the traditional particle swarm algorithm and sparrow search algorithm. Specifically, the improved algorithm successfully identifies an optimal path that effectively avoids obstacles and threat areas. This outcome serves as evidence supporting the enhanced algorithm's enhanced robustness and feasibility in UAV path planning [5].

2.5. Application of t-distribution in cyber-physical fusion system

The application of t-distribution in the domain of particle swarm was undertaken by researchers Liu Zilai and Wang Jianmin. In response to the challenges of low efficiency and the inability to satisfy multiple quality of service (QoS) demands of users in real-time task scheduling within cyber-physical fusion systems (CPS), this study presents an enhanced particle swarm optimization (PSO) algorithm for real-time task scheduling (referred to as t-PSO) that utilizes an adaptive t-distribution. The approach incorporates the adaptive t-distribution variation mechanism to enhance convergence speed and prevent the algorithm from being trapped in local optima, building upon the conventional particle swarm operation (PSO). Additionally, the fitness function is configured to perform task scheduling by considering the metrics of task completion time, total task cost, and service quality. The job scheduling simulation experiment is ultimately conducted using both the particle swarm optimization and Cauchy modified particle swarm method (Cauchy-PSO). The findings indicate that, when subjected to identical experimental settings, the t-PSO algorithm exhibits superior overall performance compared to the other two algorithms. Specifically, the t-PSO approach demonstrates significantly better performance in terms of job completion time, total task cost, and service quality [6].

2.6. Application of t-distribution in engineering and arithmetic optimization algorithm

Scientist Xia Yinxiang conducted a research study in the field of engineering, specifically focusing on the reliability of concrete using the t-distribution. This study examines the reliability analysis of concrete strength in a sample of buildings from the construction project in a small farmland water conservancy key county in Ninghua County in 2016. It establishes a random variable representing the concrete strength, which follows a t-distribution, and conducts a sample analysis to assess the reliability of concrete strength. The study also aims to provide an overall perspective on the application of concrete in the project [7]. Furthermore, a group of researchers consisting of Zheng Tingting, Liu Sheng, and Yie Xu conducted a study on arithmetic optimization utilizing the t-distribution. This study proposes an arithmetic optimization algorithm (t-CAOA) that addresses the issues of sluggish convergence speed and susceptibility to local optima commonly observed in arithmetic optimization algorithms (AOAs). The suggested t-CAOA incorporates adaptive t-distribution variation and dynamic boundary strategy enhancement to enhance its performance. The incorporation of the adaptive t-distribution variation strategy has been shown to enhance the diversity and quality of the population, leading to improved convergence speed of the algorithm. Additionally, the optimization process of AOA can be further enhanced by introducing a dynamic boundary strategy for the cosine control factor, which effectively balances the global exploration and local exploitation capabilities of AOA [8].

2.7. Application of t-distribution in economy

Scientist Zong Haoqian utilized the t-distribution within the framework of a stochastic volatility model to analyze the dynamics of the stock market. This paper commences by presenting the developmental procedure of the financial volatility model, along with an overview of the current research status and the utilization of the Markov Chain Monte Carlo (MCMC) method and Bayesian theory for estimating model parameters. Subsequently, the paper conducts a KS test and kurtosis test on stock market returns, revealing that the returns of the A-share index and Hong Kong Hang Seng index exhibit a pronounced and heavy-tailed distribution. The SV-T model is derived by extending the standard random fluctuation model to incorporate a t-distributed error component for the yield series, based on the aforementioned properties. The concept of risk and return coexistence is examined, with particular attention given to risk compensation. The SV-MT model is derived using the t-distribution. Additionally, the SV-leverage-T model is developed by taking into account the "leverage effect" of the stock market and utilizing the t-distribution. Furthermore, the Gibbs algorithm and Metropolis-Hastings algorithm inside the Markov Chain Monte Carlo (MCMC) framework are employed for parameter estimation and numerical computation of the model using WinBUGS software. Additionally, the Deviance Information Criterion (DIC) values are acquired for each model in order to determine the most suitable model. The DIC criterion is utilized to construct a random wave model that incorporates the "leverage effect" using a t-distribution. The ideal model for both A-share and Hong Kong stock markets is SV-leverage-T, as stated in reference [9].

