THE SINGLE PROCESS LAW:
A STUDY IN NONLINEAR REGRESSION

by


This research was supported in part by the Revolving Research Fund of the Institute of Statistics, of the Consolidated University of North Carolina through a grant from the General Education Board for research in basic statistical methodology.

Institute of Statistics
Mimeo Series No. 235
July, 1959
ABSTRACT


Some important differences between linear and nonlinear regression models are described using a geometric device, due to R. A. Fisher, of representing all possible samples as points in an n-dimensional observation space. For illustrative purposes the two-dimensional case is considered in some detail. Differences between the classes of models noted are (1) orthogonality of sums of squares (represented as square distances), (2) distribution of sums of squares, (3) existence of sufficient statistics. Essentially nonlinear models (i.e., those not capable of being linearized) are characterized by non-orthogonality, unusual distribution of sums of squares, and unavailability of sufficient statistics. It is pointed out that essentially nonlinear models are most often encountered in science and should therefore be given due consideration. Attempts to study the properties of these nonlinear models are hampered by the fact that although there is only one kind of linear model there is an infinite variety of nonlinear ones and hence there is a need for the definition of classes of nonlinear models. The most interesting ones will have simply interpretable differential equations since the power of description of these latter is greater. An interesting family is given by the expression \( y = \alpha + \beta (x - \gamma \delta)^5 + \epsilon \) which arises from the differential equation \( \frac{d \gamma}{d \delta} = \delta (\gamma - \lambda)/(\delta - \gamma \delta). \) Since by taking \( \delta \) to be various values we obtain special processes such as the "exponential process", "linear process", "parabolic process", "inverse-square process", etc.
The name single process law is suggested for the generic model. After a discussion of analytic properties of the special cases certain statistical problems concerned with discriminating the special cases are taken up. Computational schemata for obtaining maximum likelihood estimates and asymptotic confidence limits are set out. In order to illustrate the model-building potential of the single process law, as well as to illustrate the estimation procedure, an original theory of middle and long-distance track records is sketched and numerically fitted to the world records for 1957. It is concluded that "the excess of average speed over the asymptotic value is inversely proportional to the $3/7$ power of the time of the race."

It is possible that the asymptotic confidence limits will not be satisfactory for small samples.

Three exact tests of significance and their associated confidence limits are described. It is thought that one of these, the parabolic test, has not been recorded previously and that this test may prove to be useful. Application to the track record example has been made. Review of some important results in linear path regression is made in order to conveniently extend the results to the nonlinear situation in which the single process law describes each connection in the causal network. Computational methods are described which are simple generalizations of the univariate techniques. The surprising fact is brought out that problems of identification are the same in the linear case and in the particular nonlinear case studied and for this reason certain estimation algorithms are also valid for the nonlinear system. A sampling investigation has been carried out in the univariate situation in order to explore the properties of the maximum likelihood estimates of the indicial or
exponential parameter $\delta$. In particular, comparison of the finite variances and the asymptotic ones have been made so as to judge the adequacy of the asymptotic methods. It appears that, with a 2% or smaller coefficient of variation at the center of the curve and with at least ten equidistantly spaced observations, we may trust the large sample techniques. Larger rates of error will not permit resolution of the special forms of the single process law and hence are of little practical interest to the model-builder. Brief presentations of a cubic approximation valid for large $\gamma\delta$ and a simplified estimator based upon a finite difference approximation for the derivative are given in the appendices.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>3.0 LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>4.0 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>4.1 Objectives of the Study</td>
<td>1</td>
</tr>
<tr>
<td>4.2 Some Remarks on the Differences between Linear and Nonlinear Models</td>
<td>2</td>
</tr>
<tr>
<td>5.0 THE SINGLE PROCESS LAW</td>
<td>9</td>
</tr>
<tr>
<td>5.1 Need for a General Approach</td>
<td>9</td>
</tr>
<tr>
<td>5.2 The Single Process Law</td>
<td>9</td>
</tr>
<tr>
<td>5.3 Discussion of the Curves</td>
<td>11</td>
</tr>
<tr>
<td>5.4 A Geometric Criterion of Curvature</td>
<td>13</td>
</tr>
<tr>
<td>5.5 A Tabular Summary</td>
<td>16</td>
</tr>
<tr>
<td>6.0 ESTIMATION BY MAXIMIZING THE LIKELIHOOD</td>
<td>18</td>
</tr>
<tr>
<td>6.1 Introductory Remarks</td>
<td>18</td>
</tr>
<tr>
<td>6.2 Maximizing the Likelihood</td>
<td>18</td>
</tr>
<tr>
<td>6.3 The Method of First Derivatives</td>
<td>19</td>
</tr>
<tr>
<td>6.4 An Iterative Scheme for the Single Process Law</td>
<td>21</td>
</tr>
<tr>
<td>6.5 Restrictions on the Iteration</td>
<td>25</td>
</tr>
<tr>
<td>6.6 A Numerical Example</td>
<td>26</td>
</tr>
<tr>
<td>7.0 SUMMARY OF ASYMPTOTIC RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>7.1 Asymptotic Variances</td>
<td>28</td>
</tr>
<tr>
<td>7.2 Asymptotic Confidence Limits</td>
<td>30</td>
</tr>
<tr>
<td>8.0 THE THEORY OF MIDDLE AND LONG DISTANCE TRACK RECORDS:</td>
<td></td>
</tr>
<tr>
<td>AN EXAMPLE OF THE SINGLE PROCESS LAW</td>
<td></td>
</tr>
<tr>
<td>8.1 Uses of the Single Process Law</td>
<td>31</td>
</tr>
<tr>
<td>8.2 A Theory of Track Records</td>
<td>31</td>
</tr>
<tr>
<td>8.3 Fitting the Theory to World Record Data</td>
<td>34</td>
</tr>
<tr>
<td>9.0 SOME EXACT TESTS AND CONFIDENCE LIMITS</td>
<td>37</td>
</tr>
<tr>
<td>9.1 Introduction</td>
<td>37</td>
</tr>
<tr>
<td>9.2 A Test of Assigned Nonlinear Parameters in the Case of Replication</td>
<td>37</td>
</tr>
<tr>
<td>9.3 A Test Based upon the Taylor's Series Expansion</td>
<td>39</td>
</tr>
<tr>
<td>9.4 The Parabolic Test for Nonlinearity</td>
<td>41</td>
</tr>
<tr>
<td>9.5 Confidence Limits for the Exponential Parameter in the Track Record Example Based upon the Parabolic Test.</td>
<td>42</td>
</tr>
</tbody>
</table>
1.0 TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0 LINEAR PATH REGRESSION</td>
<td>46</td>
</tr>
<tr>
<td>10.1 The Origins of Path Regression Analysis</td>
<td>46</td>
</tr>
<tr>
<td>10.2 The Structural Equations</td>
<td>46</td>
</tr>
<tr>
<td>10.3 The Reduced Structural Equations</td>
<td>48</td>
</tr>
<tr>
<td>10.4 The Equations of Identification</td>
<td>49</td>
</tr>
<tr>
<td>10.5 Estimation in the Case of Under- and Just-identified Schemes</td>
<td>50</td>
</tr>
<tr>
<td>10.6 Classification of Path Schemata</td>
<td>52</td>
</tr>
<tr>
<td>10.7 Some Examples of Particular Path Schemata</td>
<td>54</td>
</tr>
<tr>
<td>11.0 NONLINEAR PATH REGRESSION</td>
<td>57</td>
</tr>
<tr>
<td>11.1 The Structural Equations for Pathways Following the Single Process Law</td>
<td>57</td>
</tr>
<tr>
<td>11.2 Identification</td>
<td>59</td>
</tr>
<tr>
<td>11.3 Some Examples</td>
<td>61</td>
</tr>
<tr>
<td>11.4 The Gaussian Iterant for Maximum Likelihood Estimators</td>
<td>63</td>
</tr>
<tr>
<td>12.0 A SAMPLING INVESTIGATION</td>
<td>66</td>
</tr>
<tr>
<td>12.1 Objectives</td>
<td>66</td>
</tr>
<tr>
<td>12.2 The Samples</td>
<td>66</td>
</tr>
<tr>
<td>12.3 Estimation when ( \mu ) is Known</td>
<td>67</td>
</tr>
<tr>
<td>12.4 Estimation when ( \mu ) is Not Known</td>
<td>69</td>
</tr>
<tr>
<td>12.5 Discussion and Summary</td>
<td>71</td>
</tr>
<tr>
<td>13.0 SUMMARY AND CONCLUSIONS</td>
<td>73</td>
</tr>
<tr>
<td>14.0 LIST OF REFERENCES</td>
<td>75</td>
</tr>
<tr>
<td>15.0 APPENDIX A. A POLYNOMIAL APPROXIMATION</td>
<td>77</td>
</tr>
<tr>
<td>16.0 APPENDIX B. A SIMPLIFIED ESTIMATOR</td>
<td>79</td>
</tr>
</tbody>
</table>
2.0 LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Some special forms of the single process law</td>
<td>11</td>
</tr>
<tr>
<td>5.2</td>
<td>Constants for special extinction curves</td>
<td>17</td>
</tr>
<tr>
<td>8.1</td>
<td>Some world track records for 1957</td>
<td>34</td>
</tr>
<tr>
<td>9.1</td>
<td>Values of t for various hypotheses</td>
<td>43</td>
</tr>
<tr>
<td>10.1</td>
<td>Classification of path schemata</td>
<td>53</td>
</tr>
<tr>
<td>12.1</td>
<td>Thirteen curves chosen for investigation showing</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>computed values of eta</td>
<td></td>
</tr>
<tr>
<td>12.2</td>
<td>Estimated biases and variances of the maximum likelihood estimate for delta</td>
<td>70</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Diagram showing analysis of distances for a linear model</td>
<td>5</td>
</tr>
<tr>
<td>4.2</td>
<td>Diagram showing analysis of distances for a nonlinear model</td>
<td>7</td>
</tr>
<tr>
<td>5.1</td>
<td>Special extinction curves</td>
<td>12</td>
</tr>
<tr>
<td>5.2</td>
<td>Relationship between the $\lambda$ criterion and the exponential parameter $\delta$</td>
<td>15</td>
</tr>
<tr>
<td>9.1</td>
<td>Determination of confidence limits for the track record example using the parabolic test method</td>
<td>44</td>
</tr>
</tbody>
</table>
4.0 INTRODUCTION

4.1 Objectives of the Study

This monograph reports the results of a study dealing with a class of nonlinear regression models called herein the single process law or the log-exponential model. There have been several specific objectives of the study. First of all, we wished to explore the elementary analytic geometry of the curves. How do they behave? How can they be characterized? Section 5.0 is devoted to considerations of this sort. A second objective was to describe and record methods and formulae for obtaining the maximum likelihood estimates and their asymptotic confidence limits. Especial regard was to be given to the exponential parameter which serves as an index of the particular class member, that is, to the parameter which might be used for model-building. Such methods and formulae are found in Sections 6.0 and 7.0. A third objective was to investigate by Monte Carlo procedures the properties of the maximum likelihood estimates for a small sample. In particular we wished to find out how bad the asymptotic variance might be for purposes of describing confidence limits about the exponential parameter. Can we use asymptotic results for the purpose of empirical model selection? This part of the study is described in Section 12.0. A fourth and final original objective was to extend the method of path regression analysis to cover the case in which each casual pathway behaves according to the single process law. After a brief review of the linear results in Section 10.0 such an extension is given in Section 11.0.
In addition to these four original objectives several other points concerning the single process law were investigated. An original theory of middle and long distance track records has been developed in order to illustrate the model building potentiality of the single process law. This theory is outlined in Section 8.0. Some exact tests of significance and confidence regions for the single process law are presented in Section 9.0. A cubic approximation to the single process law and a simplified estimator based upon finite differences are briefly described in Appendices A and B, respectively.

4.2 Some Remarks on the Differences between Linear and Nonlinear Models

Regression models may be conveniently grouped into three classes:

(1) linear, (2) pseudo-nonlinear, and (3) essentially nonlinear. In the first case the unknown parameters each occur to the first power and the errors are simply additive. In the second case these conditions may not hold but linearity can be achieved either by reparameterization or by transformation of the variables. These changes must be such as to either leave an additive error still additive or to produce additivity when the error was not originally so. There remains a large class of models which are neither linear nor can be made linear by any convenient device. These constitute the third class and these models may be termed essentially nonlinear in order to connote the fundamental nature of the nonlinearity. Unhappily, most physical and biological processes are best described by these essentially nonlinear models. The reasons why this fact is unfortunate may be seen by resort to an explanatory device due to R. A. Fisher (1939).
Let us consider, for example, two observations $y_1$ and $y_2$ obeying the simple linear model

$$y_i = \beta x_i + \varepsilon_i$$  \hspace{1cm} (4.1)

$i = 1, 2$ and where the $\varepsilon_i$ are normally and independently distributed with expectation zero and constant variance $\sigma^2$. The two-dimensional space made up of all possible pairs of observations may be termed the observation space. The results of any particular two-valued sample produces a pair $(y_1, y_2)$ which may be represented as a point, the observation point in the observation space. Now in the linear example being considered we also have the expected values

$$E(y_1) = \beta x_1$$
$$E(y_2) = \beta x_2$$

and eliminating $\beta$ between them we get

$$E(y_2) = \frac{x_2}{x_1} E(y_1).$$  \hspace{1cm} (4.2)

Thus, the expected locus of all observation points, no matter what the true value of $\beta$ may be, is given by the straight line (4.2). For any estimate of $\beta$, say $\hat{\beta}$, we have estimated observations

$$\hat{y}_1 = \hat{\beta} x_1$$
$$\hat{y}_2 = \hat{\beta} x_2$$

and $(\hat{y}_1, \hat{y}_2)$ falls on the line (4.2). The maximum likelihood estimate of $\beta$ is equivalent to the least squares estimate and is found by minimizing the square distance

$$\text{SSE} = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2.$$
This minimum sum of squares is clearly obtained by taking the perpendicular to (4.2) passing through \((y_1, y_2)\). See Figure 4.1. The estimated point \((\hat{y}_1, \hat{y}_2)\) is then given by the intersection of (4.2) and the perpendicular. Then \(\hat{\beta} = \frac{\hat{y}_1}{x_1} = \frac{\hat{y}_2}{x_2}\).

