"The whole of Science is nothing more than a refinement of everyday thinking."

Albert Einstein
Physics and Reality
4.1 Project Progress Data

A project is fundamentally problems solving activity. In order to construct a progress model of a project, analysis of progress data is required. The pattern of progress can be obtained from the life cycle data i.e. by recording events (number of problems solved and decision taken) as a function of time. From this data success/progress rate may be computed and plotted. On the basis of the plot and from various informations, judgement and statistical tests a characteristic, behaviour model may be developed which should be representative of the projects of the similar nature. The parameters of the model are estimated from the graph or may be computed using statistical principles of estimation. In this chapter, we shall analyse project progress data that will provide a basis for formulating mathematically Rayleigh-Norden Model for general analysis.

4.2 Analysis of Progress Data

Progress data is generally obtained from the progress reports of projects under execution. This data is usually presented as a plot of success density function or progress rate as a function of time.
The data we are dealing with are a sequence of times to success in solving problems and decision making, but the success density function and the progress rate are continuous variables. We first compute a discrete success density function and progress rate from the data. It can be shown that these discrete functions approach the continuous functions in the limit as the number of data becomes large and the interval between success times approaches zero.

Suppose a data describes a set of $N$ problems to be solved when the project starts at time $t=0$. As time elapses, events occur (problems are solved and decisions taken) and at any time $t$ the number of successes or events is $s(t)$ and number of remaining events $n(t)$. Obviously,

$$s(t) + n(t) = N \quad \ldots \quad (4.1)$$

Now, we define two functions as follows:

### 4.2.1 Success Density Function

It is defined over the time interval $\Delta t$ and is given by the ratio of the number of events occurring in the interval to the size of the original population of unsolved problems, divided by the length of the time interval:

$$s(t) = \left[ \frac{s(t+\Delta t) - s(t)}{\Delta t} \right] / N$$

$$d = \left[ \frac{n(t) - n(t+\Delta t)}{\Delta t} \right] / N \quad \ldots \quad (4.2)$$

It is a measure of the overall speed at which successes are achieved.


4.2.2 Progress Rate

Progress rate over the interval $\Delta t$ is defined as the ratio of the number of events occurring in the interval to the average* number of events during the interval divided by the length of the time interval:

$$g(t) = \frac{\left[ n(t) - n(t+\Delta t) \right]}{\Delta t}$$

where

$$\bar{n}(t) = \frac{\left[ n(t) + n(t+\Delta t) \right]}{2} \quad \ldots \ldots \ (4.3)$$

It is a measure of the instantaneous speed of the success.

Let us consider a project consisting of 1000 events. The total duration of the project was 18 weeks. The number of events ($s$) that occur during each week ($\Delta t$) is recorded. The results obtained are tabulated as shown in Tab. 4.1. Since the no. of events occurred during a particular interval only was noted at the end of the interval (or at the beginning of the next interval), the values of $s$ are entered between two values of $t$ as shown in col.(2). Cumulative events ($S$) or achievements at the end of the interval is given in col.(3). Col.(4) gives the no. of remaining events ($F$) at the end of each time interval.

*population at the beginning or close of the interval may also be taken.
<table>
<thead>
<tr>
<th>Time t</th>
<th>Number of Cumulative Problems solved or events s</th>
<th>Number of Successes/Successes</th>
<th>Problems remaining unsolved F</th>
<th>Success/Objectivity Progress Rate s</th>
<th>g</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.040</td>
<td>0.0408</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>960</td>
<td>0.051</td>
<td>0.0546</td>
<td>-</td>
<td>0.960</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>909</td>
<td>0.061</td>
<td>0.0637</td>
<td>-</td>
<td>0.909</td>
</tr>
<tr>
<td>3</td>
<td>61</td>
<td>848</td>
<td>0.070</td>
<td>0.0861</td>
<td>-</td>
<td>0.848</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>778</td>
<td>0.079</td>
<td>0.1070</td>
<td>-</td>
<td>0.778</td>
</tr>
<tr>
<td>5</td>
<td>79</td>
<td>699</td>
<td>0.091</td>
<td>0.1393</td>
<td>-</td>
<td>0.699</td>
</tr>
<tr>
<td>6</td>
<td>91</td>
<td>608</td>
<td>0.094</td>
<td>0.1676</td>
<td>-</td>
<td>0.608</td>
</tr>
<tr>
<td>7</td>
<td>94</td>
<td>514</td>
<td>0.095</td>
<td>0.2036</td>
<td>-</td>
<td>0.514</td>
</tr>
<tr>
<td>8</td>
<td>95</td>
<td>419</td>
<td>0.109</td>
<td>0.2990</td>
<td>-</td>
<td>0.419</td>
</tr>
<tr>
<td>9</td>
<td>109</td>
<td>310</td>
<td>0.081</td>
<td>0.3005</td>
<td>-</td>
<td>0.310</td>
</tr>
<tr>
<td>10</td>
<td>81</td>
<td>229</td>
<td>0.061</td>
<td>0.3073</td>
<td>-</td>
<td>0.229</td>
</tr>
<tr>
<td>11</td>
<td>61</td>
<td>168</td>
<td>0.053</td>
<td>0.3746</td>
<td>-</td>
<td>0.168</td>
</tr>
<tr>
<td>12</td>
<td>53</td>
<td>115</td>
<td>0.036</td>
<td>0.3711</td>
<td>-</td>
<td>0.115</td>
</tr>
<tr>
<td>13</td>
<td>36</td>
<td>79</td>
<td>0.031</td>
<td>0.4881</td>
<td>-</td>
<td>0.079</td>
</tr>
<tr>
<td>14</td>
<td>31</td>
<td>48</td>
<td>0.019</td>
<td>0.4935</td>
<td>-</td>
<td>0.048</td>
</tr>
<tr>
<td>15</td>
<td>19</td>
<td>29</td>
<td>0.011</td>
<td>0.4681</td>
<td>-</td>
<td>0.029</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>18</td>
<td>0.010</td>
<td>0.7692</td>
<td>-</td>
<td>0.018</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>8</td>
<td>0.008</td>
<td>2.0000</td>
<td>-</td>
<td>0.008</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>0</td>
<td>0.008</td>
<td>2.0000</td>
<td>-</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Tab. 4.1 - Progress Data and computation of success density, progress rate and objectivity.
Based on this data \( s(t) \) and \( g(t) \) are computed* and given in col.(5) and (6). In col.(7) we have remaining objective (O) which is defined as follows:

4.2.3 Remaining Objective (Target) or Objectivity (O)

This is the ratio of the remaining events (unsolved problems) at any given time \( t \) to the total initial population. In the beginning of the project objectivity factor is 1. As the project progresses, more and more problems are solved and decisions taken with the result this factor decreases rapidly to 0. Mathematically,

\[
O(t) = \frac{N - s(t)}{N} = \frac{n(t)}{N} \quad \text{........ (4.4)}
\]

Since \( s(t) \) is a density function, we can define success and objectivity cumulative distribution functions as follows-

\[
S(t) = \int_0^t s(\xi) \, d\xi \quad \text{........ (4.5)}
\]

\[
O(t) = 1 - S(t) = 1 - \int_0^t s(\xi) \, d\xi \quad \text{.... (4.6)}
\]

where \( \xi \) is a dummy variable of integration.

Cumulative success (Progressivity) and remaining work (Objectivity) are complementary functions. They are associated with the probability of success and failure respectively.

The data of Tab. 4.1 is plotted as Fig. 4.1.

*Generally, it is not possible to quantify a project in events, so milestones may be fixed and the time elapsed in achieving these milestones is recorded. From this data also \( s(t) \) and \( g(t) \) may be computed similarly.
4.3 Objectivity, Progress Rate and Success Density

Progress rate was defined as ratio of no. of events in a certain interval to the average population during the interval or alternatively,

\[
g(t) = \frac{\text{No. of remaining events at } t - \text{events at } (t+\Delta t)}{\Delta t \cdot \text{Average population during interval}}
\]

Dividing N & D by total initial population \( N \) -

\[
g(t) = \frac{[O(t) - O(t+\Delta t)]}{\Delta t \cdot \frac{[O(t) + O(t+\Delta t)]}{2}}
\]

\[
= 2x \frac{[O(t) - O(t+\Delta t)]}{\Delta t \cdot \frac{[O(t) + O(t+\Delta t)]}{2}}
\]

\[
\text{In the limit, when } \Delta t \to 0, O(t) \to O(t+\Delta t)
\]

\[
g(t) = \lim_{\Delta t \to 0} \frac{O(t) - O(t+\Delta t)}{\Delta t \cdot O(t)}
\]

\[
= \frac{1}{O(t)} \frac{dO(t)}{dt}
\]

or - \( g(t) = \frac{d[\log O(t)]}{dt} \)

Integrating:

\[
\int_{\xi}^{t} g(\xi)d\xi = \log O(t) + C
\]

\[
\log O(t) = - C - \int_{0}^{t} g(\xi)d\xi
\]

Taking exponential of both sides:

\[
O(t) = e^C e^{-\int_{0}^{t} g(\xi)d\xi}
\]
In the beginning of the project when $t=0$, Objectivity factor is $1$, hence $C=0$

$$O(t) = e^{-\int_0^t g(\xi) d\xi} \quad \cdots \cdots \cdots \ (4.9)$$

Similarly, success function can be expressed as

$$S(t) = \frac{\text{No. of remaining events at } t - \text{No. of remaining events at } t+\Delta t}{\Delta t \cdot \text{total population}}$$

$$\begin{align*}
&= \frac{1}{\Delta t} \left[ \frac{n(t)}{N} - \frac{n(t+\Delta t)}{N} \right] \\
&= \frac{O(t) - O(t+\Delta t)}{\Delta t} \\
&\text{In the limit-} \\
S(t) &= \lim_{\Delta t \to 0} \frac{O(t) - O(t+\Delta t)}{\Delta t} \\
&= \frac{dO(t)}{dt} \\
&= g(t) \cdot O(t) \quad \text{(substituting from 4.8)}
\end{align*}$$

$$S(t) = g(t) \cdot e^{-\int_0^t g(\xi) d\xi} \quad \cdots \cdots \cdots \ (4.10)$$

### 4.4 Success Models

We have observed that from progress data analysis we can obtain objectivity, progressivity, progress rate and other information. Obviously, behaviour characteristics exhibited by one class of projects differ from those exhibited by another class.
class of projects. In order to compare different behaviour characteristics and also to draw general conclusion from the behaviour pattern of similar projects, a mathematical model representing the progress characteristics becomes necessary.

