THRESHCHRONIC
Concept Proposal of a Fuzzy Bayesian Tool for Health Support

by

Rares SUVAILA [1,2*]

[1]National Institute for Marine Geology and Geoecology GeoEcoMar, Bucharest, Romania
[2] Horia Hulubei National Institute for Physics and Nuclear Engineering, Magurele, Romania

Technical paper
https://doi.org/10.2298/NTRP2501073S

A multitude of chronic diseases are not yet classified as being stochastic or deterministic. The purpose of this work was to provide an adaptative software tool which enables detection of thresholds for the cause-consequence relationships with a maximum degree of confidence given the available statistics. This tool is meant to be tested on radioactivity datasets, for which the thresholds exist and are well-known, then to be applied to other datasets on radioactivity and microplastics (environment), antibiotics (pharmacological factors), and nitrites (food industry). The main algorithm for this supporting program for health research was built as an unitary procedure for all situations, additional covariance analysis on multi-factor cases will be performed, together with a scheme which explicates the importance of transfer factors for the data recorded on various substances (*e.g.* water-fish-human transfer for radioactivity and microplastics). The key elements are the non-biased datasets available and the algorithms for conditional probability calculations and potential thresholds determination. One very important action to take is separating the accurate and verifiable data from the plethora of biased information from the repositories.

Key words: stochastic process, deterministic process, environmental influence, fuzzy logic, bayesian statistics, correlation analysis

INTRODUCTION

Medical and hazard data are being recorded in a well-organized manner and stored systematically, at least since WWII, in order to evaluate various parameters of interest for human health and to have a better prediction for the evolution of the latter. During the last years, an enormous amount of data has been accounted and many studies of the correlations between natural or anthropogenic influencing factors have been performed. To cite a few examples of the areas we intend to investigate, recent and relevant examples are categorized as radioactivity, food, drugs, ecology (especially anthropogenic, emergent, such as microplastics) and their convolutions.

All classic (frequentist) approaches are based on the most commonly used definition of probabilities, namely $P_i = n_i / n_t$ with $t \to \infty$, where i is the event of interest, n_i – the number of events of interest, and n_t – the total number of events. In other words it is axiomatic that the number of iterations of the same experiment could get as high as we need in order to consider with a good approximation that correction factors of the residual probabilities is 0 for any given set of events i_1, i_2, ... i_x, no matter how big x gets.

How about if we chose to analyze significant nuclear accidents? Actually, there were three of them (Three Miles Island [1] – 1979, Chernobyl [2] – 1986, Fukushima [3] – 2011), the worst accidents in the history of commercial nuclear power facilities. The first two were due to a convolution of design/security failures and human errors. The third was caused by a convolution of natural disasters and the incapacity to anticipate a chain of consequences. It is questionable whether one or more of them was a Black Swan [4] type event, as the faults could have been introduced in the risk analysis as non-impossible case scenarios from the beginning (non Black Swan), which would have avoided the disasters, on the other hand, we are (or at least should be) now aware of the lack of ability of human kind when it comes to apprehend its limitations – as our perception of similar events at a different scale makes them look fundamentally different, which is false in the a posteriori analysis. This implies possible Black Swans which originate in the non-exhaustive set of alternate hypotheses considered, so the likelihoods lose their meaning. Anyway, in order to perform (conditional) probability analysis on such a set of catastrophic events, all we wish is to not make the total number of events grow. Not to mention a SARS epi-

* Corresponding author, e-mail: rares.suvaila@geoecomar.ro

demic/pandemic can be mathematically considered as an event, and equivalent cases may be found in any domain. It's obviously better to adapt the mathematical work frame than to keep on looking for better statistics by growing numbers. Here comes the bayesian approach [5] into play: conditional probabilities with prior and posterior determined by likelihood and evidence. Instead of the frequentist approach, we rather use the conditional probabilities formula in the bayesian approach: *Posterior* = = *Likelihood x Prior* / *Evidence*. This way a probability is assigned to each hypothesis, instead of testing the hypothesis as a boolean proposition. This is what we intend to apply in order to investigate the distributions, performing statistic tests, and finally try to close the door on the debate for stochastic versus deterministic type effects generated by different substances. This needs to be validated for given degrees of confidence, assumption validity and model acceptance. Our intention is to investigate if a threshold exists between individual to collective level, and if yes, by what kind of sampling it is reasonable to treat such properties using fuzzy logic for a given population in order to categorize effects as being stochastic or deterministic [6]: this is the role of THRESHCHRONIC, a novel software under development.