Let us now consider some null hypothesis, for example, \(\beta = 0\). We thus get the hypothetical observations

\[
\hat{y}_{10} = 0 \\
\hat{y}_{20} = 0
\]

The square distance between the estimated point and the hypothesized point is obviously

\[
SSR = (\hat{y}_1 - \hat{y}_{10})^2 + (\hat{y}_2 - \hat{y}_{20})^2
\]

and the square distance between the observation point and the hypothesized point is

\[
SST = (y_1 - \hat{y}_{10})^2 + (y_2 - \hat{y}_{20})^2
\]

The three distances \(\sqrt{SSE}, \sqrt{SSR},\) and \(\sqrt{SST}\) constitute the sides of a right triangle with \(\sqrt{SST}\) as hypotenuse. The Pythagorean theorem then holds giving

\[
SST = SSR + SSE
\]

We see that the sum of squares for the distance between observation and hypothesis is decomposed into two orthogonal components, one corresponding to the distance between estimate and hypothesis and the other between estimate and observation. It is also true that both SSR and SSE are sums of squared normal deviates when the null hypothesis is true and hence the usual distribution theory is applicable. The linearity of (4.2) further
Fig. 4.1 Diagram showing analysis of distances for a linear model.
accounts for the fact that our estimates are sufficient since under
these circumstances the information concerning $\beta$ depends only upon SSE.

In the case of an essentially nonlinear model the situation is
entirely different. The model

\[ y_i = \rho i + e_i, \quad i = 1, 2 \]  

and $e_i$ as before, is illustrated in Figure 4.2. We have

\[ E(y_1) = \rho_1 \]
\[ E(y_2) = \rho_2 \]

and

\[ E(y_2) = E(y_1)^{x_2/x_1} \]  \hspace{1cm} (4.3)

If $x_2/x_1 = 2$ the expected locus of observation points is a parabola.

Now we consider the null hypothesis $\rho = 0$. Then

\[ ^\wedge \ \hat{y}_{10} = 0 \]  
\[ ^\wedge \ \hat{y}_{20} = 0 \]

We note (1) SST \neq SSR + SSE, (2) SSR and SSE are not sums of squared
normal deviates under the null hypothesis and hence the usual distribu-
tion theory would not apply even if the lack of orthogonality were
allowed for, and (3) a sufficient estimate for $\rho$ does not exist since
the precision of any estimate must depend not only upon SSE but also
upon the curvature of (4.3). Knowledge of this curvature would have
to be furnished by appropriate ancillary statistics in the absence of
direct knowledge of $\rho$. 
Fig. 4.2 Diagram showing analysis of distances for a nonlinear model
The remarks in this section apply generally to all linear and nonlinear models and to samples of any size. We see that nonlinear models differ in a quite fundamental way from linear models. It is indeed unfortunate that the majority of models encountered in well-developed areas of research belong to the class that we term essentially nonlinear.
5.0 THE SINGLE PROCESS LAW

5.1 Need for a General Approach

As suggested in Section 4.0 there is a complete and satisfying theory of statistical inference in the case of linear models and, by transformation, in the case of pseudo-nonlinear models. On the contrary, no such complete theory exists for essentially nonlinear models. To a certain extent every nonlinear model is a unique problem and must be studied de novo when inferences are to be drawn regarding a particular algebraic form. It is apparent, therefore, that there is considerable advantage to be gained in studying models of some generality although the degree of generality is limited by the great amount of labor involved in investigating complex systems with the primitive methods now available. Systems of models satisfying simple differential equations are preferable to other systems as objects of study since their power of physical description is greater.

5.2 The Single Process Law

The above considerations motivated the present investigation. Thus, a system of models was sought which

1) was somewhat more general than that usually applied in particular situations;

2) was not too complex nor contain too many parameters for practical investigation;

3) and was generated by a simple meaningful differential equation.

After inspection of a number of candidates the following model was
selected as being promising:

\[ y = \alpha + \beta(x-y) + \varepsilon \]  

(5.1)

where \( y \) and \( x \) are observables; \( \mu, \beta, \gamma, \) and \( \delta \) are constant parameters and \( \varepsilon \) is a random element - normally distributed with expectation zero and constant variance \( \sigma^2 \). Thus, this is a classical regression model of the essentially nonlinear variety. This model results from integrating the differential equation

\[ \frac{d\gamma}{d\varepsilon} = \delta \frac{\gamma - \alpha}{\varepsilon - \gamma \varepsilon} \]  

(5.2)

and replacing \( \gamma \) by \( y-\varepsilon \) and \( \varepsilon \) by \( x \). The nature of the process is determined by the proportionality constant \( \delta \). By inspection it may be seen that the process is linear if \( \delta = 1 \), is quadratic if \( \delta = 2 \), is hyperbolic if \( \delta = -1 \), and follows the inverse square law if \( \delta = -2 \). In addition the process tends toward the exponential limit

\[ y = \alpha + \beta e^{-x/\gamma} + \varepsilon \]  

(5.3)

as \( \delta \) tends toward infinity. This series of models determined by a single parameter is best depicted in terms of a parameter \( \delta^{-1} \) which is the inverse of \( \delta \). See Table 5.1.
Table 5.1. Some special forms of the single process law

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>Model</th>
<th>Algebraic Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>Rectangular hyperbola</td>
<td>$y = \alpha + \frac{\beta}{x^{\gamma}} + \epsilon$</td>
</tr>
<tr>
<td>-0.5</td>
<td>Inverse square law</td>
<td>$y = \alpha + \frac{\beta}{(x+2\gamma)^2} + \epsilon$</td>
</tr>
<tr>
<td>0</td>
<td>Exponential</td>
<td>$y = \alpha + \beta e^{-x/\gamma} + \epsilon$</td>
</tr>
<tr>
<td>0.5</td>
<td>Quadratic parabola</td>
<td>$y = \alpha + \beta (x-2\gamma)^2 + \epsilon$</td>
</tr>
<tr>
<td>1.0</td>
<td>Straight line</td>
<td>$y = \alpha + \beta (x-\gamma) + \epsilon$</td>
</tr>
</tbody>
</table>

The more general case $\delta > 0$ is a generalization of the regression analogue of the well known "allometric equation". For convenience the general model (5.1) will be referred to henceforth as the single process law. This choice of terms is not meant to imply that all single physical processes can be subsumed so simply by equation (5.1); however, it is felt that some generality should be suggested in the chosen name. An alternative name could be the log-exponential model since (5.1) can be written in the form

$$y = \alpha + \beta e^{\delta \log(x-\gamma)} + \epsilon$$

5.3 Discussion of the Curves

There is a smooth transition from one special case of the single process law to another. This fact is suggested by Figure 5.1. Observe that the exponential law which has a single asymptote is the transition curve between the "hyperbolas" with two asymptotes and the "parabolas"
with no asymptote. "Curvature" progressively increases as one proceeds from the straight line through the parabolas, the exponential, and upward through the various hyperbolas. The hyperbolas and the exponential are monotonically increasing or monotonically decreasing functions. The parabolas, of course pass through a maximum or minimum; however, greatest interest is centered in either the monotonically increasing segment of the curve as in "growth" processes or in the monotonically decreasing segment as in "decay" or "extinction" processes. The present discussion is limited to those curves for which \( \frac{d\gamma}{d\xi} < 0 \) and \( \frac{d^2\gamma}{d\xi^2} \geq 0 \) since the other cases can be generated from those studied by linear transformations of the axes. That is, only the extinction or decay form of the family is considered explicitly, no loss of generality being engendered thereby. For convenience we will take \( x - \gamma \xi \) to be always greater than or equal to zero.

5.4 A Geometric Criterion of Curvature

We have seen that the parameter \( s \) measures the tendency to curve. It would be convenient to be able to approximately pick out the proper \( s \) (and hence the proper member of the family) by some form of graphical analysis. Unfortunately, the parameter \( s \) is singularly unsuited for this purpose. However, a simple transformation exists which provides a convenient index of curvature and which has a simple graphical interpretation. We have the functional equation

\[
\gamma = \alpha + \beta (\xi - \gamma \xi)^5
\]
corresponding to (5.1). When $\xi = 0$ the "initial" value of $\eta$ is given by

$$\eta_0 = \alpha + \beta (-\gamma_6)^0.$$ 

The minimum value of $\eta$ is given by

$$\eta_{\min} = \alpha.$$ 

The value of $\xi$ corresponding to $(\eta_0 - \alpha)/2$ may be termed the half-life and is given by

$$\xi_{50} = \gamma_6 (1 - 2^{-1/6}).$$ 

Similarly, the value of $\xi$ corresponding to $(\eta_0 - \alpha)/4$ may be called the three-quarters-life and is given by

$$\xi_{75} = \gamma_6 (1 - 4^{-1/6}).$$ 

Let us now define a new criterion $\lambda$ for the curves found by dividing the three-quarters-life by the half-life. Thus

$$\lambda = \frac{\xi_{75}}{\xi_{50}} = \frac{1 - 4^{-1/6}}{1 - 2^{-1/6}}.$$ (5.4)

Inverting, we find

$$x^{-1} = -\frac{\log_{10}(\lambda - 1)}{\log_{10}^2} = -\log_2(\lambda - 1).$$ (5.5)

The relationships (5.4) and (5.5) are charted for the various curves in Figure 5.2.
Fig. 5.2: Relationship between the λ criterion and the exponential parameter $\delta$.
We now can determine the particular curve (8) from a plot of its course. A more general criterion is readily found to be

\[ \lambda(p_1, p_2) = \frac{\phi(p_1)}{\phi(p_2)} = \frac{1 - q_1^{1/6}}{1 - q_2^{1/6}} \]

where \( \phi(p_1) \) is the value of \( \phi \) corresponding to any fractile-life and \( \phi(p_2) \) corresponds to any other fractile-life; and \( p_1 + q_1 = p_2 + q_2 = 1 \).

5.5 A Tabular Summary

We will terminate this presentation by giving a table (Table 5.2) in which a number of important characteristics are summarized for special extinction curves. It is seen that the parameter \( \gamma \) appears as the "point of extinction" for the straight line, the "three-quarters life" for the parabola, and the "half-life" for the rectangular hyperbola. The "integration constant" \( \beta \) doubles as the "initial velocity" for the straight line and as the "initial level" for the exponential.
### Table 5.2. Constants for special extinction curves

<table>
<thead>
<tr>
<th>Curve of Curve</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$\gamma$</th>
<th>Asymptote</th>
<th>Initial level</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
<th>Initial Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight Line</td>
<td>1</td>
<td>$&lt;0$</td>
<td>$&gt;0$</td>
<td>$-\beta\gamma$</td>
<td>$\gamma/4$</td>
<td>$\gamma/2$</td>
<td>$2\gamma$</td>
<td>$\gamma$</td>
<td>$1.5$</td>
<td>$0$</td>
<td>$&lt;0$</td>
<td>$&gt;0$</td>
<td>$-D\gamma$</td>
<td>$2\gamma$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
</tr>
<tr>
<td>Parabola</td>
<td>2</td>
<td>$&gt;0$</td>
<td>$&gt;0$</td>
<td>$-2\gamma$</td>
<td>$\frac{b}{b}\gamma$</td>
<td>$(\frac{1}{2}\sqrt{3}-2)\gamma$</td>
<td>$(\sqrt{2}-2)\gamma$</td>
<td>$2\gamma$</td>
<td>$\gamma$</td>
<td>$3.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
<td>$3.8284$</td>
<td>$-\frac{1}{3}V$</td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>$\infty$</td>
<td>$&gt;0$</td>
<td>$&gt;0$</td>
<td>$-\gamma$</td>
<td>$\frac{b}{b}\gamma$</td>
<td>$\gamma/3$</td>
<td>$\gamma$</td>
<td>$3\gamma$</td>
<td>$\gamma$</td>
<td>$3.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
<td>$3.8284$</td>
<td>$-\frac{1}{3}V$</td>
<td></td>
</tr>
<tr>
<td>Inverse Square Law</td>
<td>$-2$</td>
<td>$&gt;0$</td>
<td>$&gt;0$</td>
<td>$-2\gamma$</td>
<td>$\frac{b}{b}\gamma$</td>
<td>$(\frac{5}{3}-2)\gamma$</td>
<td>$(\sqrt{2}-2)\gamma$</td>
<td>$2\gamma$</td>
<td>$\gamma$</td>
<td>$3.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
<td>$3.8284$</td>
<td>$-\frac{1}{3}V$</td>
<td></td>
</tr>
<tr>
<td>Rectangular Hyperbola</td>
<td>$-1$</td>
<td>$&gt;0$</td>
<td>$&gt;0$</td>
<td>$-\gamma$</td>
<td>$\frac{b}{b}\gamma$</td>
<td>$\gamma/3$</td>
<td>$\gamma$</td>
<td>$3\gamma$</td>
<td>$\gamma$</td>
<td>$3.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
<td>$3.8284$</td>
<td>$-\frac{1}{3}V$</td>
<td></td>
</tr>
<tr>
<td>Inverse Square Law</td>
<td>$-2/3$</td>
<td>$&gt;0$</td>
<td>$&gt;0$</td>
<td>$-2\gamma$</td>
<td>$\frac{b}{b}\gamma$</td>
<td>$(\frac{5}{3}-2)\gamma$</td>
<td>$(\sqrt{2}-2)\gamma$</td>
<td>$2\gamma$</td>
<td>$\gamma$</td>
<td>$3.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
<td>$3.8284$</td>
<td>$-\frac{1}{3}V$</td>
<td></td>
</tr>
<tr>
<td>Root Law</td>
<td>$-1/2$</td>
<td>$&gt;0$</td>
<td>$&gt;0$</td>
<td>$-\gamma$</td>
<td>$\frac{b}{b}\gamma$</td>
<td>$\gamma/3$</td>
<td>$\gamma$</td>
<td>$3\gamma$</td>
<td>$\gamma$</td>
<td>$3.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.0$</td>
<td>$-\frac{1}{2}a$</td>
<td>$2.1414$</td>
<td>$-\frac{1}{3}V$</td>
<td>$3.8284$</td>
<td>$-\frac{1}{3}V$</td>
<td></td>
</tr>
</tbody>
</table>

---

*a The curves all have the following characteristics in common:
1. Finite, positive initial value of $\gamma(\gamma = 0)$
2. $d\gamma/d\gamma > 0$ for all values of $\gamma$
3. $\gamma < 0$ for all values of $\gamma$
4. $\min \gamma = \epsilon$ except for the straight line values for which we will ignore beyond the point $\gamma = \gamma, \gamma = \delta$.

