5. Bayesian Analysis of Grading Pattern

5.1 Introduction

Under the model considered earlier and the t.p.m. (2.3.2), the analysis of results based on real life data is sometimes difficult, because proper information on individuals are not always recorded, in the sense that past data are not always available. So the main problem arises in estimating \( p \) / \( s \), the transition probabilities.

Here, we suspect that \( p \) / \( s \) are not always fixed. These transition probabilities \( p \) / \( s \) which are obtained from examination data are influenced by many factors namely, (i) infrastructure of the institution, (ii) economic conditions of the individual, (iii) family environment and many other socio-economic problems. So in such cases there is enough reason to suspect that \( p \) / \( s \) are random variables. In that case Bayesian statistical analysis is advisable.

As discussed earlier before going to find the Baye’s estimators of \( p \) / \( s \)’s, based on different priors such as uniform prior, beta prior etc. we obtained the M.L. estimators of \( p \) / \( s \)’s from individual or micro unit data. Inference problems such as estimation and hypothesis testing have been considered by several authors viz., Anderson and Goodman (1957), Collin (1974) etc. not only because of their theoretical interests, but also for their applications in diverse fields.
Here in this chapter, the maximum likelihood estimators of the transition probabilities are obtained in Section 2. Section 3 deals with Baye’s estimators of \( p_{ij} \)'s under uniform and beta distributions as prior distributions. In Section 4, the Baye’s estimators of the transition probability \( p_{ij} \)'s are obtained, taking prior information as suggested by the data set. Again in Section 5 Bayesian analysis of examination results are carried out using real life data. In fact the prior information about the parameters changes the distribution of the parameters to a great extent. This can be seen in the last section of this chapter. The prior and the posterior distribution of the parameters are represented graphically.

5.2 Maximum likelihood estimators of transition probabilities

Here a time homogenous Markov chain is considered with finite states \( S = \{0,1,2\} \) and having t.p.m. as considered in (2.3.2), where,

\[
p_{ij} = p_{r}(X_n = j | X_{n-1} = i) ; \quad i,j = 0,1,2
\]

let, \( n_{ij} = \) the number of direct transition from state \( i \) to \( j \).

\[
n_{ii} = \sum_{j=0}^{2} n_{ij}
\]

\[
n_{ij} = \sum_{i=0}^{2} n_{ij} ; \quad i,j = 0,1,2
\]
Since there is striking similarity between a sample from a Markov chain and one from multinomial trials, the logarithm of the likelihood function can be put as

\[
L(p_y) = c + \sum_{i=0}^{2} \sum_{j=0}^{2} n_y \log p_y
\]

where \( c \) contains terms independent of \( p_y \)’s

or,

\[
L(p_y) = c + \sum_{i=0}^{2} \sum_{j=0}^{1} n_y \log p_y + \sum_{i=0}^{2} n_{y2} \log(1 - \sum_{j=0}^{1} p_y)
\]

\[
\therefore \sum_{j=0}^{2} p_y = 1
\]

Let \( r \) be the specified value of \( i \), therefore the MLEs of \( p_{rj} \) are given by,

\[
\frac{\partial L(p_{rj})}{\partial p_{rj}} = 0, \quad j=0,1,2
\]

\[
\frac{n_{rj}}{p_{rj}} - \frac{n_{r2}}{1 - \sum_{j=0}^{1} p_{rj}} = 0, \quad j=0,1
\]

Again if, \( s \) is the specified value of \( j \),

\[
\frac{n_{rs}}{p_{rs}} = \frac{n_{sr}}{p_{sr}} = \frac{n_{r2}}{1 - \sum_{j=0}^{1} p_{rj}}
\]

\[
1 - \sum_{j=0}^{1} p_{rj} = \frac{n_{r2}}{n_{rs}} \quad \cdots \cdots (5.2.1)
\]
and \( p_{ij} = \frac{n_{ij}}{n_{rs}} p_{rs} \) \hspace{1cm} \ldots \ldots \quad (5.2.2)

Now adding (5.2.2) over \( j \) and adding to (5.2.1) we get,

\[
1 = \frac{\sum_{j=0}^{2} n_{ij}}{n_{rs}} p_{rs}
\]

\[
\therefore \quad \hat{p}_{rs} = \frac{n_{rs}}{\sum_{j=0}^{2} n_{ij}}
\]

In general,

\[
\hat{p}_{y} = \frac{\sum_{j=0}^{2} n_{yj}}{n_{r0}} = \frac{n_{y}}{n_{r0}}
\]

In the following sections this estimate of \( p_{y} \) obtained by M.L. estimation method is compared with other Baye's estimators of \( p_{ij} \) obtained under different prior probabilities.

