Fractals As Triggers For Exploratory Statistical Analysis of Clinical Pharmacological Data

S.K. Kulishov

O.M. Iakovenko

Fractals As Triggers For Exploratory Statistical Analysis of Clinical Pharmacological Data

S.K. Kulishov & O.M. Iakovenko

Department of Internal Medicine No 1, HSEIU "Ukrainian Medical Stomat. Academy", Poltava, Ukraine;
Daily Establishment, Institute of Endocrinology and Metabolism AMS, Kyiv, Ukraine
E-mail : kulishov@meta.ua, alexandrya@meta.ua

Abstract - Proposed and tested an algorithm of using principles of Cantor, von Koch sets for exploratory fractals clinical pharmacological data analysis. The algorithm is based on the grouping data, formation of categorical variabilities in the form of subgroups as iteration process as for receiving Cantor, von Koch sets. It boils down to: selection of informative numerical dependent variabilities; transformation these informative numerical dependent variabilities to new categorical variabilities; formation categorical variabilities in the form of subgroups as a result of an iterative process as for Cantor, von Koch sets; statistical analysis of the data; determination of the distribution of variabilities; transformations that may be normalize from non-normal data; ANOVA - analysis of variance parametric data or nonparametric equivalent of ANOVA - Kruskal-Wallis testing; formulation of the conclusion. Our algorithm of using Cantor, von Koch sets principles for Exploratory Fractals Data Analysis of clinical pharmacological data will help maximize insight, uncover underlying structure, extract important variables, develop models and determine optimal factor settings.

Keywords - fractals, clinical pharmacological analysis.

I. Prerequisites of fractal and anti-fractal data transformation as the basis, triggers for exploratory statistical analysis

The essence of good prescribing is to pick the most appropriate drug for the disease in question, taking pathophysiology in account [1]. Drugs act on a wide variety of targets: receptors, transport processes, enzymes, by others miscellaneous effects [1]. Following administration, disposition of drugs in the body is determined by drug absorption, distribution, metabolism and excretion [1]. Taken together, these processes define pharmacokinetics of drug. Drug therapy monitoring, also known as Therapeutic Drug Monitoring (TDM), is a means of monitoring drug levels in the blood [1]. Because so many different factors influence blood drug levels, the following points should be taken into consideration during TDM: the age and weight of the patient; the route of administration of the drug; the drug's absorption rate, excretion rate, delivery rate, and dosage; other medications the patient is taking; other diseases the patient has; the patient's compliance regarding the drug treatment regimen; and the laboratory methods used to test for the drug [1]. TDM is a practical tool that can help the physician provide effective and safe drug therapy in patients who need medication [1].

Monitoring can be used to confirm a blood drug concentration level that is above or below the therapeutic range, or if the desired therapeutic effect of the drug is not as expected. If this is the case, and dosages beyond normal then have to be prescribed, TDM can minimize the time that elapses [1].

TDM is important for patients who have other diseases that can affect drug levels, or who take other medicines that may affect drug levels by interacting with the drug being tested [1]. Therapeutic drug monitoring refers to the individualisation of dosage by maintaining plasma or blood drug concentrations within a target range (therapeutic range, therapeutic window) [1]. There are two major sources of variability between individual patients in drug response [1]. These are variation in the relationship between: dose and plasma concentration (pharmacokinetic variability); drug concentration at the receptor and the response (pharmacodynamic variability) [1]. Several methods have been developed to improve the prediction of individual dose requirements based on sparse data for individual patients [1]. These are based either on calculation of clearance and volume of distribution from one or a few timed drug concentrations, or by a Bayesian feedback method [1]. This latter method is based on differences between 'typical' population

parameter values and those predicted for the individual patient from measured drug concentrations [1].

Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

1. maximize insight into a data set;

2. uncover underlying structure;

3. extract important variables;

4. detect outliers and anomalies;

5. test underlying assumptions;

6. develop parsimonious models; and

7. determine optimal factor settings.

Most EDA techniques are graphical in nature with a few quantitative techniques. The reason for the heavy reliance on graphics is that by its very nature the main role of EDA is to open-mindedly explore, and graphics gives the analysts unparalleled power to do so, enticing the data to reveal its structural secrets, and being always ready to gain some new, often unsuspected, insight into the data. In combination with the natural pattern-recognition capabilities that we all possess, graphics provides, of course, unparalleled power to carry this out.

For classical analysis, the sequence is from problem to data, model, analysis, conclusions

For EDA, the sequence is from problem to data, analysis, model, conclusions. For Bayesian, the sequence is problem, data, model, prior distribution, analysis, conclusions.