---

$b \gamma = \gamma(1 - \frac{1}{\sqrt{6}}), \gamma + q = 1$.

c $\gamma(\infty) = 6\gamma, \gamma > 0$.

d $\gamma = \frac{(2/\sqrt{3})^2 - 1}{(2/\sqrt{3}) - 2/\sqrt{6}), \frac{b}{b} = \frac{\log 2}{\log(\lambda - 1)}$ for $-\infty < b \gamma < +\infty, \gamma > 1.$
6.0 ESTIMATION BY MAXIMIZING THE LIKELIHOOD

6.1 Introductory Remarks

In practice, one may be confronted with a situation like this: data are available which are believed to follow the single process law (5.1); however, the parametric value of the exponent (i.e., \( \delta \)) governing the nature of the process is unknown. Thus, \( \delta \) as well as the other parameters must be estimated from the data. Alternatively, \( \delta \) may be known so that estimation is confined to \( \lambda \), \( \beta \), and \( \gamma \). A further common special case is the one in which the limit \( \lambda \) is known; in particular, it may often be zero. In other cases the product \( \gamma \beta \) may be known. This chapter is devoted to presentation of a procedure for finding the maximum likelihood estimates of the various combinations of parameters which may be unknown in practice. A single computational method is found for all cases.

6.2 Maximizing the Likelihood

For a series of \( N \) observations the likelihood function is given by

\[
e^L = \frac{1}{(2\pi\delta^2)^{N/2}} \exp \left\{ -\frac{1}{2\delta^2} \sum \left[ y - \lambda - \beta(x-\gamma \delta) \right]^2 \right\} \tag{6.1}
\]

and thus the "log likelihood" is

\[
L = -\frac{N}{2} \log(2\pi\delta^2) - \frac{1}{2\delta^2} \sum \left[ y - \lambda - \beta(x-\gamma \delta) \right]^2 \tag{6.2}
\]

where summation is over the \( N \) observations in the sample. As is commonly pointed out the likelihood has a maximum at the same point in parametric
space as does the log likelihood so that it is sufficient to maximize the log likelihood. In the usual way we locate the maximum by differentiating $L$ with respect to each unknown parameter in turn and setting the resulting expressions equal to zero. The equations obtained by equating the partial derivatives to zero are nonlinear and usually cannot be explicitly solved for the maximizing values of the parameters, the "maximum likelihood estimates." Several devices have been described (for example, see Will, 1936; Whitaker and Robinson, 1944; Deming, 1943) for finding the maximizing values. In general these devices either entail some form of interpolation in the parametric space or else some form of successive approach to the maximum point using derivatives or differences. The origins of the former approach date back to antiquity whereas the latter approach was first employed by Newton. Gauss appears to have been the first to employ the second approach for the purpose of maximizing the likelihood. In the case of regression R. A. Fisher's (1925) important "method of scoring" is equivalent to Gauss' procedure. Fisher's method is rich with meaning in terms of the general theory of estimation but usually involves the tedious algebra of determining expectations of certain variates. Gauss' method, generally speaking, is simpler to apply but is to be regarded merely as a computational device. Of course, once the equivalence of the two methods is recognized in a special case Gauss' process may be interpreted in terms of the Fisherian Information concept.

6.3 The Method of First Derivatives

There are many ways to describe Gauss' approach to the maximization problem. The following has some merits:
Let us suppose that we have the regression model

\[ y = f(\beta, x) + \varepsilon \quad (6.3) \]

where \( \beta \) is a vector of unknown parameters and \( \varepsilon \) is the random component supposed to be normally and independently distributed with expectation zero and unknown variance \( \sigma^2 \). If maximum likelihood estimates \( \hat{\beta} \) are substituted for the parameters \( \beta \) we have

\[ y = f(\hat{\beta}, x) + \varepsilon \quad (6.4) \]

where \( \varepsilon \) is an "estimate" of \( \varepsilon \). Then the estimate of the observation \( y \) is given by

\[ \hat{y} = f(\hat{\beta}, x) \quad (6.5) \]

By the rule for the total derivative we get

\[ d\hat{y} = \hat{y}_1 db_1 + \hat{y}_2 db_2 + \ldots \quad (6.6) \]

where \( \hat{y}_1 = \partial \hat{y}/\partial b_1 \). Then to a first approximation

\[ d\hat{y} = \hat{y}_{10} db_1 + \hat{y}_{20} db_2 + \ldots \quad (6.7) \]

where \( \Delta \hat{y} = y - \hat{y}_0 \), \( \Delta b_1 = b_1 - b_{10} \), \( \Delta b_2 = b_2 - b_{20} \), \ldots.

The zero subscript indicates that trial values for the undetermined estimates have been interjected. Thus, we get

\[ \hat{y} = \hat{y}_0 + \Delta b_1 \hat{y}_{10} + \Delta b_2 \hat{y}_{20} + \ldots \quad (6.8) \]

and

\[ y = \hat{y}_0 + \Delta b_1 \hat{y}_{10} + \Delta b_2 \hat{y}_{20} + \ldots + \varepsilon \quad (6.9) \]

Now (6.9) is an ordinary linear regression equation and estimates may be found in the usual way. From the linear regression estimates may be
found improved values for the original nonlinear regression estimates. Thus, the process is iterated finding successive improvements. It may be seen from (6.9) that trial values for estimates which are linear in (6.4) are not needed to prime the iterative process. This point is illustrated in the next section.

6.4 An Iterative Scheme for the Single Process Law

The single process law is of the form (6.3) and hence we may write down from inspection the linearized form analogous to (6.9). Before doing this we note that the linearization is not unique. There is nothing in the derivation of (6.9) which specifies the definition of the parameters. Thus, we could reparameterize (5.1) as follows: Let \( \mu = \gamma \delta \), then

\[
y = \alpha + \beta (x - \mu)^5 + \varepsilon \tag{6.10}
\]

and we could use (6.9) to specify an iterative scheme. Again, we could let \( \rho = 1/5 \) so that

\[
y = \alpha + \beta (x - \mu)^{1/\rho} + \varepsilon \tag{6.11}
\]

Obviously, there are an infinity of possible forms and for each (6.9) would indicate a particular iterative scheme. Of course, the final estimates are invariant regardless of the particular definition of the parameters used, so long as there is convergence. However, the rate of convergence, the tendency to diverge, and the amplitude of the oscillations may be rather drastically affected by the particular form of the parameters chosen. Some preliminary investigation suggests that use of the parameters \( \gamma \) and \( \rho \) may be better than some other forms, especially in the neighborhood of the rectangular hyperbola. No definitive evidence
for this statement is offered but this form will be used in the present presentation and has been used in the sampling study described in Section 12.0. The form of the single process law embodying \( \gamma \) and \( \rho \) is

\[
y = \alpha + \beta \left( x - \frac{\alpha}{\rho} \right)^{1/\rho} + \epsilon. \tag{6.12}
\]

Thus,

\[
\hat{y} = a_0 + b_0 \left( x - \frac{\alpha}{\rho} \right)^{1/\rho}. \tag{6.13}
\]

Then, we have by differentiating with respect to \( a_0 \), \( b_0 \), \( c_0 \), and \( r_0 \)

\[
\hat{y}_{10} = 1
\]

\[
\hat{y}_{20} = \left( x - \frac{\alpha}{\rho} \right)^{1/r_0} \tag{6.14}
\]

\[
\hat{y}_{30} = -\frac{b_0}{r_0^2} \left( x - \frac{\alpha}{\rho} \right)^{1/r_0 - 1}
\]

\[
\hat{y}_{40} = b_0 \left\{ \frac{c_0}{r_0} \left( x - \frac{\alpha}{\rho} \right)^{1/r_0 - 1} - \frac{1}{r_0^2} \left( x - \frac{\alpha}{\rho} \right)^{1/r_0} \ln \left( x - \frac{\alpha}{\rho} \right) \right\}.
\]

Now, let

\[
z_1 = \left( x - \frac{\alpha}{r_0} \right)^{1/r_0} = (x - m_0)^{d_0}
\]

\[
z_2 = \left( x - \frac{\alpha}{r_0} \right)^{1/r_0 - 1} = (x - m_0)^{d_0 - 1} \tag{6.15}
\]

\[
z_3 = \left( x - \frac{\alpha}{r_0} \right)^{1/r_0} \ln \left( x - \frac{\alpha}{r_0} \right) = (x - m_0)^{d_0} \ln (x - m_0).
\]
So, substituting (6.15) into (6.14) and then substituting the result into (6.9) we get

\[ y = B_0 + B_1 Z_1 + B_2 Z_2 + B_3 Z_3 + \varepsilon \tag{6.16} \]

where

\[ B_0 = a_1 \]
\[ B_1 = b_1 \]
\[ B_2 = -\frac{b_0}{r_0^2} (c_1 - c_0) + \frac{b_0}{r_0^3} (r_1 - r_0) \]
\[ B_3 = -\frac{b_0}{r_0^2} (r_1 - r_0) \]

The entire procedure is now as follows:

1. Trial values \( m_0 = c_0/r_0 \) and \( d_0 = 1/r_0 \) are somehow obtained, perhaps by a graphical method;
2. The transforms (6.15) are obtained for each data point;
3. Then \( y \) is regressed on the "new" variables \( Z_1, Z_2, \) and \( Z_3 \) to obtain estimates of \( B_0, B_1, B_2, \) and \( B_3, \) and from these latter improved estimates of \( m \) and \( d \) are obtained from the relations

\[ m_1 = \frac{d_0 B_1 - B_2 - m B_3}{d_0 B_1 - B_3} \tag{6.17} \]
\[ d_1 = \frac{d_0^2 B_1 - d_0 B_2}{d_0 B_1 - B_3} ; \]
4. Steps (2) and (3) are repeated using the improved "trial" values \( m_1 \) and \( d_1 \) to obtain still better values \( m_2 \) and \( d_2, \)
These latter are used to obtain $m_3$ and $d_3$ and so the process is iterated until convergence.

After the final iteration we have

\[ a = B_0 \]
\[ b = B_1 \]
\[ c = c_1 - \frac{1}{d_1} (c_1 B_3 + \frac{B_2}{d_1}) \]
\[ d = \frac{d_1 d_2}{d_1 d_1 - B_3} \]
\[ n = cd \]
\[ r = 1/d \]

In order to obtain $B_0$, $B_1$, $B_2$, and $B_3$ we have to solve the normal equations

\[ Z'ZB = Z'y \]

for $B$ where

\[
Z = \begin{pmatrix}
1 & Z_{11} & Z_{12} & Z_{13} \\
1 & Z_{21} & Z_{22} & Z_{23} \\
\vdots & \vdots & \vdots & \vdots \\
1 & Z_{n1} & Z_{n2} & Z_{n3}
\end{pmatrix}, \quad
B = \begin{pmatrix}
B_0 \\
B_1 \\
B_2 \\
B_3
\end{pmatrix}, \quad
y = \begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_n
\end{pmatrix}.
\]

Thus

\[ B = (Z'Z)^{-1} Z'y. \]
The matrix product $Z'Z$ is found to be

$$
\begin{pmatrix}
N & \sum z_1 & \sum z_2 & \sum z_3 \\
\sum z_1 & \sum z_2^2 & \sum z_1 z_2 & \sum z_1 z_3 \\
\sum z_2 & \sum z_2 z_1 & \sum z_2^2 & \sum z_2 z_3 \\
\sum z_3 & \sum z_3 z_1 & \sum z_3 z_2 & \sum z_3^2 
\end{pmatrix}
$$

It is apparent that if $\delta$ were known the last row and column of $Z'Z$ should be deleted. Similarly if $\gamma$ and $\delta$ were known the last two rows and columns should be deleted yielding the ordinary straight line regression case. If $\gamma$ were known the next to last row and column should be deleted. If $\lambda = 0$ is known then the first row and column should be deleted but if $\lambda$ is known and non-zero then $y' = y - \lambda$ is substituted in (6.20) and the first row and column of $Z'Z$ deleted as before. Therefore, one computational method suffices for fitting the general single process law and for fitting the various special cases as well.

6.5 Restrictions on the Iteration

In some cases curvature in the parametric space may cause grossly exaggerated corrections to the trial estimates to be indicated. Such large corrections if applied may often lead the iterative process outside of the region of convergence. Several courses may be taken to prevent overly large oscillations. A simple remedy is to place bounds upon the solution so that whenever a value is indicated beyond the bounds the boundary value itself is used for the next iteration. If the process
converges to one or the other of the boundary values then the bounds must be widened. Another course has been suggested by Wegstein (1958). Instead of using \( p_i \) in the \( i + 1 \)st iteration we use the "damped" estimate

\[
p_i^* = u p_{i-1} + v p_i
\]  

(6.21)

where \( p_i \) is the estimate of some parameter after the \( i \)th iteration and \( u + v = 1 \). Experience has shown the need for some such restriction on the solution. Other types than those mentioned are of course possible. The device (6.21) has been used successfully in the carrying out of the sampling study described in Section 12.0.

In some cases it may be convenient to "fix" \( \mu \) and converge for \( \delta \) only. By taking several values of \( \mu \) the best estimates may be found by interpolation.

