test

by Test Test

Submission date: 11-Jan-2024 11:16PM (UTC-0600)
Submission ID: 2269775416
File name: userfile (296.72K)
Word count: 3643
Character count: 20061

Salahaddin University – Erbil

College of Administration and Economics

High Education: MSc.

Department: Statistics and Informatics

Subject: Non parametric

First Semester

Kolmogorove-Smirnov test

Maria Saber Ibrahim

Email:

aest15918@student.su.edu.krd

Supervision:

Asst. Prof. Dr. Nazeera Sedeek Kareem

2023-2024

ABSTRACT

The KS test is a key statistic tool for checking if an observed distribution and the theory of probabilities match each other. This test is named after Andrey Kolmogorov and Nikolai Smirnov. It's widely used in science areas like astronomy, physics, economics and biology. This work aims to give a full explanation about the Kolmogorov test, including its basic ideas, real use and results. The test is about finding the biggest difference between a theory's distribution and real-life results. By comparing, the scientists can see if a group of data matches with what is expected in theory (Berger, Vance W., & YanYan Zhou). "Kolmogorov-smirnov test: "Intro. The KS test is very useful and has some big benefits, most importantly its ability to handle different sizes and shapes of distributions because it doesn't need any particular guesses about the characteristics of a population. It also responds to changes in distribution and position, making it a useful tool for researchers & statisticians. This study looks at the main steps of the test. It talks about making guesses, calculating a special number from these, and figuring out if they matter or not for judging things well. It looks into how results need to be understood, describing situations where scientists can agree or disagree with the null hypothesis.It also gives words of caution and limits. This simple talk looks at the big parts of testing, like coming up with ideas, doing math tests and finding numbers. These important things help us make decisions. This looks at how we have to understand the outcomes. It talks about situations when scientists can decide if they should agree or disagree with a no-effect idea, along with warnings and tips. This review talks about changes to the KS test, like multi- and two-sample versions. These improvements make it more helpful in different situations where you need to analyze things. In short,

this work is a helpful guide for all types of workers. It teaches them about the Kolmogorov-Smirnov test and how to use it in stats analysis and testing guesses.

KEYWORDS

"Goodness-of-fit, statistical test, distribution comparison, independence check."

INTRODUCTION

The Kolmogorov-Smirnov test, sometimes called the Kolmogorov test is a basic statistical method that helps to check if data observed, and the related probability distribution fit together well. This test is named after two mathematicians, Nikolai Smirnov & Andrey Kolmogorov. It's a big method in areas like biology, engineering, and economics. Scientists want to check the ideas about spreading their results or facts that back up everything they know. The K-S test is used to check if a set of data fits into a chance model or an assumed plan. It studies the largest difference between the total function of distribution (CDF) proposed in theory and how data is collected. These are compared using statistical methods known as CDFs to better understand real-world observations. The exam works out a number, usually shown by the letter "D," to show this biggest space.

When the real and imagined patterns are more different, that means a bigger disparity. The nude idea says that the data finds for this test should be taken from given shape. It's like a guessing game when information comes from rolling dice or splitting them in certain ways which are already decided upon by someone beforehand so they can see if things match up closely enough, showing not much difference between what was hoped for and reality checking inside method of making these decisions. So, a small number of D shows that the observed data matches well with what was expected. This confirms if there's no need to reject any idea (null hypothesis) because mostly used words in English include up to 2000 ones for basic discussions and understanding complex issues more simply & effectively friendly ways do exist available nowadays too provide better initial checking results split these details further A high value of D means that there's a big difference between the sample data and what is expected.This supports not believing or choosing the null hypothesis, and instead picking another distribution as the right answer. The Kolmogorov-Smirnov test is very easy to use and works for lots of things. That's why experts like Dr. Fasano, Mr Giovanni, and Alberto Franceschini often use it when they do their research (Fasano et al., The astronomical society journal). 1. 1987 was the year of something good, which we now call "statistics". It's useful for looking at a lot of information because you can use it in many forms normal like what happens every day, fast-rising like how things wear out over time and stay all even spread everywhere else more changes come up too! It's important to remember, though, that the test might be hard on size of samples. This could happen more for small sets of data where we use a bit change from what was expected and still get statistical meaning. In short, the Kolmogorov-Smirnov test is still a big tool in studying numbers. This because it offers true means to see how accurate real data

agrees with expected distribution patterns. This helps scientists see if suggested versions are right for the group of data they're testing with.

