Parametric Tests of Equality of Several Univariate Frequency Distributions/Several Contingency Tables and Several Markov Chains/Several Transition Frequency Matrices

Mian Arif Shams Adnan
Department of Mathematical Sciences, Ball State University, Muncie, IN 47304

Abstract
The asymptotic tests for the equality of several frequency distributions and several markov chains have been developed. The tests of the equality of several contingency tables or frequency matrices and several transition frequency matrices have been discussed. Examples are cited for all cases.

Key Words: Chi-square Matrix ; Matrix of p-values ; Multi-level multi-variate vector.

1. Introduction

A frequency table is a summarized grouping of data into mutually exclusive classes and the number of occurrences in all cells. Each cell in the table contains the count of the occurrences of values within a particular interval or group. A bivariate joint frequency distribution is represented as a two way contingency table or matrix where the total row and total column report the marginal frequencies or marginal distributions and each cell of the body of the table refers the joint frequencies.

The term contingency table was first coined by Karl Pearson (1904). Numerous authors including Fisher (1922, 1925, 1935, 1962), Neyman (1928), Yates (1934, 1984), Wilks (1935), Deming and Stephan (1940), Barnard (1945, 1947, 1949, 1979), Pearson (1947, 1900, 1904), Lancaster (1949, 1969), Chernoff (1954), Lindley (1956), Bennett and Hsu (1960), Birnbaum (1962), Plackett (1964, 1977), Grizzle (1967), Boschloo (1970), Gail and Gart (1973), Kempthorne (1978), Gokhale and Kullback (1978), Berkson (1978), Mehta and Patel (1980), Upton (1982), Bartlett (1984), Goodman (1984), Suissa and Shuster (1985), Little (1989), Cox and Snell (1989), Agresti (1990, 2002, 2005), Gray (1990), Greenland (1991), Muse (1992), Berger and Boos (1994), Berger (1996), Kou and Ying (1996), Behseta (2005), Cheng (2008), Falay (2007), Klugkist (2010), Cho (2011) left their research works on contingency tables. No test has been developed so far for testing the equality of several contingency tables or joint frequency distributions or marginal frequency distributions. The authors aim to develop new asymptotic test statistics for checking the similarity or dissimilarity among the individual (cell) frequencies, marginal frequencies and overall discrepancy of several populations. Section 2 demonstrates the methods and methodology and the subsequent section displays a real life application of the proposed tests. The tertiary section draws the conclusion.

A stochastic process or random process is a collection of random variables that represents the evolution of some physical process through the change of time, state or space. There are several (often infinitely many) directions in which the process may evolve. In case of discrete time, a stochastic process amounts to a sequence of random variables known as a time series (for example Markov chain). And the other is a random field, whose domain is a region of space, or random function whose arguments are drawn from a range of continuously changing values. One approach to stochastic processes treats them as functions of

one or several deterministic arguments whose values (outputs) are random variables: non-deterministic (single) quantities which have certain probability distributions. Random variables corresponding to various times (or points, in the case of random fields) may be completely different. Although the random values of a stochastic process at different times may be independent random variables, in most commonly considered situations they exhibit complicated statistical correlations. Assessing these correlations can be evaluated by means of knowing transitions which express the changes of state of the system and the probabilities associated with various state-changes are called transition probabilities. Markov chain, due to Andrey Markov, is a mathematical system that undergoes transitions from one state to another, between a finite or countable number of possible states. It is a random process characterized as memoryless stating the conditional probability distribution for the sequence in the system at the next step (and in fact at all future steps) depending only on the current stat, and not additionally on the state at previous steps. So, a Markov Chain is completely characterized by the set of all states and transition probabilities. By convention, we assume all possible states and transitions have been included in the definition of the Markov processes in such a way that there is always a next state and the process goes on forever. Thus, Markov chains have many applications as statistical models of real-life processes.

Checking the discordance of two Markov Chains is a preliminary step of finding the mobility of any system over the change of time or place or other dimension(s). It is also a primary stage of comparing multiple Markov Chains. Unfortunately, comparison of Makov Chains is due to very few authors. Muse et al (1992) proposed a likelihood ratio test for testing the equality of evolution rates. Tan *et al* (2002) developed a Markov-chain-test for time dependence and homogeneity using likelihood ratio test statistic. Dannemann *et al* (2007) proposed a method of testing the equality of transition parameters based on transition probabilities and likelihood ratio test statistic that simply gives the significant dissimilarity of the total transition but not that of the individual transition. Falay, B. (2007) described intergenerational income mobility by testing the equality of opportunity due to knowing the comparison of East and West Germany using a transition matrix having positive and negative elements. Bartolucci, F. *et al* (2009) demostarted the use of a multidimensional extension of the latent Markov model using a multidimensional two parameter logistic model where they developed likelihood ratio test based on log of the ratio of transition probabilities. Cho, J. S *et al* (2011) expresses a test of equality of two unknown positive definite matrices with an application of information matrix testing. Hillary, R. M. (2011) proposed a Bayesian method of estimation the growth transition matrices. Altug, S *et al* (2011) showed the cyclical dynamics of industrial production and employment over developed and developing countries using the by Tan *et al* and first passage time analysis. Recently a new statistical method of Pair-wise and Multiple sequence alignment has been developed by Adnan *et al* (2012, 2011). It accomplishes not only an overall decision of the significant similarity/dissimilarity but also the similarity/dissimilarity of all possible individual and group wise transitions that help the biotechnologists to quickly identify the portion of the total infrastructure of the entire transitions that is significantly differing from that of the other sequence(s) and detect the core fact(s) for possible differences between bio-organisms.

However, there is no test for the equality of multiple transition probability matrices. The present study aims to improve the comparison method of multiple transition probability matrices considering the more analysis of transition probabilities of the multiple sampled transition probability matrices. The author addresses an idea of using the difference among multiple transition probabilities of the multiple transition probability matrices which will ensure three advantages at least. Firstly, it will find the degree of disorderness among all possible individual and groupwise transition probabilities of states of multiple Markov chains; and secondly, will reduce the incompleteness of comparison among the multiple chains from the several unknown populations. Thirdly, it clearly identifies the portion of the total infrastructure of the entire transition that is significantly differing from those of the other chains.

