**DOI: http://doi.org/10.32792/utq.jceps.10.01.01**

# Effective of Renyi Entropy in Markov Basis for Independent Model

**Ali Talib Mohammed[1]** **Maysaa Jalil Mohammed[2]**

[1]ali.t.m@ihcoedu.uobaghdad.edu.iq, [2]maysaa.j.m@ihcoedu.uobaghdad.edu.iq

[1,2]Department of Mathematics, College of Education For Pure Science ( Ibn Al Haitham), University of Baghdad, Iraq

**Received 23/4/2023, Accepted 29/5/2023, Published 11/6/2023**



**Abstract**

The general concept of algebraic statistics is to employ algebra tools to provide a better view of the structure of statistical problems and to contribute to finding solutions. The algebraic and statistical mixture represented by the method of Markov basis for the independent model (MBIM) on the contingency table is considered highly efficient to study and analyze the target problem which is Rheumatoid arthritis. The verification of the proposed method is carried out by taking advantage of the effect of the entropy property by using the law of Renyi entropy to show which of the alternate matrices that exist within the fiber elements are more independent than others.

**Keywords**: Contingency table, Markov Basis for independent Model, Renyi Entropy.

## 1. Introduction

As the name implies, rheumatoid arthritis is a chronic inflammatory disease that affects not only the joints but also other parts of the body. There are several body systems that are at risk from this disease, including the skin, eyes, lungs, heart, and blood vessels. The immune system attacks the human body accidentally in rheumatoid arthritis, an autoimmune disorder. This disease affects most cases of both sexes between the ages of 40 - 60 years. The incidence of rheumatoid arthritis increases in people whose family members have been exposed to this disease. On the other hand, doctors confirm that this disease is not considered a genetic disease. The main interest in statistics is to understand and analyze the relationships and dependence between variables and this is available through organizing the data represented by the variables in the two-way contingency table. (Koch 1988) represented Rheumatoid arthritis as a two-way contingency table with factors of Sex, Treatment, and Improvement. (Aoki, Hara, and Takemura 2012) defined Markov basis as a set of moves that are used to construct through any fiber a Markov Chain. Many relationships between variables have been studied by contingency tables through Markov basis. (Yoshida 2000) explained the importance of the relationships that the moves and binomials have in the Toric ideal and that contribute to reaching alternatives. (Rényi 1961) was first introduced of Renyi family measure information's which is, in fact, Shannon extension. (Maysaa J Mohammed et al. 2016) applied MBIM on a smoking problem in the form of a contingency table and benefit from Shannon entropy as via information method. (Maysaa Jalil Mohammed and Mohammed 2020) analyzed agriculture data with approaches of MBIM on the contingency table and Tsallis Entropy. (Ali Talib Mohammed 2019) they discussed how to estimate appropriate functions from statistical data, especially non-parametric data, as well as how to ensure the data are independent. (Ali T. Mohammed et al. 2022) they discussed how to deal with complex data and the information attached to it through statistical analysis.. This paper presents an analysis of two-way contingency table data of Rheumatoid arthritis using MBIM to find the alternatives to the real baseline arthritis data matrix and then test these alternatives to find the most independent alternative through the Renyi entropy information law.

## 2. Proposed Methods

The objective of this study is to depend on the contingency table, relationship/correlations between the factors. It is possible to view the count of each cell in an array as the outcome of a multinomial probability distribution since all possible combinations of factor labels produce cells in an array. It is a matrix with positive cells; the sum of all cells in all rows equals the sum of cells in all columns. In the case of two-way contingency tables with i rows and J columns, we also denote the sample space as $X = \{1,...,I\} \times \{1,...,J\}$. It is a matrix with two conditions; the first one is the sum of the rows and columns are the same. The second is the entries of the contingency table are positive integers. The mechanism of finding alternative matrices is summarized in the following algorithm steps:
Step (1). Order the data as in Table 1, Arthritis treatment data (Koch 1988) which represents the relationship between the Improvement and treatment Sex.

