

Several Methods of Determining the Continuous or Discrete Distribution

Sergey Akimov¹, Olga Kvan^{2,*}

¹Postgraduate of Management and Informatics in Technical Systems Department, Faculty of Informational Technology, Orenburg State University

²Bioelementology institution of Orenburg State University, Orenburg State University, 460018, Orenburg, 13 Pobeda avenue

*Corresponding Author: kwan111@yandex.ru

Copyright © 2013 Horizon Research Publishing All rights reserved.

Abstract The article is about the problem of calculating the probability of discreteness and continuity sample from the general totality. There is a definition of discreteness. The main task of research is the definition of continuity or discreteness of unknown data. We consider the existing methodology as a method of finding the frequency of repetition of individual values variants of totality under test. The presented procedure is mathematically described. The basic disadvantage of this procedure: this procedure has great difficulties in interpreting the results. Based on the foregoing, the task of creating an algorithm determining the continuous or discrete becomes very important. The new algorithm is also based on the search for a match in the data array. However, now we use not only the array, but the quantity of changes between two successive values. To do it we need a sorting procedure of array from a minimum value to maximum one. In addition, we introduce the concept of "step" as a minimum amount of change between two values in the discrete series. An iterative method for detecting the matches in the array and defining the identity of the changes of the neighboring values is proposed in the article. Thus we have obtained three key values that define the continuity or discreteness. It has been found empirically that each of these values change its sensitivity based on the number of observations in the array. We also identified factors, which usage as (dependence on the number of values in the data) helps to attribute data array to the continuous or discrete distribution.

Keywords Discreteness, Continuity, Interval, Algorithm

take an uncountable set totality close to each other means are called continuous. The vast majority of real physical and theoretical objects, which condition is characterized by only macroscopic physical quantities (temperature, pressure, velocity, acceleration, current, electric and magnetic fields, etc.) has a continuity property [2]. Mathematical structures which describe such objects must be also continuous. That's why apparatus of differential and integral equations is used for the model description of such objects.

Objects, variables of which can take some almost always a finite number of known means in advance, said to be discrete. The unit of mathematical logic (logic functions, algorithmic languages, etc.) is a basis of a formalized description of discrete objects. In connection with the development of digital computers discrete methods of analysis are widespread and are also used for describing continuous and research facilities [3].

In connection with all mentioned above, there is a question: if there is any possibility to distinguish continuous data and discrete one?

If the researcher has data and he doesn't not have any information about its nature - that question is very relevant, because in the face of uncertainty properties of the distribution may be unknown, so we need a method to determine the continuity and discontinuity.

The aim is that there is a certain set of unknowns, and there is no information about their source, characteristics, and any characteristics that would indicate it belongs to a given set of anything. There is a need to define one parameter of the totality, namely its belonging to a continuous or discrete series.

1. Introduction

Discontinuity (discreteness) is spatial- temporal delimitation of the elements, object states; continuity - the relationship (interdependence) of the elements and conditions of the object [1].

In the modern sense of all the objects, variables, which can

2. Materials and Methods

The most common way is to find a repetition frequency of $f(x)$ values of individual options study totality [4].

This procedure can be mathematically described in the following way. Suppose I - repetition number x_i in the aggregate study values $Z(x_i)$. Then, by coincidence $C(x_i) =$

$I(x_i) - 1$. The total number of matches $C = \sum C(x_n)$.

If the maximum value $\max C(x)$ is not too large, it is fair to say that the distribution is continuous. The algorithm for determining the discontinuity (continuity) will be reduced to the following steps: finding the largest number of repetitions sign $\max C(x)$; assessment of this value. Based on these data it is possible to make the assumption about continuity or discreteness of test values. However, this method has so great difficulties in interpreting the results, because it is necessary to calculate the threshold number of repetitions according to which it is possible to assess which category applies to this totality.

The studies were performed in the laboratory department of Management and Informatics in Technical Systems Department of the Orenburg State University, using a random number generator program Mathcad 15. One thousand continuous and discrete data sets with varying amounts of research (from 10 to 10,000) were generated. We tested the proposed algorithm using the generated data sets.

An iterative technique and sort data were used as a mathematical basis of the algorithm.

3. Results and Discussion

Let us assume that a discrete mean is always changed to the same mean (the so-called "step"), it means that the difference between two adjoining results $x_2 - x_1 = \text{const}$. We denote the result of the operation $x_2 - x_1 = a$, and the difference between any two variables $x_m - x_n = a(m-n)$ [5].