2. Methods and Methodology for Several Contingency Tables

With an aim of finding a test for comparing several contingency tables, let us demonstrate our method assuming that we have *m* population contingency tables or matrices from *m* populations and let the hypothesis be

$$H_0: N_1 = N_2 = \cdots N_m$$

$$\Rightarrow H_0: \begin{pmatrix} N_{111} & N_{121} \cdots N_{1c1} \\ N_{211} & N_{221} \quad N_{2c1} \\ & \vdots \\ N_{r11} & N_{r21} \quad N_{rc1} \end{pmatrix} = \begin{pmatrix} N_{112} & N_{122} \cdots N_{1c2} \\ N_{212} & N_{222} \quad N_{2c2} \\ & \vdots \\ N_{r12} & N_{r22} \quad N_{rc2} \end{pmatrix} = \cdots = \begin{pmatrix} N_{11m} & N_{12m} \cdots N_{1cm} \\ N_{21m} & N_{22m} \quad N_{2cm} \\ & \vdots \\ N_{r1m} & N_{r2m} \quad N_{rcm} \end{pmatrix}$$

$$\therefore H_0: P_1 = P_2 = \cdots . = P_m$$

$$\Rightarrow H_0: \begin{pmatrix} P_{111} & P_{121} \cdots P_{1c1} \\ P_{211} & P_{221} \quad P_{2c1} \\ & \vdots \\ P_{r11} & P_{r21} \quad P_{rc1} \end{pmatrix} = \begin{pmatrix} P_{112} & P_{122} \cdots P_{1c2} \\ P_{212} & P_{222} \quad P_{2c2} \\ & \vdots \\ P_{r12} & P_{r22} \quad P_{rc2} \end{pmatrix} = \cdots = \begin{pmatrix} P_{11m} & P_{12m} \cdots P_{1cm} \\ P_{21m} & P_{22m} \quad P_{2cm} \\ & \vdots \\ P_{r1m} & P_{r2m} \quad P_{rcm} \end{pmatrix}.$$

where, the $N_l \quad (\forall\, l = 1,2,\dots,m)$ is the population frequency matrix or contingency table of the l^{th} population; P_l is the population probability matrix or contingency table of the l^{th} population such that $P = (p_{ijl})_{r\times c}$, , where $p_{ijl} = \frac{N_{ijl}}{N_{..}}$ whereas N_{ijl} is the population frequency of the $(i,j)^{\text{th}}$ element of the population frequency matrix $N_{..l}$ of the l^{th} population and $N_{..l} = \sum_{i=1}^{r}\sum_{j=1}^{c} N_{ijl}\,; \forall\, i = 1,2,\dots,r; j = 1,2,\dots,c$. *k* sample contingency tables from each of the *m* population joint frequency distributions (a total of *k* samples are collected from each population) have been collected and on the basis of these samples we want to test whether they come from the same population. After collecting k sample-frequency matrices or tables from each of the *m* populations, the maximum likelihood estimators of the probability matrices are obtained as $\hat{P}_l = (\hat{p}_{ijl})_{r\times c}$ where $\hat{p}_{ijl} = \frac{n_{ijl}}{n_{..l}}$ whereas n_{ijl} is the average frequency of the $(i,j)^{\text{th}}$ element of the average frequency matrix $n_{..l}$ constructed from *k* sample-frequency tables drawn from the l^{th} population. Here, $n_{..l} = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ijl}\,;\ \forall\, i = 1,2,\dots,r;\ j = 1,2,\dots,c.$

For large $n_{..l}$ the asymptotic distribution of each element of transition probability matrices, according to the Central Limit Theorem, are distributed as normal such that

$$\hat{p}_{ijl} \underset{\sim}{\overset{n_{..l}\to\infty}{}} N\left(p_{ijl}, \frac{p_{ijl}\ (1-p_{ijl})}{k n_{..l}}\right).$$

$$\therefore \sum_{l=1}^{m} \frac{\left(\hat{p}_{ijl} - \bar{p}_{ij.}\right)^2}{\frac{\bar{p}_{ij.}(1-\bar{p}_{ij.})}{k n_{..l}}} \sim \chi^2_{(m-1)} \forall\, i = 1,2,\dots,r; j = 1,2,\dots,c.$$

$$\text{where } \bar{p}_{ij.} = \frac{n_{ij1}\hat{p}_{ij1} + \cdots + n_{ijm}\hat{p}_{ijm}}{n_{ij1} + \cdots + n_{ijm}};\ \forall\, i = 1,2,\dots,r; j = 1,2,\dots,c.$$

However, we obtain an element-chi-square-matrix $\chi 2$ of the following form

$$\chi 2 = \begin{pmatrix} \sum_{l=1}^{m} \frac{(\hat{p}_{11l} - \bar{p}_{11.})^2}{\frac{\bar{p}_{11.}(1-\bar{p}_{11.})}{kn_{..l}}} & \sum_{l=1}^{m} \frac{(\hat{p}_{12l} - \bar{p}_{12.})^2}{\frac{\bar{p}_{12.}(1-\bar{p}_{12.})}{kn_{..l}}} & \cdots & \sum_{l=1}^{m} \frac{(\hat{p}_{1cl} - \bar{p}_{1c.})^2}{\frac{\bar{p}_{1c.}(1-\bar{p}_{1c.})}{kn_{..l}}} \\ \sum_{l=1}^{m} \frac{(\hat{p}_{21l} - \bar{p}_{21.})^2}{\frac{\bar{p}_{21.}(1-\bar{p}_{21.})}{kn_{..l}}} & \sum_{l=1}^{m} \frac{(\hat{p}_{22l} - \bar{p}_{22.})^2}{\frac{\bar{p}_{22.}(1-\bar{p}_{22.})}{kn_{..l}}} & \cdots & \sum_{l=1}^{m} \frac{(\hat{p}_{2cl} - \bar{p}_{2c.})^2}{\frac{\bar{p}_{2c.}(1-\bar{p}_{2c.})}{kn_{..l}}} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{l=1}^{m} \frac{(\hat{p}_{r1l} - \bar{p}_{r1.})^2}{\frac{\bar{p}_{r1.}(1-\bar{p}_{r1.})}{kn_{..l}}} & \sum_{l=1}^{m} \frac{(\hat{p}_{r2l} - \bar{p}_{r2.})^2}{\frac{\bar{p}_{r2.}(1-\bar{p}_{r2.})}{kn_{..l}}} & \cdots & \sum_{l=1}^{m} \frac{(\hat{p}_{rcl} - \bar{p}_{rc.})^2}{\frac{\bar{p}_{rc.}(1-\bar{p}_{rc.})}{kn_{..l}}} \end{pmatrix}$$

$$\therefore \chi 2 = \begin{pmatrix} \chi^2_{11} & \cdots & \chi^2_{1c} \\ \vdots & \ddots & \vdots \\ \chi^2_{r1} & \cdots & \chi^2_{rc} \end{pmatrix}.$$

The above matrix of chi-squares can also be called as element-chi-square-matrix. From this matrix we basically can test four types of hypotheses which are as follows:

(*i*) H_0: $p_{ij1} = \ldots = p_{ijm}$; or, the hypothesis of testing the equality of the each (*i*,*j*)[th] individual probabilities of the m population probability matrices $P_1, P_2, \ldots, P_m$.