Table (1): The relationship between improvement and the treatment sex

| | Treatment Improvement | | | |
|---|---|---|---|---|
| Sex | None | Some | Marked | Total |
| Female | 6 | 5 | 16 | 27 |
| Male | 7 | 2 | 5 | 14 |
| Total | 13 | 7 | 21 | N= 41 |

Step (2). Using the following hypothesis to test independence:
$H_0$: Relationship between the improvement and treatments sex are independent.
$H_1$: Relationship between the Improvement and treatments sex are dependent.
This hypothesis is testing with following Chi square formula (Zibran 2007):

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

With $X^2((r-1)(c-1))$ , where $r = No.rows$ and $c = No.columns$ . Moreover, the expected frequency is given as:

$$E_{r,c} = \frac{n_r \times n_c}{n}$$

Step (3). using moves to obtain Markov basis (MB):

$m_{ij} = \{+1 \quad (i,j) = (i_1, j_1), (i_2, j_2) \quad -1 \quad (i,j) = (i_1, j_2), (i_2, j_1) \quad 0 \quad otherwise$ Where, $MB(i_1, i_2, j_1, j_2) = \{m_{ij}\}$.

Step (4). $MB = \{m_1, m_2, m_3, m_4, m_5, m_6\}$ (Maysaa Jalil Mohammed and Mohammed 2020)

$m_1 = [0\ 1\ -1\ 0\ -1\ 1\ ]$ $\quad m_2 = [1\ 0\ -1\ -1\ 0\ 1\ ]$ $\quad m_3 = [1\ -1\ 0\ -1\ 1\ 0\ ]$

$m_4 = [-1\ 1\ 0\ 1\ -1\ 0\ ]$ $\quad m_5 = [-1\ 0\ 1\ 1\ 0\ -1\ ]$ $\quad m_6 = [0\ -1\ 1\ 0\ 1\ -1\ ]$

Step (5). Add the elements of Markov basis to the original matrix to find the elements of the fiber which is all the alternative matrix of the original contingency table for example add $A + m_1 = [6\ \ 7\ \ 6\ 1\ 15\ 6\ ] = B$, then stop after the cells become negative numbers as follows:

| | | | | | |
|---|---|---|---|---|---|
| $A$ = [6 7 5 2 16 | $B$ = [6 7 6 1 15 | $C$ = [6 7 6 1 15 6 | $D$ = [7 6 4 3 16 | E= [8 5 3 4 16 | F= [9 4 2 5 16 5 |
| G= [10 3 1 6 16 5 | H= [11 2 0 7 16 5 | I = [7 6 5 2 15 6 ] | J = [8 5 5 2 14 7 | K = [9 4 5 2 13 8 | L= [10 3 5 2 12 9 |
| M= [11 2 5 2 11 1 | N= [12 1 5 2 10 1 | O= [7 6 5 2 15 6 ] | P= [5 8 5 2 17 4 | Q= [4 9 5 2 18 3 | R= [3 10 5 2 19 2 |
| S= [2 11 5 2 20 1 | T= [1 12 5 2 21 | U= [5 8 6 1 16 5 ] | V= [4 9 7 0 16 5 | W= [6 7 4 3 17 4 | X= [6 7 3 4 18 3 ] |
| Y= [6 7 2 5 19 2 ] | Z= [6 7 1 6 20 1 | $A_1$= [6 7 0 7 21 0 ] | $B_1$= [7 6 6 1 14 7 | $C_1$= [8 5 6 1 13 8 | $D_1$= [9 4 6 1 12 9 ] |
| $E_1$= [10 3 6 1 11 1 | $F_1$= [11 2 6 1 10 | $G_1$= [12 1 6 1 9 | $H_1$= [13 0 6 1 8 1 | $I_1$= [5 8 7 0 15 6 | $J_1$= [7 6 7 0 13 8 ] |
| $K_1$= [8 5 7 0 12 9 ] | $L_1$= [9 4 7 0 11 1 | $M_1$= [10 3 7 0 10 1 | $N_1$= [11 2 7 0 9 1 | $O_1$= [12 1 7 0 8 1 | $P_1$= [13 0 7 0 7 14 |
| $Q_1$= [8 5 7 0 12 9 ] | $R_1$= [9 4 4 3 14 7 | $S_1$= [10 3 4 3 13 8 | $T_1$= [11 2 4 3 12 | $U_1$= [12 1 4 3 11 | $V_1$= [13 0 4 3 10 11 |
| $W_1$= [7 6 3 4 17 4 ] | $X_1$= [7 6 2 5 18 3 | $Y_1$= [7 6 1 6 19 2 ] | $Z_1$= [7 6 0 7 20 1 | $A_2$= [9 4 3 4 15 6 | $B_2$= [10 3 3 4 14 7 |
| $C_2$= [11 2 3 4 13 8 | $D_2$= [12 1 3 4 12 | $E_2$= [13 0 3 4 11 1 | $F_2$= [8 5 2 5 17 4 | $G_2$= [8 5 1 6 18 3 | $H_2$= [8 5 0 7 19 2 ] |
| $I_2$= [10 3 2 5 15 6 | $J_2$= [11 2 2 5 14 | $K_2$= [12 1 2 5 14 7 | $L_2$= [13 0 2 5 13 | $M_2$= [9 4 1 6 17 4 | $N_2$= [9 4 0 7 18 3 ] |
| $O_2$= [11 2 1 6 15 6 | $P_2$= [12 1 1 6 14 | $Q_2$= [13 0 1 6 13 8 | $R_2$= [10 3 0 7 17 | $S_2$= [12 1 0 7 15 | $T_2$= [13 0 0 7 14 7 |
| $U_2$= [4 9 6 1 17 4 ] | $V_2$= [3 10 7 0 17 | $W_2$= [5 8 4 3 18 3 ] | $X_2$= [5 8 3 4 19 2 | $Y_2$= [5 8 2 5 20 1 | $Z_2$= [5 8 1 6 21 0 ] |
| $A_3$= [4 9 4 3 19 2 ] | $B_3$= [3 10 6 1 18 | $C_3$= [2 11 7 0 18 3 | $D_3$= [4 9 3 4 20 1 | $E_3$= [4 9 2 5 21 0 | $F_3$= [2 11 6 1 18 3 |
| $G_3$= [1 12 7 0 18 3 | $H_3$= [3 10 4 3 20 | $I_3$= [3 10 3 4 21 0 | $J_3$= [1 12 6 1 20 1 | $K_3$= [0 13 7 0 20 | $L_3$= [2 11 4 3 21 0 |
| $M_3$= [0 13 6 1 21 0 | $N_3$= [1 12 7 0 19 | | | | |