6.6 A Numerical Example

The following data may be used to illustrate the iterative process:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>162.47</td>
</tr>
<tr>
<td>2</td>
<td>108.10</td>
</tr>
<tr>
<td>3</td>
<td>81.12</td>
</tr>
<tr>
<td>4</td>
<td>66.11</td>
</tr>
<tr>
<td>5</td>
<td>55.06</td>
</tr>
<tr>
<td>6</td>
<td>49.37</td>
</tr>
<tr>
<td>7</td>
<td>39.72</td>
</tr>
<tr>
<td>8</td>
<td>33.81</td>
</tr>
<tr>
<td>9</td>
<td>30.90</td>
</tr>
<tr>
<td>10</td>
<td>30.73</td>
</tr>
</tbody>
</table>
Trial values $m_0 = -1.00$ and $d_0 = -1.00$ were tried and the following results for successive iterations were found

<table>
<thead>
<tr>
<th>Iteration</th>
<th>m</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>1</td>
<td>-0.23</td>
<td>-0.56</td>
</tr>
<tr>
<td>2</td>
<td>-0.38</td>
<td>-0.69</td>
</tr>
<tr>
<td>3</td>
<td>-0.43</td>
<td>-0.67</td>
</tr>
<tr>
<td>4</td>
<td>-0.44</td>
<td>-0.67</td>
</tr>
</tbody>
</table>

Then for the final iteration we have the estimated relationship

\[ y = -18.92 + 231.5(x + 0.44)^{-0.67} \quad (6.22) \]

Rarely in our experience has it been found necessary to compute more than six iterations. Of course, over-conservative use of the device described in the last section may produce a very slow rate of convergence.
7.0 SUMMARY OF ASYMPTOTIC RESULTS

7.1 Asymptotic Variances

In this chapter is presented the large sample results relating to the variances and distributions of the maximum likelihood estimates. The solution of the normal equations was given in Section 6.0 to be

\[ B = (Z'Z)^{-1} Z'y. \] (7.1)

The asymptotically unbiased estimate of \( \sigma^2 \) is as usual given by

\[ s^2 = \frac{1}{N-1} \left\{ \bar{y}' \bar{y} - B'Z' \bar{y} \right\} \] (7.2)

or

\[ s^2 = \frac{1}{N-1} \left\{ \sum y^2 - B_0 \sum y - B_1 \sum Z_1 y - B_2 \sum Z_2 y - B_3 \sum Z_3 y \right\}, \] (7.3)

Let us denote by \( c_{ij} \) the element in the \( i \)th row and \( j \)th column of \((Z'Z)^{-1} \). It has been well known since Gauss that the variance \( \sigma^2_x \) of any function of the \( B_i \)'s (\( f \) say) is asymptotically given by

\[ \sigma^2_x = \sigma^2 \sum_{i=0}^{3} \sum_{j=0}^{3} c_{ij} f_i f_j \] (7.4)

where \( f_i = \partial f / \partial B_i \) and \( f_j = \partial f / \partial B_j \). We may use as the estimated variance of the function

\[ s^2_x = s^2 \sum_{i=0}^{3} \sum_{j=0}^{3} c_{ij} f_i f_j \] (7.5)
By applying (6.9) to the form (6.10) of the single process law we obtain the limiting relations

\[ a = B_0 \]
\[ b = B_1 \]
\[ d = d_0 + B_3/b \]
\[ m = m_0 - B_2/\beta d \quad (7.6) \]

Thus, from (7.5) we obtain the asymptotic estimates

\[ s_a^2 = c_{10} s^2 \]
\[ s_b^2 = c_{11} s^2 \]
\[ s_m^2 = c_{22} s^2/b^2 d^2 \]
\[ s_d^2 = c_{33} s^2/b^2 \quad (7.7) \]

Similarly, for \( c = m/d \) we find

\[ s_c^2 = \left( \frac{c_{22}}{b^2 d^2 m} - \frac{2c_{23}}{b^2 d} + \frac{m^2 c_{33}}{b^2 d^4} \right) s^2 \quad (7.8) \]

and for \( r = 1/d \)

\[ s_r^2 = c_{33} s^2/b^2 d^4 \quad (7.9) \]
Finally, we find the asymptotic estimate of the variance of a predicted \( y \) for given \( x \):

\[
s^2(\hat{y} \mid x) = s^2 \sum_{i=0}^{3} \sum_{j=0}^{3} c_{ij} Z_i Z_j
\]  

(7.10)

where \( Z_i \) and \( Z_j \) are evaluated for the given \( x \).

### 7.2 Asymptotic Confidence Limits

Maximum likelihood estimates are, under suitable conditions, asymptotically normally distributed with expectation equal to the parametric values and variances given by (7.4). This fact may be utilized to obtain asymptotic confidence limits useful under certain circumstances. We have

\[
a - t_n s_a \leq \hat{y} \leq a + t_n s_a
\]

\[
b - t_n s_b \leq \beta \leq b + t_n s_b
\]

\[
\vdots
\]

\[
\hat{y} \mid x - t s(\hat{y} \mid x) \leq \gamma \mid x \leq \hat{y} \mid x + t s(\hat{y} \mid x)
\]

(7.11)

with 1-\( \alpha \) confidence where \( t_n \) is student's \( t \) for the \( n \) level of significance with \( N-n \) degrees of freedom. Sir Ronald Fisher (1939) has said "... that the ordinary tests will still be sufficiently exact when the radii of curvature of the spaces concerned are all large compared with the distances." Thus, it is quite possible that these "asymptotic" results will suffice even for "small" samples in some cases. Section 12.0 is devoted to a sampling exploration of the adequacy of the asymptotic results for certain cases.
8.1 Uses of the Single Process Law

It is anticipated that the single process law will be quite useful in empirically describing monotonic "growth" or "decay" curves. Indeed, the special cases are in wide use for this purpose. However, the fact that the differential equation of the single process law is of an easily interpretable form suggests that it will lend itself readily to the theoretical model-building process. As an illustration of this point we proceed to derive a theory to describe the form of the world record curve for middle and long distance track events. The general line of reasoning is after A. V. Hill (1927).

8.2 A Theory of Track Records

We define the following symbols:

\[ y \] The total energy (measured as volume of oxygen) excess over resting required to run a distance in a certain time at a given constant speed.

\[ x \] The distance run.

\[ t \] The time to run \( x \).

\[ \dot{x} \] The constant speed.

\[ x_0 \] The greatest speed for which the energy (oxygen) required is not greater than the energy available for an indefinitely long period of time. Thus, at speed \( x_0 \) the energy required is equal to the maximum amount of oxygen which can be brought into the body from the atmosphere.

\[ r \] The maximum rate at which oxygen can be brought into the body from the atmosphere.

\[ D \] The maximum tolerable oxygen debt beyond which complete exhaustion is manifest.
To a first approximation we may say that

$$d_r = \left[r + b(\dot{x} - \dot{x}_0)^k\right]dt, \quad \dot{x} \geq \dot{x}_0$$  (8.1)

where $b$ and $k$ are arbitrary constants of scale. Integrating we get

$$y = rt + b(\dot{x} - \dot{x}_0)^kt.$$  (8.2)

The constant of integration is obviously zero since no excess energy is required to run a race of zero duration. Equation (8.2) can be used to describe the experimental data shown in Figure 11 of Hill (1927). Hill measured directly the excess oxygen required by an athlete to run for given periods of time at constant speeds. This experiment, of course, does not correspond to actual contest conditions, but does serve to illustrate the relation (8.2) which was not given by Hill. The meaning of the relation is clear. If a runner runs at his own asymptotic speed $\dot{x}_0$ then the excess energy required is obviously proportional to the time run where the proportionality constant is $r$; but if he runs faster than $\dot{x}_0$ then the energy required will be increased monotonically with the divergence of his speed from the asymptotic speed.

Now, the speed cannot be increased indefinitely since there is an upper bound on energy available. This upper bound is given by

$$y_{\text{max}} = rt + D.$$  (8.3)

For all but the shorter races (say, less than 400 meters) the time lost in attaining optimum speed is negligible and it can be seen that the best time for a race can be obtained by maintaining a nearly constant speed throughout the race. It may be supposed that records are broken when a
runner with a great amount of available energy is able to select that constant speed which equates the maximum available energy to the required energy. In this case

\[ rt + D = rt + b(x - \dot{x}_0)^k t \]  \hspace{1cm} (8.4)

which upon cancelling \(rt\) from both sides and rearranging gives

\[ \dot{x} = \dot{x}_0 + \left(\frac{b}{D} t\right)^{-1/k} \]  \hspace{1cm} (8.5)

relating speed to time run. The constants \(\dot{x}_0\), \(D\), \(b\), and \(k\) are constants pertaining to the individual runner. \(D\) is the maximum oxygen debt the runner can tolerate and \(b\) and \(k\) relate to the mechanical efficiency of the runners body, i.e., to the amount of work required to overcome the viscosity of the muscles. \(\dot{x}_0\) is a function of the rate at which oxygen can be gotten to the tissues. Relation (8.5) could be used to study the progress of an individual runner in his period of training; however, our interest at present is in characterizing the world record curve. Although \(\dot{x}_0\), \(D\), \(b\), and \(k\) are personal parameters it may be supposed that for world record holders \(\dot{x}_0\) and \(D\) approach the human maximum for all and that \(b\) and \(k\) are similarly close to the human minimum for all and that therefore it is appropriate to treat \(\dot{x}_0\), \(D\), \(b\), and \(k\) as constant parameters for all record holders as a first approximation. Obviously (8.5) is a special case of the single process law

\[ \dot{\gamma} = \alpha + \beta (\gamma - \gamma_0)^6 \]  \hspace{1cm} (8.6)
where
\[
\begin{align*}
\eta &= \dot{x} \\
\xi &= \dot{\xi}_0 \\
\gamma &= 0
\end{align*}
\]

Thus (8.6) is equivalent to
\[
\eta = \dot{x} + \beta \xi. \quad (8.7)
\]

8.3 Fitting the Theory to World Record Data

Data for world track records were taken from Hansen (1958). These data are presented in the first 3 columns of Table 8.1. The average speed for each record has been computed and is presented in column 4 of this table.

Table 8.1. Some world track records for 1957a

<table>
<thead>
<tr>
<th>Distance</th>
<th>Time</th>
<th>Average Speed</th>
<th>Predicted Average Speeds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Units</td>
<td>Meters</td>
<td>Seconds</td>
<td>Meters/Second</td>
</tr>
<tr>
<td>220 yd.</td>
<td>201.17</td>
<td>20.0</td>
<td>10.06</td>
</tr>
<tr>
<td>400.</td>
<td>45.2</td>
<td></td>
<td>8.85</td>
</tr>
<tr>
<td>800.</td>
<td>105.7</td>
<td></td>
<td>7.57</td>
</tr>
<tr>
<td>1000.</td>
<td>139.0</td>
<td></td>
<td>7.19</td>
</tr>
<tr>
<td>1 mi.</td>
<td>1609.35</td>
<td>237.8</td>
<td>6.77</td>
</tr>
<tr>
<td>2000.</td>
<td>302.2</td>
<td></td>
<td>6.62</td>
</tr>
<tr>
<td>3000.</td>
<td>472.8</td>
<td></td>
<td>6.35</td>
</tr>
<tr>
<td>5000.</td>
<td>616.8</td>
<td></td>
<td>6.12</td>
</tr>
<tr>
<td>10000.</td>
<td>1710.4</td>
<td>5.85</td>
<td>5.80</td>
</tr>
<tr>
<td>20052.</td>
<td>3600.0</td>
<td>5.57</td>
<td>5.60</td>
</tr>
</tbody>
</table>

The events which had clearly inferior average speeds as well as the shorter races in which "starting" plays a major role have been omitted. We have assumed a regression model, derived from (8:5), of the form

\[ x = x_0 + \left( \frac{b}{D} - t \right)^{-1/k} + \varepsilon \]  

and fitted this model by the methods of Section 6.0. In order to employ the iterative method a trial value of \( \delta = -1/k \) is required. This is quickly provided by the use of the \( \lambda \) criterion. From a plot of the above data we estimate the "3/4 life" to be approximately 540 seconds and the "1/2 life" to be about 80 seconds. This gives a value of 6.75 for \( \lambda \) and referring to the nomogram relating \( \lambda \) to \( \delta \) (Figure 5.2) we estimate \( \delta \) to be about -0.4. This value we then take as our trial estimate. By iteration we find the maximum likelihood estimate of \( \delta \) to be -0.426. The maximum likelihood estimate of \( \lambda \) is 5.04 and that for \( \beta \) is 18.34. Thus, we have the fitted curve

\[ x = 5.04 + 18.34 t^{-0.426} \]  

Predicted values for the average speed are given in Column 5 of Table 8.1 along with the observed average speeds. It will be noted that the fit is rather good.

For esthetic reasons one may prefer to replace the decimal estimate by a simple rational fraction. The fraction \(-3/7\) is quite close numerically to the estimate \(3/7 = 0.4285\) and may be used instead. In this case the law of track records (8.9) can be stated thus: "The excess of
average speed over the asymptotic value is inversely proportional to the 3/7th power of the time. In short, we may say that we have demonstrated that middle and long distance track records obey the "inverse 3/7th law."

Hypotheses other than that $\delta = -3/7$ are, of course, possible. The class of tenable hypotheses may be found by constructing confidence limits about the unknown parameter. Three methods of constructing exact confidence limits are discussed in the next section and numerical results are obtained for the present example. We will see that the hypothesis $\delta = -1/2$ is not acceptable.
9.0 SOME EXACT TESTS AND CONFIDENCE LIMITS

9.1 Introduction

The expression exact test is used to denote a test which rejects the null hypothesis at precisely that relative frequency corresponding to the type I error for the test. These tests are to be distinguished from asymptotic tests which generally reject the null hypothesis at a somewhat higher relative frequency than the assigned type I error. It must not be concluded that in a particular case an exact test is necessarily to be preferred over an asymptotic one since the available exact tests may suffer from deficiencies such as inappropriateness for a particular purpose or inadequate power. In some cases it may be that a non-probabilistic rejection rule, such as the one based upon relative likelihood, is to be preferred over available probability tests.

In the present chapter we consider three exact tests of significance and the corresponding confidence statements.