(*ii*) H_0: $(p_{i11} \quad p_{i21} \quad \cdots \quad p_{ic1}) = \cdots = (p_{i1m} \quad p_{i2m} \quad \cdots \quad p_{icm})$; or, the hypothesis of checking the equality of the *i*[th] row probability vector or frequency distribution for all populations. Actually, it tests the equity of the frequentness of the *i*[th] variable of the first category over all intervals of the second category of m population contingency tables. Indeed the equality of the frequency distribution of the *i*[th] variable of the 1[st] category is tested over m populations. That is, m (types of) frequency distributions are being tested whether equal or not for same variable. So, over a variable the equity of m frequency distributions drawn from m populations is being tested.

(*iii*) H_0: $[p_{1j1} \quad p_{2j1} \quad \cdots \quad p_{rj1}] = \ldots = [p_{1jm} \quad p_{2jm} \quad \cdots \quad p_{rjm}]$; or, the hypothesis of checking the equality of the j[th] column vector for all populations. Indeed, it tests the equity of the frequentness of the j[th] variable of the second category over all variables of the first category of m population contingency tables. The frequency distribution of the j[th] variable of the 2[nd] category is tested whether equal or not over m populations.

(*iv*) H_0: $P_1 = P_2 = \cdots . = P_m$; or the hypothesis of testing the equity of the total contingency table or matix for one population is significantly varying to that of the other populations. It tests the similarity of m populations where each of the m populations has joint frequency distributions over *rc* cells or whether the m types of sample-joint frequency distributions or matrices or tables are drawn from same population.

For the aforementioned tests for m populations, the concern test statistics are given below respectively.

(i) Test of equality of m [(i,j)[th]] cell frequencies: Comparing each χ^2_{ij} with the tabulated $\chi^2_{(m-1,.\propto)}$ of $(m-1)$ degree of freedom,

(ii) Test of equality of m [i[th] variable's] marginal frequency distributions: Comparing each $\sum_j \chi^2_{ij}$ with the tabulated $\chi^2_{(c(m-1),.\propto)}$ of $c(m-1)$ degrees of freedom,

(iii) Test of equality of m [j[th] variable's] marginal frequency distributions: Comparing each $\sum_i \chi_{ij}^2$ with the tabulated $\chi^2_{(r(m-1),\propto)}$ of $r(m-1)$ degrees of freedom,

(iv) Test of equality of m joint frequency distributions**:** Comparing Chi-squares' matrix sum $= \chi_{11}^2 + \cdots + \chi_{1c}^2 + \cdots + \chi_{r1}^2 + \cdots + \chi_{rc}^2$ with the tabulated $\chi^2_{(rc(m-1),\propto)}$ of $rc(m-1)$ degrees of freedom.

3. Methods and Methodology for Several Transition Probability Matrix/Markov Chains

Now, let the stochastic process is $\{X(t), t \in T\}$, then for each value of t, $X(t)$ is a random variable. So, the process is a sequence of outcomes for discrete states and time space. These outcomes may be dependent on earlier ones in the sequence. A Markov chain is collection of random variables $X(t)$ (where the index runs through 0, 1, ...) having the property that, given the present, the future is conditionally independent of the past. So, the stochastic process $\{X_n, n \geq 0\}$ is called a Markov chain, if for $j, k, j_1, \ \ldots \ j_{n-1} \in J$

$$\Pr[X_n = k \mid X_{n-1} = j, X_{n-2} = j_1, \ \ldots \ X_0 = j_{n-1}] = \Pr[X_n = k \mid X_{n-1} = j] = P_{jk}$$

The outcomes are called the states of the Markov Chain; if X_n has the outcome j $(i.e., \ X_n = j)$ the process is said to be at state j at n^{th} trial. The conditional probability $P[X_{n+1} = j | X_n = i] = P_{ij}$ is known as transition probability referring the probability that the process is in stat i and will be in state j in the next step and the transition probability P_{ij} satisfy the properties $(i)\ P_{ij} \geq 0 \ \ and \ \ (ii) \sum_j P_{ij} = 1$ for the transition probability matrix $P = \left[P_{ij} \right] \ \forall \ i, \ j = 1, 2, \cdots, n$.

Here, two states i and j are said to be communicate state if each is accessible from the other, it is denoted by $i \leftrightarrow j$; then there exist integer m and n such that $P_{ij}^{(n)} > 0$ and$P_{ij}^{(m)} > 0$. If state i communicate with state j and state j communicate with state k then state i communicate with state k.

3.1 Proposed method

With an aim of developing a test procedure of testing the equality of several transition probability matrices or several evolutionary rates from several Markov chains or several sequences, let us demonstrate our method assuming that we have several population transition frequency matrices or several population transition probability matrices or several Markov chains each of which having *r* states and let the hypothesis be

$$H_0: N_1 = N_2 = \cdots = N_m$$

$$\Rightarrow H_0: \begin{pmatrix} N_{111} & N_{121} & \cdots & N_{1r1} \\ N_{211} & N_{221} & & N_{2r1} \\ & \vdots & & \\ N_{r11} & N_{r21} & & N_{rr1} \end{pmatrix} = \begin{pmatrix} N_{112} & N_{122} & \cdots & N_{1r2} \\ N_{212} & N_{222} & & N_{2r2} \\ & \vdots & & \\ N_{r12} & N_{r22} & & N_{rr2} \end{pmatrix} = \cdots = \begin{pmatrix} N_{11m} & N_{12m} & \cdots & N_{1rm} \\ N_{21m} & N_{22m} & & N_{2rm} \\ & \vdots & & \\ N_{r1m} & N_{r2m} & & N_{rrm} \end{pmatrix}$$

$$H_0: P_1 = P_2 = \cdots . = P_m;$$

$$\therefore H_0: \begin{pmatrix} p_{111} & p_{121} & \cdots & p_{1r1} \\ p_{211} & p_{221} & & p_{2r1} \\ & \vdots & & \\ p_{r11} & p_{r21} & & p_{rr1} \end{pmatrix} = \begin{pmatrix} p_{112} & p_{122} & \cdots & p_{1r2} \\ p_{212} & p_{222} & & p_{2r2} \\ & \vdots & & \\ p_{r12} & p_{r22} & & p_{rr2} \end{pmatrix} = \cdots = \begin{pmatrix} p_{11m} & p_{12m} & \cdots & p_{1rm} \\ p_{21m} & p_{22m} & & p_{2rm} \\ & \vdots & & \\ p_{r1m} & p_{r2m} & & p_{rrm} \end{pmatrix}.$$

where, N_l ($\forall \ l = 1,2, \ldots, m$)) is the population transition frequency matrix of the l^{th} population such that $N_l = (n_{ijl})_{r \times r}$; P_l is the population transition probability matrix of the l^{th} population such that $P_l =$