5.3 \textbf{Bayesian Analysis with proper prior}

In this section the Baye's estimators of \( p_{yj} \)'s are obtained under different prior distributions, such as uniform distribution in 0 to 1, beta
distributions and so on and they are compared with the corresponding M.L.E.'s with the help of real life data.

Case I:

When the prior distribution is uniform in 0 to 1.

In this case $g(p_{ij}) = 1, \ 0 \leq p_{ij} \leq 1, \ i, j = 0, 1, 2$

Since there is similarity between a sample from a Markov model and one from multinomial trials we get the posterior distribution as,

$$H(p_{ij} | n_{ij}) = \frac{\sum_{i=0}^{2} \sum_{j=0}^{2} p_{ij}^{n_{ij}}}{\sum_{i=0}^{2} \sum_{j=0}^{2} p_{ij}^{n_{ij}} dp_{ij}}$$

where, $c = \frac{(n)!}{\prod_{i=0}^{2} \prod_{j=0}^{2} (n_{ij})!}$

Hence the Baye’s estimators of $p_{00}$ is,

$$\hat{p}_{00} = \frac{\sum_{i=0}^{2} \sum_{j=0}^{2} P_{01}^{n_{01}} P_{02}^{n_{02}} P_{10}^{n_{10}} P_{11}^{n_{11}} P_{12}^{n_{12}} P_{20}^{n_{20}} P_{21}^{n_{21}} P_{22}^{n_{22}} 1^{p_{00}^{n_{00}+1} dp_{00}}}{\sum_{i=0}^{2} \sum_{j=0}^{2} P_{01}^{n_{01}} P_{02}^{n_{02}} P_{10}^{n_{10}} P_{11}^{n_{11}} P_{12}^{n_{12}} P_{20}^{n_{20}} P_{21}^{n_{21}} P_{22}^{n_{22}} 1^{p_{00}^{n_{00}} dp_{00}}}

where \( P_{ij} \)'s are equal except for \( P_{ij} \) \( i = 0, j = 0 \)

$$= \frac{n_{00} + 1}{n_{00} + 2}$$
Again since the multinomial family of distribution is complete, we can write that,
\[ \hat{p}_{02} = p_{02} = 0 \]

which is clear from the t.p.m. (2.3.2)

\[ \therefore \hat{p}_{01} = 1 - \hat{p}_{00} \]

Again,
\[ \hat{p}_{11} = \frac{n_1 + 1}{n_1 + 2} \]

Again, as before since the family is complete

\[ \hat{p}_{11} = 1 - c \]

\[ 1 - c = \frac{n_1 + 1}{n_1 + 2} \]

\[ c = 1 - \frac{n_1 + 1}{n_1 + 2} \]

\[ = \frac{1}{n_1 + 2} \]

\[ \therefore \hat{p}_{10} = \frac{c}{2} = \frac{1}{2(n_1 + 2)} \]

\[ \hat{p}_{12} = \frac{c}{2} = \frac{1}{2(n_1 + 2)} \]
Similarly, \[ \hat{p}_{21} = \frac{n_{21}+1}{n_{21}+2} \]

And for the completeness of the multinomial family of distributions, we can estimate \( p_{20} \) as,

\[ \hat{p}_{20} = p_{20} = 0 \]

\[ \therefore \hat{p}_{22} = 1 - \hat{p}_{21} = 1 - \frac{n_{21}+1}{n_{21}+2} = \frac{1}{n_{21}+2} \]

Case II

In this case the prior distribution is considered as a beta distribution of first type,

i.e., \[ g(p_y) = \frac{1}{\beta(l,m)} p_y^{l-1}(1-p_y)^{m-1} , 0 \leq p_y \leq 1 \]

where \( l \) and \( m \) are fixed and can be fixed at will.

The posterior distribution of \( p_{ij} \) in this case will be,
So the Baye's estimator of $p_{00}$ will be,

$$\hat{p}_{00} = \frac{\beta(l + n_{00} + 1, m)}{\beta(l + n_{00}, m)} = \frac{l + n_{00}}{l + m + n_{00}}$$

Since the multinomial family of distributions is complete, we can estimate $p_{02}$ as,

$$\hat{p}_{02} = p_{02} = 0$$

$$\therefore \hat{p}_{01} = 1 - \hat{p}_{00} = \frac{m}{l + m + n_{00}}$$

Again Baye's estimator of $p_{11}$ is,

$$\hat{p}_{11} = \frac{l + n_{11}}{l + m + n_{11}}$$

$$\Rightarrow 1 - c = \frac{l + n_{11}}{l + m + n_{11}}$$

$$\Rightarrow c = \frac{m}{l + m + n_{11}}$$
\begin{align*}
\therefore \hat{p}_{10} = \hat{p}_{12} &= \frac{c}{2} = \frac{1}{2} \left( \frac{m}{l + m + n_{11}} \right) \\

