ISSN 1684-8403
Journal of Statistics
Volume 23, 2016. pp. 86-101

Estimation of Transition Probability Matrix in Continuous Evaluation Process of School Examination

Goutam Saha[1]

Abstract

This paper is about estimation of transition probability matrix using proportion data available from continuous evaluation process of student's performance in school final examination. Since the available data do not provide individual transition information over time, the method of Maximum Likelihood fails. Weighted Least Square method is applied to obtain Estimates of transition probabilities using real life data. The transition probabilities are estimated using Iteratively Re-Weighted Least Square (IRLS) method where errors are correlated.

Keywords

Transition probability matrices, Iteratively re-weighted least square method, Quadratic programming problem, M-estimator, Correlated errors

1. Introduction

Teaching-learning is a complex process which varies widely from place to place depending upon economic, social environment, culture, language etc. The goal of assessment process is to monitor student's performance and to observe their progress in the spirit of continuous improvement. Hence, to understand the grading pattern and analyze the performance of the student's even beyond the classroom, scientific analysis of the pattern is more than necessary. Modeling of assessment pattern represents the behavior of teaching-learning process. In general models are simpler than the system they describe, but are useful in understanding and predicting systems behavior.

[1] Assistant Professor of Statistics, M.B.B. College, Tripura, India
Email: goutam.stat@gmail.com

The transition probability matrix (t.p.m), corresponding to Markov chain needs detailed and individual information about the transition among the grades which are used for estimation of transition probability matrix. However, estimating transition probability matrix is relatively straight-forward process. More generally, let n_{ij} is the number of individual in the state j at time k provided they were in state i at time $(k\text{-}1)$, then P_{ij} may be estimated as $\frac{n_{ij}}{\sum n_{ij}}$ and hence transition probability matrix may be estimated. These estimators are consistent but Biased, however Bias goes to zero as sample size increases (Goodman 1957).

Unfortunately, in real life situation it rarely happens. The best information available is an aggregate ratio or proportion showing the percent of total observations in a particular grading category at a point of time. In this case, method of Maximum Likelihood fails. Suppose, that instead of observing actual count of transitions for different grade points, only it may observe the aggregate proportions corresponding to different grades at different time point. For Such type of data, there is a scope of applying Iteratively Re-Weighted Least Square method for estimating the transition probability matrix.

Here, transition probability matrix (t.p.m), corresponding to Markov chain may use where detailed and individual information about the transitions among the grades are available. The estimation of parameters P_{ij} for $i, j = 0, 1, 2$ are carried out with real life data. In this paper, the assumption that, detailed and individual information are available is dropped and hence Iteratively Re-Weighted Least Square method is applied.

A brief review of literature is included in the next Section. Section 3 is about the data source and Section 4 contains the model and methodology, whereas Section 5 contains transition matrices when individual information of transitions is unknown. Section 6 includes the conditions for invertibility of time and states when errors are finite and correlated. Estimation of transition probability by IRLS method is presented in Section 7. The paper ends with numerical analysis of Weighted Least Square Estimates.

2. Literature Review

Estimation of transition probability using method of Maximum Likelihood is not new. The key paper in this area perhaps is Goodman (1957). Method of estimating P_{ij} and procedure of verifying the properties of estimates is well exhibited by Bhat (2000). But, if the sample consists of only the number of individuals or proportions of individuals in each state at varying times, these methods cannot be used any longer. Instead, Weighting Least Square or method of minimum Chi-square may be used. The other areas where IRLS is applicable, is described below.

Stirling (1984) worked on Iteratively Re-Weighted Least Squares for models with a linear part which shows that, the algorithm converges considerably faster than the EM (Expectation-Minimization) algorithm. Pires et. al, (1999) finds a Robust Iteratively Re-Weighted Least-Square method for power system state estimation through givens rotations. An orthogonal implementation through givens rotations to solve estimators based on non-quadratic criteria, are defined based on the Robust Statistical Theory. On the other hand, a Weighted Least-Squares algorithm for estimation and visualization of relative latencies in event-related functional obtained by Calhoun et. al., (2000). Whereas, estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration was studied by Welton and Ades in 2005. The paper Least-Squares Estimation of transition probabilities from aggregate data by Kalbfleisch and Lawless (1984), deals with the problem of estimating P on the basis of aggregate data which record only the numbers of individuals that occupy each of the k states at different times.

To find the minimal element can be identified via linear programming algorithms, Daubechies et al. (2008) studied an alternative method of determining the limit of an Iteratively Re-Weighted Least Squares algorithm in the paper entitled as Iteratively Re-Weighted Least Squares minimization for sparse recovery. Also for computation of Robust Regression Oleary (1990) used Iteratively Re-Weighted Least Square method.