$(p_{ijl})_{r\times r}$, where $p_{ij}=\frac{N_{ijl}}{N_{i.l}}$ whereas N_{ijl} is the population transition frequency of the $(i,j)^{\text{th}}$ element of the l^{th} population transition frequency matrices N_l and $N_{i.l}=\sum_{j=1}^{r}N_{ijl}\,;\forall\, i,j=1,2,\dots,r$.

k pairs of sample sequences from *m* populations (a total of *k* sample-sequences are collected from each population) have been collected and on the basis of these samples we want to test whether they come from the same population. After collecting *k* sample-sequences we obtain *k* transition frequency matrices from each of the *m* populations. The maximum likelihood estimators of the transition relative frequency or probability matrices are obtained as $\hat{P}_l=(\hat{p}_{ijl})_{r\times r}$ where $\hat{p}_{ijl}=\frac{n_{ijl}}{n_{i.l}}$ whereas n_{ijl} is the average frequency of the $(i,j)^{\text{th}}$ element of the average transition frequency matrix n_l constructed from *k* sample-transition frequency matrices drawn from the l^{th} population. Here, $n_{i.l}=\sum_{j=1}^{r}n_{ijl}\,;\forall\, i,j=1,2,\dots,r$.

For large $n_{i.l}$ the asymptotic distribution of each element of transition probability matrices, according to the Central Limit Theorem, are distributed as normal such that

$$\hat{p}_{ijl}\ \sim N\left(p_{ijl},\frac{p_{ijl}\ (1-p_{ijl})}{kn_{i.l}}\right).$$

$$\therefore\sum_{l=1}^{m}\frac{\left(\hat{p}_{ijl}-\bar{p}_{ij.}\right)^2}{\frac{\bar{p}_{ij.}(1-\bar{p}_{ij.})}{kn_{i.l}}}\sim\chi^2{}_{(m-1)}\forall i,j=1,2,\dots,r;$$

$$\text{where }\bar{p}_{ij.}=\frac{n_{i.1}\hat{p}_{ij1}+\cdots+n_{i.l}\hat{p}_{ijl}}{n_{i.1}+\cdots+n_{i.l}};\ \forall\, i,j=1,2,\dots,r.$$

However, we obtain an element-chi-square-matrix $\chi 2$ of the following form

$$\chi 2=\begin{array}{c} \\ 1\\ 2\\ \\ r\end{array}\begin{array}{c}\begin{array}{cccc}1 & \mathbf{2} & & \mathrm{r}\end{array}\\ \begin{pmatrix}\sum_{l=1}^{m}\frac{(\hat{p}_{11l}-\bar{p}_{11.})^2}{\frac{\bar{p}_{11.}(1-\bar{p}_{11.})}{kn_{1.l}}} & \sum_{l=1}^{m}\frac{(\hat{p}_{12l}-\bar{p}_{12.})^2}{\frac{\bar{p}_{12.}(1-\bar{p}_{12.})}{kn_{1.l}}} & \cdots & \sum_{l=1}^{m}\frac{(\hat{p}_{1rl}-\bar{p}_{1r.})^2}{\frac{\bar{p}_{1r.}(1-\bar{p}_{1r.})}{kn_{1.l}}}\\ \sum_{l=1}^{m}\frac{(\hat{p}_{21l}-\bar{p}_{21.})^2}{\frac{\bar{p}_{21.}(1-\bar{p}_{21.})}{kn_{2.l}}} & \sum_{l=1}^{m}\frac{(\hat{p}_{22l}-\bar{p}_{22.})^2}{\frac{\bar{p}_{22.}(1-\bar{p}_{22.})}{kn_{2.l}}} & \cdots & \sum_{l=1}^{m}\frac{(\hat{p}_{2rl}-\bar{p}_{2r.})^2}{\frac{\bar{p}_{2r.}(1-\bar{p}_{2r.})}{kn_{2.l}}}\\ \cdots & \cdots & & \cdots\\ \sum_{l=1}^{m}\frac{(\hat{p}_{r1l}-\bar{p}_{r1.})^2}{\frac{\bar{p}_{r1.}(1-\bar{p}_{r1.})}{kn_{r.l}}} & \sum_{l=1}^{m}\frac{(\hat{p}_{r2l}-\bar{p}_{r2.})^2}{\frac{\bar{p}_{r2.}(1-\bar{p}_{r2.})}{kn_{r.l}}} & \cdots & \sum_{l=1}^{m}\frac{(\hat{p}_{rrl}-\bar{p}_{rr.})^2}{\frac{\bar{p}_{rr.}(1-\bar{p}_{rr.})}{kn_{r.l}}}\end{pmatrix}\end{array},$$

$$\therefore\chi 2=\begin{pmatrix}\chi^2_{11} & \cdots & \chi^2_{1r}\\ \vdots & \ddots & \vdots\\ \chi^2_{r1} & \cdots & \chi^2_{rr}\end{pmatrix}.$$

The above matrix of chi-squares can also be called as element-chi-square-matrix. From this matrix we basically can test three types of hypotheses which are as follows:

(*i*) H_0: $p_{ij1}=p_{ij2}=\dots=p_{ijm}$; or, the hypothesis of testing the equality of the each individual ((i,j)$^{\text{th}}$) transition probability of the multiple (m) population transition probability matrices $P_1,P_2\ ,\dots.,P_m$ for all values of $i,j=1,2,\dots,r$.

(*ii*)H_0: $(p_{i11}\ \ p_{i21}\ \ \cdots\ \ p_{ir1})=(p_{i12}\ \ p_{i22}\ \ \cdots\ \ p_{ir2})=\cdots=(p_{i1m}\ \ p_{i2m}\ \ \cdots\ \ p_{irm})$; or, the hypothesis of checking the equality of the i-th row vector of all population transition probability matrices $P_1,P_2\ ,\dots.,P_m$ for all values of $i=1,2,\dots,r$.. Actually, it tests the equity of the frequentness of the transition of the random movement of multiple population sequences from each state to all states.

(*iii*) $H_0: P_1 = P_2 = \cdots . = P_m$; or the hypothesis of testing the equity of the total transitions for all population sequences. It tests the similarity of multiple population sequences or whether the *m* sample sequences are drawn from same population.

For the aforementioned tests the concern test statistics are given below respectively.

(i) Comparing each χ^2_{ij} ($\forall\, i,j = 1,2,\ldots,r$) with the tabulated $\chi^2_{(m-1,.\propto)}$ of (m-1) degree of freedom,

(ii) Comparing each $\sum_{j=1}^{r} \chi^2_{ij}$ ($\forall\, i = 1,2,\ldots,r$) with the tabulated $\chi^2_{[\mathrm{r(m-1)-1},.\propto]}$ of [r(m-1)-1] degrees of freedom,

(iii) Comparing Chi-squares' matrix sum $= \chi^2_{11} + \cdots + \chi^2_{1r} + \cdots + \chi^2_{r1} + \cdots + \chi^2_{rr}$ with the tabulated $\chi^2_{(r(r\mathrm{m}-\mathrm{r}-1),.\propto)}$ of $[r(r\mathrm{m} - \mathrm{r} - 1)]$ degrees of freedom.