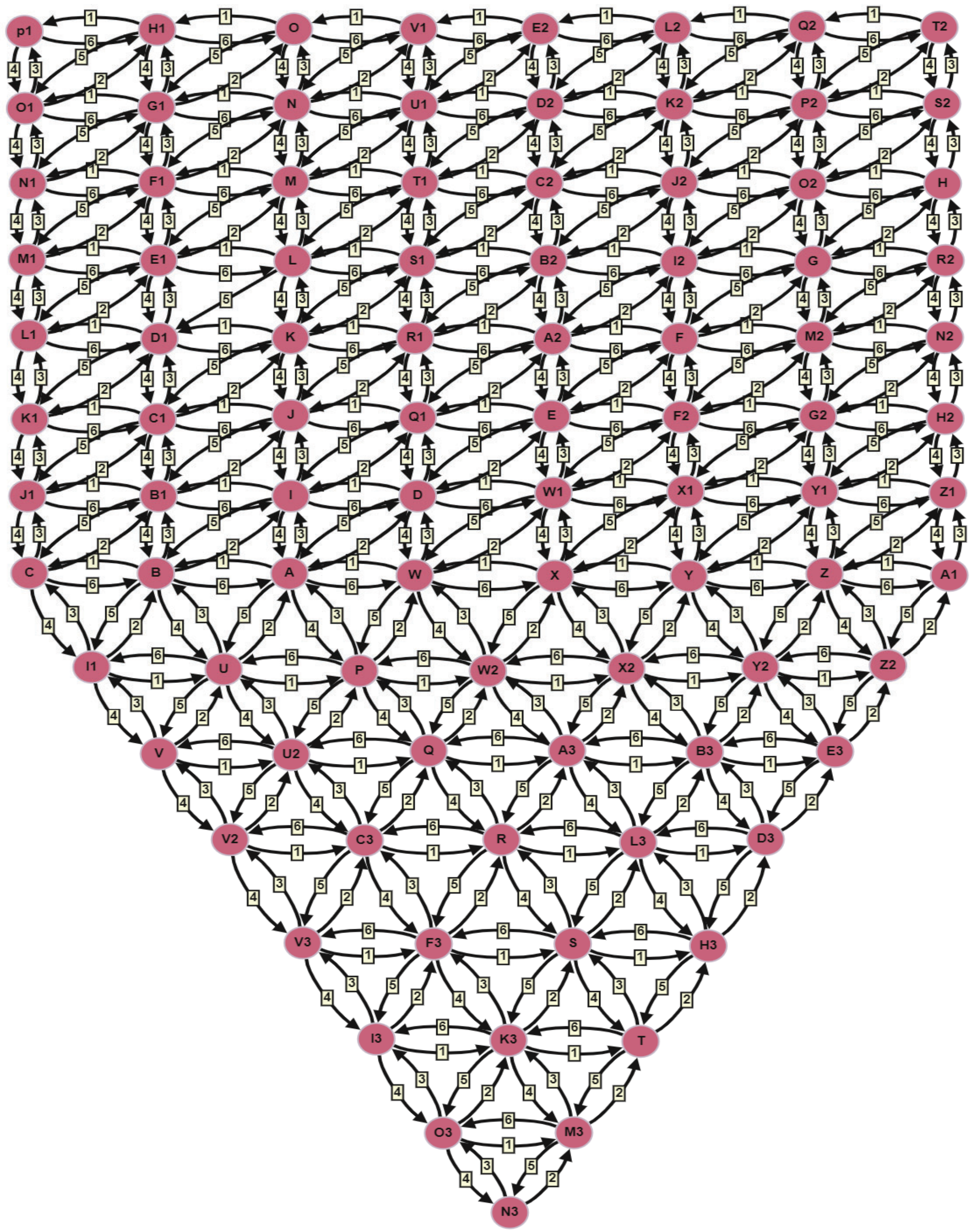


Figure (1). data fiber elements matrix

Step (6). Application of Renyi entropy law to measure independence, according to the following: (Rényi 1961)

$$H_\alpha = \frac{1}{1-\alpha} \, log\left(\sum_{i=1}^{n} \quad p_i^\alpha\right)$$

$H_\alpha$ is define as Renyi entropy value of order $(\alpha = 2), P_i$, is the prpbablity of the cells of contingeny table