3. Data Source

For calculation of transition probabilities, data collected from 25 schools with student's size 538 as a sample. The schools are selected from in an around Agartala, the capital city of Tripura - a state of North-East India. The marks of 538 students are collected from higher secondary school records. The marks are also converted to grades according to the definition mentioned in the next Section of this paper.

In this survey, from a total population of 65 schools in Agartala, offering both matriculation (HSLC) and higher secondary (HSSLC) (science stream) courses, a sample of 25 schools are randomly selected and data related to examination scores are collected for analysis. The results of the students in both the examinations are assumed to be influenced by the variables gender, medium of instructions (English, Bengali), type of schools (boy's, girl's and co-educational), category of schools (Govt., non-Govt.), location of schools (urban, rural) and boards of examinations viz. Tripura Board of Secondary Education (TBSE), Central Board of Secondary Education or in short CBSE and ICSE i.e., Indian Certificate of School Examination and data corresponding to the above mentioned variables are also collected.

4. Model and Methodology

4.1. Assumptions:

- A student is evaluated at regular interval of time and each time point the grade of the student is recorded.
- The present grade of the student is dependent on immediate previous grade.
- The range [0, 100] of scores 'S' is partitioned into three non-overlapping sets viz. S_0, S_1, and S_2 such that, $S_0 = [M_0, 100]$, $S_1 = [M_1, M_0)$ and $S_2 = (M_1, 0]$, where $0 \leq M_1 < M_0 \leq 100$. M_0 & M_1 are the lower limits of the top and middle grade respectively.
- The grade of a student is assumed to be i if $S \in S_i$ for i = 0, 1, 2.

Now,
Let, $X_n = i$ for i = 0, 1, 2 be the grade of a student at the n^{th} test.
where, n = 1, 2, 3 ....
The consequence of assumptions 1 to 4 states that $\{X_n,\ n = 1, 2, ...\}$ follows a Markov chain with state space $\$ = \{0, 1, 2\}$, transition probability matrix $P = (P_{ij})$, initial distribution (P_0, P_1, P_2), such that $(P_0 + P_1 + P_2) = 1$, and $P_r(X_0 = i) = P_i$ with $0 < P_i, P_{ij} < 1$ for i, j = 0,1,2. where X_0 = initial states of the chain.

As far as the grades of individual student's are available at different point of time they may be probabilistically analyzed. Statistical estimation of transition probabilities P_{ij} may also be obtained by using method of Maximum Likelihood (Goodman, 1957).

The transition matrices may be widely used to explain the performance of student's with respect to the change of grades over time. These matrices provide a succinct way to describe the evolution of grades using Markov transition probability model. Estimating transition probability matrix is relatively straight forward process, if it is observed that the sequence of states for each individual unit under consideration i.e, if individual transitions are observed. More generally, if n_{ij} be the number of individual in the state *j* at time *k* provided they ware in state *i* at time (k-1), then p_{ij} may be estimated as $\frac{n_{ij}}{\sum n_{ij}}$ and hence transition probability matrix may be estimated.

5. Transition Matrices when Individual Transitions Unknown

As mentioned already, the estimation of transition matrices is relatively easy when individual transition information is available over time. The best information available is an aggregate ratio or proportion showing the percent of total observations in a particular grading category at a point of time. In this case, method of Maximum Likelihood fails.

Suppose that, instead of observing actual count of transitions for different grade points, it is only observed that the aggregate proportions i.e., $y_j(k)$ and $y_i(k\text{-}1)$ which represents proportion of observations corresponding to the grade points *j* and *i* at the time point's *k* and *(k-1)*, respectively.
where,
$\mathrm{y}_{\,j}(k) = \frac{n_j(k)}{N}$; $n_j(k)$ = number of student's getting grade *j* at the k^{th} test,
j = 0, 1, 2 and k = 1, 2, ……, n.

Hence a stochastic relation relating to the actual and estimated occurrence of $y_j(k)$ can be written as

$$\mathrm{y}_{\,j}(k) = \sum_{i=0}^{2} \mathrm{y}_{\,i}(k-1)P_{ij} + \mathrm{u}_{\,j}(k) \qquad (5.1)$$

where,
j = 0, 1, 2, and k = 1, 2, ..., n.

