A digital multi-informational stochastic dynamic medical decision-making model and the application

Xiaonan Xiao
Xiamen University Tan Kah Kee College, Zhangzhou 363105 China

Corresponding author and e-mail: Xiaonan Xiao, xiaoxn@xujc.com

Abstract. As we all know, big data and intelligent computing have been widely used in information science, life science, computer science and intelligent control, intelligent robots, vehicle networking, space technology, marine development and other fields, especially in the future life science and medical field. The results of the discussion show that the intersection and fusion research of big data analysis, mathematics, computational science and life science in the 21st century is getting more tightly. We can understand life science from a new perspective by using big data and Intelligent Computing technology. Thus a new research model is being used to shift the application of mathematics from non-life to life. Furthermore, it greatly strengthens the mutual penetration and connection of big data analysis, mathematics, computational science and life science. It also greatly accelerates the research process of modern life science.

1. Data thinking and intelligent thinking in the information age

The intraclass correlation coefficient is useful for measuring the degree of dependence of observations within a group. In particular, it has been used for measuring the degree of familiar resemblance on either animal or human populations; It may also be used to measure the degree of resemblance within geographical units: for instance, in a study on cerebral aging, it is of interest to know whether there is a resemblance with respect to cognitive performance between subjects living in the same parish. In psychology, it has also a central role in the studies of reliability of tests [1-4].

The intraclass correlation coefficient is generally defined in the framework of a normal linear random effect model.

\[ X_{ij} = \mu + a_i + e_{ij} \] (1)

Where \( \mu = E(X_{ij}) \), the group effects \( \{a_i\} \) and the residual effects \( \{e_{ij}\} \) are independently normally distributed with mean 0 and variance \( \sigma_a^2 \) and \( \sigma_e^2 \) respectively. The variance of \( X_{ij} \) is \( \sigma^2_s = \sigma_a^2 + \sigma_e^2 \) and the intraclass correlation is defined as

\[ \rho_a = \frac{\sigma_a^2}{\sigma_s^2} \] (2)

For binary data, Donald and Donner have defined the intracluster correlation coefficient based on the following random effect model: \( X_{ij} \) are Bernoulli variables with probability \( \rho_i \); the \( \rho_i \)'s are
themselves random variables, identically and independently distributed with expectation $\pi$ and variance $\text{var}(\rho)$. The intracluster correlation coefficient is defined as

$$\rho_b = \frac{\text{var}(\rho)}{\pi(1-\pi)} \quad (3)$$

In this paper we give a general definition of the intraclass correlation coefficient $\rho$ which is valid for a general class of models and which reduces to $\rho_c$ in case of the linear model and to $\rho_b$ in Donald and Donner’s model. For the exponential family of distributions, we propose to test the hypothesis “$\rho = 0$” by the score test, which appears to be invariant over this family. This test is extended to adjust for explanatory variables [5-7].

2. Probability method for early diagnosis of disease
If there is a strong correlation between disease $A$ and factor $B$, and the positive rate of factor $B$ in healthy people is $\alpha\%$, and the positive rate of factor $B$ in patient $A$ is $\beta\%$ ($\beta > \alpha$ and $\beta$ close to 1). If the diagnosis is based on whether there is a factor $B$, there will be false positives $\alpha\%$, false negatives $1-\beta\%$. If the pre-clinical probability of diagnosing a patient's illness $A$ can be counted $P_0$, then when the patient has factor $B$, the probability of diagnosing the illness $A$ is:

$$P(A | B^+) = \frac{1}{1 + \frac{1-P_0}{P_0} \frac{\alpha}{\beta}}$$

When the patient does not carry a factor $B$, the probability of being diagnosed as the $A$ illness is:

$$P(A | B^-) = \frac{1}{1 + \frac{1-P_0}{P_0} \frac{1-\alpha}{1-\beta}}$$

Therefore because $P(A | B^+) > P_0, P(A | B^-) > 1-P_0$, it can be explained whether the examination really carries a factor $B$. Then the reliability of diagnosis can be improved.