Table 1. Entropy values for some alternative matrix

| Table | $H_\alpha$ | Table | $H_\alpha$ | Table | $H_\alpha$ |
|---|---|---|---|---|---|
| A | 0.62898427375 | B | 0.64236893952 | C | 0.64692850351 |
| D | 0.63339095708 | E | 0.62897061783 | F | 0.61591330422 |
| G | 0.59513983845 | H | 0.56755631679 | I | 0.65153758425 |
| J | 0.66566108843 | K | 0.67047326489 | L | 0.66566108848 |
| M | 0.65153644572 | N | 0.62897061785 | O | 0.59922734611 |
| P | 0.46363987501 | Q | 0.56375502793 | R | 0.5239997284 |
| $S$ | 0.48127473034 | $T$ | 0.43669259768 | $U$ | 0.61597435324 |
| $V$ | 0.59513983843 | $W$ | 0.60751961676 | $X$ | 0.579163978727 |
| $Y$ | 0.51123795371 | $Z$ | 0.50706602462 | $A1$ | 0.46589986876 |
| $B1$ | 0.66090164924 | $C1$ | 0.67047326493 | $D1$ | 0.65047326482 |
| $E1$ | 0.66090164924 | $F1$ | 0.6403689385 | $G1$ | 0.61597330428 |
| $H1$ | 0.62026266735 | $I1$ | 0.62459481779 | $J1$ | 0.66090164924 |
| $K1$ | 0.66566108845 | $L1$ | 0.66090164922 | $M1$ | 0.6469285035 |
| $N1$ | 0.62460030691 | $O1$ | 0.59513983847 | $P1$ | 0.55998678503 |
| $Q1$ | 0.65153644576 | $R1$ | 0.66090164918 | $S1$ | 0.66090164926 |
| T1 | 0.65153644576 | U1 | 0.63339095609 | V1 | 0.60751966761 |
| W1 | 0.60751961678 | X1 | 0.57526019035 | Y1 | 0.53803875225 |
| Z1 | 0.49721393145 | A2 | 0.642689952 | B2 | 0.64692850521 |
| C2 | 0.64236893952 | D2 | 0.62897061788 | E2 | 0.6075196676 |
| F2 | 0.61597330423 | G2 | 0.56375502796 | H2 | 0.5239997284 |
| $I2$ | 0.66566108843 | $J2$ | 0.62459481777 | K2 | 0.60335369051 |
| L2 | 0.59109044332 | M2 | 0.4387706445 | N2 | 0.54523220007 |
| O2 | 0.59922734611 | P2 | 0.595113983847 | Q2 | 0.58310319325 |
| R2 | 0.55998672245 | S2 | 0.56755631682 | T2 | 0.55998672245 |
| U2 | 0.583310319324 | V2 | 0.55998672243 | W2 | 0.57526019036 |

| | | | | | |
|---|---|---|---|---|---|
| X2 | 0.54523203505 | Y2 | 0.51040035562 | Z2 | 0.47197956111 |
| A3 | 0.53803858996 | B3 | 0.54523226008 | C3 | 0.52055975398 |
| D3 | 0.50706602462 | E3 | 0.47198465456 | F3 | 0.53096251455 |
| G3 | 0.50375709827 | H3 | 0.48996562046 | I3 | 0.46577979352 |
| $J3$ | 0.44669324165 | K3 | 0.43201478247 | L3 | 0.45398023257 |
| M3 | 0.41466343279 | N3 | 0.47815590557 | | |

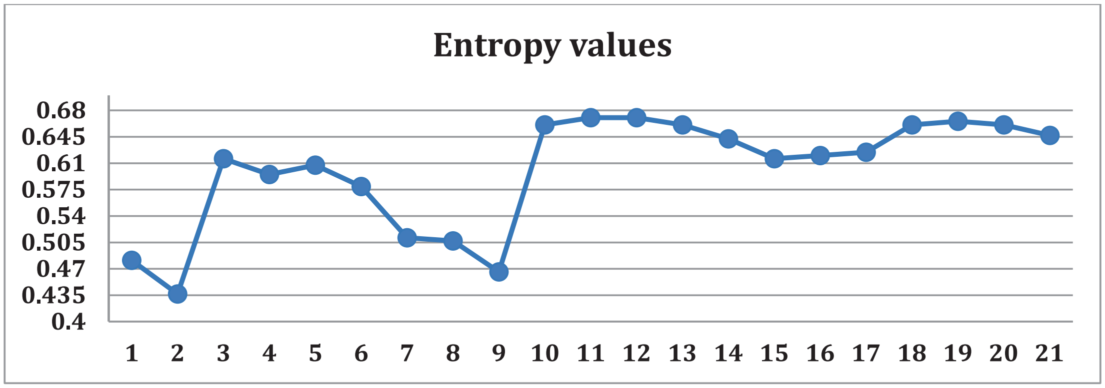


Figure 2. The entropy values for alternative tables

$(S, T, U, V, W, X, Y, Z, A1, B1, D1, E1, F1, G1, H1, I1, J1, K1, L1, M1)$ with original matrix$A$. Note for this table, value of $H_\alpha$ of matrix C1is higher than other matrix. Therefore, the matrix C1 is more independent comparing with other alternative matrix.

