Journal of Applied Science and Engineering Innovation, Vol.6, No.4, 2019, pp. 181-183
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

Statistical Analysis of Medical Genetic Information Prediction

Xiao Xiaonan
Xiamen University Tan Kah Kee College, Zhangzhou, Fujian,China
xiaoxn@xujc.com

Abstract*:* The medical genetic information system is a complex information system that needs to be deeply studied in the comprehensive application of medicine, pharmacology, biological science, chemistry, information science, mathematics and computer science. It has far-reaching research significance and application prospects. The research on prediction and decision-making of medical genetic information system has become a major topic in mathematics and medical workers in recent years. Thus based on a new idea of digital medical genetic information optimization decision-making, this paper extends non-digital medical genetic information prediction to digital medical genetic information prediction in the generalized information measure space, and makes full use of probabilistic and statistical methods to make pre-decision on early diagnosis of diseases, carcinogenicity of chemical substances, genetic counseling and paternity testing. The problems were discussed and studied in depth. It provides a new theory, new approach and new method for further improving the optimal prediction level of medical genetic information system and effectively enhancing the pre-decision benefit of medical genetic information system.

Keywords Medical genetic information system; Probabilistic statistical method; Prediction; decision benefit; Optimization level

INTRODUCTION

Probabilistic methods have a wide range of applications in the prediction of medical genetic information, especially in the early diagnosis of diseases, carcinogenicity test of chemical substances, genetic disease counseling and paternity testing, etc.. [Choi, *et. al.*, 2007] This method is used to quantitatively describe the probability of occurrence of an event, so that people understand the results of the prediction.[Distinguin, *et. al.*, 2013]; [Kochubey, *et. al.*, 2014] It provides effective data and information for the relevant personnel to consider and implement countermeasures.[Zeng, *et. al.*, 2019]

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 β close to 1). If the diagnosis is based on whether there is a factor B, there will be false positives $\alpha\%$, false negatives $1-\beta\%$. [Jokipii, *et. al.*, 2011]; [Roger, *et. al.*, 2011] 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: $P(A \mid B^+) = \dfrac{1}{1+\dfrac{1-P_0}{P_0}\dfrac{\alpha}{\beta}}$

When the patient does not carry a factor B, the probability of being diagnosed as the A illness is [De, *et. al.*, 2012]:

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

Therefore because

$P(A \mid B^+) > P_0, P(\overline{A} \mid B^-) > 1-P_0$, it can be explained whether the examination really carries a factor B. Then the reliability of diagnosis can be improved [Behrman, *et. al.*, 2012].

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 $\alpha_i^+ = P(M_i^+ \mid C^+)$, Specificity $\alpha_i^- = P(M_i^- \mid C^-)$, Result group $M = \{M_1, M_2, \mathrm{L}\ , M_n\}$, 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

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

and we set

$$\alpha=P(C^-)/P(C^+),\gamma=\frac{P(M_1\mid C^+)\cdot P(M_2\mid C^+)\mathrm{L}\ P(M_n\mid C^+)}{P(M_1\mid C^-)\cdot P(M_2\mid C^-)\mathrm{L}\ P(M_n\mid C^-)}$$

so $P(C^+\mid M)=\frac{\gamma}{\gamma+\alpha}$. When $a\le\alpha\le b$, we get the estimated value of $P(C^+\mid M)$ which is $\hat{P}(C^+\mid M)=\frac{\gamma}{2}\left(\frac{1}{\gamma+a}+\frac{1}{\gamma+b}\right)$. (Note: The value a,b is deduced from toxicological statistics). The carcinogenicity of a chemical can then be quantitatively evaluated [Meng, *et. al.*, 2007].

GENETIC COUNSELING

A woman asks for genetic counseling, and her (A) uncle and her mother's (B 's) uncle both have Z-linked recessive genetic disease. It is known that she has a sister (C) and a normal son (D). A has a healthy brother (E). A hopes to know what the risk of recurrence of hereditary diseases in her unborn children (F) is ?

Assuming

$$B_1=\{B^+\},B_2=\{B^-\},C_1=\{C^+\},C_2=\{C^-\},D=\{C\ \ has\ \ a\ \ healthy\ \ son\ \ D\},$$
$$E=\{B\ \ has\ \ a\ \ healthy\ \ son\ \ E\}\text{,Since}$$

$$\begin{aligned}P(DE)&=P(B_1)P(DE\mid B_1)+P(B_2)P(DE\mid B_2)\\&=P(B_1)P(D\mid B_1)P(E\mid B_1)+P(B_2)P(D\mid B_2)P(E\mid B_2)\\&=P(B_1)[P(C_1)P(D\mid C_1)+P(C_2)P(D\mid C_2)]P(E\mid B_1)+P(B_2)P(D\mid B_2)P(E\mid B_2)\\&=\frac{1}{2}\left(\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot1\right)\cdot\frac{1}{2}+\frac{1}{2}\cdot1\cdot1=\frac{11}{16}\end{aligned}$$

We can get $P(B_1\mid DE)=\frac{3}{16}\Big/\frac{11}{16}=\frac{3}{11}$, $P(F^+\mid DE)=\frac{1}{4}\cdot\frac{3}{11}=\frac{3}{44}\approx6.8\%$. This indicates that the probability of the woman's unborn child (F) carrying the pathogenic gene (reappearance risk) is 6.8%.