Also, $\mathrm{u}_{\,j}(k)$ are correlated random errors.

$$\therefore \mathrm{y}_{\,j}(k) = \mathrm{y}_{\,0}(k-1)P_{0j} + \mathrm{y}_{\,1}(k-1)P_{1j} + \mathrm{y}_{\,2}(k-1)P_{2j} + \mathrm{u}_{\,j}(k)$$

Now in particular, for j = 0, 1, 2.

$$\mathrm{y}_{\,0}(k) = \mathrm{y}_{\,0}(k-1)P_{00} + \mathrm{y}_{\,1}(k-1)P_{10} + \mathrm{y}_{\,2}(k-1)P_{20} + \mathrm{u}_{\,0}(k)$$

$$\mathrm{y}_{\,1}(k) = \mathrm{y}_{\,0}(k-1)P_{01} + \mathrm{y}_{\,1}(k-1)P_{11} + \mathrm{y}_{\,2}(k-1)P_{21} + \mathrm{u}_{\,1}(k)$$

$$y_2(k) = y_0(k-1)P_{02} + y_1(k-1)P_{12} + y_2(k-1)P_{22} + u_2(k) \qquad (5.2)$$

where,
P_{ij}'s are the transition probabilities to be estimated, obtained from the transition probability matrix $P = (P_{ij})_{3\times 3}$

The above set of eq. (5.2) may be written in compact form as
$y_j(k) = X(k-1)P + U_j(k)$; j = 0, 1, 2 and k = 1, 2, ….., n.

where,
$y_j(k) = \left(y_0(k) \;\; y_1(k) \;\; y_2(k)\right)$ and $P = (P_{ij})$ for i, j = 0, 1, 2 is a 3×3 matrix with 9 elements and

$$X_j(k-1) = \begin{pmatrix} y_0(0) & y_1(0) & y_2(0) \\ y_0(1) & y_1(1) & y_2(1) \\ y_0(2) & y_1(2) & y_2(2) \\ \vdots & \vdots & \vdots \\ y_0(n-1) & y_1(n-1) & y_2(n-1) \end{pmatrix}$$

Also,
$U_j(k) = (u_0(k) \;\; u_1(k) \;\; u_2(k))$ for k = 1, 2, ………, n.

Now expanding the above eq. (5.2) for all values of k, we get a system of equation as
$Y = (\, y_0(1) \;\; y_1(1) \;\; y_2(1) \quad y_0(2) \;\; y_1(2) \;\; y_2(3) \ldots y_0(n)\; y_1(n)\; y_2(n)\,)$
$P = (p_{00} \;\; p_{01} \;\; p_{20} \quad p_{01} \;\; p_{11} \;\; p_{21} \quad p_{02} \;\; p_{12} \;\; p_{22})$
$X = \text{diag}(x_0^{(k-1)} \;\; x_1^{(k-1)} \;\; x_2^{(k-1)})$
$U = (u_0^{(1)}\; u_1^{(1)}\; u_2^{(1)} \quad u_0^{(2)}\; u_1^{(2)}\; u_2^{(2)} \ldots u_0^{(n)}\; u_1^{(n)}\; u_2^{(n)})$

Hence, the system of equation may be written as

$$Y = XP + U \qquad (5.3)$$

where,
U is an 3n×1 vector of errors.

Assuming that $U \sim N(0, \sigma^2 I_n)$ leads to the familiar Ordinary-Least-Squares.

However, let us assume more generally that, $U \sim N(0, W)$, where the error-covariance matrix W is symmetric and positive-definite. Different diagonal entries in W correspond to non-constant error variances; while non-zero off-diagonal entries correspond to correlated errors. Suppose, for the time-being, that W is known. Then, the log-likelihood for the model is

$$\log_e L(P) = -\frac{1}{2}\log_e 2\pi \ -\frac{1}{2}\log_e \left|W\right| -\tfrac{1}{2}(Y-XP)' W^{-1}(Y-XP)$$

maximized by the Generalized-Least-Squares (GLS) Estimator of P.

$$\therefore \ \text{P(GLS)} = \left(X'W^{-1}X\right)^{-1} X'W^{-1}Y$$

with covariance matrix

$$\text{V(PGLS)} = \left(X'W^{-1}X\right)^{-1}$$

For example, when W is a diagonal matrix of (generally) unequal error variances, then P(GLS) is just the Weighted-Least-Squares (WLS) Estimator.

Hence, the variance co-variance of U $_j(k)$ is given by

$$W\ (k) = \begin{pmatrix} \text{var}(u_0(k)) & \text{cov}(u_0(k)u_1(k)) & \text{cov}(u_0(k)u_2(k)) \\ & \text{var}(u_1(k)) & \text{cov}(u_1(k)u_2(k)) \\ & & \text{var}(u_2(k)) \end{pmatrix} \tag{5.4}$$

for k = 1, 2, ……....., n.