## 3. Result

The purpose of study is to construe the contingency statistical data through algebra, represented by the Markov basis, to obtain new data close to the original, to help researchers in understanding the relationships that link variables. For independence, the test used is $X^2(\ 2)$ (with degree of freedom equal to 2), to test the independence of the variables that represent the acceptability of the effect of active treatment for each of the sexes under study. Under significance level (0.05) and $X^2(2)$ value of the original contingency table is $(X_c^2(2) = 3.00847)$ and the Chi-square table value is $(X_t^2(2) = 5.99)$. It's clear that $(X_c^2(2) < X_t^2(2))$ which is mean that the variables are independent.

Arthritis data studied for 41 people of both sexes related to the effectiveness of the active treatment protocol. Through the proposed method, 92 alternative tables were produced for the original table, each of them represents the original table through the relationships between the variables in the original terms, as the alternative tables with the original are linked to each other through moves $MB = \{m_1, m_2, m_3, m_4, m_5, m_6\}$, and as shown in Figure 1, The determination of the alternative contingency table closest to the original contingency table is done through the application of Renyi entropy in terms of independence (Gray 2011), table (1) shows the Renyi entropy values for 20 alternative contingency tables with the original one (Dong 2016).

## 4. Conclusion

It can be concluded that the proposed method has proven excellent effectiveness through the applications of Markov foundations in contingency tables to find contingency alternatives through which to arrive at the best representative alternative to the origin through the information values of independence provided by Renyi Entropy.

## References

[1] Aoki, Satoshi, Hisayuki Hara, and Akimichi Takemura. 2012. 199 *Markov Bases in Algebraic Statistics*. Springer Science & Business Media.

[2] Dong, Xi. 2016. "The Gravity Dual of Rényi Entropy." *Nature communications* 7(1): 1–6.

[3] Gray, Robert M. 2011. *Entropy and Information Theory*. Springer Science & Business Media.

[4] Koch, G G. 1988. "Clinical Efficacy Trials with Categorical Data." *Biopharmaceutical statistics for drug development*.

[5] Mohammed, Ali T., Mayssa J. Mohammed, Muna D. Salman, and Rabha W. Ibrahim. 2022. "The Inverse Exponential Rayleigh Distribution and Related Concepts." *Italian Journal of Pure and Applied Mathematics* 47: 852–61.

[6] Mohammed, Ali Talib. 2019. "Density Estimation for Right Censored Data Using Hybrid Breslow and Semi-Symmetric Wavelet." *Journal of Engineering and Applied Sciences* 14(1): 242–46.

[7] Mohammed, Maysaa J et al. 2016. "A New Mathematical Evaluation of Smoking Problem Based of Algebraic Statistical Method." *Saudi journal of biological sciences* 23(1): S11–15.

[8] Mohammed, Maysaa Jalil, and Ali Talib Mohammed. 2020. "Analysis of an Agricluture Data Using Markov Basis for Independent Model." In *Journal of Physics: Conference Series*, IOP Publishing, 12071.

[9] Rényi, Alfréd. 1961. "On Measures of Entropy and Information." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, University of California Press, 547–62.

[10] Yoshida, Ruriko. 2000. "Connectivity of Fibers with Positive Margins in Multidimensional Contingency Tables." *Harmony of Gröbner Bases and the M odern Industrial Society*.

[11] Zibran, Minhaz Fahim. 2007. "Chi-Squared Test of Independence." *Department of Computer Science, University of Calgary, Alberta, Canada*: 1–7.