\text{And Bayes estimator of } p_{21} \text{ is,} \\
\hat{p}_{21} &= \frac{l + n_{21}}{l + m + n_{21}} \\

\text{Again from the completeness of multinomial family of distributions we get from t.p.m. (2.3.2)} \\
\hat{p}_{20} &= p_{20} = 0 \\
\hat{p}_{22} &= 1 - \hat{p}_{21} \\
&= \frac{m}{l + m + n_{21}} \\

\text{Maximum likelihood estimates of the transition probabilities from real life data} \\
\text{Table : 5.3.1} \\
\begin{array}{cccccccc}
<table>
<thead>
<tr>
<th>p_{00}</th>
<th>p_{01}</th>
<th>p_{02}</th>
<th>p_{10}</th>
<th>p_{11}</th>
<th>p_{12}</th>
<th>p_{20}</th>
<th>p_{21}</th>
<th>p_{22}</th>
</tr>
</thead>
<tbody>
<tr>
<td>.300</td>
<td>.697</td>
<td>.0028</td>
<td>.079</td>
<td>.837</td>
<td>.084</td>
<td>0</td>
<td>.375</td>
<td>.625</td>
</tr>
</tbody>
</table>
\end{array}
\end{align*}
Bayes estimators of $p_i$'s obtained from real life data

Case I: Uniform prior

Table: 5.3.2

<table>
<thead>
<tr>
<th>$p_{00}$</th>
<th>$p_{01}$</th>
<th>$p_{02}$</th>
<th>$p_{10}$</th>
<th>$p_{11}$</th>
<th>$p_{12}$</th>
<th>$p_{20}$</th>
<th>$p_{21}$</th>
<th>$p_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.995</td>
<td>.005</td>
<td>0</td>
<td>.0012</td>
<td>.997</td>
<td>.0012</td>
<td>0</td>
<td>.983</td>
<td>.016</td>
</tr>
</tbody>
</table>

Case II: $\beta_{1}(1,m)$ prior

Table: 5.3.3

<table>
<thead>
<tr>
<th>$l=1, m=1$</th>
<th>$p_{00}$</th>
<th>$p_{01}$</th>
<th>$p_{02}$</th>
<th>$p_{10}$</th>
<th>$p_{11}$</th>
<th>$p_{12}$</th>
<th>$p_{20}$</th>
<th>$p_{21}$</th>
<th>$p_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.995</td>
<td>.004</td>
<td>0</td>
<td>.0012</td>
<td>.997</td>
<td>.0012</td>
<td>0</td>
<td>.984</td>
<td>.0161</td>
<td></td>
</tr>
</tbody>
</table>

| $l=1, m=2$ | .991     | .009     | 0        | .0025    | .995     | .0025    | 0        | .968     | .031     |

| $l=1, m=3$ | .986     | .013     | 0        | .0035    | .993     | .0035    | 0        | .953     | .047     |

| $l=1, m=5$ | .977     | .023     | 0        | .0065    | .987     | .0065    | 0        | .910     | .089     |

| $l=1, m=10$| .955     | .045     | 0        | .012     | .976     | .012     | 0        | .859     | .140     |

| $l=1, m=20$| .915     | .085     | 0        | .0236    | .953     | .0236    | 0        | .753     | .247     |

| $l=2, m=10$| .955     | .045     | 0        | .012     | .976     | .012     | 0        | .861     | .139     |

| $l=3, m=20$| .915     | .085     | 0        | .0236    | .953     | .0236    | 0        | .753     | .247     |

| $l=4, m=30$| .879     | .121     | 0        | .0345    | .931     | .0345    | 0        | .681     | .319     |

| $l=10, m=1$| .995     | .005     | 0        | .001     | .998     | .001     | 0        | .986     | .014     |

| $l=30, m=4$| .983     | .016     | 0        | .005     | .990     | .005     | 0        | .957     | .042     |
5.4 Bayesian analysis with improper prior

From above it is clear that prior distributions considered in Case I and in Case II do not give a clear picture of the estimate of $p_{ij}$ and we end up with the estimates of $p_{ij}$ which are not comparable with the maximum likelihood estimates of the transition probability $p_{ij}$'s. So we proceed by taking different prior distributions for each of these $p_{ij}$'s, as data also suggests us to do the same.

Looking at the data set collected from different educational institutions, it may be observed that probability of excellence that is $p_{00}$ is a very sensitive parameter, depending upon various factors like practice, intelligence, understanding, presentation, capacity and some other socio-economic factors.