Main Body

Certainly! A study of books looks at the history, uses and changes in statistical research that affect the Kolmogorov test. People don't always agree on this change too. This paper will talk about how important the Kolmogorov-Smirnov (KS) test is. It will also discuss if it's good or bad, different changes made to meet its problems in other subjects and new things added that make this better at those issue as well. Without a doubt! A study of literature includes looking at the history and use of a Kolmogorov test. It also talks about challenges it had or was improved for work with stats in research science writings. This article will talk about the importance of Kolmogorov-Smirnov test. It will also cover its pluses and minuses, different versions created to fix issues in other fields too.

Historical Background Information

The main ideas of the test were first created by famous math person Andrey Kolmogorov in a very important article about probability theory from early 1900s. Nikolai Smirnov came up with a good way to check if real-world data matches the idea we have in our heads. This led us to create something called KS test based on Kolmogorov's work.

The Kolmogorov-Smirnov Test

The KS test contrasts two cumulative distributions operates: the fake pattern thought to show real data creation, and the true one that comes from real-life facts. The test number, usually marked as "D", calculates the largest difference between both patterns of overall distribution. If the outcomes we see are just like what was expected, the test number gets smaller (Lopes RH, Reid ID, Hobson PR).

Uses in Different Areas of Life

The KS test is helpful in many jobs. By looking at how closely money information matches to theory rules, it aids in finding the level of risk and costing assets properly in finance business. It helps study how life evolves by matching observed character patterns with possible future expectations. Moreover, it helps in deciding if the ways of making things are meeting agreed-upon quality checking rules.

Advantages and Disadvantages

The KS test can be used to study many types of distribution without making any guesses about the population. This is because it doesn't need special information like most tests do, which makes it useful for different things! It's very helpful when you can't use much data or some rules about the information are broken. With less data, the test might not be strong enough and may work worse for different spreads of numbers or on things other than place like size shape.

Progress and Change

The first KS test had problems, so scientists fixed them and made it better. The Lilliefors test modifies the KS tests for normality testing in a way. It's like changing them just a little bit to make sure they work better with things that follow this normal pattern of numbers or shapes called "normality". For example, tests such like the Anderson-Darling and Kuiper's are more likely to spot changes in distribution patterns.

Observations and Difficulties

Even though the KS test is quite popular and used a lot, it has faced criticism too. Several experts say that when things like size and types of data get more complicated, being sensitive to these changes can go down. Moreover, the two-way decision of accepting or not choosing between our starting point idea could make checking how well it fits too simple (Massey Jr., Frank J. 1951).

Current Developments and Future Directions

With time, complex statistic ways have increased. Nowadays they are using machine learning-based methods more often. Sometimes, these steps might be enough to fully take the place of or just add on top what's done in a KS test. Besides that, efforts to find solutions for complex problems and extensions not related to numbers keep influencing the track of evaluating how well something fits. In the end, the Kolmogorov-Smirnov test is a simple statistical method that gives you an easy way to measure how well two types of distribution fit together. It's good for checking real life data with theory ideas and can be used in many situations. Despite its problems, changes and improvements keep making it better fit for many uses. This also allows

us to use very accurate statistics in different areas of life nowadays.This look at books and articles talks about the basics, using it in different ways people use numbers. They also discuss some difficulties they had to solve or improvements others made with a special test called the Kolmogorov-Smirnov test. It also looks at how the test is used for statistic measuring and what impact it has on other learning subjects.

REVIEW

(Young, Ian T.,1977) The Kolmogorov-Smirnov test is a statistical technique used to ascertain whether two datasets differ significantly from one another. It is also commonly referred to as the KS test or just the K-S test. This test, which bears the names of Andrey Kolmogorov and Nikolai Smirnov, is a nonparametric assessment that determines if two datasets have the same underlying distribution or follow a similar distribution by comparing their cumulative distribution functions (PDFs). The KS test is widely used because it offers a flexible way to compare distributions without assuming anything about the particular distributions of the datasets. This makes it useful in a variety of domains, including biology, economics, statistics, and more. Because of its adaptability, it is useful in situations where assumptions on the distribution of the data may be hazy or unclear. Knowing How to Take the Kolmogorov-Smirnov Test.