Thus for classical analysis, the data collection is followed by the imposition of a model (normality, linearity, etc.) and the analysis, estimation, and testing that follows are focused on the parameters of that model. For EDA, the data collection is not followed by a model imposition; rather it is followed immediately by analysis with a goal of inferring what model would be appropriate. Finally, for a Bayesian analysis, the analyst attempts to incorporate scientific/engineering knowledge/expertise into the analysis by imposing a data-independent distribution on the parameters of the selected model; the analysis thus consists of formally combining both the prior distribution on the parameters and the collected data to jointly make inferences and/or test assumptions about the model parameters.

The classical approach imposes models (both deterministic and probabilistic) on the data. Deterministic models include, for example, regression models and analysis of variance (ANOVA) models. The most common probabilistic model assumes that the errors about the deterministic model are normally

distributed-this assumption affects the validity of the ANOVA F tests.

The Exploratory Data Analysis approach does not impose deterministic or probabilistic models on the data. On the contrary, the EDA approach allows the data to suggest admissible models that best fit the data.

The two approaches differ substantially in focus. For classical analysis, the focus is on the model - estimating parameters of the model and generating predicted values from the model.

For exploratory data analysis, the focus is on the data - its structure, outliers, and models suggested by the data.

Classical techniques are generally quantitative in nature. They include ANOVA, t tests, chi-squared tests, and F tests.

EDA techniques are generally graphical. They include scatter plots, character plots, box plots, histograms, bihistograms, probability plots, residual plots, and mean plots.

Classical estimation techniques have the characteristic of taking all of the data and mapping the data into a few numbers ("estimates"). This is both a virtue and a vice. The virtue is that these few numbers focus on important characteristics (location, variation, etc.). The vice is that concentrating on these few characteristics can filter out other characteristics (skewness, tail length, autocorrelation, etc.) of the same population. In this sense there is a loss of information due to this "filtering" process.

The purpose of this work is using Cantor, von Koch sets principles as basis, triggers of exploratory clinical research analysis for better insight into a data.

II. The methodology of fractal data transformation for exploratory statistical analysis

The methodologies of fractal data transformation for exploratory statistical analysis have 3 steps. The first step is:

- selection of informative numerical dependent variabilities for transformation to categorical types by using Cantor, von Koch or others sets principles as basis, triggers for exploratory clinical research analysis.

Cantor set obtained from the closed interval from 0 to 1 by removing the middle third from the interval, then the middle third from each of the two remaining sets, and continuing the process indefinitely [2]. The Cantor ternary set is created by repeatedly deleting the open middle thirds of a set of line segments. One starts by

deleting the open middle third ($\frac{1}{3}$, $\frac{2}{3}$) from the interval [0, 1], leaving two line segments: [0, $\frac{1}{3}$] ∪ [$\frac{2}{3}$, 1]. Next, the open middle third of each of these remaining segments is deleted, leaving four line segments: [0, 1/9] ∪ [2/9, 1/3] ∪ [2/3, 7/9] ∪ [8/9, 1]. This process is continued ad infinitum. The Cantor set cannot contain any interval of non-zero length. In fact, it may seem surprising that there should be anything left — after all, the sum of the lengths of the removed intervals is equal to the length of the original interval. However, a closer look at the process reveals that there must be something left, since removing the "middle third" of each interval involved removing open sets (sets that do not include their endpoints). So removing the line segment (1/3, 2/3) from the original interval [0, 1] leaves behind the points 1/3 and 2/3. Subsequent steps do not remove these (or other) endpoints, since the intervals removed are always internal to the intervals remaining [2].

The second step is statistical analysis of these informative numerical variabilities:

- Determination of mean, standard error of mean, standard deviation, 95% confidence interval for mean, median, minimum, maximum, range, quartiles, interquartile range, skewness, kurtosis.

- Determination of the variabilities distribution - parametric or nonparametric by single-factor the Kolmogorov-Smirnov test; Shapiro-Wilk W test and graphical methods: frequency distribution histograms stem & leaf plots; scatter plots; box & whisker plots; normal probability plots: PP and QQ plots; graphs with error bars (Graphs: Error Bar).

The third step is:

- To choose using Cantor or von Koch sets principles for transformation informative numerical variabilities to categorical types by results of statistical analysis. More high level of standard deviation, interquartile range is better for using Cantor or von Koch sets principles.