MATERIALS AND METHODS

Tools

The tools needed are mainly the mathematical model/framework and the algorithms&code for probability calculation and simulations. First of all, we must consider the inference limitations, which are due to the impossibility of tracing events linked to the current definitions more than 100 years or so ago: radioactivity was discovered in 1896 [7] and its harmful actions acknowledged only by the end of WWI [8]. Cancer was first acknowledged in a primary manner in Ancient Egypt [9], but the proof it spreads through malignant cells came only in the mid-1800′s and efficient cures are just decades old. Dangers implied by antibiotics misuse and microplastics intake have not been seriously considered till recent alarms [10]. The limitations must be set on a clear timeline for the population of interest, from periods when the effects began, considering how many of the affected people are still alive if such is the case, to nowadays. Based on this data, we may also be able to initiate forecasts, but this is another topic, and statistic extrapolation in terms of effects is not a very good practice. Anyway, the lessons to be learned from the thresholds are just as valuable.

Then the mathematical frame has its heavy influence on the whole approach. The distribution moments [11] are very important for our work and need to be examined carefully. Of course, the mean and variance (first and second order central moments) give us a hint about the distribution's expected value and dispersion, but in order to have a robust statistical approach, one needs to evaluate superior moments as well. The third standardized moment, skewness, measures the (a)symmetry of the distribution. A positive skew implies in general an asymmetry such as the graphic representation shows the three main measures of central tendency – mode, median and mean from left to right, while a negatively skewed distribution generally implies converse display in the spectrum, *i. e.* mean, median, mode. Of course, real life data sets provide a series of exceptions which need data transformation (*e.g.* log transformation for a highly positively skewed distribution) in order to get closer to a normal distribution, which is preponderantly appropriate for statistical tests. Another practical example would be switching the use of mean to median, which is less sensitive to extreme or outlier values. The fourth standardized moment of the distribution, kurtosis, is also of considerable importance, as it provides a measure of the tails of the distribution: a greater the kurtosis implies greater deviations, which have to be seriously taken into account if we are looking for potential thresholds. The difference between stochastic and deterministic processes is not to be differentiated from the raw data by simply searching for a threshold and a constant slope: by analyzing distributions by age, gender, activity, region, then period of ingestion (short term), time of trigger (mid term) and so on, one can see it is essential to study the distribution moments carefully.

Online statistic tools [12] are nowadays present worldwide from simple calculation applications to complex interactive programming, nevertheless, the approach we propose for supporting health studies is novel, in the sense that combining bayesian and fuzzy approaches has not yet been performed (on a separate note, the application we intend to develop has nothing to do neither with bayesian health/other AI platforms, nor with any tool for diagnosis). A key element for the present work is the fuzzy logic approach for health external anthropogenic factors together with their potential consequences. Again, this is meant to be a supporting tool for health, for decision making in terms of environmental and food parameters, not medical ones. We do not interfere in the medical or diagnosis process, we just strenghten it with robust data, as the correlations we make are between external parameters and post-diagnosis conclusions. We want to have the locally reported data on substances compared to the national, continental and worldwide ones. Then follows intercomparison between local health issues interpreted as potentially conditioned by the external factors we study. Finally, we must decide how far we can go in the sampling process inside a region, which in its turn is sampled from greater areas, then from the continent. Averaging on the total population would make no sense, while the other extreme, namely individual data, is not available.

Statistic approach

Our idea is to combine the bayesian and fuzzy approaches in a tool meant to support threshold find-

ing in the influence of various parameters (first stage) and also provide covariances between anthropogenic factors influence on certain diseases (second stage). A multitude of tests was carried out with success in food and drug industry, including by means of bayesian approaches and/or fuzzy logic [13-15] prior to our work. The novelty we are bringing in as a support tool for health is the cause-effect correlation given a clear diagnostic; mathematically speaking, we are dealing with a *model Lego*. This tool is meant to demonstrate both threshold existence (if such is the case) and to establish the correlation functions between the studied parameters and the disease in question. Again, this is not a diagnose tool, but an adaptive mathematical one, which can be further used in other fields, along with the related web application. Similar tools – based on Bayesian statistics and fuzzy logic – are available, but not for the cause influence on chronicity, as we deal with the latter once diagnosed æ not with symptoms or biochemical tests.