3. Evaluation of carcinogenicity of chemicals
Quantitative analysis of the carcinogenicity of a chemical requires $n$ experiments. If the sensitivity of the test No. $i$ is $\alpha_i^+ = P(M_i^+ | C^+)$, Specificity $\alpha_i^- = P(M_i^- | C^-)$, Result group $M = \{M_1, M_2, \cdots, M_n\}$, when the result is positive, it is recorded as $M_i^+$. When the result is negative, it is recorded as $M_i^-$. Present ratio is

$$\frac{P(C^+ | M)}{P(C^- | M)} = \frac{1}{P(C^-) / P(C^+)} \frac{P(M_1 | C^+) \cdot P(M_2 | C^+) \cdots P(M_n | C^+)}{P(M_1 | C^-) \cdot P(M_2 | C^-) \cdots P(M_n | C^-)}$$

and we set

$$\alpha = \frac{P(C^-)}{P(C^+)}, \gamma = \frac{P(M_1 | C^+) \cdot P(M_2 | C^+) \cdots P(M_n | C^+)}{P(M_1 | C^-) \cdot P(M_2 | C^-) \cdots P(M_n | C^-)}$$

so

$$P(C^+ | M) = \frac{\gamma}{\gamma + \alpha} \cdot P(C^-) \cdot \frac{1}{\gamma + \alpha} + \frac{1}{\gamma + b} \cdot \gamma \cdot \frac{P(C^+)}{P(C^-) / P(C^+)}$$

(Not: The value $a, b$ is deduced from toxicological statistics). The carcinogenicity of a chemical can then be quantitatively evaluated.
4. Analysis of application examples

Now apply the above methods to diagnose acute ileus.

Acute ileus in clinic is classified into two kinds of wring ileus and pure ileus. There are 682 cases (among them D1 of 358 cases of pure ileus, D2 of 324 cases of wring ileus) in altogether 43 signs about diagnosis of acute ileus, and according to formulas (1)~(2) calculation results, the information evaluation about symptom of the ileus discernment and diagnosis(nat), which are $H(Z \mid S_{bij})$, $T(Z,S_{bij})$, $H(Z \mid S_k)$, and $T(Z,S_k)$, can be obtained.

According to the obtained information, we can depend on $T(Z,Z_t)$ value to evaluate the diagnostic value of all kinds of symptoms, and then get an arrangement in $T(Z,Z_t)$ from large to small as the following:

S1,S9,S15,S8,S7,S16,S3,S12,S10,S2,S11,S6,S13,S41,S5,S17,S14.

If subtract $T(Z,S_{ij}) < 0.01$, (nat) of S4, S5, S17 and S14 of symptoms, we establish the criterion of diagnosis 13 important symptoms of the surplus, and then we get the result of $H_1(S_{bij})$ and $H_2(S_{bij})$. Furtherly, we get the following 13 important symptoms, such as Fall ill(S1), Muscle tensity (S9), Leukocyte number (thousand, S15), Pressed pains of abdomen (S8), Chirping of instestine (S7), Neutral leukocyte (% , S16), Abdomen-ache intensity (S3), Pulse(frequency/min, S12), Counterbounce pain(S10), Abdomen-ache character(S2), Abdomen tumor(S11), the history of operated abdomen(S6), Blood pressure(S13,Pa), the statistics of H Numerical Value Applying for Discernment and Diagnosis, which are $H_1(S_{bij})$, $H_2(S_{bij})$ and $H_3(S_{bij})$, can be obtained as well.

After that, calculate H1 and H2, according to formula (3) on the basis of a group of 13 symptoms, which the patients are to be diagnosed, if H1>H2, the diagnosed patient contracts the disease D1 (pure ileus); if H1<H2, the patient contracts the disease D2(wring ileus).

On condition that only two diseases, also firstly calculate

$\Delta H(S_{ij}) = H_1(S_{ij}) - H_2(S_{ij})$, $k = 1,2,3,\ldots,r$, $j = 1,2,\ldots,n_k$

after that, according to the patients to be diagnosed of r specific symptoms (this case r=13), we calculate: $\Delta H = \sum_{k=1}^{r} \Delta H(S_{ij})$, if $\Delta H < 0$, it is disease D1, if $\Delta H > 0$, it is disease D2.

Mr zhang, a 63-year-old male patient, is a worker. His 13 symptoms are implicitly recorded as following: Yesterday he was working on the daytime. Two hours later after supper his belly button ached suddenly and at the same time he was sick and vomited, but he could still endure the occasional pain. Last year, he had his appendix amputated; in the physical exam, he was all right in general. His blood pressure is 130/85 mmHg, and blood pulse is 75 times per minute. His belly muscle felt soft and there was a little pain when they were pressed and when he jumped, but there was no tumour. The whistle voice of his intestines could be heard occasionally. The result of chemical exam showed that leukocyte number was 5800, 80% was neutral.

On the basis of the proceeding data, we can calculate the 13 symptoms of $H_1(S_{ij})$ and $H_2(S_{ij})$. After adding up them, we know $H_1 = 4.250$ (nat), $H_2 = 4.261$ (nat) or $\Delta H = -0.432$ (nat). Therefore, we think that the patient has pure clogged intestines D1, and this conclusion is in accordance with the final diagnosis of the hospital.