Algorithm of exploratory data analysis using the technology of the iteration process as for receiving Cantor set:

- Selection of informative numerical dependent variabilities;

- Transformation these informative numerical dependent variabilities to categorical variabilities;

- Formation categorical variabilities in the form of subgroups with the maximum, median and minimum values;

- Formation of a new categorical variabilities in the form of sub-subgroups with maximum, median and minimum values in the subgroups with the highest and lowest values (closed ring)

- Formation categorical variabilities in the form of subgroups as a result of an iterative process as for Cantor set;

- Statistical analysis of the data;

- Determination of the variabilities distribution - parametric or nonparametric by single-factor the Kolmogorov-Smirnov test; Shapiro-Wilk W test and graphical methods: frequency distribution histograms stem & leaf plots; scatter plots; box & whisker plots; normal probability plots: PP and QQ plots; graphs with error bars (Graphs: Error Bar);

- Transformations that may be normalize of non-normality data: If residuals have a right skew, should apply a square-, a cube- or fourth-root, a logarithmic, and an inverse transformation to data. If residuals have a left skew, should raise to the second, third or fourth power, an exponential transformation to data;

- ANOVA - Analysis of Variance, with variations depending on the linear nature of variability. Method of multiple comparison groups Tukey HSD, Scheffe, Bonferroni if deviations were homogeneous for the test Levene, and in the absence of homogeneity - the criteria Tamhane's T2, Games-Howell;

- Nonparametric equivalent of ANOVA / MANOVA - Kruskal-Wallis test;

- Formulation of a conclusion based on statistical analysis.

Algorithm of exploratory data analysis using the technology of the iteration process as for receiving von Koch set:

- Selection of informative numerical dependent variabilities;

- Transformation these informative numerical dependent variabilities to categorical variabilities;

- Formation categorical variabilities in the form of subgroups with the maximum, median and minimum values;

- Formation of a new categorical variabilities in the form of sub-subgroups with maximum and minimum values in the subgroup with median values (closed ring);

- Formation categorical variabilities in the form of subgroups as a result of an iterative process as for von Koch set;

- Statistical analysis of the data;

- Determination of the variabilities distribution - parametric or nonparametric by single-factor the Kolmogorov-Smirnov test; Shapiro-Wilk W test and graphical methods: frequency distribution histograms stem & leaf plots; scatter plots; box & whisker plots; normal probability plots: PP and QQ plots; graphs with error bars (Graphs: Error Bar);

- Transformations that may be normalize of non-normality data: If residuals have a right skew, should apply a square-, a cube- or fourth-root, a logarithmic, and an inverse transformation to data. If residuals have a left skew, should raise to the second, third or fourth power, an exponential transformation to data;

- ANOVA - Analysis of Variance, with variations depending on the linear nature of variability. Method of multiple comparison groups Tukey HSD, Scheffe, Bonferroni if deviations were homogeneous for the test Levene, and in the absence of homogeneity - the criteria Tamhane's T2, Games-Howell;

- Nonparametric equivalent of ANOVA / MANOVA - Kruskal-Wallis test;

- Formulation of a conclusion based on statistical analysis.

III. Some illustrations of graphic modeling, programming of using fractal transformation of data as the basis for exploratory statistical analysis

Examples of the results of the algorithm for using principles of Cantor, von Koch sets for Exploratory Data Analysis of clinical research data are presented by syndromes of the consumption as a common biological principle of desadaptation and diagnosis of the viable, stunned, hibernation myocardium and cardiac protective precondition on the basis of modeling of biochemical, instrumental, mathematical, coronaro-ventricularal data [3,4,5,6,7].

Examples of the results of the algorithm for using principles of Cantor set for Exploratory Data Analysis of clinical research data are presented by determination of the dependence between tolerance to stress, inflammation syndrome, coronary and myocardial failure, arterial pressure, ventricular arrhythmias in the patients with coronary heart disease and essential hypertension [8,9]. Fractal analysis of monitoring glycemic profile and arterial pressure helped with diagnosis of the sum negative effects in the patients with 1st diabetes mellitus in combination with essential hypertension [10]. Daily changes in glucose levels,

blood pressure, heart rate are the basis of individual diagnosis of pathogenic mechanisms, the correction of the treatment of the patients with diabetes mellitus in combination with essential hypertension [10].

Transformations of variables may be by recode and compute procedures. The recode procedure is typically used with transformations involving categorical variables. It is the best option when we want to create a categorical distinction based on an existing numeric variable. Using the technologies of the iteration process as for receiving Cantor, von Koch or others sets for transformations of variables may be by recode procedure. If new variables that consist from subgroups similar as the Cantor, von Koch or others sets will have non-normal distribution we use the compute procedure. This procedure allows the analyst to perform mathematical operations on variables.

The methodologies of graphic modeling, programming of the algorithm for using principles of Cantor sets for Exploratory Data Analysis of clinical research data are presented on visual programming language "Dragon" [11] (fig.1).

IV. CONCLUSIONS:

Our algorithm of using principles of Cantor, von Koch sets for Exploratory Data Analysis of clinical research data will help maximize insight into a data set; uncover underlying structure; extract important variables; develop models and determine optimal factor settings.

REFERENCES

[1] Nicki R.Colledge, Brian R. Walker, Stuart H. Ralston "Davidson's Principles & Practice of Medicine", 21st Edition, Churchill Livingstone Elsevier, London, 2010, p. 17-38.