Brief history and context of the proposal

Historically, the recorded start of probability approach belongs to Aristotle, 322 BC in the Art of Rhetoric, probabilities were mostly based on a philosophical approach and the strongest connection for their application was with law (through the notion of the extent to which guilt is implied by evidence) till modern times as described by J. M. Garrett in The History of Probability Theory. The first deeply mathematical exchange, which led the translation of pre-existing ideas into what we call now (quantitative) probability theory belongs to Pascal and Fermat in the 17th century.

Thomas Bayes derived his theorem in the early 1760's, publishing his work *An Essay towards solving a Problem in the Doctrine of Chance*s in 1763. The conditional probability approach he used is of critical interest to statisticians nowadays, although it has been marginalized for two centuries.

Then, Simeon Denis de Poisson derived for the first time the distribution now applied to the count rates in most of the nuclear processes today, without having any clue of the importance his probability distribution would have in modern science. This goes back to his 1837 book (Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités) treating probabilities for civil and criminal judgement – one of his last published works.

Meanwhile, worldwide development in the field of probabilities lead to significative improvement in terms of methods and algorithms which now seem evident, though a long and laborious effort was necessary to reach this point. Also, during the last decades progress in programming provided a variety of solutions – most of them very user friendly – for analyzing huge amounts of data in short time. However, our approach is to take a step back from that environment where a multitude of tools dictate the rules as dice are thrown whenever a new dataset comes into play, at least in terms of how to treat the problem. After all, we're doing it the bayesian way. To use in-house developed software. The main reason behind this is, we do not know if the distributions we study do have thresholds at all, in first place. So even if it is one of the oldest tricks in the world, we bring in some novelty for the speed era approach: attention to details, expecting an increase in terms of correlations.

THRESHCHRONIC objective

The THRESHCHRONIC was designed to analyse data and distributions in order to provide information on their correlations, potential threshold existence and related consequences. Practically, this means determining the parameters involved in a potential cause- -effect relationship, study the covariances between them, identifying the threshold if it exists and providing a mathematical function for their evolution, as a single cause effect and as a convolution of multiple factors. In our case the causes are the substances under investigation, the effect is a chronic disease, and the convolutions are given by the conditional probabilities which result from a multi-factor environment.

This tool in question must be demonstrated as being able to deal with any data – representing any precursors for a similar class of effects – in order to find thresholds and correlations, and to display the co-evolution of data in terms of conditional functions. This is strongly supported by the prior results: our previous work demonstrated the feasibility of statistic hypothesis testing for quantitative determination of the elements of interest in various distributions – mainly radioactive material [16]. Here we have a very important issue to be considered, as the most important aspects in the paper are linked to the distribution itself: the electromagnetic radiation recorded with a detection chain in the lab is subject of a Poisson distribution of the number of events in time, just as the number of hospital admissions, drug prescriptions or nitrite consumption for the vast majority of the population. For example, the number of hospital admissions per year in England is equal to the number of photons detected without trouble for the dead time corrections for a Hyper Pure Germanium detector in a few seconds. By associating an energy (namely a radioisotope) from the spectra to a disease, we can simulate the admissions by type of disease, each type being associated to a radionuclide. Consequences of certain major diseases can be regarded alike members of a natural radioactive series (*e.g.* uranium, thorium) and the correlations are straightforward. Also, the coincidence peaks (mean-

ing energies from different transitions summing up in a certain region of the spectra because the excited states have lifetimes considerably lower the detector time resolution) can be treated as unexpected events in the spectra, and subsequently affected by other weights in terms of probability, in order to emulate real life scenarios: we can obtain simulations of medical situations by studying the spectra of a simple calibration source for acquisition times beyond comparison to those of grabbing data, at zero costs.