4.4.1 Weibull Model

In general, the progress rate can be expressed as follows:

\[ g(t) = \alpha t^\beta \]  \hspace{1cm} \text{(where } \beta > -1) \quad \text{......... (4.11)}

The time integral of \( g(t) \) becomes:

\[ \int_0^t g(\xi)d\xi = \int_0^t \alpha \xi^\beta d\xi = \alpha \xi^{\beta+1}/(\beta+1) \quad \text{......... (4.12)} \]

The objectivity, progressivity and success density may be expressed as:

\[ O(t) = \exp \left[ -\alpha t^{\beta+1}/(\beta+1) \right] \]
\[ S(t) = 1 - \exp \left[ -\alpha t^\beta / (\beta+1) \right] \quad \text{......... (4.13)} \]
\[ s(t) = \alpha t^\beta \exp \left[ -\alpha t^\beta / (\beta+1) \right] \]

The progress rate function \( g(t) \) consists of two parameters \( \alpha \) and \( \beta \). If these are selected appropriately, a variety of situations arise (Shoo83). The drawback is that this is a two parameter model which means a greater difficulty in sketching the results and increased difficulty in estimating the parameters.

If \( \beta = 0 \), \( g(t) = \alpha \) \hspace{1cm} \text{Constant Progress Model}

If \( \beta = 1 \), \( g(t) = \alpha t \) \hspace{1cm} \text{Linearly Increasing Progress Model}

4.4.2 Rayleigh Model

This particular case is of utmost interest to us, because software engineers (personnel) learn to solve software
problems with an effectiveness which increases linearly during each cycle i.e. familiarity with the problem at hand leads to greater insight and success. This $g(t) = \alpha t$ is a particular linear case of generalised family of learning curves $g = \alpha t^b$. So the previous equations can now be written as:

$$
\int_0^t g(\xi) d\xi = \alpha t^2/2 - \alpha t/2
$$

$$
O(t) = e^{-\alpha t/2}
$$

$$
y \equiv S(t) = (1 - e^{-\alpha t/2})
$$

$$
y' = s'(t) = \alpha t e^{-\alpha t/2}
$$

$$
y'' = \alpha (1 - \alpha t) e^{-\alpha t/2}
$$

MTTS* = $\sqrt{T/2\alpha}$

4.5 Summary

In this chapter we have shown how we can model software project progress by Rayleigh curve (Fig. 4.2). We shall be using this model subsequently. Summarily, this curve has following mathematical characteristics:

*For a population of N problems with success times $t_1, t_2, ..., t_n$ mean time to success is given by-

$$
MTTS = \frac{1}{N} \sum_{i=1}^{N} t_i = \int_0^\infty 0(t) dt = \int_0^\infty e^{-\alpha t/2} dt
$$

$$
= \frac{\sqrt{T/2}}{2 \sqrt{\alpha/2}} = \sqrt{T/2\alpha}
$$
1. It is determined by two parameters i.e. scale factor($\bar{\Phi}$) (total area of the curve) and shape factor($\alpha$) (initial slope of the curve).

2. It starts with a value = 0 at $t = 0$, reaches a peak and decays exponentially.

3. The value of the function is maximum at $t = \sqrt{1/\alpha} = \tau$ called the 'development time'.

4. The slope of $y'$ i.e. $y'' = \alpha$ at $t = 0$

   $y'' = 0$ at $t = \tau$

$y''$ being increasingly negative after the peak is passed, reaches a negative maximum at $t = \sqrt{3/\alpha}$ (inflexion point) and then increases asymptotically toward zero.

The theoretical basis of this approach is that software development is basically a problem solving effort and decision making is the exhaustion process. The various development activities partition the problem space into subspaces corresponding to various phases in the life cycle. The following logical assumptions are made (Nord70):

1. The number of problems to be solved in developing a software product is finite but unknown.

2. Problem solving effort by software personnel creates an environment (SEE) and makes technological impact on the project.

3. Developmental activities like information gathering, thinking possible solutions, identifying alternatives all consume time. Any decision, due to such deliberations is an event and transforms one of the unsolved problems into a solved problem (success).
4. Occurrence of these events is independent and random and depends on the expertise of personnel.

5. Applied manpower effort requirement is proportional to the number of problems 'ripe' for a solution.
FIG. 4.1 - GRAPHICAL PLOTS FOR DATA OF TAB. 4.1
FIG. 4.2 - PROJECT PROGRESS MODELLED BY RAYLEIGH CURVE