9.2 A Test of Assigned Nonlinear Parameters in the Case of Replication

Let us take \( \mu = \gamma \delta \) as in previous chapters. Then the single process model is of the following form for the \( j \)th replication of the level of \( x \):

\[
Y_{ij} = \alpha + \beta(x_i - \mu)\delta + \epsilon_{ij} \tag{9.1}
\]

\( i = 1, 2, \ldots, k; \ j = 1, 2, \ldots, r \). But now suppose that we substitute some other values of the parameters (say \( \mu_o \) and \( \delta_o \)) for \( \mu \) and \( \delta \) then (9.1) would have to be replaced by

\[
Y_{ij} = \alpha + \beta_i + \epsilon_{ij} \cdot \tag{9.2}
\]
Now, if we wish to test the null hypothesis

\[ H_0 : (\mu = \mu_0) \cap (\delta = \delta_0) \]  

(9.3)

we can test for "departures from linearity" in the usual way. Let 

\[ z_i = (x_i - \mu_0) \delta_0. \]

Then the sum of squares of residuals about the fitted linear model (9.1) is given by

\[ \text{SSE}' = S_{yy} - S_{zy}^2/S_{zz} \]  

(9.4)

where

\[ S_{yy} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^2 - (\sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij})^2/kn \]

\[ S_{zy} = \sum_{i=1}^{k} z_i \sum_{j=1}^{n} y_{ij} - (\sum_{i=1}^{k} z_i)(\sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij})/k \]  

(9.5)

\[ S_{zz} = n \sum_{i=1}^{k} z_i^2 - n(\sum_{i=1}^{k} z_i)^2/k. \]

Similarly, the residual sum of squares about the fitted model (9.2) is given by

\[ \text{SSE} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^2 - (\sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij})^2/n \]  

(9.6)

Then the null hypothesis (9.3) is tested by the variance ratio

\[ F = \frac{\text{SSE}' - \text{SSE}}{\text{SSE}} \frac{k(n - 1)}{k - 2}. \]  

(9.7)

where \( F \) follows the F-distribution with \((k-2)\) and \(k(n-1)\) degrees of freedom when \( H_0 \) is true.
The test (9.7) is well known in the linear case but not so well known in the nonlinear case. The nonlinear application has been considered previously by Hurst\(^1\) (1958).

A joint confidence region for the nonlinear parameters \(\mu\) and \(\delta\) (and hence also for \(\gamma\) and \(\delta\)) may be found by inverting (9.7) to give

\[
SSE' = SSE + F_n(SSE) \tag{9.8}
\]

where \(F_n\) is the tabular value of \(F\) for the \((1-\alpha)\)th confidence coefficient. Now \(SSE'\) is calculated for various combinations of \(\mu_0\) and \(\delta_0\) until the confidence region

\[
SSE' \leq SSE_n \tag{9.9}
\]

is delimited.

9.3 A Test Based upon the Taylor's Series Expansion

In many cases only a single observation has been made at each level of \(I\). E. J. Williams (1959) has suggested an exact test of the null hypothesis (9.3) obtained from the expansion of the model (9.1) about the point \((\mu_0, \delta_0)\). Thus, we replace the model (9.1) by the approximate linearized form of Section 6.0

\[
y_1 = B_0 + B_1 I_1 + B_2 Z_2I + B_3 Z_3 I + \varepsilon_1 \tag{9.10}
\]

See (6.16) for derivation. The coefficients \(B_2\) and \(B_3\) are proportional to the differences between \(\mu_0\) and \(\mu\) and between \(\delta_0\) and \(\delta\). Hence a test of the null hypothesis

\[
H_0 : B_2 = B_3 = 0
\]

---

\(^1\) Personal communication
is approximately equivalent to a test of the null hypothesis (9.3).

Linear regression theory supplies the test for (9.11). If \( \hat{B}_1 \), \( \hat{B}_2 \) and \( \hat{B}_3 \) are the minimum variance unbiased estimates then the residual sum of squares for (9.10) is given by

\[
SSE = S_{yy} - \hat{B}_1 S_{ly} - \hat{B}_2 S_{2y} - \hat{B}_3 S_{3y}
\]  

(9.12)

where

\[
S_{yy} = \sum_{i=1}^{k} y_i^2 - \left( \sum_{i=1}^{k} y_i \right)^2 / k
\]

(9.13)

\[
S_{ly} = \sum_{i=1}^{k} z_{hi} y_i - \left( \sum_{i=1}^{k} z_{hi} \right) \left( \sum_{i=1}^{k} y_i \right) / k
\]

(9.14)

\( h = 1, 2, 3 \). When \( H_0 \) is true, (9.10) reduces to (9.1) and the residual sum of squares is given by

\[
SSE' = S_{yy} - S_{ly} S_{11}/S_{11}
\]

(9.14)

where

\[
S_{11} = \sum_{i=1}^{k} z_{ii}^2 - \left( \sum_{i=1}^{k} z_{ii} \right)^2 / k
\]

(9.15)

Then the variance ratio for testing (9.11) is

\[
F = \frac{SSE' - SSE}{SSE} \frac{k - 4}{2}
\]

(9.16)

and \( F \) has the \( F \)-distribution with 2 and \((k-4)\) degrees of freedom when \( H_0 \) is true. We may claim to have an exact test of an approximately equivalent hypothesis.

A confidence region may be found by calculating \( F \) for various values of \( \mu_0 \) and \( \sigma_0 \) and determining the region such that
where $F_n$ is the tabular value of $F$ for a confidence coefficient $(1-\alpha)$.

It may be supposed that the test (9.16) is more powerful than the test (7.7) for small departures from $\mu_0$ and $\delta_0$. However, this point has not been investigated.

In the case of replication at each point we can replace the denominator of test (9.16) by the sum of squares (9.6).

### 9.4 The Parabolic Test for Nonlinearity

An approach similar to that of the last section but differing in details is given below.

If $\mu = \mu_0$ and $\delta = \delta_0$ then (9.1) may be written

$$y_1 = \lambda + \beta Z_{10} + \varepsilon_1$$  \hspace{1cm} (9.18)

for the case of single replication. However, in the case that the null hypothesis is not true, then instead of (9.18) we would have

$$y_1 = f(Z_{10}) + \varepsilon_1.$$  \hspace{1cm} (9.19)

Expanding (9.19) in a Maclaurin series and preserving terms through the quadratic, we get

$$y_1 = B_0 + B_1 Z_{10} + B_2 Z_{10}^2 + \varepsilon_1.$$  \hspace{1cm} (9.20)

Now to a first approximation the null hypothesis

$$H_0 : B_2 = 0$$  \hspace{1cm} (9.21)

is equivalent to (9.3). A test of (9.21) is well known and is easily performed. If $B_1$ and $B_2$ are the minimum variance unbiased estimates then
the residual sum of squares about the predicted curve corresponding to (9.20) is given by

\[ SSE = \sum_{i=1}^{k} z_{i0}^2 y_i - \left( \sum_{i=1}^{k} z_{i0}^2 \right) \left( \frac{\sum_{i=1}^{k} y_i}{k} \right) / k \]  

(9.23)

where \( \sum_{i=1}^{k} z_{i0}^2 y_i \) and \( \sum_{i=1}^{k} z_{i0}^2 \) are given by (9.13) and \( S_{2y} \) is given by

\[ S_{2y} = \frac{1}{k} \sum_{i=1}^{k} z_{i0}^2 y_i - \frac{1}{k} \left( \sum_{i=1}^{k} z_{i0}^2 \right) \frac{1}{k} \sum_{i=1}^{k} y_i / k \]  

The test statistic is

\[ F = \frac{\text{SSE}^1 - \text{SSE}}{\text{SSE}} \frac{k-3}{1} \]  

(9.24)

where \( F \) has the \( F \)-distribution with 1 and \( (k-3) \) degrees of freedom and \( \text{SSE}^1 \) is obtained from (9.14). This test too may be described as an exact test of an approximately equivalent hypothesis.

Again, the confidence region is given by

\[ F \leq F_{\alpha} \]  

(9.25)

In some cases, it may be desirable to retain terms higher than the second in the series expansion. In any case the numerical work involved in performing the test is relatively light and that involved in determining the confidence region quite heavy.

9.5 Confidence Limits for the Exponential Parameter in the Track Record Example Based upon the Parabolic Test

Values of \( F \) given by the formula (9.24) have been computed for various values of the exponential parameter \( \delta \) for the example of world track records. The somewhat more convenient values of \( t = \pm \sqrt{F} \) are reported in Table 9.1.
Table 9.1. Values of \( t \) for various hypotheses

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.6</td>
<td>5.43</td>
</tr>
<tr>
<td>-0.5</td>
<td>2.89</td>
</tr>
<tr>
<td>-0.46</td>
<td>1.81</td>
</tr>
<tr>
<td>-0.426</td>
<td>0.67</td>
</tr>
<tr>
<td>-0.4</td>
<td>-0.20</td>
</tr>
<tr>
<td>-0.36</td>
<td>-1.52</td>
</tr>
<tr>
<td>-0.3</td>
<td>-3.50</td>
</tr>
</tbody>
</table>

Under the null hypothesis (9.21) \( t \) has student's \( t \)-distribution with 7 degrees of freedom. The tabular value of \( t \) with 7 degrees of freedom at the 5% significance level is found to be 2.365. By graphical interpolation we find the 95% confidence statement

\[-0.48 \leq \delta \leq -0.33.\tag{9.25}\]

The details of the interpolation are shown in Figure (9.1). It is apparent from these data that the simple hypothesis \( \delta = -0.5 \) is not to be accepted at the 5% level of significance. Likewise, the hypothesis \( \delta = -0.33 \) is to be rejected. Of all simple rational fractions only \(-3/7\) comes at all close to the maximum likelihood estimate although several others (notably \(-2/5\)) would not be rejected at the assigned level.

In closing we note that the parabolic confidence limits converge, not to the maximum likelihood estimate (\( \delta = -0.426 \)), but rather to the "zero curvature" estimate (\( \delta = -0.415 \)). This phenomenon often occurs when confidence limits are not fiducial limits, as has been pointed out by Fisher (1936). Both estimates are consistent so that it can be seen that our parabolic limits have the asymptotic "enclosure" or convergence
Fig. 9.1 Determination of confidence limits for the track record example using the parabolic test method
property while for small samples, although not possessing this property, they do produce exact probability inequalities of confidence. Similar remarks apply to the limits based upon the test of Williams (9.16).
10.1 The Origins of Path Regression Analysis

Sewall Wright (1921) conceived of the notion and term of path regression analysis. The notion has been much extended and developed in the econometric literature. See especially Hood and Koopmans (1953) and Wold (1956). Wright himself has come to favor the use of standardized variates, the analysis of which he refers to as "path analysis". Summaries of Wright's views may be found in his review papers, Wright (1934, 1954). Tukey (1954) has argued against the use of standardized variates even when these have a rational basis, i.e., when the "independent" variables have normal distributions. When the independent variables are "fixed", i.e., have no probability distribution, then standardization as Wright advocates is meaningless. For elementary presentations of the regression method, see Kempthorne (1957) and Turner and Stevens (1959).

In this section will be presented a recapitulation of the main results of linear path regression analysis. An extension to the non-linear case using the single process law will be described in the next section.

10.2 The Structural Equations

Let us suppose that we have p "secondary" variables which are casually related to one another and to a set of q "primary" or "controllable" variables. We imagine that the q primary variables (denoted $\bar{a}, \bar{b}, \ldots, \bar{q}$) have values which can be chosen at will by the
experimenter. The $p$ secondary variables (denoted by $\gamma_1, \gamma_2, \ldots, \gamma_p$) are then supposed to be completely determined. We further assume that all rates of change are constant. Thus, the matrix of partial derivatives $\gamma_{ij}$ is $H$ where

$$
H = \begin{bmatrix}
\gamma_{1a} & \gamma_{2a} & \cdots & \gamma_{pa} \\
\gamma_{1b} & \gamma_{2b} & \cdots & \gamma_{pb} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{1q} & \gamma_{2q} & \cdots & \gamma_{pq} \\
\gamma_{11} & \gamma_{21} & \cdots & \gamma_{p1} \\
\gamma_{12} & \gamma_{22} & \cdots & \gamma_{p2} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{1p} & \gamma_{2p} & \cdots & \gamma_{pp}
\end{bmatrix}
$$

and where the "path regression coefficients", $\gamma_{ij}$, are constant. Solution of (10.1) yields the $p$ "structural equations" 

$$
\gamma_1 = \lambda_1 + \gamma_{1a} a + \gamma_{1b} b + \cdots + \gamma_{1q} q + \gamma_{11} \gamma_2 + \gamma_{12} \gamma_3 + \cdots + \gamma_{1p} \gamma_p \\
\gamma_2 = \lambda_2 + \gamma_{2a} a + \gamma_{2b} b + \cdots + \gamma_{2q} q + \gamma_{21} \gamma_1 + \gamma_{22} \gamma_3 + \cdots + \gamma_{2p} \gamma_p \\
\vdots \\
\gamma_p = \lambda_p + \gamma_{pa} a + \gamma_{pb} b + \cdots + \gamma_{pq} q + \gamma_{p1} \gamma_1 + \gamma_{p2} \gamma_2 + \gamma_{p3} \gamma_3 + \cdots
$$

(10.2)
where \( \phi_1, \phi_2, \ldots, \phi_p \) are constants of integration. These equations may be written in matrix notation as

\[
\begin{bmatrix}
\phi_1 & \phi_2 & \cdots & \phi_q & 0 & \phi_2 & \cdots & \phi_p \\
\phi_2 & \phi_3 & \cdots & \phi_q & \phi_2 & 0 & \cdots & \phi_p \\
\vdots & \vdots & & \vdots & \vdots & \ddots & & \vdots \\
\phi_p & \phi_p & \cdots & \phi_q & \phi_1 & \phi_2 & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
1 \\
\phi \phi_a \\
\phi \phi_b \\
\vdots \\
\phi \phi_p
\end{bmatrix}
= 
\begin{bmatrix}
\eta_1 \\
\eta_2 \\
\vdots \\
\eta_p
\end{bmatrix}
\]

or

\[
(A_1, A_2)
\begin{pmatrix}
\phi \\
\eta
\end{pmatrix}
= \eta .
\] (10.3)

Therefore, the "structure" of the causal system is completely described by (10.3).

10.3 The Reduced Structural Equations

From (10.3) we get by multiplication

\[
A_1 \phi + A_2 \eta = \eta
\] (10.4)
and

\[ \eta - A_2 \eta = A_1 \xi \]

\[ (I - A_2) \eta = A_1 \xi \]

If \((I - A_2)\) is non-singular we get the "reduced structural equations"

\[ \eta = (I - A_2)^{-1} A_1 \xi \]  \hspace{1cm} (10.5)

Thus, the secondary variables \(\eta\) are expressed as linear functions of the primary variables \(\xi\) and it is seen that the necessary and sufficient condition for this to be possible is that \((I - A_2)\) be non-singular.