So, the variance co-variance of U $_j(k)$ for all k is

$$W = diag(W(1) \quad W(2) \quad . \quad . \quad . \quad W(n))$$

6. Conditions for Invertibility of Time and States when Errors are Finite and Correlated

According to the assumptions in the model, the grading pattern follows Markov chain and proportion showing the percentage of total observations in a particular grading category observed at a point of time is dependent on previous observation. Here, time and states both are not independent of previous observations.

Results: The errors have finite non constant variances and are correlated, if the following conditions are satisfied:

- $BP < I$
- $BP + B^1P^1 - BP\,B^1P^1 < I$ is true
- where $P = (P_{0j}\ \ P_{1j}\ \ P_{2j})$ which is a 1×3 matrix, for j = 0, 1, 2 with $(P_{ij}) = (P_{i0}\ \ P_{i1}\ \ P_{i2})$ for i = 0, 1, 2.
- $B = (B_{0j}^{-1}\ \ B_{1j}^{-1}\ \ B_{2j}^{-1})$ is a 1×3 matrix, for j = 0, 1, 2 with $(B_{ij}^{-1}) = (B_{i0}^{-1}\ \ B_{i1}^{-1}\ \ B_{i2}^{-1})$ for i = 0, 1, 2 and $B^1 = (B_{ij}^{-1}\ \ B_{2j}^{-1}\ \ B_{3j}^{-1}) = (P_{1j}\ \ P_{2j}\ \ P_{3j})$ for j = 0,1,2

6.1. Condition for Finite Variance of Errors: With $0 < P_{ij} < 1$, $\sum_j P_{ij} = 1$, I = diag(1,1,1) and *B* is a matrix of operators.

6.2. Condition for Existence of Correlation Between U_i and U_j: Since *U* are correlated random variables, then, $E(U_i^{(k)}\,U_j^{(k)}) \neq 0$ for i , j = 0, 1, 2.

7. Estimation of Transition Probability by Re-Weighted Least-Square Method

The estimate of *P* may be obtained by minimizing

$$(Y-XP)'W^{-1}(Y-XP)$$

where,

$$W_{ii}^{-1} = \frac{\sum_{j\neq i} W(i)}{\sum_j \left|W(j)\right|}$$

Subject to the condition that, $0 \le P_{ij} \le 1$ and $\sum P_{ij} = 1$.

The minimization of *(Y-XP)W⁻¹(Y-XP)* with respect to *P* leads to a Quadratic Programming Problem (QPP). The method of IRLS is used to solve certain optimization problems. It solves the objective function of the form as

$$\arg\min(P) \sum W(P)\left|Y_i - f_i(P)\right|^2$$

by an iterative method in which each step involves solving a problem of the form

$$P^{t+1} = \arg\min(P) \sum W\left(P^t\right)\left|Y_i - f_i\left(P\right)\right|^2$$

IRLS is used to find the Maximum Likelihood Estimates of a Generalized Linear Model and in Robust Regression to find an M-estimator (Maximum Likelihood Estimator).

Since the sequence $\{y_j^{(k)}\}$, j = 0, 1, 2 follows Trinomial Distribution for k = 0, 1, ….., n. Then the elements of *W* may be given by

$$W_{ii}(k) = \frac{y_i(k)(1 - y_i(k))}{N_k} \quad \text{for i = 0, 1, 2. and k = 1, 2, ….., n.}$$

and

$$W_{ij}(k) = \frac{-y_i(k)\, y_j(k)}{N_k} \quad \text{for i, j = 0, 1, 2.}$$

Hence, the Weighted Least Square (WLS) Estimate is given by

$$\hat{P} = (X' W^{-1} X)^{-1} X' W^{-1} Y \tag{7.1}$$

Here, $\hat{P}$ may be obtained by using Quadratic Programming Problem or Iteratively Re-Weighted Least Square method under the situation viz.

- $W_{ij}(k) = 0$, when i $\neq$ j = 1, i.e., i = j
- $W_{ij}(k) = 0$, when i $\neq$ j = σ_i^2 i.e., i = j = 0, 1, 2.
- $W_{ij}(k) \neq 0$, when i $\neq$ j = σ_i^2, i.e., i = j = 0, 1, 2.

Here, the second method i.e., Iteratively Re-Weighted Least Squares method is applied, suggested by Beoten and Tukey (1974). The procedure is as follows:

Step 1: Select an initial vector using random number Table, p1 = (p_{ij}^1), where,
i, j = 0, 1, 2. Subject to the condition that, $0 \leq P_{ij}^1 \leq 1$ and $\sum P_{ij} = 1$.