[2] Christopher Shaver "An Exploration of the Cantor Set", Rockhurst University .09, James and Elizabeth Monahan Summer Research Fellowship, Summer 2008, MT4960: Mathematics Seminar, Spring, 2009, 19 p.

[3] Bobrov V.A., Kulishov S.K. "The adaptive ischemic and reperfusion syndromes in the patients with ischemic heart disease: mechanisms, diagnosis, substantiation of therapy", Poltava: Dyvosvit, 2004, 240 p.

[4] Kulishov S.K, Shtompel P.S. "Mathematical modeling, programming algorithms for diagnosis of ischemic adaptation and reperfusion syndromes at patients with coronary heart disease", Materials of All-Ukrainian Congress of Ist "Medical informatics and biological

informatics and cybernetics" with international participation, Proceedings, 23-26 June 2010, Kyiv, 2010, p. 230.

[5] Kulishov S.K, Shtompel P.S. "Selection principles of fractal sets for differential analysis of the systemic hemodynamic response to the test with compression-decompression of extremities at the patients with coronary heart disease", Materials of All-Ukrainian Congress of Ist "Medical informatics and biological informatics and cybernetics" with international participation, Proceedings, 23-26 June 2010, Kyiv, 2010, p. 231.

[6] Kulishov S.K., Iakovenko O.M., Tretiak N.G. "Clinical thinking training as a derivative of system and antisystem comparison, precondition to increase creativity of medical students, physicians", Proceedings of the ICL conference" (Hasselt, Belgium), 15.09.- 17.09.2010, the Kassel University Press, 2010, pp. 337-343.

[7] Kulishov S., Iakovenko O.M. "Solving clinical problems using system and anti-system comparison, graphic modeling","Innovative Medicine and Biology", Canadian International Monthly Reviewed Journal (CIJIMB), ISSN 1925-2188; 2011, No 3, pp. 30-42.

[8] Kulishov S., Vakulenko K., Latoha I. "Myocardial electrical instability as the derivative of inflammation consumption of antiiflammatory factors syndrome, changes in geometry depolarization of atria, ventricles in the patients with coronary heart disease", Absract book of "Rhythm 2011", Congress, 2011, p. 39.

[9] Kulishov S., Kudria I. "Stress and inflammation as determinative factors in the patients with coronary heart disease and essential hypertension", Final programme& Abstract Book of the European Conference, IInd edition "Heart, Vessels & Diabetes", 3-5 November, Athens, Greece, 2011, p. 50.

[10] Kulishov S., Iakovenko O.M. "Individual diagnostic value of monitoring blood glucose and arterial pressure in the patients with diabetes mellitus and essential hypertension: a case report", Final programme& Abstract Book of the European Conference. IInd edition "Heart, Vessels & Diabetes", 3-5 November, Athens, Greece, 2011, p. 51.

[11] Parondzhanov V.D. "How to improve the work of the mind. Algorithms without programming - it's easy!" M.: Delo, 2001, 360 p.
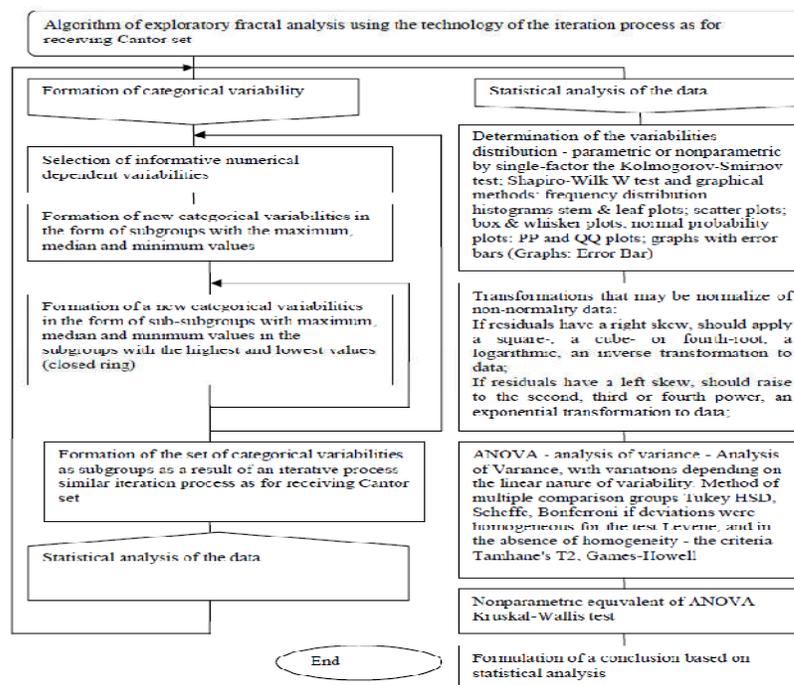.

Fig. 1. Scheme of the Cantor set algorithm for Exploratory Data Analysis of clinical pharmacological data
.