2.8. Application of t-distribution in image segmentation

To enhance the precision and resilience of pavement crack image segmentation, a novel approach is introduced. This approach utilizes the student's t-distribution mixture model (TMM) as a replacement for the conventional Gaussian mixture model (GMM) in pavement crack image segmentation. The proposed methodology initially employs the K-Means algorithm for the purpose of conducting the initial image segmentation. To address the limitations of the K-Means algorithm, researchers have proposed utilizing the outcomes of the Fireworks Algorithm (FA) as the initial cluster centers for the K-Means method. The initial segmentation is performed using the K-Means algorithm, and the resulting segmentation is utilized as the initial value for the TMM parameter solution. Subsequently, the Expectation-Maximization algorithm (EM) is employed, wherein the final parameter value of the model is obtained by iteratively executing the E step and the M step. Finally, the image segmentation is completed by applying Bayes' formula. Ultimately, by conducting a comparison between simulated images and real images, the findings indicate that the proposed method exhibits superior accuracy in segmentation and enhanced stability. The user's text is already academic and does not require any rewriting. Moreover, a group of researchers consisting of Xu Chao, Zhan Tianming, Kong Lingcheng, and Zhang Hui have developed a resilient hierarchical fuzzy algorithm that utilizes Student's t-distribution for the purpose of picture segmentation [10]. This study presents a novel hierarchical fuzzy C-means method that enhances the robustness of the conventional fuzzy C-means algorithm against picture noise and outliers. Based on this premise, a more adaptable function is proposed, wherein the distance function is considered as a sub-student t-distribution function. Enhance the versatility and flexibility of the hierarchical model. The technique presented in this research work has the potential to be used to many algorithm implementations that rely on FCM models, hence enhancing their resilience.

The experimental findings demonstrate that the efficacy of the newly introduced hierarchical fuzzy C-means method, as shown in this research work, is certainly substantial [11].

2.9. Application of t-distribution in biology

Yuan Rui, a scientist, utilized the t-distribution in their investigation of gene network rewiring. This research introduces a novel approach to infer the underlying network disturbances under various scenarios by utilizing the multivariate student t-distribution and proposing a more robust difference graph model. Due to its heavy-tailed structure, the multivariate distribution exhibits more resilience to anomalies compared to normal distributions. Simultaneously, the fused lasso penalty is employed to account for the homogeneity of the network across various situations. The present study employs the expectation maximization algorithm as a means to address the optimization problem at hand. The simulation results indicate that our approach has superiority over the comparative method. The methodology outlined in this scholarly article was utilized to examine several subtypes of breast cancer and glioblastoma. Through this analysis, several significant factors contributing to the disruption of the motivational network were identified [12].

3. Conclusion

In summary, the t-distribution is a valuable statistical tool that finds application in various statistical analyses. Currently, this distribution is utilized across various disciplines, including mathematics, engineering, computer science, and biology, among others. However, due to the increased need for mathematical proficiency in scientific research, it is imperative to explore further applications of the t-distribution. The increased adoption of this distribution in the future would facilitate scientific research, rendering it more accessible and streamlined.

References

[1] ZHU Xi, Helen, YI Cancan. Mechanical fault classification method for improved clustering of stochastic neighbor embeddings in improved T-distribution [J]. Mechanical Design and Manufacturing,2023(03):5-10.DOI:10.19356/j.cnki.1001-3997.20230216.005.)

[2] Bi Xiuchun, Yang Haofeng. Bayesian deep learning model based on t-distribution and its application[J]. Computer Systems and Applications,2022,31(11):330-338.DOI:10.15888/j.cnki.csa.008787.)

[3] LI Yongyi, ZHANG Jianmei,LIAN Wei. Adaptive slime mold optimization algorithm based on t-distribution variation [J]. Journal of Shanxi Datong University(Natural Science Edition),2022,38(06):40-44.)

[4] HUANG Zhen, WU Jun. Application of variational Bayesian UKF algorithm based on student's t-distribution in unmanned ship alignment [J]. Chinese Journal of Sensors and Actuators,2022,35(10):1340-1347.)

[5] LI Nan, XUE Jiankai,SHU Huisheng. UAV track planning based on adaptive t-distribution variant sparrow search algorithm[J]. Journal of Donghua University(Natural Science Edition),2022,48 (03):69-74.DOI:10.19886/j.cnki.dhdz.2021.0223.)

[6] LIU Zilai, WANG Jianmin. Improved real-time task scheduling algorithm of particle swarm based on adaptive t-distribution[J]. Automation & Instrumentation in Chemical Industry,2020,47(05):393-397+424.)

[7] Xia Yinxiang. Reliability analysis of concrete strength based on t distribution [J]. Water Conservancy Science and Technology,2020(02):53-55.)

[8] ZHENG Tingting, LIU Sheng, YE Xu. Arithmetic optimization algorithm for improved adaptive t-distribution and dynamic boundary strategy [J]. Application Research of Computers,2022,39(05):1410-1414.DOI:10.19734/j.issn.1001-3695.2021.09.0428.)

[9] Zong Haoqian. Stock market stochastic fluctuation model based on t-distribution [D]. Guangxi Normal University, 2018.)

DOI: 10.54254/2753-8818/9/20240753

[10] Lu Yinju, Duan Mingyi, Li Zuzhao, et al. Improved image segmentation method of pavement cracks based on t-distribution mixing model [J]. Highway,2021,66(04):40-47.)
[11] XU Chao, ZHAN Tianming, KONG Lingcheng, et al. Robust hierarchical fuzzy algorithm based on student's t-distribution and its application in image segmentation[J]. Acta Electronica Sinica,2017,45(07):1695-1700.)
[12] YUAN Rui. Research on gene network perturbation analysis method based on multivariate student t-distribution [D]. Central China Normal University,2020.DOI:10.27159/d.cnki. ghzsu.2020.000108.)