Step 2: Calculate $s_1 = median\left|u_i - median(u_i)\right|$

Then, Obtain $s = \dfrac{s_1}{0.6745}$

Step 3: Obtain

(i) $\left(Y - X P^1\right)$

and

(ii) $\left(Y - X P^1\right)^{/}\left(Y - X P^1\right)$

Step 4: Obtain a_{i0} = $$\frac{\left\{ {(Y_i - X_i P^1)}/{s} \right\}^{/} W^{-1} \left\{ {(Y_i - X_i P^1)}/{s} \right\}}{\left\{ {(Y_i - X_i P^1)}/{s} \right\}}$$

where,

- W =(I) errors are $N\,(0,\ \sigma_k^2 = 1)$.
- W = σ_k^2(I) errors are $N\,(0,\ \sigma_k^2 \neq 0)$
- W = σ_k^2(I) errors are N $(0,\ W_k^2)$

A_0 is matrix of the correction factor as $A_0 = \text{diag}\,(a_{10}, a_{20}, \ldots a_{n0})$ with n = 24.

Step - 5 (Case - I): Calculate

$$P^{(2)} = \left(X^{/} A_0 X\right)^{-1} X^{/} A_0 Y$$

Check

(i) $(IP^1 - P^{(2)} I)$

and

(ii) $\left(Y - X P^1\right)^{/}\left(Y - X P^1\right)$

This procedure goes on till we select that value *P* for which

$\left(Y - X P^1\right)^{/} W^{-1}\left(Y - X P^1\right)$ is minimum.

Step - 5 (Case - II): Calculate

$$P^2 = \left(X^{/} A_0 W^{-1} X\right)^{-1} (X^{/} A_0 W^{-1} Y)$$

Check $\left(Y - X P^1\right)^{/} W^{-1}\left(Y - X P^1\right)$

This procedure goes on till we select that value *P* for which

$\left(Y - X P^1\right)^{/} W^{-1}\left(Y - X P^1\right)$ is minimum.

Step - 5 (Case - III): Calculate

$$P^2 = \left(X^{/} A_0 W_{ij}^{-1} X\right)^{-1}\left(X^{/} A_0 W_{ij}^{-1} Y\right)$$

Check $\left(Y - X P^1\right)^{/} W_{ij}^{-1}\left(Y - X P^1\right)$

This procedure goes on till we select that value *P* for which

$\left(Y - X P^1\right)^{/} W_{ij}^{-1}\left(Y - X P^1\right)$ is minimum.

8. Numerical Analysis of Weighted Least Square Estimates

The transition probability matrix $\hat{P}$ are obtained by using Re-Weighted Least Square method for Case - I, Case - II and Case – III in step - 5. Here, C / C++ programming is used for this purpose. Results are presented in Table 1, 2 and 3.

For demonstration purpose, the Bar Diagrams (Figure 1 and Figure 2) of estimated probabilities for the final sets of $\hat{P}$ = (P_{00}, p_{01}, P_{02}, P_{10}, P_{11}, P_{12}, P_{20}, P_{21}, P_{22}) presented for the parameter 'medium of instruction' (for $\sigma_k^2 = 1$, $\sigma_k^2 \neq 0$ and correlated cases). Here, (P_{00}, p_{01}, P_{02}, P_{10}, P_{11}, P_{12}, P_{20}, P_{21}, P_{22}) are labelled as 1, 2, .., 9 along the horizontal axes.

The least square values for the parameter 'medium of instruction' are presented in Table 4.

9. Discussion

From the Figure 1, it has been observed that, the estimated transition probability varies considerably under all the three models. Whereas, Figure 2 shows that almost equal values of transition probability is attained by the students who transit from middle grade to lower grade. This also a clear indication that, considering the same situation variability is more for English medium cases comparing to Bengali medium.

Figure 3 shows that, minimum value of Least-Square is attained at the 3rd iteration for case – II, whereas, minimum value of Least-Square for case – III is attained at the 4th iteration. This result also differentiates the two medium of instructions in two different cases i.e., case – I and case - II.

Again for case – I, where $\sigma_k^2 = 1,$ minimum value of Least-Square is attained at 4th iteration for both the cases, which is apparent from the Table 4.

Hence, from Table 1, Table 2 and Table 3, it is clear that, when detailed individual information about transitions is not available, estimates $\hat{P}_{ij}$ differ from those where individual information is available.

Acknowledgements

The author acknowledges and extends his heartfelt gratitude to the editors and anonymous reviewers of the *Journal of Statistics (J. Stat)* for their careful reading of this paper and constructive suggestions.