10.4 The Equations of Identification

Let \(B\) be a \(p \times (q+1)\) matrix of regression coefficients as follows:

\[
B = \begin{bmatrix}
\beta_1 & \beta_{1a} & \beta_{1b} & \cdots & \beta_{1q} \\
\beta_2 & \beta_{2a} & \beta_{2b} & \cdots & \beta_{2q} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\beta_p & \beta_{pq} & \beta_{pb} & \cdots & \beta_{pq}
\end{bmatrix}
\]

Now we will state the "identification equations" as follows:

\[ B = (I - A_2)^{-1} A_1 \]  \hspace{1cm} (10.6)

The reduced structural equations (10.5) may then be written

\[ \eta = B \xi \]  \hspace{1cm} (10.7)

the structural equations for an ordinary multiple regression system.
We note that it may or may not be possible to solve (10.6) for $(A_1, A_2)$ uniquely or at all. If it is not possible to find a solution at all, the system is said to be "under-identified". If a unique solution exists the system is said to be "just-identified", and if more than a single solution exists, the system is said to be "over-identified".

If we expand the right hand side of (10.6) we will find, in general, that some of the $\beta$'s are zero as a consequence of the casual restrictions on the $\gamma$'s (in particular, some of the casual pathways will be absent and the corresponding $\gamma$'s will be zero). Let us agree to omit all those $\gamma$'s from the rows of (10.7) having zero $\beta$'s as coefficients. Then, we would replace (10.7) by

$$
\begin{align*}
\eta_1 &= \beta_1 \gamma_1 \\
\eta_2 &= \beta_2 \gamma_2 \\
&\vdots \\
\eta_p &= \beta_p \gamma_p
\end{align*}
$$

(10.8)

where each $\beta_i$ and $\gamma_i$ is of order less than or equal to $q+1$.

10.5 Estimation in the Case of Under- and Just-identified System

The under- and just-identified cases are of especial interest because in these cases regression type estimators are available. We consider two possibilities:
(1) All $\beta$'s are non-null and
\[
\begin{align*}
\bar{X}_j &= \gamma_j + \varepsilon_j \\
\bar{X}_j &= \gamma_j \\

j &= 1, 2, \ldots, n.
\end{align*}
\]
Then by substitution (10.7) becomes
\[
\bar{X}_j = B\bar{X}_j + \varepsilon_j \\
(10.9)
\]
or
\[
\bar{X}_j = X_j B' + \varepsilon_j. \\
(10.10)
\]
For $n$ observations we have
\[
\begin{bmatrix}
\bar{X}_1 \\
\bar{X}_2 \\
\vdots \\
\bar{X}_n
\end{bmatrix} =
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix} B' + 
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_n
\end{bmatrix}
\]
or
\[
Y = XB' + E. \\
(10.11)
\]
Now, if the $n \times p$ elements of the "error" matrix are uncorrelated and have constant variance $\sigma^2$ then the minimum variance unbiased estimator is given by the well known expression
\[
\bar{B} = (X'X)^{-1} X'Y. \\
(10.12)
\]
If in addition $E$ is distributed as a multivariate normal distribution, then, of course, $\bar{B}$ is the sufficient estimator for $B'$. In the case of just-identification $B'$ may be replaced by
\(^B \) in (10.6) and estimates of \((A_1, A_2)\) solved for. In the case of under-identification, it will not be possible to solve for all elements of \((A_1, A_2)\) explicitly unless a sufficient number of a priori constraints can be assumed.

(2) More often than not restrictions on \(A_1\) and \(A_2\) will force some \(\beta\)'s to be zero and then we must omit the corresponding \(x\)'s from the estimation equations. In this case a separate regression analysis must be performed for each \(y\) variable.

Hence, we get the individual estimators

\[
\hat{\beta}_1 = (x_1' x_1)^{-1} x_1' x_1 \\
\hat{\beta}_2 = (x_2' x_2)^{-1} x_2' x_2 \\
\vdots \\
\hat{\beta}_p = (x_p' x_p)^{-1} x_p' x_p .
\]  

The remarks following (10.12) concerning distribution and identification apply similarly to (10.13).

For treatment of the over-identified case see Hood and Koopmans (1953). Refer also to Turner and Stevens (1959).

10.6 Classification of Path Schemata

There are several types of casual relationships which are conveniently distinguished. First of all there is the situation in which a single secondary variable (or response variable) is determined by one or more primary variables. This is the case of "ordinary" regression, the single structural equation corresponding to that of the usual multiple regression model. A second case is the situation in which two or more
secondary variables are jointly determined by some or all of a common set of primary variables. We term this case the case of "joint" regression. In a third situation there is a chain of cause and effect leading from one or more primary variables to a secondary variable and then on to still another "secondary" variable. This case we term the case of "chain" regression. A final type involves cycles of causation or "feedback" from one response variable to another and back again after possibly passing through one or more other secondary variables. This is the case of "cyclic" regression. All of these cases individually or in combination have been synoptically treated in the previous sections of this chapter. Here we wish only to note that the several cases are distinguished by having different kinds of \((I - A_2)\) matrices. For this reason we term \((I - A_2)\) the "classification" matrix. The type of regression is identified as in Table 9.1.

<table>
<thead>
<tr>
<th>Regression Type is</th>
<th>((I - A_2)) is</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinary</td>
<td>1</td>
</tr>
<tr>
<td>Joint</td>
<td>I</td>
</tr>
<tr>
<td>Chain</td>
<td>triangular</td>
</tr>
<tr>
<td>Cyclic (feedback)</td>
<td>non-triangular</td>
</tr>
</tbody>
</table>

It is evident that any scheme not involving feedback will possess a triangular classification matrix. In this case the identification equations (10.6) are of a particularly simple variety and algorithms have been devised by Wright and his followers for writing down these equations.
10.7 Some Examples of Particular Path Schemata

We will consider several very simple cases of linear casual networks, nonlinear analogues of which we will consider in the next chapter.

(1) The case of two primary variables and one secondary variable.

We represent the flow of cause and effect by a "path diagram" as follows:

\[ \begin{align*}
    a \rightarrow \lambda_a \rightarrow \eta_1 \\
    b \rightarrow \lambda_b \rightarrow \eta_1 
\end{align*} \]

The structural equation is

\[ \eta_1 = \lambda_a + \lambda_a a + \lambda_b b \]  \hspace{1cm} (10.14)

or

\[ \begin{pmatrix}
    \lambda_a \\
    \lambda_b \\
    0
\end{pmatrix}
\begin{pmatrix}
    1 & a & b
\end{pmatrix}
= \begin{pmatrix}
    \eta_1
\end{pmatrix} \]

and

\[ (\beta_1, \beta_{la}, \beta_{lb}) = (1) (\lambda_a, \lambda_b) \]  \hspace{1cm} (10.15)

from (10.6). Thus, we get the obvious result that if one obtains the usual multiple regression estimates \( \hat{\beta}_1, \hat{\beta}_{la}, \hat{\beta}_{lb} \) by regressing \( \eta_1 \) on \( a \) and \( b \) then we would estimate the path coefficients
by equating according to (10.15):

\[ \hat{\beta}_1 = \hat{\beta}_1 \]
\[ \hat{\beta}_{la} = \hat{\beta}_{la} \]
\[ \hat{\beta}_{lb} = \hat{\beta}_{lb} \]

(2) A case in which both joint and chain regression occurs.

Consider the path diagram

\[
\begin{array}{c}
\psi_a \xrightarrow{\alpha_{la}} \eta_1 \\
\psi_b \xrightarrow{\alpha_{lb}} \eta_2 \\
\end{array}
\]

Proceeding as before we obtain

\[
I - A_2 = \begin{pmatrix} 1 & 0 \\ -\hat{\beta}_{21} & 1 \end{pmatrix}
\]

and according to (10.6) we get identification equations

\[
\begin{pmatrix} \beta_1 & \beta_{la} & \beta_{lb} \\ \beta_2 & \beta_{2a} & \beta_{2b} \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_{la} & \hat{\beta}_{lb} \\ \hat{\beta}_2 & 0 & \hat{\beta}_{2b} \end{pmatrix}
\]

\[
= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_{la} & \hat{\beta}_{lb} \\ \hat{\beta}_2 & 0 & \hat{\beta}_{2b} \end{pmatrix}
\]

\[
= \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_{la} & \hat{\beta}_{lb} \\ \hat{\beta}_2 & 0 & \hat{\beta}_{2b} \end{pmatrix}
\]

Now (10.16) is solved for the path coefficients and the \( \beta \)'s are replaced by regression estimates to give estimates of the path coefficients.
(3) An example of feedback. The path diagram of a simple case of cyclic regression is given below.

\[ \begin{align*}
\gamma_a & \xrightarrow{\alpha_{1a}} \eta_1 \\
\gamma_b & \xrightarrow{\alpha_{2b}} \eta_2 \\
\end{align*} \]

We have according to (10.6)

\[
\begin{pmatrix}
\beta_1 & \beta_{1a} & \beta_{1b} \\
\beta_2 & \beta_{2a} & \beta_{2b}
\end{pmatrix} = \begin{pmatrix}
1 & -\alpha_{12}^{-1} & \alpha_{1a} & 0 \\
-\alpha_{21} & 1 & 0 & \alpha_{2b}
\end{pmatrix}
\]

\[
\frac{1}{1-\alpha_{12} \alpha_{21}} \begin{pmatrix}
1 & -\alpha_{12}^{-1} & \alpha_{1a} & 0 \\
\alpha_{21} & 1 & 0 & \alpha_{2b}
\end{pmatrix}
\]

\[
\frac{1}{1-\alpha_{12} \alpha_{21}} \begin{pmatrix}
1 + \alpha_{12} \alpha_{21} & \alpha_{1a} & \alpha_{12} \alpha_{2b} \\
\alpha_{21} & 1 + \alpha_{12} \alpha_{21} & \alpha_{21} \alpha_{1a} & \alpha_{2b}
\end{pmatrix}
\]

Regression estimates of all path coefficients are then easily found.
11.0 NONLINEAR PATH REGRESSION

11.1 The Structural Equations for Pathways Following the Single Process Law

The ideas of path analysis presented in the last chapter can be extended to the case of nonlinear causal relationships between variables. In this chapter we will study the situation in which the relationships can each be represented by the single process law. The general approach is valid for other relationships as well.

As in the linear case we describe the structure of a hypothetical network of causal pathways by a system of partial differential equations. In the linear case each of the partial derivatives are constant — the constants being termed "path coefficients". In the non-linear case the analogues of the path coefficients (i.e., the partial derivatives) are not constant.

Let the matrix of partial derivatives be given, as before, by

\[
H' = \begin{bmatrix}
\gamma_{1a} & \gamma_{2a} & \ldots & \gamma_{pa} \\
\gamma_{1b} & \gamma_{2b} & \ldots & \gamma_{pb} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{1q} & \gamma_{2q} & \ldots & \gamma_{pq} \\
\gamma_{11} & \gamma_{21} & \ldots & \gamma_{pl} \\
\gamma_{12} & \gamma_{22} & \ldots & \gamma_{p2} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{1p} & \gamma_{2p} & \ldots & \gamma_{pp}
\end{bmatrix}
\]
and a matrix of rate coefficients be

\[ \Delta^t = \begin{bmatrix}
\delta_{1a} & \delta_{2a} & \ldots & \delta_{pa} \\
\delta_{1b} & \delta_{2b} & \ldots & \delta_{pb} \\
\vdots & \vdots & \ddots & \vdots \\
\delta_{1q} & \delta_{2q} & \ldots & \delta_{pq} \\
1 & \delta_{21} & \ldots & \delta_{p1} \\
\delta_{12} & 1 & \ldots & \delta_{p2} \\
\vdots & \vdots & \ddots & \vdots \\
\delta_{1p} & \delta_{2p} & \ldots & 1
\end{bmatrix} = \begin{bmatrix}
\Delta \xi \\
\Delta \eta
\end{bmatrix} \]

and diagonal matrices be

\[ D_\xi = \begin{bmatrix}
\xi_a & 0 & \ldots & 0 \\
0 & \xi_b & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \xi_q
\end{bmatrix} \quad D_\mu = \begin{bmatrix}
\mu_a & 0 & \ldots & 0 \\
0 & \mu_b & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \mu_q
\end{bmatrix} \]

\[ D_\eta = \begin{bmatrix}
\eta_1 & 0 & \ldots & 0 \\
0 & \eta_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \eta_p
\end{bmatrix} \quad D_\chi = \begin{bmatrix}
\chi_1 & 0 & \ldots & 0 \\
0 & \chi_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \chi_p
\end{bmatrix}. \]
Let us now consider the multivariate analogue of the single process law given by

\[
\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1} \\ \mathbf{H}_{2} \end{bmatrix} = \begin{bmatrix} (D_{1} - D_{\mu})^{-1} & \Delta_{1} \\ (D_{2} - D_{\mu})^{-1} & \Delta_{2} \end{bmatrix} \tag{11.1}
\]

The analogy in form to the univariate case is evident. Compare with equation (5.2).

A solution of (11.1) is

\[
\gamma_{1} = \alpha + \beta_{1}(\frac{q}{\mu})^{6}a_{1}(\frac{q}{\mu})^{6}b_{1} \cdots (\frac{q}{\mu})^{6}q_{1}(1)(\gamma_{2} - \gamma_{1})^{6}2 \cdots (\gamma_{p} - \gamma_{1})^{6}p \\
\gamma_{2} = \alpha + \beta_{2}(\frac{q}{\mu})^{6}a_{2}(\frac{q}{\mu})^{6}b_{2} \cdots (\frac{q}{\mu})^{6}q_{2}(1) \cdots (\gamma_{p} - \gamma_{2})^{6}p \\
\vdots \\
\gamma_{p} = \alpha + \beta_{p}(\frac{q}{\mu})^{6}a_{p}(\frac{q}{\mu})^{6}b_{p} \cdots (\frac{q}{\mu})^{6}q_{p}(1) \cdots (\gamma_{2} - \gamma_{p})^{6}p \cdots (1) \tag{11.2}
\]

Thus, equations (11.2) are the "structural equations" for the causal network arising from allowing each pathway to individually follow the single process law.

11.2 Identification

If one eliminates the \(\gamma\)'s from the right hand side of (11.2) in the spirit of the preceding chapter to form "reduced equations", the problem of identification arises once more. Generally, one does not place restrictions on the scale factors \(a\) or \(\mu\) when describing the causal network. When a particular pathway is non-existent, this fact is represented by setting the appropriate exponent \(a\) equal to zero. Hence, we will not be surprised to discover that the problem of identification
primarily concerns these exponents, although the constants of integration
($\beta$'s), signifying initial conditions of the system, also may be involved.