Therefore, we assume that $p_{00}$ follows improper prior distribution,

$$
g(p_{00}) = p_{00}^\theta_1 (1 - p_{00})^\theta_2, \quad \theta_1, \theta_2 \in \mathbb{R}
$$

$$
0 \leq p_{00} \leq 1
$$

where, $\theta_1$ is much higher compared to $\theta_2$.

Therefore, the posterior distribution of $p_{00}$ is,

$$
H(p_{00} \mid n_{00}) = \frac{\prod_{i=0}^{2} \prod_{j=0}^{2} p_{ij}^\eta_i p_{00}^\eta_j (1 - p_{00})^{\eta_2} \int_0^{p_{00}} p_{00}^\eta_1 (1 - p_{00})^\eta_2 dp_{00}}{P_{01}^{n_{01}} P_{02}^{n_{02}} P_{10}^{n_{10}} P_{11}^{n_{11}} P_{20}^{n_{20}} P_{21}^{n_{21}} P_{22}^{n_{22}}}
$$
Therefore, the Baye's estimator of $p_{00}$ is obtained as,

$$\hat{p}_{00} = \frac{\beta(n_{00} + \theta_1 + 2, \theta_2 + 1)}{\beta(n_{00} + \theta_1 + 1, \theta_2 + 1)}$$

$$= \frac{n_{00} + \theta_1 + 1}{n_{00} + \theta_1 + \theta_2 + 2}$$

$$\hat{p}_{02} = 0$$

And $$\hat{p}_{01} = 1 - \hat{p}_{00} = \frac{\theta_2 + 1}{n_{00} + \theta_1 + \theta_2 + 2}$$

Again, keen observation of the data reveals that the state 1 is a recurring state and the probability of retaining in the middle grade is higher than the other two grades. Therefore it was found reasonable to assume that $p_{11}$ follows uniform distribution in 0 to 1, i.e. the prior distribution in this case is,

$$G(p_{11}) = 1, \quad 0 \leq p_{11} \leq 1$$

Hence the Baye's estimator of $p_{11}$ is,

$$\hat{p}_{11} = \frac{n_{11} + 1}{n_{11} + 2}$$
As shown in case I of this section.

\[ \hat{p}_{10} = \hat{p}_{22} = \frac{1}{2(n_{11} + 2)} \]

Again our observation conveys us that the parameter \( p_{22} \) is rarely influenced by some factors. Therefore, we assume \( g(p_{22}) \) to be,

\[ g(p_{22}) = p_{22}(1 - p_{22})^\theta ; \quad 0 \leq p_{22} \leq 1. \]

\[ \theta \in R \]

Hence the posterior distribution of \( p_{22} \) is,

\[ G(p_{22} \mid \theta) = \frac{c \prod_{i=0}^{2} \prod_{j=0}^{2} p_{ij}^{n_{ij}} p_{22}^{(1-p_{22})^\theta}}{c \prod_{i=0}^{2} \prod_{j=0}^{2} p_{ij}^{n_{ij}} p_{22}(1-p_{22})^\theta \, dp_{22}} \]

Hence the Bayes estimator of \( p_{22} \) is.

\[ \hat{p}_{22} = \frac{n_{22} + 2}{n_{22} + \theta + 3} \]

As before for the completeness of multinomial distribution.
As before for the completeness of multinomial distribution,

$$\hat{p}_{20} = 0$$

$$\hat{p}_{21} = 1 - \hat{p}_{22}$$

$$= \frac{\theta + 1}{n_{22} + \theta + 3}$$
Bayes estimators of the transition probabilities $p_{ij}$, $i, j = 0, 1, 2$

(i) Bayes’s estimator of $p_{0i}$, $i=0,1,2$

(a) When $\theta_1 > \theta_2$

Table 5.4.1

<table>
<thead>
<tr>
<th>$\theta_1$, $\theta_2$</th>
<th>$p_{00}$</th>
<th>$p_{01}$</th>
<th>$p_{02}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 0</td>
<td>0.995</td>
<td>0.005</td>
<td>0</td>
</tr>
<tr>
<td>3, 0</td>
<td>0.995</td>
<td>0.005</td>
<td>0</td>
</tr>
<tr>
<td>5, 1</td>
<td>0.990</td>
<td>0.010</td>
<td>0</td>
</tr>
<tr>
<td>10, 1</td>
<td>0.991</td>
<td>0.009</td>
<td>0</td>
</tr>
<tr>
<td>20, 1</td>
<td>0.991</td>
<td>0.009</td>
<td>0</td>
</tr>
<tr>
<td>5, 2</td>
<td>0.986</td>
<td>0.014</td>
<td>0</td>
</tr>
<tr>
<td>5, 4</td>
<td>0.977</td>
<td>0.023</td>
<td>0</td>
</tr>
<tr>
<td>20, 4</td>
<td>0.979</td>
<td>0.021</td>
<td>0</td>
</tr>
<tr>
<td>10, 5</td>
<td>0.973</td>
<td>0.027</td>
<td>0</td>
</tr>
<tr>
<td>10, 8</td>
<td>0.961</td>
<td>0.039</td>
<td>0</td>
</tr>
<tr>
<td>15, 8</td>
<td>0.962</td>
<td>0.038</td>
<td>0</td>
</tr>
<tr>
<td>10, 9</td>
<td>0.957</td>
<td>0.043</td>
<td>0</td>
</tr>
<tr>
<td>20, 10</td>
<td>0.955</td>
<td>0.045</td>
<td>0</td>
</tr>
</tbody>
</table>
(b) When $\theta_1 < \theta_2$