PRELIMINARY AND EXPECTED RESULTS

Prior work

The preliminary results which strenghten our approach are theoretical, in the field of statistics [17] – on one hand, and experimental [18] (data analysis and methods) – on the other. The *within unit uniformity* article cited was the trigger to this approach, as we realised the concept of *within unit* can be regarded simultaneously at micro and macroscopic scales – consequently a sample can always be a hotspot of a batch and the activity detected in a few seconds for natural/calibration nuclides reproduce the statistics on considerable periods in the health system, particularly regarding prescriptions. Consequently, alternate hypothesis testing has a very different meaning when it's applied to properties distributed by means of fuzzy logic within different populations, at different scales. From our perspective, to complete and bring up to date the short chronicle from the previouos section, we must mention a few events:

- the *Bayesian thinking as Statistics in Medicine* was launched in 1982 [19]
- the first fuzzy logics approach in medicine emerged in the 1980′s [20], but applications are more recent,
- the first bayesian and fuzzy rules applied together in medicine was in 2014 [21], but for medical decision support, not external factor influence analysis.
- the first tool for such applications goes back to 2018 [22], and
- the first approach combining fuzzy logic and bayesian statistics for external/environmental parameter influence to our knowledge is the current project.

All those contributions were of major impact for bayesian and/or fuzzy analysis, which have their principles deeply seeded in our primary algorithms. To some extent, our work here is a logical continuation of prior work in the fields of statistics, health, radiation protection and environment, which responds to nowadays needs in order to cathegorize, adapt and understand models, including covariance and consequence analysis.

Numerical simulations of factors as the one we treat have also been performed, but most of them for chemical, biological, radiological and nuclear purposes [23], and not necessarily applying the mathematical framework we make use of. Nevertheless, their results support our idea, in the sense that the individual contributions of those factors are explained and leaves the path clear for our development.

THRESHCHRONIC expectations

To the best of our knowledge, no health support tool based on such principles and dealing with diagnosed diseases potentially caused by natural and anthropogenic external factors was recorded yet; THRESHCHRONIC (undergoing a patenting process) would be the first software able to do this. The algorithm is based on the following steps:

A. Analytical model elaboration
- Framework choice, sampling, choice of batch size
- Model development

B. Correlation analysis
- Data preparation
- Fuzzy distribution of data
- Simple covariance calculation
- Covariance calculation of combined risk factors with disease
- Likelihood ratios calculations
- Fitting correlation curves
- Interpretation: the output of the analysis is
  - Deterministic
  - Stochastic
  - Failed to determine

C. Validation
- Method validation by data sets with known types of correlations
  - Match
  - Back to step A
- Introduction of dummy data sets
  - Discover unknown correlation which completes our set of hypothesis
  - Coincidence model match (correlation without causation)
  - Failed to assign model (but no threshold)

D. Negative feedback loop
- If unknown correlations arise from dummy data set
  - Complete our set of hypothesis
  - Add to model & reiterate from Phase A
- If no unknown correlations appear, Threshchronic is ready to be used.

If no unknown correlations arise at step D, then the model is good enough to be used for any disease and risk factors. The conclusions to be drawn are (in)existence of the threshold, failure to determine (bad statistics or absence of correlations). Though, it is hard to believe the absence of correlations would appear while trying to link chronic diseases with recog-

nized risk factors. Once the thresholds are determined, this may help defining limits for consumption which would be more adapted in the sense of lowering the risks.

Critical points

An activity of crucial importance is the separation of probabilities and likelihoods from the database we obtain, as this has been subject of serious confusions and led to misinterpretation of many results. While probability refers to possible results, likelihood refers to hypotheses. Possible results for a given iteration of an experiment are generally mutually exclusive and exhaustive, the resulting being one (and only one) of the possible outcomes. Anyway, the probabilities that attach to the possible results must sum to 1 (this is of course valid for independent events). On the other hand, hypotheses are generally neither mutually exclusive, nor exhaustive. Unlike an experimental dataset, which is given by randomness, but does not exclude any physically possible outcome, a set of hypotheses to which likelihoods refer to is given by our imagination.

We cannot be sure of imagining all alternate hypothesis (of which some may be nested within others) or of having a correct estimation of the experimental data impact on the likelihoods. Consequently, it's not a good practice to think in terms of summing and normalizing as for the case of probabilities, as the likelihoods related to our hypothesis do not have meaning by themselves: it's the likelihood ratios, or relative likelihoods that do have meaning.