4. Real Life Examples for Contingency Tables

Suppose we have two contingency 2×2 tables as those in the example of the page 521 of the book entitled "Handbook of Parametric and Nonparametric Statistical Procedures" by David. J. Sheskin are given as

	Not a biter	*Mild biter*	*Flagrant biter*
Mice	20	16	24
Guinea pigs	19	11	50

	Not a biter	*Mild biter*	*Flagrant biter*
Mice	100	56	44
Guinea pigs	19	11	50

The problem is to gauge whether the two contingency tables show significant dissimilarity, to assess, for example, whether they have a common joint distribution or bivariate distribution that is whether two bivariate samples come from same population bivariate distribution. If the samples were generated at random from two populations, we like to use our proposed statistical method for assessing the similarity of two population joint frequency distributions. Due to a quick unavailability of the replicates of two types of bivariate samples from the book of Sheskin, we are assuming that, after observing 30 pairs of bivariate samples (30 bivariate samples have been drawn from each population bivariate population) from two population bivaraite populations, we have obtained the two average frequency tables or average frequency matrices. So, the sample bivariate mean frequency tables or matrices are

	Not a biter	*Mild biter*	*Flagrant biter*
Mice	20	16	24
Guinea pigs	19	11	50

	Not a biter	*Mild biter*	*Flagrant biter*
Mice	100	56	44
Guinea pigs	19	11	50

Therefore the average relative frequency tables or average probability tables or matrices are

	$Not\ a\ biter$	$Mild\ biter$	$Flagrant\ biter$
$Mice$	0.14	0.11	0.17
$Guinea\ pigs$	0.14	0.08	0.36

	$Not\ a\ biter$	$Mild\ biter$	$Flagrant\ biter$
$Mice$	0.36	0.20	0.16
$Guinea\ pigs$	0.07	0.04	0.18

The averages transition probability matrices result as follows

$$\text{The chi square matrix} = \begin{pmatrix} 144.64 & 21.71 & 0.57 \\ 14.50 & 4.86 & 100.45 \end{pmatrix},$$

$$p \text{ value matrix} = \begin{pmatrix} 2.57 \times 10^{-33} & 3.16 \times 10^{-6} & 0.45 \\ 0.13 \times 10^{-3} & 0.027 & 1.22 \times 10^{-23} \end{pmatrix}.$$

The tabulated value of Chi – square at 1% level of significance with 1 degree of freedom is 6.634897. There is one calculated value for each of the 6 chi-square test statistics for 6 types of cells in the matrix of chi-squares. For the first cell (mice, not a biter), the calculated value (= 144.64) of chi-square test statistic is greater than the tabulated value (= 6.634897) which means the null hypothesis

$$H_0: p_{mice,\ not\ a\ biter} = q_{mice,\ not\ a\ biter}$$

is rejected at 1 percent level of significance with p value 2.57×10^{-33}. So, we conclude that the joint probability of two populations for the joint occurrence of mice with not a biter is dissimilar and we denote the dissimilarity by a notation "DS". Again for the joint frequentness (mice and flagrant biter), the null hypothesis

$$H_0: p_{mice,\ flagrant\ biter} = q_{mice,\ flagrant\ biter}$$

is not rejected at the same level of significance. It can be inferred that the frequentness of contemporarily happening of mice with no biter for two population joint distributions is similar and we denote similarity by a notation "S". So the resultant decision matrix for the 6 various cells is given below:

$$\text{the resultant decision matrix} = \begin{pmatrix} DS & DS & S \\ DS & S & DS \end{pmatrix}.$$

Moreover, the calculated value of overall chi – square, the sum of all individual chi-squares of the chi-squares' matrix sum, is obtained as 286.74. Therefore, the null hypothesis $H_0: P_{2\times3} = Q_{2\times3}$ of the equality of joint probability matrix of two population joint probability distribution is rejected at 1 % level of significance (since the tabulated value of the chi-squares matrix sum with 5 degrees of freedom is 15.09). So, with an overall point of view it can be concluded that the two population joint distributions are dissimilar or do not belong to the same bivariate distribution. Even though, the row similarity and column similarity can be measured here. The sum of chi- squares for the 1st, 2nd and 3rd columns are calculated as 159.15, 26.58 and 101.02 respectively. The tabulated value of the column wise sum of chi-squares with 2 degree of freedom is 9.21 at 1 % level of significance. So, all columns are dissimilar for two population joint distributions, that is, 1st column of the one category and that of the same category for the two populations are dissimilar and so forth. Similar results have been found in case of marginal probabilities for all (two) rows over two populations. So, the marginal frequencies of one category over various intervals in one population is dissimilar to those of the same category over the same intervals in the another population. The dissimilarity between all row-wise marginal probabilities, column-wise marginal probabilities and maximum cell probabilities of the two joint frequency matrices is also a potential evidence of ensuring the conclusion that the two bivariate populations are dissimilar.

5. An Application of Several Markov Chains to Multiple DNA Sequence Alignment

Since it is stated that in a Markov process all possible states and transitions have been assumed in such a way that there is always a next state and the process goes on forever; the characteristics of the DNA, the basic genetic material in living organisms and having a double standed-helical structure each of which is consisting of very long sequence from four letters/alphabets (nucleotides), *a, g, c,* and *t* (for adenine, guanine, cytosine, and thymine, respectively); sequence that undergoes the change within any population over the course of many generations, as random mutations arise and become fixed in the population can easily be treated as a Markov Chain. It is useful for discovering functional, structural, and evolutionary information in biological sequences. Obtaining the best possible or so-called optimal alignment is important to discover this information. Sequences that are very much alike, or "similar" in the parlance of sequence analysis, probably have the same function, be it a regulatory role in the case of similar DNA molecules, or a similar biochemical and three dimensional structure in the case of proteins. Additionally, if two sequences from different organisms are similar, there may have been a common ancestor sequence, and the sequences are then defined as being homologous. The alignment indicates the changes that could have occurred between the two homologous sequences and a common ancestor sequence during evolution. So, a common gauge is to check whether the two sequences show significant similarity, to assess, for example, whether they have a remote common ancestor. As a result, sequence alignment is one of the most important techniques to analyze biological system.