(Justel, Ana, Daniel Peña, and Rubén Zamar,1997) The primary goal of the KS test is to determine which of two datasets' empirical cumulative distribution functions (ECDFs) has the highest vertical difference. The fraction of observations in a dataset that are less than or equal to a given value is represented by the ECDF. The KS test compares these empirical distributions to find the largest difference, which measures how similar or dissimilar the datasets are.

The KS test comes in two varieties: One-Sample KS Test: This variation determines if a single dataset follows a certain distribution (such as the normal distribution) by comparing it to the distribution. The Two-Sample KS Test determines if two independent datasets are taken from the same underlying distribution by comparing them.

(Büning, 2002) They have a variety of statistical tests for the well-known two- sample dilemma, including parametric, non-parametric, and robustified tests that account for various forms of alternatives and underlying distributions. If the results are assumed to be regular, the t-test and the F-test are consistently the most effective tests for determining the equality

between two means and two variances, respectively, and it is well established that the t-test is resilient to non-normal distributions, in comparison to the F-test, which is highly insensitive to non- normal data. For general alternatives, where the distribution functions of the X- and Y-variables can have different forms, the Kolmogorov-Smirnov and Cramér- von Mises tests seem to be the most common. These experiments are robust in the presence of symmetric and long-tailed distributions. Additionally, they propose modifying all experiments by using sufficient weight functions in order to achieve a higher power than the classical equivalents for short-tailed and right- skewed distributions. The aim of this paper is to provide a detailed analysis of the robustness and strength of both of these tests for position and scale alternatives of equivalent and different shapes of the underlying X- and Y-variables distributions, assuming small, medium, and long tails as well as right-skewed distributions. The latter example, in particular, is critical in mathematical work. These measurements are compared to the t-test and other position tests, such as the Welch test, as well as the Levene scale test. Naturally, there is no definite winner among the studies; nevertheless, for symmetric distributions of the same form, Lepage style tests are preferred; for distributions of different shapes, Cramer-von Mises type tests are preferred. A variant of the Kolmogorov-Smirnov test can be used with highly right-skewed distributions.

(Aslam, 2019) The Kolmogorov-Smirnov (K-S) tests were popularly used for data interpretation based on conclusions of determined findings in the dataset. Existing K-S experiments for one and two samples cannot be applied whether the results include neutrosophical observations, determined by or under the complex method. In this article, they suggest the generalisation under neutrosophical statistics of the current K-S experiments. The suggested experiments are classified as Kolmogorov-Smirnov (NK-S) neutrosophical tests. They present the steps and procedures required to carry out the proposed experiments. The paper gives an example and benefits of the proposed NK-S experiments.

(Massey Jr, F. J. ,1951)

Important Ideas for the Kolmogorov-Smirnov Test: Alternative Hypothesis (H_1) and Null Hypothesis (H_0)

Hypothesis Null (H_0): supposing that the datasets are drawn from an identical distribution.

H_1, or the alternative hypothesis, postulates that the datasets originate from various distributions.

Statistic of Test (D) The greatest vertical difference between the two ECDFs is represented by the KS test statistic (D). The supremum of the absolute differences between the two ECDFs is used to calculate it.

(Wilcox, R. R. ,1997)

P-value and Critical Value

Critical Value: Based on the selected significance level (α) and sample size, critical values are derived from the KS distribution tables.

P-value: To ascertain the statistical significance of the test, the probability (p-value) connected to the test statistic is contrasted with the selected significance level. A lower p-value denotes more robust evidence opposing the null hypothesis**.**

(Wang, Fanggang, and Xiaodong Wang,2010)

Procedure for Conducting the Kolmogorov-Smirnov Test

Construct Hypotheses: Given the type of comparison being done, define the null and alternative hypotheses.

Collect Data: Assemble the datasets for comparison, making sure they are impartial and accurate representations of the populations being studied.

Determine the empirical cumulative distribution functions (ECDFs) for every dataset.