Similarities with radioactive decay

The similarity between the fuzzy presence of nuclides in a sample and chemicals in a population must be treated prudently, as the actual sampling is of great importance. Studies on the within-unit statistic approach versus within group have to be optimized in terms of sample size: the number of samples that do have a chance to get passed the threshold (if it exists) is a function of how low we can go with sampling the respective population. Consequently, the sampling process is another very important activity for enabling threshold finding for the concerned substances – if we distribute in a fuzzy manner all the intake on the whole population the approach makes no sense. The existence of a threshold qualifies the process as being deterministic, so that under the threshold dose the factor of interest causes no effects, but does cause them above that dose, and the incidence rate increases sharply (*e. g.* tissue effects from ionizing radiation). The absence of a threshold qualifies the process for being stochastic, which assumes the effects appear depending on the dose level, above a *spontaneous incidence* (the incidence of the disease is not 0 for a 0 dose, and it may be related to other factors), *e. g.* cancer caused by ionizing radiation. Here we must note the same factor may have both deterministic and stochastic effects on the same organism. This way, taking into account all acknowledged possible covariances, we can get to the desired result, namely a robust tool able to separate the two kinds of disease from the statistical perspective and characterize them further, with the respect to all the possible outcomes, with or without coevolution.

CONCLUSIONS

This work responds to a present need in terms of disease classification, namely the need to design, build and demonstrate a tool meant to determine whether the studied substances do have a threshold for producing cancer or other chronic diseases to the population under investigation. Also, this tool is conceptualized as a correlation instrument for (multi)cause-effect evolution. This enables defining warning limits per person, population or area for various compounds or products, and also plot the effects against the amount of substances involved. The instrument we are building is not a medical one or a health appliance as such, but rather a supporting tool for correlations in various fields, from food to radiation protection. We are also working on a scheme for the covariances between factors of different nature, which can lead to a commonly built threshold associated to the convolution which generates chronic effects observed within a given population, assuming those covariances exist. The last supposition has though a high degree of plausibility, *e.g.* smoking (inhaling at least ^{234}Th, ^{226}Ra, ^{210}Pb, ^{214}Bi, ^{228}Ac, ^{40}K, and ^{210}Po) and working in an anthropogenic radioactive environment – both factors exceeding the natural radioactive dose. The ultimate purpose of this tool is to enable discriminating the stochastic and deterministic nature of any disease, based on simple datasets with the respect to the precursors.

ORCID NO

R. Suvaila: 0000-0001-6463-8685

REFERENCES

[1] Upton, A. C., Health Impact of the Three Mile Island Accident, *Annals of the New York Academy of Sciences, 365* (1981), 1, pp. 63-75
[2] Saenko, V., *et al*., The Chernobyl Accident and its Consequences, *Clinical Oncology, 23* (2011), 4, pp. 234-243
[3] Koo, Y. H., Yang, Y. S., Song, K.W., Radioactivity Release from the Fukushima Accident and its Consequences: A Review, *Progress in Nuclear Energy, 74* (2014), July, pp. 61-70