Suppose we have three small DNA sequences such as those in the book of 'Statistical Methods in Bioinformatics' by Ewens, W. *et al* (2004), 30 pairs of sample sequences from same species have been considered. The average transition frequency matrices cum average transition probability matrices (one average transition probability matrix has been obtained from the 30 sample sequences accessed first population, another average transition probability matrix form 30 sample sequences of second population and the third average transition probability matrix from 30 sample sequences collected from the third population) are estimated as follows:

$$\hat{P}_1 = \begin{array}{c} \\ a \\ t \\ c \\ g \end{array} \begin{array}{c} \begin{array}{cccc} a & t & c & g \end{array} \\ \begin{pmatrix} 0.19 & 0.17 & 0.16 & 0.47 \\ 0.20 & 0.03 & 0.22 & 0.56 \\ 0.38 & 0.34 & 0.19 & 0.09 \\ 0.27 & 0.11 & 0.29 & 0.33 \end{pmatrix} \end{array}; \hat{P}_2 = \begin{array}{c} \\ a \\ t \\ c \\ g \end{array} \begin{array}{c} \begin{array}{cccc} a & t & c & g \end{array} \\ \begin{pmatrix} 0.34 & 0.21 & 0.26 & 0.19 \\ 0.11 & 0.15 & 0.26 & 0.49 \\ 0.22 & 0.39 & 0.28 & 0.11 \\ 0.18 & 0.25 & 0.13 & 0.45 \end{pmatrix} \end{array}; \hat{P}_3 = \begin{array}{c} \\ a \\ t \\ c \\ g \end{array} \begin{array}{c} \begin{array}{cccc} a & t & c & g \end{array} \\ \begin{pmatrix} 0.09 & 0.29 & 0.32 & 0.30 \\ 0.14 & 0.13 & 0.33 & 0.40 \\ 0.27 & 0.32 & 0.32 & 0.10 \\ 0.14 & 0.14 & 0.30 & 0.42 \end{pmatrix} \end{array}$$

We first want to observe the properties of three average transition probability matrices to judge the comparability of them as well as the samples. As such the following calculations have been performed.

5.1 Comparability of the three Matrices

From the transition probability graphs of the matrix $\hat{P}_1$ we can conclude that it's all the states are recurrent because all the states are accessible to each other and they are communicating class and the number of states is finite. The matrices $\hat{P}_2$, $\hat{P}_3$ give the same result. The random walks for the three types of sequences have been observed from where the suspect of the difference among the sequences is evident. The Eigen values and vectors of the transition probability matrices have been observed. One of the Eigen values of the 2nd matrix and two of the Eigen values of the 1st as well as 3rd matrices are negative whereas the maximum Eigen values of the three matrices are 1.010, 0.944 and 0.922 respectively. So we can say that there is difference among the transition probabilities of the tree types of samples. Determinant of the matrices are -0.002, -0.007 and 0.001. The ranks of them are same (loosely 4) which is a sign of justification of comparing the three matrices. The stationary probabilities are given as the solution of the equations $\pi_1 = 0.19\pi_1 + 0.20\pi_2 + 0.38\pi_3 + 0.27\pi_4$, $\pi_2 = .17\pi_1 + 0.03\pi_2 + 0.34\pi_3 + 0.11\pi_4$, $\pi_3 = 0.16\pi_1 + 0.22\pi_2 + 0.19\pi_3 + 0.29\pi_4$, $\pi_4 = 0.47\pi_1 + 0.56\pi_2 + 0.09\pi_3 + 0.33\pi_4$ and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$.

Similarly for the second sample we get five equations solving those we obtain the solutions of the stationary probabilities. For the first types of samples the limiting probabilities are 0.26, 0.16, 0.22, 0.35; for the second types of samples 0.20, 0.25, 0.22, 0.33 and for the third types of samples 0.17, 0.22, 0.32, 0.29 respectively. To test the hypothesis of equality of the stationary probabilities for the samples the null hypothesis can be expressed as

$$H_0: \pi_{i1} = \pi_{i2} = \pi_{i3}$$

where, π_{i1}, π_{i2} and π_{i3} ($\forall\, i = 1, 2, 3, 4$) are the stationary probabilities of *i*th state for the 1st , 2nd and 3rd average transition probability matrices respectively. The test statistic for the aforementioned test is

$$\sum_{l=1}^{3} \frac{(\hat{\pi}_{il} - \bar{\pi}_{i.})^2}{\frac{\bar{\pi}_{i.}(1-\bar{\pi}_{i.})}{kn_{i.l}}};$$

$\forall i = 1, 2, 3, 4$; where $\bar{\pi}_{i.} = \frac{\pi_{i1}n_{i.1} + \pi_{i2}n_{i.2} + \pi_{i3}n_{i.3}}{n_{i.1} + n_{i.2} + n_{i.3}}$.

which is distributed as chi-square with (3-1) degree of freedom. The result of equality tests gives the p-values of the aforementioned chi-square statistic as 0.133, 0.248, 0.048 and 0.392. As such at 1% level of significance the limiting probabilities for the same state for the three types of samples are similar. So, for the long run the randomness visit of the population sequence to the individual state or nucleotide is similar for all states over the three populations. Therefore, from the aforementioned results it seems to us that the three matrices are compare able.

5. 2 Proposed Approach

According to the alternative approach, the chi-square matrix and the p-value matrix for obtained from three average transition probability matrices will be:

$$\chi 2 = \begin{array}{c} \\ a \\ t \\ c \\ g \end{array} \begin{array}{c} \begin{array}{cccc} a & t & c & g \end{array} \\ \begin{pmatrix} 24.92 & 7.10 & 10.72 & 28.46 \\ 4.70 & 10.38 & 5.06 & 6.60 \\ 10.89 & 2.11 & 7.49 & 0.51 \\ 13.00 & 19.33 & 27.19 & 7.56 \end{pmatrix} \end{array}$$

The tabulated value of chi – square at 1% level of significance with 2 degree of freedom is 9.21. There is one calculated value for each of the 16 chi-square test statistics for 16 types of transitions in the matrix of chi-squares. For the first transition (from adenine to adenine), the calculated value (= 24.92) of chi-square test statistic is greater than the tabulated value (= 9.21) which means the null hypothesis $H_0: p_{aa1} = p_{aa2} = p_{aa3}$ is rejected at 1 percent level of significance. So, we conclude that the probability of three population sequences for the transition from adenine to adenine is not similar and we denote the dissimilarity by a notation "DS". Again for the transition (from thymine to adenine), the null hypothesis $H_0: p_{ta1} = p_{ta2} = p_{ta3}$ is accepted at the same level of significance with a *p*- value of 0.10. It can be inferred that the frequentness of three population sequences for the transition from thymine to adenine is similar and we denote the similarity by a notation "S". So the resultant decision matrix for the 16 various transitions is given below:

$$\text{the resultant decision matrix} = \begin{pmatrix} DS & S & DS & DS \\ S & DS & S & S \\ DS & S & S & S \\ DS & DS & DS & S \end{pmatrix}.$$

Moreover, the calculated value of overall chi – square, the sum of all individual chi-squares of the chi-squares' matrix sum, is obtained as 186.009. Therefore, the null hypothesis $H_0: P_1 = P_2 = P_3$ of the equality of the entire transition probability matrices of three population sequences is rejected at 1 % level of significance (since the tabulated value of the chi-squares matrix sum with 27 degrees of freedom is 46.96). So, with an overall point of view it can be concluded that the two population sequences are dissimilar or do not belong to the same ancestor. Moreover, the row similarity can be found here. The sum

of chi- squares for the 1st, 2nd 3rd and 4th rows are calculated as 71.19, 26.73, 21.01 and 76.09 respectively. The tabulated value of the row wise sum of chi-squares with 7 degree of freedom is 18.48 at 1 % level of significance. So, all rows are significantly varying among themselves for the three population sequences. The dissimilarity among all of the rows of the three transition probability matrices is also a potential evidence of ensuring the conclusion that the three population sequences are dissimilar.