Table: 5.4.2

<table>
<thead>
<tr>
<th>$\theta_1, \theta_2$</th>
<th>$p_{00}$</th>
<th>$p_{01}$</th>
<th>$p_{02}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0, 1$</td>
<td>.990</td>
<td>.010</td>
<td>0</td>
</tr>
<tr>
<td>$0, 2$</td>
<td>.986</td>
<td>.014</td>
<td>0</td>
</tr>
<tr>
<td>$0, 5$</td>
<td>.973</td>
<td>.027</td>
<td>0</td>
</tr>
<tr>
<td>$1, 2$</td>
<td>.986</td>
<td>.014</td>
<td>0</td>
</tr>
<tr>
<td>$1, 5$</td>
<td>.972</td>
<td>.028</td>
<td>0</td>
</tr>
<tr>
<td>$1, 10$</td>
<td>.951</td>
<td>.049</td>
<td>0</td>
</tr>
<tr>
<td>$2, 5$</td>
<td>.973</td>
<td>.027</td>
<td>0</td>
</tr>
<tr>
<td>$2, 10$</td>
<td>.952</td>
<td>.048</td>
<td>0</td>
</tr>
<tr>
<td>$3, 8$</td>
<td>.960</td>
<td>.040</td>
<td>0</td>
</tr>
<tr>
<td>$3, 10$</td>
<td>.952</td>
<td>.048</td>
<td>0</td>
</tr>
<tr>
<td>$4, 9$</td>
<td>.956</td>
<td>.044</td>
<td>0</td>
</tr>
<tr>
<td>$5, 10$</td>
<td>.952</td>
<td>.048</td>
<td>0</td>
</tr>
</tbody>
</table>
(c) when $\theta_1 = \theta_2$

Table: 5.4.3

<table>
<thead>
<tr>
<th>$\theta_1, \theta_2$</th>
<th>$p_{00}$</th>
<th>$p_{01}$</th>
<th>$p_{02}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1, 1$</td>
<td>0.990</td>
<td>0.010</td>
<td>0</td>
</tr>
<tr>
<td>$2, 2$</td>
<td>0.986</td>
<td>0.014</td>
<td>0</td>
</tr>
<tr>
<td>$3, 3$</td>
<td>0.982</td>
<td>0.018</td>
<td>0</td>
</tr>
<tr>
<td>$4, 4$</td>
<td>0.978</td>
<td>0.022</td>
<td>0</td>
</tr>
<tr>
<td>$5, 5$</td>
<td>0.973</td>
<td>0.027</td>
<td>0</td>
</tr>
<tr>
<td>$6, 6$</td>
<td>0.969</td>
<td>0.031</td>
<td>0</td>
</tr>
<tr>
<td>$7, 7$</td>
<td>0.965</td>
<td>0.035</td>
<td>0</td>
</tr>
<tr>
<td>$8, 8$</td>
<td>0.961</td>
<td>0.039</td>
<td>0</td>
</tr>
<tr>
<td>$9, 9$</td>
<td>0.957</td>
<td>0.043</td>
<td>0</td>
</tr>
<tr>
<td>$10, 10$</td>
<td>0.953</td>
<td>0.047</td>
<td>0</td>
</tr>
</tbody>
</table>

(ii) Bayes estimators of $p_{ij}$, $i = 1, 2$

$j = 0, 1, 2$

Table: 5.4.4

<table>
<thead>
<tr>
<th>$p_{10}$</th>
<th>$p_{11}$</th>
<th>$p_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0012</td>
<td>0.997</td>
<td>0.0012</td>
</tr>
<tr>
<td>θ</td>
<td>( p_{20} )</td>
<td>( p_{21} )</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>.0192</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>.029</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>.037</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>.047</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>.056</td>
</tr>
</tbody>
</table>

### 5.5 Bayesian Analysis of Results from life data

**Case – I**

We have considered the results of six different institutions which are named as \( C_1, C_2, C_3, C_4, C_5 \) and \( C_6 \). We have selected a student at random from one of these institutions, which is unknown to us, but the selected student comes out to be in grade ‘0’ in the result of the final examination. We are interested to know from which institution that selected student has come.