Let us transform (11.2) by taking logarithms. We set

\[ \gamma_1' = \log(\gamma_1 - \gamma) \quad \gamma_a' = \log(\gamma_a - \mu_a) \quad A_1 = \log \beta_1 \]
\[ \gamma_2' = \log(\gamma_2 - \gamma) \quad \gamma_b' = \log(\gamma_b - \mu_b) \quad A_2 = \log \beta_2 \]
\[ \vdots \quad \vdots \quad \vdots \]
\[ \gamma_p' = \log(\gamma_p - \gamma) \quad \gamma_q' = \log(\gamma_q - \mu_q) \quad A_p = \log \beta_p \]

and $A_{ij} = \delta_{ij}$. Then (11.2) becomes

\[ \gamma_1' = A_1 + A_1 \gamma_a' + A_1 \gamma_b' + \ldots + A_1 \gamma_q' + A_2 \gamma_1' + \ldots + A_p \gamma_p' \]
\[ \gamma_2' = A_2 + A_2 \gamma_a' + A_2 \gamma_b' + \ldots + A_2 \gamma_q' + A_2 \gamma_1' + \ldots + A_p \gamma_p' \]
\[ \vdots \]
\[ \gamma_p' = A_p + A_p \gamma_a' + A_p \gamma_b' + \ldots + A_p \gamma_q' + A_p \gamma_1' + A_p \gamma_2' + \ldots \]

Now the transformed structural equations (11.3) are identical with the
 corresponding linear structural equations (10.2) of the last chapter.
Since this is true, we can apply all of the results of the earlier
chapter regarding identification and reduction directly to the present
problem.
11.3 Some Examples

In this section we will consider the same path diagrams as considered in the linear case.

(1) The "multiple regression" analogue with diagram as follows

\[
\begin{align*}
\xi_a & \rightarrow \eta_1 \\
\eta_1 & \rightarrow \eta_2 \\
\eta_1 & \rightarrow \eta_2
\end{align*}
\]

The structural equation is then

\[
\eta_1 = \gamma_1 + \beta_1(\xi_a - \mu_a)\delta_{1a} (\xi_b - \mu_b)\delta_{1b}
\]  \hspace{1cm} (11.4)

(2) The second example with diagram

\[
\begin{align*}
\xi_a & \rightarrow \eta_1 \\
\eta_1 & \rightarrow \eta_2 \\
\eta_1 & \rightarrow \eta_2
\end{align*}
\]

and structural equations:

\[
\begin{align*}
\eta_1 &= \gamma_1 + \beta_1(\xi_a - \mu_a)\delta_{1a} (\xi_b - \mu_b)\delta_{1b} \\
\eta_2 &= \gamma_2 + \beta_2(\xi_b - \mu_b)\delta_{2b} (\eta_1 - \gamma_1)\delta_{21}
\end{align*}
\]  \hspace{1cm} (11.5)

yields by substitution or by the methods of the last chapter the reduced structural equations

\[
\begin{align*}
\eta_1 &= \gamma_1 + \beta_1(\xi_a - \mu_a)\delta_{1a} (\xi_b - \mu_b)\delta_{1b} \\
\eta_2 &= \gamma_2 + \beta_2\delta_{21}(\xi_a - \mu_a)\delta_{21}\delta_{1a} (\xi_b - \mu_b)\delta_{2b} + \delta_{21}\delta_{1b}
\end{align*}
\]  \hspace{1cm} (11.6)
(3) The feedback example

\[ \eta_1 = \eta_1 \left( \xi - \mu_a \right)^{\delta_1} \left( \eta_2 - \eta_1 \right)^{\delta_2} \]

produces structural equations

\[ \eta_1 = \eta_1 + \beta_1 \left( \xi_a - \mu_a \right) \delta_1 (\eta_2 - \eta_1)^{\delta_2} \]

\[ \eta_2 = \eta_2 + \beta_2 \left( \xi_b - \mu_b \right) \delta_2 (\eta_1 - \eta_2)^{\delta_1} \] (11.7)

and reduced structural equations

\[ \eta_1 = \eta_1 + \beta_1 \left( \xi_a - \mu_a \right)^{\delta_1} \left( \eta_2 - \eta_1 \right)^{\delta_2} \]

\[ \eta_2 = \eta_2 + \beta_1 \left( \xi_a - \mu_a \right)^{\delta_1} \left( \eta_2 - \eta_1 \right)^{\delta_2} \] (11.8)

Comparison of the results for these three examples with the corresponding linear results will serve to illustrate the analogies. The important point to note is that, under suitable conditions, reduced structural equations can be found, even in this nonlinear case, such that each secondary variable can be expressed as an explicit function of only the primary variables.
11.4 The Gaussian Iterant for Maximum Likelihood Estimators

In considering estimation of parameters for the linear case, we restricted our attention to systems in which reduced equations could be found and in which a state of under- or just-identifications existed. We similarly confine our consideration to these situations for the nonlinear case.

After reduction we have \( p \) independent nonlinear regression equations, each of the form

\[
\hat{y} = B_0 + B_1(x_a - M_a)^{D_a} (x_b - M_b)^{D_b} \ldots (x_q - M_q)^{D_q}.
\]

(11.9)

As before, we linearize by expanding in a Taylor's series and neglect terms beyond those containing the first derivatives. First, we must obtain trial estimates \( M_{a0}, M_{b0}, \ldots, M_{q0}, D_{a0}, D_{b0}, \ldots, D_{q0} \). We then compute

\[
X_{10} = (x_a - M_{a0})^{D_{a0}} (x_b - M_{b0})^{D_{b0}} \ldots (x_q - M_{q0})^{D_{q0}}
\]

\[
X_{20} = X_{10}/(x_a - M_{a0})
\]

\[
X_{30} = X_{10}/(x_b - M_{b0})
\]

\[
\vdots
\]

\[
X_{q+1,0} = X_{10}/(x_q - M_{q0})
\]

\[
X_{q+2,0} = X_{10} \log(x_a - M_{a0})
\]

\[
X_{q+3,0} = X_{10} \log(x_b - M_{b0})
\]

\[
\vdots
\]

\[
X_{2q+1,0} = X_{10} \log(x_q - M_{q0})
\]

(11.10)

for all observational vectors.
In addition, we define

\[ B_2 = -B_{10} D_{ao}(M_a - M_{ao}) \quad B_{q+2} = B_{10}(D_a - D_{ao}) \]
\[ B_3 = -B_{10} D_{bo}(M_b - M_{bo}) \quad B_{q+3} = B_{10}(D_b - D_{bo}) \]
\[ \vdots \]
\[ B_{q+1} = -B_{10} D_{qo}(M_q - M_{qo}) \quad B_{2q+1} = B_{10}(D_q - D_{qo}) \]  \hspace{1cm} (11.11)

Now we have by use of Taylor's series

\[ \hat{y} = \hat{y}_o + \hat{y}_{B_{10}}(B_o - B_{oo}) + \hat{y}_{B_{10}}(B_1 - B_{10}) + \hat{y}_{M_{ao}}(M_a - M_{ao}) \]
\[ + \hat{y}_{M_{bo}}(M_b - M_{bo}) \]
\[ \vdots \]
\[ + \hat{y}_{M_{qo}}(M_q - M_{qo}) \]
\[ + \hat{y}_{D_{a0}}(D_a - D_{ao}) \]  \hspace{1cm} (11.12)
\[ + \hat{y}_{D_{b0}}(D_b - D_{bo}) \]
\[ \vdots \]
\[ + \hat{y}_{D_{qo}}(D_q - D_{qo}) \]

and by substitution we obtain the linear equation

\[ \hat{y} = B_o + B_1 x_{10} + B_2 x_{20} + \cdots + B_{2q+1} x_{2q+1} \]  \hspace{1cm} (11.13)
as an approximate substitute for (11.9). By the usual methods of linear regression we now obtain the estimates $B_0, B_1, \ldots, B_{2q+1}$ and from (11.11) we get the improved estimates $M_a, \ldots, D_q$. These can now be used as our "trial" estimates and the entire process repeated until convergence.

If any subset of parameters is known a priori then, as in the simple regression case, the appropriate $X$'s are simply omitted and the analysis carried out without them. For example, if it is known that all processes have vertical asymptotes zero then we omit $X_{20}$ through $X_{q+1,0}$.

When there is more than a single response variable the application of (11.13) separately to each provides multiple estimates for the vertical asymptotes $M_a, M_b, \ldots, M_q$. In order to provide unique, efficient estimates of these values we fit, instead of (11.13), the family of hyperplanes

$$Y_i = B_{0i} + B_{1i}X_{1i} + B_{2i}X_{20i} + \ldots + B_{q+1i}X_{q+1,0i} + B_{q+2,i}X_{q+2,0i} + \ldots + B_{2q+1,i}X_{2q+1,0i}$$

(11.14)

where $i = 1, 2, \ldots, p$ and $B_2, B_3, \ldots, B_{q+1}$ are common for all $i$.

Linear regression theory provides the minimum variance unbiased estimators for fitting (11.14).

Asymptotic tests and confidence limits are found as before by applying linear results to the final iteration. Exact tests of the type discussed in Section 9.0 are also applicable to the general case of this chapter, but, of course, the numerical difficulties are much increased.
12.0 A SAMPLING INVESTIGATION

12.1 Objectives

In the case of models linear in the unknown parameters and with normally distributed errors, maximum likelihood estimators of the coefficients possess a number of desirable properties. Among these are **expectation unbiasedness** and **minimum variance**, properties possessed by estimates taken from samples of any size whatsoever. Maximum likelihood estimators for models nonlinear in the unknown parameters acquire these two properties only asymptotically. Not much is known about the behavior of the estimators for small samples in the nonlinear case. For large samples we expect the asymptotic results to approximately hold. Is it safe to rely upon the asymptotic results for small numbers of observations, say ten or twenty? The objectives of the study reported in this chapter are to explore the behavior of the maximum likelihood estimator for the exponential parameter $\delta$ occurring in the single process law (5.1) and to compare the results with those produced by use of the asymptotic theory of Section 7.0.

12.2 The Samples

Thirteen structural equations of the form

$$\eta = \kappa + \beta (\gamma - \gamma_0)^\delta$$

(12.1)

were selected for study. These thirteen equations correspond to thirteen distinct points in the two-dimensional nonlinear parameter space $(\gamma, \delta)$ and were chosen so as to more or less completely cover the useful parameter region. For convenience the constant $\kappa$ is taken in every case to
be zero and the scale factor $\beta$ is chosen such that $\gamma = 50$ when $\xi = 5.5$. Values of $\gamma$ have been computed for the ten values $\xi_1=1, \xi_2=2, \ldots, \xi_{10}=10$. Thus the thirteen curves intersect midway along their computed course. The selected sets of parameters and computed points are shown in Table 12.1 for each of the thirteen curves.

Simulated data were prepared by generating 98 sets of ten normal deviates for each of the thirteen curves. The 12,740 normal deviates were obtained by use of the IBM 650 Program developed in the N. C. State College computation laboratory (1958). These normal deviates were added to the computed values of Table 12.1 so as to give 98 sets of points for each of thirteen curves. That is, 1274 simulated experiments were obtained, 98 replications for each point in the parameter space. The convention in choice of the scale factor $\beta$ implies that the coefficient of variation is 2% for all curves at the midpoint $\xi = 5.5.

12.3 Estimation when $\mu$ is Known

Maximum likelihood estimates were obtained for all curves in the sample under the assumption of known $\mu$ where as before $\mu = \gamma \delta$. Analysis of the estimates of the exponential parameter $\delta$ has been made. In particular, evaluation of the estimated bias

$$\text{bias (d) = } \hat{\delta} - \delta$$  \hspace{1cm} (12.2)

---


3 Due to a mishap in card handling several cases had to be discarded. Five cases were omitted from each of curves No. 4 and No. 7 and one case omitted from each of curves No. 10 and No. 11. Thus 1262 simulated experiments were analyzed.
Table 12.1. Thirteen curves chosen for investigation showing computed values of \( \eta \)

<table>
<thead>
<tr>
<th>CURVE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>( \delta )</td>
<td>-1/2</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>+3</td>
<td>-1</td>
<td>-4</td>
<td>+4</td>
<td>-2</td>
<td>+4</td>
<td>+2</td>
<td>+2</td>
</tr>
<tr>
<td>1</td>
<td>117.26</td>
<td>162.90</td>
<td>651.60</td>
<td>180.50</td>
<td>327.26</td>
<td>867.33</td>
<td>106.25</td>
<td>164.19</td>
<td>110.10</td>
<td>112.30</td>
<td>117.41</td>
<td>113.20</td>
<td>102.04</td>
</tr>
<tr>
<td>2</td>
<td>118.92</td>
<td>108.33</td>
<td>514.24</td>
<td>125.35</td>
<td>207.16</td>
<td>599.44</td>
<td>85.00</td>
<td>122.07</td>
<td>280.10</td>
<td>92.12</td>
<td>118.74</td>
<td>118.34</td>
<td>88.89</td>
</tr>
<tr>
<td>3</td>
<td>67.70</td>
<td>81.25</td>
<td>169.62</td>
<td>92.09</td>
<td>131.12</td>
<td>276.06</td>
<td>70.83</td>
<td>92.63</td>
<td>183.78</td>
<td>75.31</td>
<td>94.47</td>
<td>95.86</td>
<td>76.64</td>
</tr>
<tr>
<td>4</td>
<td>58.63</td>
<td>65.00</td>
<td>99.43</td>
<td>70.51</td>
<td>89.17</td>
<td>114.73</td>
<td>60.71</td>
<td>71.56</td>
<td>114.73</td>
<td>63.28</td>
<td>74.13</td>
<td>75.74</td>
<td>64.31</td>
</tr>
<tr>
<td>5</td>
<td>52.44</td>
<td>51.17</td>
<td>62.07</td>
<td>55.71</td>
<td>60.36</td>
<td>72.51</td>
<td>53.12</td>
<td>56.15</td>
<td>67.25</td>
<td>53.92</td>
<td>57.26</td>
<td>57.99</td>
<td>54.88</td>
</tr>
<tr>
<td>6</td>
<td>44.87</td>
<td>46.43</td>
<td>40.72</td>
<td>45.12</td>
<td>41.60</td>
<td>33.83</td>
<td>47.22</td>
<td>44.67</td>
<td>36.30</td>
<td>46.49</td>
<td>43.45</td>
<td>42.60</td>
<td>45.35</td>
</tr>
<tr>
<td>7</td>
<td>55.32</td>
<td>40.62</td>
<td>24.47</td>
<td>32.29</td>
<td>29.15</td>
<td>14.57</td>
<td>42.50</td>
<td>35.98</td>
<td>17.51</td>
<td>40.50</td>
<td>32.30</td>
<td>29.59</td>
<td>36.74</td>
</tr>
<tr>
<td>8</td>
<td>44.46</td>
<td>36.11</td>
<td>19.64</td>
<td>31.31</td>
<td>20.74</td>
<td>5.68</td>
<td>38.64</td>
<td>29.31</td>
<td>7.17</td>
<td>35.60</td>
<td>23.45</td>
<td>18.91</td>
<td>29.02</td>
</tr>
<tr>
<td>10</td>
<td>37.08</td>
<td>29.51</td>
<td>10.60</td>
<td>23.02</td>
<td>10.93</td>
<td>0.57</td>
<td>32.69</td>
<td>20.02</td>
<td>4.45</td>
<td>28.12</td>
<td>11.31</td>
<td>4.73</td>
<td>16.33</td>
</tr>
</tbody>
</table>
and of the estimated variance

\[ s^2(d) = n \sum d^2 - (\sum d)^2 / n(n-1) \]  

(12.3)

has been made. The results are shown in Table 12.2. Confidence limits have been obtained for the bias by use of Student's t-distribution. The upper and lower 95% points are given in the table. We note that in three of the cases biases are significantly different from the asymptotic expectation of zero. However, even in these cases the bias is quite small, being in every case less than 0.3%. It appears that under the conditions assumed here, bias is of little consequence in the region of the parameter space studied.