6. Advantages

The credence of the proposed tests for the equality of several population joint frequency distributions are evident from the given real life examples. The *p* values of the proposed tests for the equality of the marginal row frequency distributions or column frequency distributions over several populations are 0. This certain difference is very much due to the difference among row-wise marginal probability distributions and the column wise marginal probability distributions. The results seem to be appreciating since maximum of the cell frequencies vary among populations. Besides, the test for comparing the several population joint frequency distributions prescribed a *p* value of 0 which means that several population joint distributions are not similar and the given samples are not drawn from the same population.

Besides, the credence of the proposed test for the equality of multiple transition probability matrices are evident from the given real life example, since it is observed that the *p* - value of the proposed test is close to zero (since the *p*-values for the chi-square test is 10^{-25}) indicating bold rejection of the null hypothesis of the equality of the transition probability matrices whereas the samples were really drawn from three different populations. Therefore, the performance of the proposed method seems to be better.

The proposed approach for comparing multiple transition probability matrices gives not only an overall decision of the significant dissimilarity/ similarity of the individual paired transitions but also the significant dissimilarity/similarity of all possible transitions. It clearly identifies the possible similarity or dissimilarity among multiple population sequences. The current method specifically detects for which transition(s) the overall dissimilarity for the multiple population Markov chains is being evident. This idea of more specification can help the biotechnologist to quickly detect the core fact of the possible difference among bio-organisms more easily and more efficiently.

Maximum literature in multiple sequence alignment for quantifying the disorderness of multiple sequences in case of alignment algorithm has been suffering from either calculating superficial gap penalty or obtaining unsatisfactory accuracy. There is no requirement of the treatment of the gaps in the sequence in this method. The accuracy of the proposed method is also satisfactory enough due to its assurance of the optimum *p*-value.

The authors also checked the results of the proposed test with those obtained by combining the 3 tests of equality of two transition probability matrices (3 pair for three populations) for the aforementioned samples (30 sample sequences drawn from each of the three populations). The 3 pair-wise tests test better (since the equality of the entire transition probability matrices of the three population sequences is rejected with a lower *p*-value of 10^{-36}). However, the proposed multiple test will be more amiable since it requires relatively less effort and time.

Concluding Remarks

Joint frequency distributions or contingency tables have been widely being studied by numerous authors since the early ages of statistics. Unfortunately, the discordance of them has not yet been studied so far with any test. The proposed tests ensemble the individual, group wise and overall pattern of the frequencies of one population whether significantly differing from those of other population of any discipline. It has extensive applications in many disciplines like agricultural, biological, pharmaceutical, business and environmental statistics. Advanced multiple test for the equality of bivariate frequency distributions for the several populations can be the further scope of the proposed heuristic.

Transition Probability Matrices have been widely being studied by numerous authors since the childhood of evolutionary statistics. Unfortunately, the discordance of them has not yet been studied with greater effort. The proposed test ensembles the individual, group wise and overall pattern of the transition frequencies of one population whether significantly differing from those of other populations.

Any inquiry and proof(s) of the mathematical development of the tests can be accessible from the authors.

References

Adnan, M. A. S., Shamsuddin, M. (2012). An Advanced Statistical Method of Multiple Sequence Alignment. In JSM Proceedings. Genetic Epidemiology and Genomics. Statistics in Epidemiology Section. Alexandria, VA: American Statistical Association, 3222 - 3236.

Adnan M. A. S., Moinuddin, M, Roy, S, Jaman, R. (2011). An Alternative Approach of Pair-wise sequence Alignment. Proceedings. JSM 2011, American Statistical Association, p2941 - 2951.

Altug, S, Tan, B. and Gencer, G (2011). Cyclical Dynamics of Industrial Production and Employment: Markov Chain-based Estimates and Tests. Working Paper 1101, January 2011, TÜSİAD-KOÇ UNIVERSITY ECONOMIC RESEARCH FORUM, Rumeli Feneri Yolu 34450 Sarıyer/Istanbul.

Agresti, A. (1990) (2002). Categorical Data Analysis. Wiley, New York.

Agresti, A., Klingenberg, B. (2005) Multivariate tests comparing binomial probabilities with application to safety studies for drugs. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54 (4) 691-706.

Barnard, G. A. (1945). A new test for tables. Nature 156, 177.

Barnard, G. A. (1947). Significance tests for tables. Biometrika 34, 123-138.

Barnard, G. A. (1949). Statistical inference. J. Roy. Statist. Soc. Ser. B 11, 115-139.

Barnard, G. A. (1979). In contradiction to J. Berkson's dispraise: conditional tests can be more efficient. J. Statist. Plann. Inference 3, 181-187.

Bartlett, M. S. (1984). Discussion on tests of significance for 2x2 contingency tables (by F. Yates). J. Roy. Statist. Soc. Ser. A 147, 453.

Behseta, S and Kass, R. E. (2005). Testing equality of two functions using BARS. Statistics in Medicine. Doi: 10. 1002/sim.2195.

Bennett, B. M. and Hsu, P. (1960). On the power function of the exact test for the contingency table. Biometrika, 47, 393-398 (correction 48 (1961), 475).

Berger, R. L. and Boos, D. D. (1994). P-values maximized over a confidence set for the nuisance parameter. J. Amer. Statist. Assoc. 89, 1012-1016.

Berger, R. L. (1996). More powerful tests from confidence interval values. Amer. Statist. 50, 314-318.

Berkson, J. (1978). In dispraise of the exact test. J. Statist. Plann. Inference 2, 27-42.