[4] Aven, T., On the Meaning of a Black Swan in a Risk Context, *Safety Science, 57* (2013), Aug., pp. 44-51
[5] Van de Schoot, R., *et al*., Bayesian Statistics and Modelling, *Nature Reviews Methods Primers, 1* (2021), 1, p. 1
[6] Edwards, A. A., Lloyd, D. C., Risks from Ionising Radiation: Deterministic Effects, *Journal of Radiological Protection, 18* (1998), 3, p. 175
[7] Radvanyi, P., Villain, J., The Discovery of Radioactivity, Reports, *Physics, 18* (2017), 9-10, pp. 544-550
[8] Lindell, B., A History of Radiation Protection, *Radiation Protection Dosimetry, 68* (1996), 1-2, pp. 83-95
[9] Sudhakar, A., History of Cancer, Ancient and Modern Treatment Methods, *Journal of Cancer Science & Therapy, 1* (2009), 2, p. 1
[10] Nature, E., The Antibiotic Alarm, *Nature, 495* (2013), 7440, p. 141
[11] Wang, X., Peng, Z., Method of Moments for Estimating Uncertainty Distributions, *Journal of Uncertainty Analysis and Applications, 2* (2014), pp. 1-10, https://doi.org/10.1186/2195-5468-2-5
[12] Ali, Z., Bhaskar, S. B., Basic Statistical Tools in Research and Data Analysis (vol 60, pg 662, 2016), *Indian Journal of Anaesthesia, 60* (2016), 10, pp. 790-790
[13] Lo, B. W., *et al*., Clinical Outcome Prediction in Aneurysmal Subarachnoid Hemorrhage Using Bayesian Neural Networks with Fuzzy Logic Inferences, *Computational and Mathematical Methods in Medicine,* (2013), DOI 10.1155/2013/904860
[14] Gürsel, G., Healthcare, Uncertainty, and Fuzzy Logic, *Digital Medicine, 2* (2016), 3, pp. 101-112
[15] Jayalakshmi, M., *et al*., Fuzzy Logic-Based Health Monitoring System for COVID'19 Patients, Computers, *Materials and Continua, 67* (2021), 2, pp. 2431-2447
[16] Suvaila, R., Stancu, E., Sima, O., On Within Sample Homogeneity Testing Using Gamma-Ray Spectrometry, *Applied Radiation and Isotopes, 70* (2021), 9, pp. 2144-2148
[17] Spiegelhalter, D. J., *et al*., An Introduction to Bayesian Methods in Health Technology Assessment, *Bmj, 319* (1999), 7208, pp. 508-512
[18] Capellari, G., Chatzi, E., Mariani, S., Structural Health Monitoring Sensor Network Optimization Through Bayesian Experimental Design, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 4* (2018), 2, p. 04018016
[19] Ashby, D., Bayesian Statistics in Medicine: a 25 Year Review, *Statistics in Medicine, 25* (2006), 21, pp. 3589-3631
[20] Ozsahin, D. U., *et al*., Fuzzy Logic in Medicine, *Biomedical Signal Processing and Artificial Intelligence in Healthcare, New York, USA, 2020,* Academic Press., pp. 153-182
[21] Zarikas, V., Papageorgiou, E., Regner, P., Bayesian Network Construction Using a Fuzzy Rule Based Approach for Medical Decision Support, *Expert Systems, 32* (2015), 3, pp. 344-369
[22] Zarikas, V., *et al*., Medical Decision Support Tool from a Fuzzy-Rules Driven Bayesian Network, In ICAART (2), 2018, pp. 539-549, https://doi.org/10.5220/0006642705390549
[23] Schumacher, J., *et al*., A Randomised Crossover Simulation Study Comparing the Impact of Chemical, Biological, Radiological or Nuclear Substance Personal Protection Equipment on the Performance of Advanced Life Support Interventions, *Anaesthesia, 72* (2017), pp. 592-597, DOI 10.1111/anae.13842

Received on April 15, 2025
Accepted on June 9, 2025

Рареш ШУВАЈЛА

THRESHCHRONIC – ПРЕДЛОГ КОНЦЕПТА ФУЗИ БАЈЕСОВСКОГ АЛАТА ЗА ПОДРШКУ ЗДРАВЉУ

Мноштво хроничних болести још увек није класификовано као стохастичко или детерминистичко. Циљ овог рада је да се обезбеди адаптивни софтверски алат који омогућава детекцију прагова за узрочно-последичне односе са максималним степеном поузданости, с обзиром на доступну статистику. Овај алат је намењен за тестирање на скуповима података о радиоактивности, за које прагови постоје и добро су познати, а затим за примену на друге скупове података о радиоактивности и микропластици (животна средина), антибиотицима (фармаколошки фактори) и нитритима (прехрамбена индустрија). Главни алгоритам за овај програм подршке за истраживање здравља изграђен је као јединствена процедура за све ситуације; биће извршена додатна анализа коваријансе на вишефакторским случајевима, заједно са шемом која објашњава значај фактора преноса за податке забележене на различитим супстанцама (на пример, пренос радиоактивности и микропластике у скупу вода-риба-човек). Кључни елементи су доступни непристрасни скупови података и алгоритми за прорачуне условне вероватноће и одређивање потенцијалних прагова. Једна веома важна акција коју треба предузети јесте одвајање тачних и проверљивих података од расположивог мноштва пристрасних информација.

Кључне речи: стохастички процес, детерминистички процес, утицај околине, фази логика, бајесовска статистика, корелациона анализа