Table 1: Estimates of transition probabilities with $\sigma_k^2 = 1$

Parameters	Medium of instruction	Gender of Students	Type of Schools	Board of Examinations
Transition Probability	Eng Beng	Male Female	Boy's Girl's Co-edu	TBSE CBSE ICSE
P_{00}	(0.386) (0.305)	(0.355) (0.257)	(0.528) (0.407) (0.372)	(0.398) (0.329) (0.389)
P_{01}	(0.443) (0.460)	(0.318) (0.436)	(0.387) (0.327) (0.283)	(0.397) (0.344) (0.037)
P_{02}	(0.233) (0.142)	(0.208) (0.392)	(0.407) (0.251) (0.306)	(0.231) (0.243) (0.433)
P_{10}	(0.036) (0.351)	(0.278) (0.347)	(0.250) (0.312) (0.536)	(0.477) (0.343) (0.212)
P_{11}	(0.468) (0.377)	(0.285) (0.312)	(0.436) (0.256) (0.319)	(0.367) (0.372) (0.661)
P_{12}	(0.299) (0.393)	(0.426) (0.201)	(0.456) (0.436) (0.403)	(0.372) (0.432) (0.326)
P_{20}	(0.579) (0.344)	(0.367) (0.397)	(0.177) (0.280) (0.009)	(0.125) (0.328) (0.398)
P_{21}	(0.089) (0.163)	(0.397) (0.251)	(0.292) (0.416) (0.397)	(0.236) (0.325) (0.301)
P_{22}	(0.468) (0.466)	(0.366) (0.407)	(0.137) (0.312) (0.290)	(0.397) (0.325) (0.241)

Table 2: Estimates of transition probabilities with $\sigma_k^2 \neq 0$

Parameters	Medium of instruction	Gender of Students	Type of Schools	Board of Examinations
Transition Probability	Eng Beng	Male Female	Boys' Girls' Co-edu	TBSE CBSE ICSE
P_{00}	(0.489) (0.351)	(0.447) (0.298)	(0.397) (0.428) (0.259)	(0.342) (0.350) (0.302)
P_{01}	(0.357) (0.118)	(0.357) (0.437)	(0.037) (0.472) (0.344)	(0.273) (0.308) (0.415)
P_{02}	(0.317) (0.302)	(0.223) (0.202)	(0.433) (0.250) (0.301)	(0.107) (0.237) (0.372)
P_{10}	(0.072) (0.323)	(0.412) (0.407)	(0.127) (0.350) (0.435)	(0.453) (0.278) (0.397)
P_{11}	(0.128) (0.407)	(0.457) (0.488)	(0.661) (0.236) (0.372)	(0.432) (0.295) (0.326)
P_{12}	(0.159) (0.397)	(0.436) (0.366)	(0.326) (0.577) (0.428)	(0.503) (0.436) (0.407)
P_{20}	(0.438) (0.322)	(0.141) (0.295)	(0.475) (0.222) (0.306)	(0.205) (0.372) (0.301)
P_{21}	(0.515) (0.475)	(0.186) (0.075)	(0.301) (0.292) (0.285)	(0.295) (0.397) (0.259)
P_{22}	(0.524) (0.301)	(0.341) (0.432)	(0.241) (0.173) (0.271)	(0.390) (0.326) (0.221)

Table 3: Estimates of transition probabilities with correlated errors

Parameters	Medium of instruction	Gender of Students	Type of Schools	Board of Examinations
Transition Probability	Eng Beng	Male Female	Boys' Girls' Co-edu	TBSE CBSE ICSE
P_{00}	(0.043) (0.623)	(0.207) (0.441)	(0.108) (0.151) (0.014)	(0.823) (0.349) (0.713)
P_{01}	(0.563) (0.484)	(0.742) (0.336)	(0.606) (0.518) (0.949)	(0.421) (0.286) (0.308)
P_{02}	(0.810) (0.137)	(0.088) (0.102)	(0.492) (0.276) (0.054)	(0.322) (0.458) (0.092)
P_{10}	(0.431) (0.095)	(0.543) (0.406)	(0.784) (0.622) (0.328)	(0.082) (0.495) (0.035)
P_{11}	(0.319) (0.197)	(0.123) (0.011)	(0.029) (0.172) (0.002)	(0.309) (0.567) (0.257)
P_{12}	(0.084) (0.399)	(0.836) (0.625)	(0.059) (0.079) (0.873)	(0.547) (0.232) (0.867)
P_{20}	(0.525) (0.282)	(0.249) (0.153)	(0.108) (0.227) (0.658)	(0.095) (0.156) (0.253)
P_{21}	(0.118) (0.319)	(0.135) (0.653)	(0.365) (0.310) (0.049)	(0.269) (0.147) (0.435)
P_{22}	(0.105) (0.463)	(0.077) (0.273)	(0.448) (0.646) (0.073)	(0.130) (0.309) (0.045)