Bartolucci, F and Trapal, I. L. S. (2010). Multidimensional latent Markov models in a development study of inhibitory control and attentional flexibility in the early childhood. Psychometrika. 75 (4), 725-743.
Bergeron, B. (2003). Bioinformatics Computing. Prentice Hall Publisher.
Bhat, U. N. (1972). Elements of Applied Stochastic Process, Wiley & Sons, Canada.
Birnbaum, A. (1962). On the foundations of statistical inference (with discussion). J. Amer. Statist. Assoc. 57, 269-326.
Boschloo, R. D. (1970). Raised conditional level of significance for the 2 × 2 table when testing the equality of probabilities. Statistica Neerlandica 24, 1-35.
Cheng, P. E., Liou. M., Aston. J. A. D., Tsai. A. C. (2008) Identification identities and testing hypotheses: Power analysis for contingency tables. Statistica Sinica, 18, 535 – 558.
Chernoff, H. (1954). On the distribution of the likelihood ratio. Ann. Math. Statist. 25, 573-578.
Cho, J. S. (2011). Testing the equality of two positive definite matrices with application to information matrix testing. Web.
Cho, J. S and White, H (2012). Testing the equality of two positive definite matrices with application to information matrix testing. Web.
Cox, D. R. and Snell, E. J. (1989). The Analysis of Binary Data. 2nd Edition. Chapman and Hall, London.
Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Statist. 11, 427-444.
Dannemann, J and Holzmann, H (2007). The likelihood ratio test for hidden Markov models in two-sample problems. Comp. Stat. & Dtat Analysis. V 52, P:1850 -1859.
Ewens, W. and Grant, G. R. (2004). Statistical Methods in Bioinformatics. Springer.
Falay, B. (2007) Intergenerational income mobility: Equality of Opportunity: A comparison of East and West Germany. EKONOMI YUKSEK LISANS PROGRAMI, Istanbul Bilgi University. 2007
Feller, W. (1968). An Introduction to Probability Theory and its Applications. 3rd Edition. Wiley, New York.
Fisher, R. A. (1922). On the interpretation of _2 from contingency tables, and the calculation of P. J. Roy. Statist. Soc. 85, 87-94.
Fisher, R. A. (1925) (5th ed., 1934; 10th ed., 1946). Statistical Methods for Research Workers, Oliver & Boyd, Edinburgh.
Fisher, R. A. (1935). The logic of inductive inference. J. Roy. Statist. Soc. Ser. A 98, 39-54.
Fisher, R. A. (1962). Confidence limits for a cross-product ratio. Austral. J. Statist. 4, 41.
Gail, M. and Gart, J. G. (1973). The determination of sample sizes for use with the exact conditional test in comparative trials. Biometrics 29, 441-448.
Gokhale, D. V. and Kullback, S. (1978). The Information in Contingency Tables. Marcel Dekker, New York.
Goodman, L. A. (1984). The Analysis of Cross-Classified Data Having Ordered Categories. Harvard University Press, Cambridge.
Gray, R. M. (1990). Entropy and Information Theory. Springer-Verlag, New York.
Greenland, S. (1991). On the logical justification of conditional tests for two-by-two contingency tables. Amer. Statist. 45, 248-251.
Grizzle, J. E. (1967). Continuity correction in the test for tables. Amer. Statist. 21, 28-32.
Haber M. (1986). An exact unconditional test for the comparative trial. Psychol. Bull. 99, 129-132.
Hillary, R. M. (2011) A New Method for Estimating Growth Transition Matrices. Biometrics. 67, 76-85.
Johnson, N. L. and Kotz, S. (1969). Discrete Distributions. Wiley, New York.
Karlin, S. and Taylor, H. M. (1975). A First Course in Stochastic Processes. Amazon.
Kempthorne, O. (1978). Comments on J. Berkson's paper "In Dispraise of the Exact Test". J. Statist. Plann. Inference 3, 199-213.
Kendall, M. G. and Stuart, A. (1979). The Advanced Theory of Statistics. Vol. 2, 4th edition. Charles Griffin, London.
Klugkist, I., Laudy, O. and Hoijtink, H. (2010). Bayesian Evaluation of Inequality and Equality Constrained Hypotheses for contingency Tables. Web. NOW-VICI-453-05-002.

Kou, S. G. and Ying, Z. (1996). Asymptotics for a table with fixed margins. Statist. Sinica 6, 809-829.
Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. Ann. Math. Statist. 22, 79-86.
Lancaster, H. O. (1949). The combination of probabilities arising from data in discrete distri- butions. Biometrika 36, 370-382, Corrig. 37, 452.
Lancaster, H. O. (1969). The Chi-squared Distributions. Wiley, New York.
Lehmann, E. L. (1986). Testing Statistical Hypotheses. 2nd Edition. Wiley, New York.
Lindley, D. V. (1956). On a measure of the information provided by an experiment. Ann. Math. Statist. 27, 986-1005.
Little, R. J. A. (1989). Testing the equality of two independent binomial proportions. Amer. Statist. 43, 283-288.
Mehta, C. R. and Patel, N. R. (1980). A network algorithm for the exact treatment of the contingency table. Comm. Statist. Ser. B 9, 649-664.
Muse, S. V. et al (1992) Testing the equality of evolutionary rates. Genetics. 1322: 269-276.
Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20, 263-274.
Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a table. Biometrika 34, 139-167.
Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. Series 50, 157-175.
Pearson, K. (1904). Mathematical contributions to the theory of evolution XIII: On the theory of contingency and its relation to association and normal correlation. Draper's Co. Research Memoirs, Biometric Series, no. 1. (Reprinted in Karl Pearson's Early Papers, ed. E. S. Pearson, Cambridge: Cambridge University Press, 1948).
Plackett, R. L. (1964). The continuity correction in tables. Biometrika 51, 327-337.
Plackett, R. L. (1977). The marginal totals of a table. Biometrika 64, 37-42.
Rao, C. R. (1973). Linear Statistical Inference and Its Applications. 2nd edition. Wiley, New York.
Ross, S. (1995). Stochastic Process. Wiley & Sons.
Santner, T. J. and Duffy, D. E. (1989). The Statistical Analysis of Discrete Data. Springer- Verlag, New York.
Suissa, S. and Shuster, J. (1985). Exact unconditional sample sizes for the binomial trial. J. Roy. Statist. Soc. Ser. A 148, 317-327.
Tan, B. and Yilmaz, K. (2002). "Markov Chain Test for Time Dependence and Homogeneity: An Analytical and Empirical evaluation," European Journal of Operational Research 137(3), 524-543. Wikipedia-Stochastic process, Markov Chain.
Upton, G. J. G. (1982). A comparison of alternative tests for the 2 × 2 comparative trial. J. Royal Statist. Soc. A 145, 86-105.
Wilks, S. S. (1935). The likelihood test of independence in contingency tables. Ann. Math. Statist. 6, 190-196.
Yalonetzky, G. (2009). Heterogeneity indices with the ratio of Pearson's goodness of fit statistics. OPHI working paper 33.
Yates, F. (1934). Contingency tables involving small numbers and the test. J. Royal Statist. Soc. Suppl. 1, 217-235.
Yates, F. (1984). Tests of Significance for contingency tables (with discussion). J. Royal Statist. Soc. A 147, 426-463.
Yule, G. U. (1911). An Introduction to the Theory of Statistics. Griffin, London.