PATERNITY TESTING

In order to determine whether a father is a real father, some blood group systems of mother, children and assumed father should be examined, and then the paternity index and the probability of the paternity can be calculated from the results and genotype frequencies to determine or exclude paternity.[Qing, *et. al.*, 2019]

Assuming that the phenotypes of a blood group system of mother, son and father are respectively M , C and F , we define the paternity index as $\mathrm{L}=\frac{\mathrm{P(C|FM)}}{\mathrm{P(C|M)}}$. We can prove $\mathrm{L}=\frac{\mathrm{P(FMC)}}{\mathrm{P(F)P(MC)}}$. The paternity probability is $W=\frac{L}{1+L}$. If n blood type systems are examined, the paternity indexes respectively are $L_1,L_2,\mathrm{L}\ ,L_n$. When the blood groups are independent, the cumulative value of paternity index is $L=L_1,L_2,\mathrm{L}\ ,L_n$. Then the probability of paternity [Harvey, *et. al.*, 2009] is

$$W=\frac{L}{1+L}=\frac{L_1\cdot L_2\mathrm{L}\ L_n}{1+L_1\cdot L_2\mathrm{L}\ L_n}$$

An example: After examination of the rh blood group system, the mother phenotype is ccDEe, and the child phenotype is CcDEe. Assume that the father is CCDee type, and the frequency of the rh gene combination is known to be cDE0.2679, cDe0.0280, cdE0.0032, cde0.0452, CDE0.0048, CDe0.6405, CdE0.0022, Cde0.0082, we can calculate first mother, Assuming that the father passes the probability of various single types, we can list all possible genotypes of the child, and determine the single type from the birth mother or the biological father. Finally, we calculate the probability $P(C\mid MF)$ that the child's phenotype comes from the mother and the hypothetical father and the probability $P(C\mid M)$ that the child's phenotype comes from the mother and the random man. We get $P(C\mid MF)=0.49997$ and $P(C\mid M)=0.32716$ respectively. So

$L=\frac{0.49997}{0.32716}=1.53$, $W=0.6047$. We examine 6 blood types. After merging $L=9.41, W=0.9039$. If the father is supposed to involve non-genetic evidence, the pre-probability of the father is 0.9, then the post-probability of paternity is $W=98.83\%>95\%$. Then basically we assume that the father is the child's biological father.

EXPECTATION

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 [Xu, et. al., 2019].

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

Behrman J R, Mitchell O S, Soo C K, et al. How financial literacy affects household wealth accumulation [J]. American Economic Review, 2012,102(3):300-304.

Choi C, Moh Y. How useful are tests for unit-root in distinguishing unit-root process from stationary but non-linear process[J]. Econometrics Journal,2007(10):82-112.

De Nicolo G, Gamba A, Lucchetta M. Capital Regulation , Liquidity Requirements and Taxation in a Dynamic Model of Banking[J]. Ssrn Electronic Journal, 2012, 12(72):5108-5122.

Distinguin I, Roulet C, Tarazi A. Bank Regulatory Capital and Liquidity: Evidence from US and European publicly traded banks [J]. Journal of Banking & Finance, 2013, 37(9):3295-3317.

Harvey D I, Leybourne S J, Taylor A M R. Unit root tesing in practice: dealing with uncertainty over the trend andinitial condition [J]. Econometric Theory, 2009(25):5879-5898.

Jokipii T, Milne A. Bank capital buffer and risk adjustment decisions[J]. Journal of Financial Stability, 2011, 7(3):165-178.

King M R. Mapping Capital and Liquidity Requirements to Bank Lending Spreads[J]. Ssrn Electronic Journal, 2010, 68(3):1-18.

Kochubey T, Kowalczyk D. The Relationship between Capital, Liquidity and Risk in Commercial Banks[J]. Dubrovnik, 2014.

Kilicr A. Testing procedure for a linear unit root against stationary ESTAR process[J]. Econometric Reviews,2011(30):274-302.

Meng, X. Wealth Accumulation and Distribution in Urban China [J]. Economic Development and Cultural Change, 2007,55(4):761-791.

Park J, Shintani M. Testing for a unit root against transitional autoregressive models[J]. Vanderbilt University Department of Economics Working Papers,2010.

Qing, Z. (2019). Research on the Impact of Carbon Emission Trading Mechanism on Power Industry Based on SWOT Analysis. Journal of Applied Science and Engineering Innovation, 6(3), 117-121.

Repullo R. Liquidity, Risk-Taking and the Lender of Last Resort[J]. International Journal of Central Banking, 2005,1(2):47-80.

Rodrigues P M. Properties of recursive trend-adjusted unit root tests[J]. Economics Letters, 2006(91):413-419.

Roger S, Vlcek J. Macroeconomic Costs of Higher Bank Capital and Liquidity Requirements[J]. Imf Working Papers, 2011, 11(103).

Xu P.; Na N.; Gao S.; Geng C., Determination of sodium alginate in algae by near-infrared spectroscopy, Desalination and Water Treatment, 168(2019)117-122.

Zeng, H. Y. (2019). Improved Particle Swarm Optimization Based on Tabu Search for VRP. Journal of Applied Science and Engineering Innovation, 6(2), 99-103.