Find the largest vertical difference (D) between the two ECDFs in order to compute the test statistic (D). Find the P-value, or critical value: Utilizing statistical tools, determine the p-value or compare the test statistic (D) to the crucial value from the KS distribution table[12].

Make a Decision: Determine whether to reject or not to reject the null hypothesis based on the p-value and the selected significance level (α).

The Kolmogorov-Smirnov Test's Applications

There are numerous fields in which the KS test finds use:

Testing for goodness-of-fit: Determining whether a dataset conforms to a particular theoretical distribution.

(Drezner, Zvi, Ofir Turel, and Dawit Zerom,2010) Finding out if two datasets are from the same distribution requires comparing distributions.

Model Validation: Assessing statistical models for suitability. Analyzing differences in production processes is known as quality control.

Benefits and Drawbacks

Benefits

The nonparametric character doesn't need any presumptions regarding the distribution of the data.

Applicability: Both discrete and continuous data can be used.

Sensitivity: The capacity to identify variations in a range of distributional parameters.

Sensitivity to sample size: Small sample sizes may have an impact on performance.

Emphasizes shape variations rather than dispersion or location.

Maybe unable to detect minute variations in distributions.

(Zhang, G., Wang, X., Liang, Y.C. and Liu, J., 2010) A useful statistical method for comparing datasets and determining goodness-of-fit is the Kolmogorov-Smirnov test. Its broad use in a variety of domains is due to its nonparametric nature and adaptability. This test provides a reliable way to ascertain whether two datasets originate from the same underlying distribution by looking at the largest difference between empirical distributions. Users should, however, be aware of its limits and proceed with caution when interpreting the results, particularly when there are small sample numbers or when minute variations in distributions are significant **.**

(Durbin, J., 1975) All things considered, the KS test continues to be a vital tool in statistical analysis, greatly assisting with model validation, distribution comparison, and hypothesis testing in a variety of fields. Its simplicity and adaptability guarantee that it will always be useful for data analysis and decision-making procedures .

In summary:

To sum up, the Kolmogorov test is still a fundamental technique in statistical analysis and is useful for determining how well actual and theoretical distributions match. Because of its adaptability and simplicity of use, it can be applied in a variety of sectors and helps practitioners and researchers make well-informed judgments based on data analysis.To guarantee accurate and trustworthy evaluations, it is crucial to take into account its limitations and investigate more sophisticated alternatives while also appreciating its merits. Given its long history and versatility, the Kolmogorov test is expected to remain an essential statistical tool for many years to come as statistical techniques continue to advance.

Conclusion

The Kolmogorov-Smirnov test is a powerful statistical tool used to assess the similarity between two probability distributions or to compare a sample distribution to a reference distribution. Its conclusion typically revolves around accepting or rejecting the null hypothesis. If the calculated test statistic exceeds the critical value at a chosen significance level, the null hypothesis (often stating that the two distributions are identical) is rejected, suggesting a significant difference between the distributions. Conversely, a test statistic below the critical value leads to the acceptance of the null hypothesis, indicating similarity between distributions.

In the context of research or analysis, the implications of the Kolmogorov-Smirnov test results can guide decision-making. When rejecting the null hypothesis, it implies that the two distributions significantly differ, prompting a need for further investigation to understand the nature and causes of this dissimilarity. On the other hand, if the test fails to reject the null hypothesis, it suggests that the distributions are similar, offering confidence in utilizing the sample distribution as a representative of the population.

Recommendations

Recommendations following the Kolmogorov-Smirnov test involve exploring the factors contributing to any identified differences between distributions, conducting additional analyses or tests to corroborate findings, and potentially adjusting methodologies or considering alternate models if significant disparities persist. Moreover, communicating the results transparently is crucial for informed decision-making in scientific research, ensuring appropriate interpretation and application of the findings.