Table 4: Least-Square values of medium of instructions

No. of Iterations	Case - I		Case - II		Case – III	
	Eng	**Beng**	**Eng**	**Beng**	**Eng**	**Beng**
1	(30.75)	(18.13)	(36732.36)	(29541.13)	(32423.12)	(22156.47)
2	(21.38)	(25.44)	(27412.66)	(49857.77)	(23797.85)	(27760.16)
3	(18.24)	(29.64)	(18309.04)	(28007.92)	(41305.45)	(47439.01)
4	(16.37)	(14.17)	(21216.23)	(47160.52)	(18342.83)	(16772.82)
5	(18.70)	(15.76)	(20488.83)	(34092.09)	(22319.72)	(84602.07)

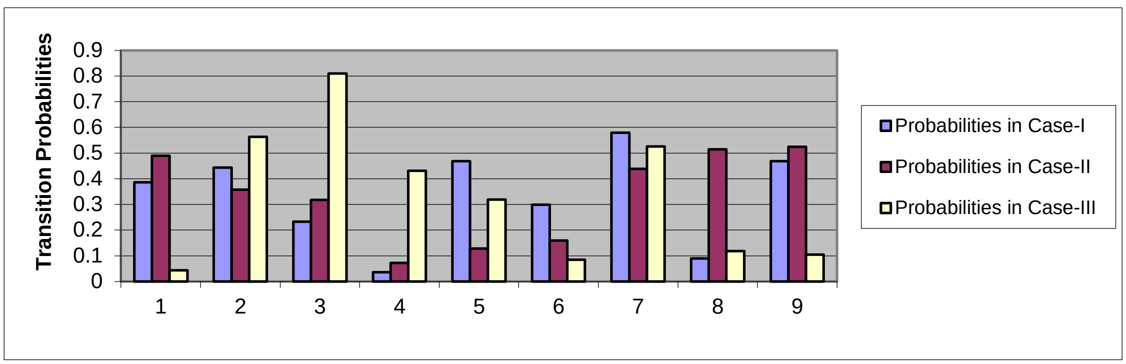

Figure 1: Estimated Probabilities for English Medium Category

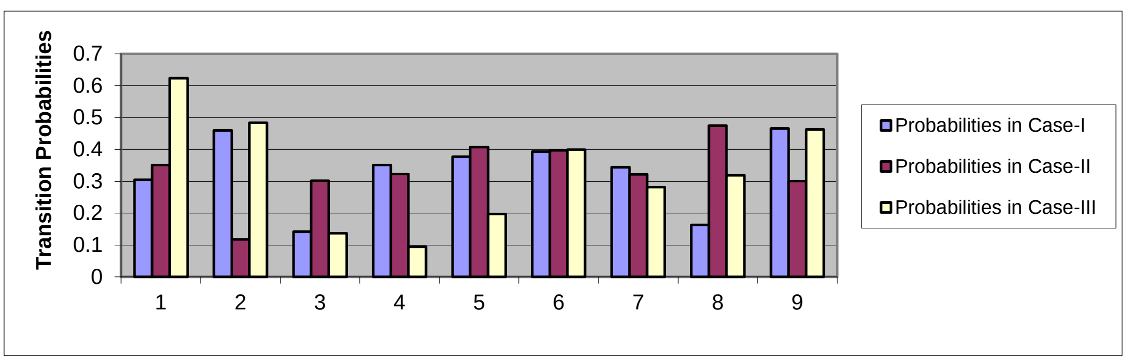

Figure 2: Estimated Probabilities for Bengali Medium Category

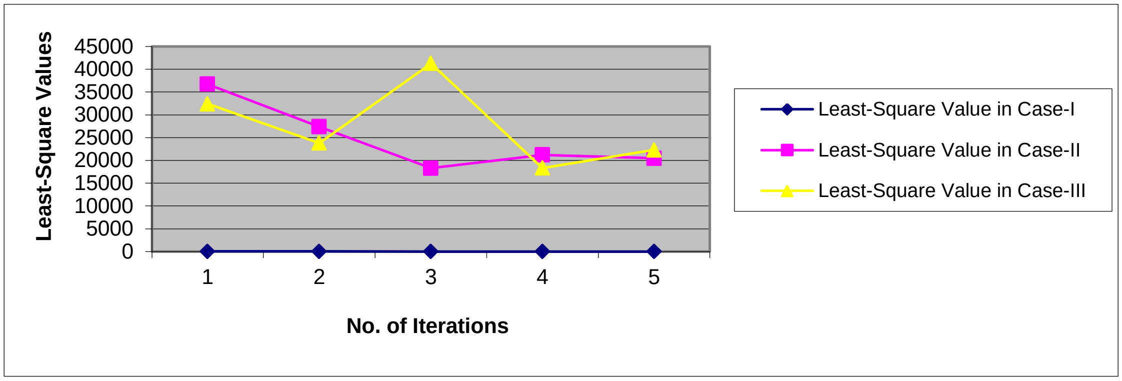

Figure 3: Comparisons of Least-Square Values for English Medium

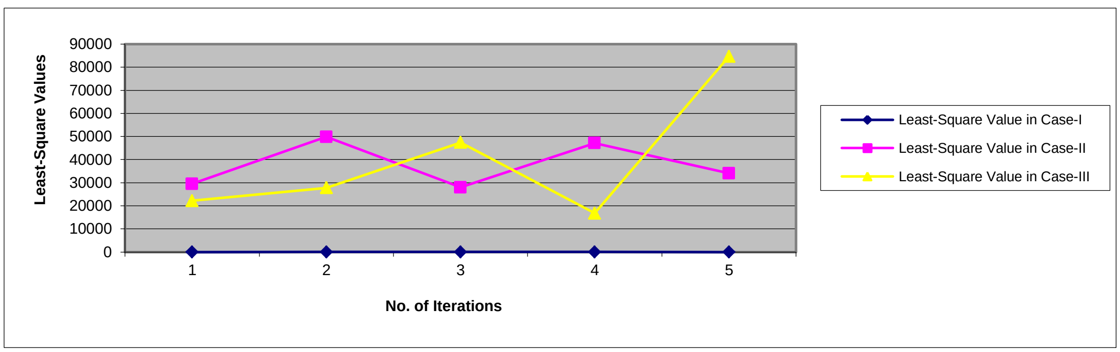

Figure 4: Comparisons of Least-Square Values for Bengali Medium

References

1. Anderson, T. W. and Goodman, L. A. (1957). Statistical Inference about Markov Chains. *The Annals of Mathematical Statistics*, **28**, 89-110.
2. Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes* Academic Press, New York.
3. Beaton, A E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147 – 185.
4. Bhatt, B. R. (2000). *Modern Probability Theory - An Introductory Text Book*. 2nd Edition, Wiley Eastern.

5. Calhoun, V., Adali, T., Kraut, M. and Pearlson, G. (2000). A Weighted Least-Squares Algorithm for estimation and visualization of Relative Latencies in Event-Related Functional MRI. *Magnetic Resonance in Medicine*, **44**, 947–954.