Confidence limits for the variance of estimate \( d \) have been computed using the \( \chi^2 \) distribution and upper and lower 95% points are reported also in Table (12.2). No case has a variance significantly greater than the asymptotic value. For ease of discussion the "finite efficiencies", found by dividing the asymptotic variances by the estimated variances, are shown. Judging by the lower 95% limit the finite efficiency may in some cases be as low as 66%; however, this is unlikely to be so and we have little reason to reject 100% as a sufficient approximation in all cases studied. Thus, the asymptotic variance would seem to provide a useful basis for construction of confidence limits under the conditions examined.

12.4 Estimation when \( \mu \) is Not Known

The constructed samples for curves 2 and 3 have been analyzed under the assumption that \( \mu \) was unknown and hence must be estimated simultaneously with \( \gamma \) and the linear parameters \( \alpha \) and \( \beta \). The results for the
Table 12.2. Estimated biases and variances of the maximum likelihood estimate for delta

<table>
<thead>
<tr>
<th>( \mu )</th>
<th>Curve</th>
<th>Parameters</th>
<th>Bias of Estimate ( \delta )</th>
<th>Variance of Estimate ( \delta )</th>
<th>Finite Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>( \delta )</td>
<td>( \mu = \delta^a )</td>
<td>lower point 9% estimate</td>
<td>upper point 9% estimate</td>
<td>lower point 9% estimate</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>-1/2</td>
<td>0</td>
<td>-.0003</td>
<td>+.0031</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-.0025</td>
<td>+.0009</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-4</td>
<td>-4</td>
<td>+.0053</td>
<td>+.0085*</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-2</td>
<td>-4</td>
<td>-.0118</td>
<td>-.0026</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-8</td>
<td>-16</td>
<td>-.0025</td>
<td>+.0051</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>+8</td>
<td>+16(-5)</td>
<td>+.0014</td>
<td>+.0035*</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>-1</td>
<td>-3</td>
<td>-.0001</td>
<td>+.0084</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>-4</td>
<td>-12</td>
<td>-.0180</td>
<td>-.0082</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>+4</td>
<td>+12(-1)</td>
<td>+.0004</td>
<td>+.0024*</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>-2</td>
<td>-3</td>
<td>-.0119</td>
<td>+.0016</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>+4</td>
<td>+20(-9)</td>
<td>-.0260</td>
<td>-.0070</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>+2</td>
<td>+12(-1)</td>
<td>-.0066</td>
<td>-.0029</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>+2</td>
<td>+16(-5)</td>
<td>-.0101</td>
<td>-.0008</td>
</tr>
</tbody>
</table>

\( ^a \) There is an equivalence between certain negative values of \( \mu \) and certain other positive values due to the convention of taking \((x - \mu)\) positive. Negative equivalents are given in parentheses for all positive \( \mu \).

\( ^b \) An asterisk signifies significant departure from asymptotic expectation zero at the 5% level.
criterion parameter $\delta$ are shown at the bottom of Table 12.2. A small but significant bias was discovered for curve 3. It is interesting to note that the bias is negative when $\mu$ is unknown but is positive when $\mu$ is known. The bias is much greater in the former case than in the latter. For unknown $\mu$ the bias is estimated to be 1.2% of $\delta$ whereas for known $\mu$ the percentage is estimated to be 0.2%.

Again the finite efficiencies do not depart significantly from the asymptotic value of 100%. As expected the asymptotic variances are considerably increased in the case that $\mu$ is not known a priori. This increase is about 50 fold for curve 2 and about 200 fold for curve 3.

Only limited experience has been obtained for the case of $\mu$ unknown but so far as these limited results go it appears that the asymptotic variance may provide a reasonable guide to the precision of the estimate of $\delta$, even when the value of $\mu$ is not specified by the model.

12.5 Discussion and Summary

Constructed curves covering the "useful" range of the parameters and consisting of a series of observations of intermediate length ($n=10$) and with a reasonable rate of error (2% at the midpoint) have been subjected to maximum likelihood analysis. In order to check on the form of the curve a shorter series than one of length 7 or 8 seems inadvisable. A length of 10 is still better. Longer series yet supply little advantage not gained by replicating the original series. An advantage of replication is that the first exact test of Chapter VI can be employed. In many cases the error rate is greater than that studied here. Replication
provides a method of reducing the intrinsic error rate to any desired level. A rate of about 2% with 10 equidistant points is required in order that \( \delta \) be determined to the nearest integer when \( \mu \) is not known a priori. Equidistant points have been used in the study since this arrangement plausibly has best average efficiency for discriminating against unspecified alternative models. If the model is known to be of the single process form then more efficient designs can be found. See Box and Lucas (1959).

The estimated sampling variances, resulting from estimating the criterion parameter \( \delta \) for each of the constructed curves, did not differ significantly from the asymptotic value. We concluded therefore that under the studied conditions, which as seen above are quite realistic ones, the asymptotic value seems a safe guide as to the precision of the criterion parameter and hence the proper curve can be selected from experimental results if these have sufficient precision and about 10 equidistant points covering the main course of the curve are employed.

The biases were of an unimportant magnitude for purposes of model-building.
13.0 SUMMARY AND CONCLUSIONS

A class of nonlinear regression models, possessing very interesting and useful analytic properties, has been described in some detail. A single exponential parameter serves to distinguish the various members of the class. Since such frequently occurring processes as the linear process, parabolic process, exponential process, and inverse-square process are members of the class it is suggested that the general model be called the single process law. The appropriateness of the name is further suggested by the simple differential equation used to generate the model. The single process law model is one that we refer to as an essentially nonlinear model, which is to say that classical linear regression techniques are not available. After a brief resume of some important distinctions between linear and nonlinear models iterative methods of obtaining maximum likelihood estimates and asymptotic confidence limits for the parameters in the single process law are developed.

Next we consider the great potentialities of the single process law for the process of model-building. This potentiality is illustrated by the development of an original theory of world track records and the theory is applied to the official records. It is suggested as a conclusion that the average excess velocity (over the asymptotic value) of a middle or long distance runner is inversely proportional to the $3/7$\textsuperscript{th} power of the time of the race.

Some tests of significance and derived confidence limits for small samples are discussed. One such test based upon a parabolic expansion appears not to have been previously reported. Confidence limits obtained from this test are applied to the track record example.
A summary of some results in linear path regression analysis and a nonlinear extension based upon the single process law are given. Computational procedures are indicated for the nonlinear case which are simple generalizations of those for the univariate situation.

A sampling study for the purpose of exploring the adequacy of asymptotic results for small samples has been carried out. Within the limitations of the study it appears that for the purpose of model-building, that is for selecting the appropriate process, the asymptotic results may well be quite adequate.

We generally conclude that the technique of the single process law is a useful and manageable tool for analyzing problems of the monotonic growth or extinction variety, whether the problem is of the simple two-dimensional type or whether complex multi-dimensional causal networks are involved.
14.0 LIST OF REFERENCES


APPENDIX A. A POLYNOMIAL APPROXIMATION

When \(|x| < |\mu|\) for all values of \(x\) required the model

\[
y = \alpha + \beta(x-\mu)^5 + \varepsilon
\]

can be expanded in a converging series

\[
y = \alpha + \beta \left\{ (-\mu)^5 + 5(-\mu)^4x + (1/2)5(5-1)(-\mu)^3x^2 + (1/6)5(5-1)(5-2)(-\mu)^2x^3 + \ldots \right\} + \varepsilon
\]

or

\[
y = A_0 + A_1x + A_2x^2 + A_3x^3 + \ldots + \varepsilon
\]

where

\[
A_0 = \alpha + \beta(-\mu)^5
\]

\[
A_1 = \beta5(-\mu)^4
\]

\[
A_2 = \frac{1}{2}\beta5(5-1)(-\mu)^3
\]

\[
A_3 = \frac{1}{6}\beta5(5-1)(5-2)(-\mu)^2
\]

\[
\vdots
\]

If \(|x| \ll |\mu|\) we may terminate the series after the cubic term. In this case we have

\[
\alpha = A_0 - A_1^2 A_2 / 2Q_1
\]

\[
\beta = \frac{A_1 Q_2}{Q_1} \left( \frac{2Q_2}{A_1 A_2} \right) A_2^2 / Q_2
\]

\[
\delta = Q_1 / Q_2
\]

\[
\mu = -A_1^2 A_2 / 2Q_2
\]
where

\[ Q_1 = 2A_2^2 - 3A_1A_3 \]
\[ Q_2 = A_2^2 - 3A_1A_3 \]

An estimate of \( \delta \) is thus given by

\[ \delta = \frac{\hat{A}_2^2 - 3\hat{A}_1\hat{A}_3}{\hat{A}_2^2 - 3\hat{A}_1\hat{A}_3} \]

where \( \hat{A}_1 \), \( \hat{A}_2 \), and \( \hat{A}_3 \) are the usual least squares estimates obtained from the cubic approximation.

The asymptotic variance for the "cubic" estimate of \( \delta \) is found to be

\[ V(\delta) = \frac{(Q_1 - Q_2)^2}{Q_2^4} \left\{ 9A_3^2\sigma_{11} + 16A_2^2\sigma_{22} + 9A_1^2\sigma_{33} - 2\mu A_2A_3\sigma_{12} + 18A_1A_3\sigma_{13} - 2\mu A_1A_2\sigma_{23} \right\} \]

where \( \sigma_{ij} \) is the covariance of \( A_i \) and \( A_j \).
APPENDIX B. A SIMPLIFIED ESTIMATOR

The differential equation for the single process law was stated to be

\[ D = \frac{d\Psi}{d\varphi} = g \frac{\varphi - \Psi}{\varphi - \Psi} \]

If values of \( \varphi \) are chosen to be equidistant (\( \Delta = 1 \) without loss of generality) then \( D \) may be estimated by passing a parabola exactly through three successive values of \( y \) and evaluating the derivative at the first. By following this procedure we arrive at the approximate relationship

\[ y_i = x_i D_i = 26y + 26y_i + \mu_i \]

for \( i = 1, 2, \ldots, n-2 \) and where \( D_i = -3y_i + \frac{1}{2}y_{i+1} - y_{i+2} \). The parameter \( \delta \) could be estimated by regressing \( y_i \) on \( y_i \) and \( D_i \). However, successive values of \( y_i \) are correlated and a proper procedure would involve these correlations. In fact the matrix of variances and covariances is found to be

\[
\Sigma = \begin{bmatrix}
A_1^2 + B_1^2 + C_1^2 & A_2 B_1 + B_2 C_1 & A_3 C_1 & 0 & 0 & \cdots \\
A_2 B_1 + B_2 C_1 & A_2^2 + B_2^2 + C_2 & A_3 B_2 + B_3 C_2 & A_4 C_2 & 0 & \cdots \\
A_3 C_1 & A_3 B_2 + B_3 C_2 & A_3^2 + B_3^2 + C_3 & A_4 B_3 + B_4 C_3 & A_5 C_3 & \cdots & \sigma^2 \\
0 & A_4 C_2 & A_4 B_3 + B_4 C_2 & A_4^2 + B_4^2 + C_4 & A_5 B_4 + B_5 C_4 & \cdots \\
0 & 0 & A_5 C_3 & A_5 B_4 + B_5 C_4 & A_5^2 + B_5^2 + C_5 & \cdots \\
& \vdots & & \vdots & & \vdots \\
& \vdots & & \vdots & & \vdots \\
\end{bmatrix}
\]
where

\[
A_i = -3x_i - 2\mu + 3
\]

\[
B_i = 4x_i - 4\mu
\]

\[
C_i = -x_i + \mu
\]

A possible estimation procedure would be to approximate \( \Sigma \) by substitution of trial values of \( \mu \) and \( \sigma \), and then to estimate the regression coefficients by generalized least squares. Hence

\[
\hat{B} = \left( x' \Sigma^{-1} x \right)^{-1} x' \Sigma^{-1} y
\]

where \( \hat{\Sigma} = (-2\hat{\mu}, 2\hat{\mu}, \hat{\mu}) \). This procedure is even more tedious than direct application of maximum likelihood methods as described in Chapter III.

A simplified procedure might have some merits. For example we could take all off-diagonal elements of \( \Sigma \) to be zero as an approximation. In this case the above procedure reduces to ordinary weighted regression analysis. Weights then are simply

\[
w_i = 1/(A_i^2 + B_i^2 + C_i^2)
\]

A still simpler approximate method is found by taking the weights to be constant. The properties of these procedures have not been investigated.