REFERENCES

1. Berger, Vance W., and YanYan Zhou. "Kolmogorov–smirnov test: Overview." Wiley statsref: Statistics reference online (2014).
2. Fasano, Giovanni, and Alberto Franceschini. "A multidimensional version of the Kolmogorov–Smirnov test." Monthly Notices of the Royal Astronomical Society 225, no. 1 (1987): 155-170.
3. Lopes RH, Reid ID, Hobson PR. The two-dimensional Kolmogorov-Smirnov test.
4. Massey Jr, Frank J. "The Kolmogorov-Smirnov test for goodness of fit." Journal of the American statistical Association 46, no. 253 (1951): 68-78.
5. Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American statistical Association, 62(318), 399-402.
6. Goodman, Leo A. "Kolmogorov-Smirnov tests for psychological research." Psychological bulletin 51.2 (1954): 160.
7. Crutcher, H. L. (1975). A note on the possible misuse of the Kolmogorov-Smirnov test. Journal of Applied Meteorology (1962-1982), 1600-1603.
8. Young, Ian T. "Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources." Journal of Histochemistry & Cytochemistry 25, no. 7 (1977): 935-941.
9. Justel, Ana, Daniel Peña, and Rubén Zamar. "A multivariate Kolmogorov-Smirnov test of goodness of fit." Statistics & probability letters 35, no. 3 (1997): 251-259
10. Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association, 46(253), 68-78.
11. Wilcox, R. R. (1997). Some practical reasons for reconsidering the Kolmogorov-Smirnov test. British Journal of Mathematical and Statistical Psychology, 50(1), 9-20.
12. Wang, Fanggang, and Xiaodong Wang. "Fast and robust modulation classification via Kolmogorov-Smirnov test." IEEE Transactions on Communications 58, no. 8 (2010): 2324-2332
13. Drezner, Zvi, Ofir Turel, and Dawit Zerom. "A modified Kolmogorov–Smirnov test for normality." Communications in Statistics—Simulation and Computation® 39, no. 4 (2010): 693-704.
14. Zhang, G., Wang, X., Liang, Y.C. and Liu, J., 2010. Fast and robust spectrum sensing via Kolmogorov-Smirnov test. IEEE Transactions on Communications, 58(12), pp.3410-3416.
15. Durbin, J., 1975. Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika, 6 2(1), pp.5-22.
16. Mason DM, Schuenemeyer JH. A modified Kolmogorov-Smirnov test sensitive to tail alternatives. The annals of Statistics. 1983 Sep 1:933-46.
17. Büning, H., 2002. Robustness and power of modified Lepage, Kolmogorov-Smirnov and Crame' r-von Mises two-sample tests. Journal of Applied Statistics, 29(6), pp.907-924.
18. Aslam, M., 2019. Introducing Kolmogorov-Smirnov tests under uncertainty: an application to radioactive data. ACS omega, 5(1), pp.914- 917.

test

ORIGINALITY REPORT

12%	7%	9%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Herbert Büning. "Robustness and power of modified Lepage, Kolmogorov-Smirnov and Crame´r-von Mises two-sample tests", Journal of Applied Statistics, 8/1/2002 Publication	3%
2	Stefano Bonnini, Livio Corain, Marco Marozzi, Luigi Salmaso. "Nonparametric Hypothesis Testing", Wiley, 2014 Publication	1%
3	pubmed.ncbi.nlm.nih.gov Internet Source	1%
4	www.researchgate.net Internet Source	1%
5	epublications.marquette.edu Internet Source	1%
6	www.tandfonline.com Internet Source	1%
7	open.uct.ac.za Internet Source	<1%

8	link.springer.com Internet Source	<1 %
9	www.coursehero.com Internet Source	<1 %
10	Submitted to South Bank University Student Paper	<1 %
11	profmsaeed.org Internet Source	<1 %
12	rsdjournal.org Internet Source	<1 %
13	Tomas Borovicka, Marcel Jirina, Pavel Kordik, Marcel Jiri. "Chapter 2 Selecting Representative Data Sets", IntechOpen, 2012 Publication	<1 %
14	doc.lagout.org Internet Source	<1 %
15	en.wikipedia.org Internet Source	<1 %
16	pure.tue.nl Internet Source	<1 %
17	"Empirical Distributions and Processes", Springer Science and Business Media LLC, 1976 Publication	<1 %

18	Muhammad Aslam. "Introducing Kolmogorov–Smirnov Tests under Uncertainty: An Application to Radioactive Data", ACS Omega, 2019 Publication	<1%
19	Shaohua Yang, Guoliang Lu, Aiqun Wang, Jie Liu, Peng Yan. "Change detection in rotational speed of industrial machinery using Bag-of-Words based feature extraction from vibration signals", Measurement, 2019 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		