Now let’s choose 166 patients who have clogged intestines and whose symptoms are all recorded, then give them a reviewing exam: 165 cases in the exam were correctly judged and only one was wrongly judged. The rate of precision was 99.4%. Obviously it was higher than compared to the result...
of Bayes’s judgement, in the Bayes’s method, 154 cases were wrongly judged and the rate of precision was 92.8%.

In this paper, the information method for statistical analysis majorization is not only an extension of stationary method for statistical analysis, but also an expansion of classic theory of information. In this paper, the author brings for the entropy of information. Meanwhile, the diagnosis of diseases provides a new way in the exploring of majorization of statistical system.

5. Prospect
The development of modern decision-making science and the deepening of decision-making research, as well as the new problems raised in decision-making practice, have prompted the decision-making methods to be mathematically, modeled and computerized, which in turn requires the continuous improvement of mathematical processing means and logical procedures, and has promoted the development of mathematics. Fuzzy decision-making is a new one. The main contribution of Xing's frontier discipline is that it integrates Fuzziness with mathematical quantitative research. Its method is not to let mathematics abandon strictness to accommodate fuzziness, but to penetrate mathematical methods into the forbidden zone with fuzzy phenomena, so as to solve some scientific decision-making problems of complex large systems and fuzzy factors. It has opened up a new road, and its influence on decision-making science is far-reaching [8-9].

With the development of the information society, the digitalization and informatization of medical decision-making and medical genetic information prediction will become the key to sustainable development of medical decision-making and medical genetic information prediction. The demand for information from statistical analysis of medical diagnosis and medical genetic prediction will increase day by day. As a valuable resource, information will be paid more and more attention by decision-making and analysis departments at all levels. It can be predicted that the entire statistical analysis process will constitute a complete information system. An important aspect of the comprehensive evaluation of pre-decision systems will be an all-round systematic analysis of the value of information. It can be seen that the collection, processing, analysis and control utilization of statistical information have very important practical significance in both theoretical research and practical application [10-13].

The statistical analysis method in medical genetic information prediction proposed in this paper is not only an extension of traditional statistical analysis methods, but also an extension of classical information theory. It will provide a new way to further explore how to enrich, improve and improve the level of digital medical decision-making and medical genetic information prediction and optimization.

References
[1] Mo Jiaqi, Yao Jingsun. A class of Singular perturbed nonlinear reaction diffusion equations with two parameters [J]. J Math, 2011, 33(2):341-346
[2] Mo Jiaqi, Liu Shude. Singularly perturbed solution for semilinear reaction diffusion equation with two parameters [J]. Applied Mathematics and Mechanics, 2009, 30(5):607-612
[3] Zhou Mingguru, Du Zengji, Wang Guangwa. The Theory of Differential Inequality in Singular Perturbation [M]. Beijing: Science Press, 2012.
[4] Ren Y, Lu S P, Xia N M. Remarks on the existence and uniqueness of the solutions to stochastic functional differential equations with infinite delay [J]. Comput Appl Math, 2008, 220:364-372
[5] Qiu Jinming, Zhang Li. F-interfere law generation and its feature recognition [J]. Journal of Systems Engineering and Electronics, 2009, 20(4):777-783.
[6] Liu Shude, Lu Shiping, Yao Jingsun, Chen Huai-jun. Singular Perturbation Boundary and Corner Layer Theory [M]. Beijing: Science Press, 2012.
[7] Guiso L, Jappelli T. Financial Literacy and Portfolio Diversification [J]. General Information, 2009,10(5):515-528.
[8] Shi Kaiquan, Yao Bingxue. Function S-rough sets and las identification [J]. Science in China Series F: Information Sciences, 2008, 51(5):499-510

[9] Chen F. D, Li Z, Huang Y. J. Note on the permanence of a completable system with infinite delay and feedback controls [J]. Nonlinear Anal. Real World Appl., 2007 8(2):680-687

[10] Yuan C G, Willian G. Approximate solutions of stochastic differential delay equations with Markovian switching [J]. Comput Appl Math, 2006, 194:207-226

[11] Dogru O, Duman O. Statistical approximation of Meyer-Konig and Zeller operators based on q-integers. Publ Math Debrecen, 2006, 68(1-2): 199-214

[12] Frauenfelder P, Schwab C, Todor R A. Finite elements for elliptic problems with stochastic coefficients [J]. Computer Methods in Applied Mechanics and Engineering, 2005, 194(2/5): 205-228

[13] Yu L, Chen G D, Chu J. Optimal Guaranteed Cost Control of Linear Uncertain Systems: LMI Approach[J]. IFAC, 1999, G-2e-21-1: 541–546.