6. Dianne, P. O. (1990). Robust Regression Computation using Iteratively Re-Weighted Least Squares. *SIAM Journals of Matrix Analysis and Application*, **11(3)**, 466-480.
7. Douglas, S. W. (1984). Iteratively Re-weighted Least Squares for models with a linear part. *Journal of Applied Statistics*, **33(1)**, 7-17.
8. Goodman, L. A. (1955). On the Statistical Analysis of Markov Chains. *The Annals of Mathematical Statistics*, **26**, 771-780.
9. Kalbfleisch, J. D. and Lawless, J. F. (1984). Least-squares estimation of transition probabilities from aggregate data. *Canadian Journal of Statistics*, **12(3)**, 169–182.
10. Nicky, J. W. and Ades, A. E. (2005). Estimation of Markov Chain transition probabilities and rates from fully and partially observed data: Uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making*, **25**, 633-645.
11. Pires, R. C., Costa, A. S. and Mili, L. (1999). Iteratively Reweighted Least Squares State Estimation through given rotations. *IEEE Transactions on Power Systems*, **14(4)**, 1499 – 1507.
13. Sarma, R. and Sarmah, P. (1999). A Stochastic modeling on grading system. *Proceedings of the Second International Conference on Operations and Quantitative Management in the Global Business Environment (ICOQM)*, *Ahmedabad*, *India*, 276–281.
14. Sarma, R. and Sarmah, P. (1999). Analysis of Results Based on Grades. *Proceedings of the Second International Conference on Operations and Quantitative Management in the Global Business Environment (ICOQM)*, *Ahmedabad*, *India*, 282–290.
15. Stirling, W. D. (1984). Iteratively Reweighted Least Squares for models with a linear part. *Applied Statistics*, **33(1)**, 7-17.