In institution \( C_1 \), 27% student comes out with Grade “0”, in \( C_2 \) the percentage is 10, in \( C_3 \) it is 20%, in \( C_4 \) it comes out to be 16%, in \( C_5 \) it is 8% and it is 3% in \( C_6 \).

Let \( D \) be the event of selecting a student getting grade “0”. Our problem is to identify the institution from which the student has come.

i.e. we want,
Pr \{\text{that the selected student getting grade "0" comes out from the institution } C_i, \}
i = 1, 2, \ldots \ldots \ldots \ldots \ldots 6; 

which we can obtain using Baye's theorem.

We assume that the initial choice of the institutions were made at 
random with equal probability.

\[ \therefore P(C_1) = P(C_2) = P(C_3) = P(C_4) = P(C_5) = P(C_6) = \frac{1}{6} = .1667 \]

These are the prior probabilities which give us the probabilities of 
selecting a institute before the data are known. Now to use Baye's theorem, the 
required data probabilities are,

\[
\begin{align*}
P_r(D | C_1) &= .27 & P_r(D | C_4) &= .16 \\
P_r(D | C_2) &= .10 & P_r(D | C_5) &= .08 \\
P_r(D | C_3) &= .20 & P_r(D | C_6) &= .03 \\
\end{align*}
\]

According to Baye's theorem,

\[
P_r(C_1 | D) = \frac{P(C_1)P(D | C_1)}{P(C_1)P(D | C_1) + \ldots + P(C_6)P(D | C_6)} \\
= .321
\]

which gives us the probability that the selected student has come from the first 
institute.
Similarly, \( P_r(C_2 \mid D) = .119 \quad P_r(C_4 \mid D) = .190 \)
\( P_r(C_3 \mid D) = .238 \quad P_r(C_5 \mid D) = .095 \)
\( P_r(C_6 \mid D) = .036 \)

These are the posterior probabilities. These computations can be arranged in the following table for comparison

Table 5.5.1

<table>
<thead>
<tr>
<th>Institution</th>
<th>Percentage of Grade “0” holder</th>
<th>Prior Probs ( P(C_i) )</th>
<th>Data Probs ( P(D \mid C_i) )</th>
<th>Posterior Probs ( P(C_i \mid D) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>.1667</td>
<td>.27</td>
<td>.321</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.1667</td>
<td>.10</td>
<td>.119</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>.1667</td>
<td>.20</td>
<td>.238</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>.1667</td>
<td>.16</td>
<td>.190</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>.1667</td>
<td>.08</td>
<td>.095</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>.1667</td>
<td>.03</td>
<td>.036</td>
</tr>
</tbody>
</table>

Sum 1.000    Sum .999

Before the knowledge of the data were available each of the institutions had the same probability .1667 of producing the student, but the knowledge of the data has changed almost all the probabilities. At this stage the uncertainty about which institute is producing the data is reflected in the six posterior probabilities in column (5). Most likely the first institute is the source
of the data, since probability is highest in this case, but there is not much
difference between uncertainties for the colleges $C_1$ and $C_3$, the probability of the
third institute producing the student cannot be ruled out.

We have extended the problem one step further by collecting some
additional data from the same institute to make a better sense of which of the six
institutions gave us the data.

Suppose, there are 10 students in the new sample and out of these 3
are in grade "0". The old posterior probabilities for the institutions are the new
prior probabilities. Data probabilities are found by using binomial distribution.

So the probability of 3 grade "0" holders and seven others becomes,

\[
\begin{align*}
    P_r(D \mid C_1) &= \binom{10}{3} (0.27)^3 (1 - 0.27)^7 = 0.261 \\
    P_r(D \mid C_2) &= \binom{10}{3} (0.10)^3 (1 - 0.10)^7 = 0.05 \\
    P_r(D \mid C_3) &= \binom{10}{3} (0.20)^3 (1 - 0.20)^7 = 0.201 \\
    P_r(D \mid C_4) &= \binom{10}{3} (0.16)^3 (1 - 0.16)^7 = 0.145 \\
    P_r(D \mid C_5) &= \binom{10}{3} (0.08)^3 (1 - 0.08)^7 = 0.034 \\
    P_r(D \mid C_6) &= \binom{10}{3} (0.03)^3 (1 - 0.03)^7 = 0.003
\end{align*}
\]
Using Baye's theorem, the posterior probabilities are obtained as,

\[ P_r(C_1 \mid D) = .497 \]

\[ P(C_2 \mid D) = .035 \]

\[ P(C_3 \mid D) = .284 \]

\[ P(C_4 \mid D) = .164 \]

\[ P(C_5 \mid D) = .019 \]

\[ P(C_6 \mid D) = .0006 \]

These posterior probabilities reveal that it is almost impossible that the student has come from the second, fifth and the sixth institutions and most likely the student has come from the first institution.