It is necessary to sort this set of data in ascending order for determining of discreteness or continuity. Next, we need to find a difference between two closest means. In carrying out this item, all the matching means disappear. After that, it is enough to count the number of zeros in the resulting data range (called "interval range"), which is equal to the number of matches of diverse values in the totality. Let's make this procedure two times more: let's sort the resulting interval range of ascending and again we find the interval range. Then again let's count the number of zeros – by a sufficiently large number of zeros we can conclude that the original set is discrete.

If we continue processing in this way, each time the number of matches (zeros) is getting larger and larger.

However, when the number of iterations is large, continuous values are also beginning to show matches (already 4-6 iteration matches vary in the range of 95-100%). Therefore, it's reasonable to stay on three iterations.

The sample size plays important role in determining continuity or discreteness. When the number of values is less than 10, results will be inaccurate, and the number of values will be reduced to one during each iteration. If the number of means is 10 and more than the dependence begins to show up. The most sensitive is the third iteration, because it is necessary to give priority. Closer to a hundred of the third iteration (as the most sensitive) the well-read match begins to appear (and, quite significant) in the notoriously continuous

distributions. That's why for means from 100 and more of the second iteration is more accurate. More than one thousand means of the second iteration ceases to give an accurate result, and the first iteration is brought to the forefront.

Considering the importance of each iteration with certain sample size, we have a basic ability to output a series of formulas for determining the stochastic discrete or continuous sampling. For this, it's necessary to assign ranks of each iteration under certain sample sizes. The coefficients of 0.1, 0.2 and 0.7 are the most common in tasks with three variables. Let's denote the first, the second, and the third iteration, accordingly, A , B , and C . For getting quantitative means in each iteration, it's necessary to divide the number of coincides by $n-1$, $n-2$ and $n-3$, respectively, where n - the number of means in the array.

Using empirical way it was determined that the best value for the ratio of 10 values is ratio $0.1A + 0.2B + 0.7C$, for the ratio of 100 means is the ratio $0.1A + 0.7B + 0.2C$ and for the ratio of 1000 and above is ratio $0.7A + 0.2B + 0.1C$.

It is necessary to smooth transition from one formula to another for improving the accuracy. It can be achieved by using the sliding means that would change according to the change in the volume of samples. For getting the second formula from the first it is necessary to reduce gradually the third iteration to the second. Mathematically, it looks like this:

$$0.1 A + (0.2 + (0.5 N/100)) B + (0.7 - (0.5 N/100)) C$$

where N - number of data sampling.

Let's make the transformation and other formulas:

$$(0.1 + (0.6 N/1000)) A + (0.7 - (0.5 N/1000)) B + (0.2 - (0.1 N/1000)) C$$

If the result is in the range 0-0.4 – this array can be attributed to the continuous distribution. If the result is from 0.6 to 1 this array can be attributed to discrete distribution. The range of 0.4-0.6 is a zone of uncertainty.

4. Conclusion

The problem of determining the continuity or discreteness studied in the article is solved by the method of ranking and consistent finding matches. Identified empirical formulas in the article give an opportunity to determine with great accuracy the continuity or discontinuity of the studied data set.

REFERENCES

- [1] Big Encyclopaedic Dictionary - [electronic resource] Access mode : <http://dic.academic.ru/dic.nsf/enc3p/243162>.
- [2] Continuous and discrete models. [Electronic resource]. Mode of access: <http://www.automationlab.ru/index.php/2011-06-26-2>

1-10-36/397-6

- [3] The concept of the models and modeling. [Electronic resource]. - Mode of access: <http://do.gendocs.ru/docs/index-338446.html>.
- [4] Iglin S.P., Kulachenko A.E., Sjarrow A.M. Research of hyperelastic orthotropic elliptical membranes. 77th Annual Meeting of the Gesellschaft für Angewandte Mathematik und Mechanik e.V. March 27th – 31st, 2006, Technische Universität Berlin. Book of Abstracts, p.388.
- [5] Akimov S.S. An integrated approach to the determination of the distribution of probability sample of the general totality. / S.S. Akimov // Actual problems of trade and economic activities and education in modern terms: electronic collection of scientific works of the Seventh International scientific practical conference dedicated to the 50th anniversary of the Orenburg branch of RSUTE . Orenburg:Orenburg branch of RSUTE , 2012. S. 468 476.