*Case – II*

Let \( P \) be the proportion of students getting grade "0" in the final examination in an institution. We know very little about \( P \). But we can say that \( P \) lies between \( \alpha \) and \( \beta \), where \( 0 \leq \alpha \leq \beta \leq 1 \).
Since nothing is known about the values of $P$, we can say that the mean of $P$ will lie at $\mu = \frac{\alpha + \beta}{2}$ and standard deviation of $\alpha$ and $\beta$ from the mean $\frac{\alpha + \beta}{2}$ is given by

$$\sigma = \left[ \frac{1}{2} \left\{ \left( \alpha - \frac{\alpha + \beta}{2} \right)^2 + \left( \beta - \frac{\alpha + \beta}{2} \right)^2 \right\} \right]^{\frac{1}{2}}$$

$$= \frac{\beta - \alpha}{2}$$

Now assuming $P$ to be a continuous variable which can take different values between 0 and 1, we can assume the prior distribution of $P$ to be a beta-distribution of the form

$$f(P) = \frac{1}{\beta(a,b)} P^{a-1}(1-P)^{b-1}, \quad 0 \leq P \leq 1$$

$$= C' P^{a-1}(1-P)^{b-1}, \quad 0 \leq P \leq 1 \quad \text{........... (5.5.1)}$$

Again, since for a beta distribution,

$$\mu = \frac{a}{a+b}$$

and,

$$\sigma^2 = \frac{\mu(1-\mu)}{a + b - 1}$$

$$\Rightarrow a = \mu \left[ \frac{\mu(1-\mu)}{\sigma^2} - 1 \right]$$
Therefore, we get,

\[
a = \frac{\alpha + \beta}{2} \left[ \frac{(\frac{\alpha + \beta}{2})(1 - \frac{\alpha + \beta}{2})}{\left(\frac{\beta - \alpha}{2}\right)^2} - 1 \right]
\]

\[
= \frac{\alpha + \beta}{2} \cdot \frac{4}{(\beta - \alpha)^2} \left[ \frac{\alpha + \beta}{2} - \frac{(\alpha + \beta)^2}{4} - \frac{(\alpha - \beta)^2}{4} \right]
\]

\[
= \frac{\alpha + \beta}{(\beta - \alpha)^2} \left[ (\alpha + \beta) - (\alpha^2 + \beta^2) \right] \quad \ldots \ldots \quad (5.5.2)
\]

And \[b = \left[1 - \frac{\alpha + \beta}{2}\right] \left[ \frac{(\frac{\alpha + \beta}{2})(1 - \frac{\alpha + \beta}{2})}{\left(\frac{\beta - \alpha}{2}\right)^2} - 1 \right]
\]

\[
= \frac{2 - (\alpha + \beta)}{2} \cdot \frac{4}{(\beta - \alpha)^2} \left[ \frac{\alpha + \beta}{2} - \frac{(\alpha + \beta)^2}{2} - \frac{(\alpha - \beta)^2}{2} \right]
\]

\[
= \frac{2 - (\alpha + \beta)}{(\alpha - \beta)^2} \left[ (\alpha + \beta) - (\alpha^2 + \beta^2) \right] \quad \ldots \ldots \quad (5.5.3)
\]

So depending upon different choices of priors, i.e., for different values of \(\alpha\) and \(\beta\), we get different posterior distributions.
Here we considered a sample of 1384 students appearing in the HS Final Examination of which 251 students passed out with grade "0". Therefore, the probability of getting 251 grade "0" holders and 1133 students receiving grades other than grade "0" will be given by,

\[ f(data|P) = \frac{1384}{251} P^{251} (1-P)^{1133}, \quad 0 \leq P \leq 1 \]

which cannot be calculated unless we know the value of P. Now posterior distribution for P using prior information as given in (5.5.1) is given by,

\[
f(P|data) = \frac{\binom{1384}{251} P^{251} (1-P)^{1133} \Gamma(P+a-1)(1-P)^{b-1}}{\int_0^1 \binom{1384}{251} P^{251+a-1} (1-P)^{1133+b-1} dP}
\]

\[ = C'' P^{250+a}(1-P)^{132+b}, \quad \ldots \ldots \quad (5.5.4) \]

where, \( C'' = \frac{1}{\beta(251+a,1133+b)} \)

Hence, Posterior mean = \( \frac{251+a}{1384+a+b} = \mu' \)

And, posterior standard deviation = \( \sqrt{\frac{\mu'(1-\mu')}{1384+a+b-1}} = \sigma' \)
Particular cases:

Case 1 $\alpha = 0$, $\beta = .25$

From the equation (5.5.2) and (5.5.3) we get,

$$a = .75 \simeq 1$$
$$b = 5.2 \simeq 5$$
$$C' = 5$$

Hence the prior distribution (5.5.1) will be,

$$f(P) = 5(1 - P)^4, \quad 0 \leq P \leq 1$$

and the posterior distribution will be given by,

$$f(P \mid \text{data}) = C''P^{25}(1-P)^{1137}, \quad 0 \leq P \leq 1$$

where,

$$C'' = \frac{1}{\beta(252, 1138)}$$

Posterior mean = .1813

And, posterior variance = .000106

Therefore, posterior standard deviation = .0103
We plot the graph of equation (5.5.5) i.e. the graph of prior distribution and equation (5.5.6), i.e. the graph of posterior distribution and we notice that the graph of the posterior distribution is very much peaked compared to the graph of prior distribution and it is symmetrical enough to be approximated by a Normal curve.

Hence we can write,

\[ P\{\mu'-1.96\sigma' \leq \mu' \leq \mu' + 1.96\sigma'\} = .95 \]

i.e., \[ P\{.1610 \leq P \leq .2016\} = .95 \]

\textit{Case II} \quad \alpha = 0, \quad \beta = .30

\[
\begin{align*}
    a & = .7 \equiv 1 \\
    b & = 3.96 \equiv 4 \\
    c' & = 4
\end{align*}
\]

Hence the prior distribution will be,

\[ f(P) = 4(1-P)^3, \quad 0 \leq P \leq 1 \]

\[ \ldots \ldots \ldots \ldots \text{(5.5.7)} \]
And the posterior distribution will be of the type

\[ f(P|\text{data}) = \frac{1}{\beta(252,1137)} P^{252}(1-P)^{1136}, \quad 0 \leq P \leq 1 \]

Posterior Mean = .1814
Posterior standard deviation = .010341

As in Case (I) here also it is seen that the graph of equation (5.5.8),
that is the graph of the posterior distribution can be approximated by a Normal
curve and hence we can write,

\[ P_r \{ .1611 \leq P \leq .2016 \} = .95 \]

**Case III** \( \alpha = 0, \beta = .35 \)

From the equation (5.5.2) and (5.5.3) we get,

\[ a = .65 \approx 1 \]
\[ b = 3.06 \approx 3 \]
\[ C' = 3 \]
Hence the prior distribution (5.5.1) will be of the form,

\[ f(P) = 3(1-P)^2, \quad 0 \leq P \leq 1 \]  \quad \ldots \quad (5.5.9)

and the posterior distribution will be given by,

\[
\frac{1}{\beta(252,1136)} P^{251} (1 - P)^{1135}, \quad 0 \leq P \leq 1 \quad \ldots \quad (5.5.9.a)
\]

Posterior Mean = .1815

Posterior standard deviation = .0103

Hence, as in Case (I) and Case (II) we can write,

\[ P \{ .1612 \leq P \leq .2018 \} = .95 \]
Case I: Prior distribution of $P$, the proportion of students getting grade 'O' in the pre-degree examination.

Fig 5.5.1: Prior distribution of $P$, the proportion of students getting grade 'O' in the pre-degree examination.
Case I: Figure 5.5.2: Posterior distribution of $P$, the proportion of students getting grade '0' in the pre-degree examination.
Case II: Prior Distribution of P, the proportion of students getting grade 'O' in pre-degree examination
Case II:
Fig 5.5.4: Posterior Distribution of $P$, the proportion of students getting grade '0' in pre-degree examination
Case III: Prior distribution of $P$, the proportion of students getting grade 'O' in the pre degree examination

Fig 5.5.5: Prior distribution of $P$, the proportion of students getting grade 'O' in the pre degree examination
Case III:
Fig 5.5.6: Posterior distribution of P, the proportion of students getting grade '0' in the pre degree examination.
Discussion

The first case study of life data helps us to predict about the institution from which a selected student has come.

Whereas in the second case study it is seen that although the prior information about the proportion of students getting grade “0” in the final examination of a particular institution was assumed to lie between 0 to 0.25 in the first case, 0 to 0.30 in the second case and 0 to 0.35 in the third case, it is clear from the graphs of the posterior distributions, fig. (5.5.2), fig. (5.5.4) and fig. (5.5.6) that our knowledge of $P$ changes and it is almost (95%) sure that the actual proportion of students getting grade “0” in the final examination, i.e. $P$, lies between 0.161 and 0.201.