REGRESSION WHEN BOTH VARIABLES ARE SUBJECT TO ERROR AND THE RANKS OF THEIR MEANS ARE KNOWN

BY

JAMES H. WARE

TECHNICAL REPORT NO. 15
NOVEMBER 20, 1969

SUPPORTED BY PUBLIC HEALTH SERVICE
GRANT USPHS-GM-14554-03

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
REGRESSION WHEN BOTH VARIABLES ARE SUBJECT TO
ERROR AND THE RANKS OF THEIR MEANS ARE KNOWN

By

JAMES H. WARE

Technical Report No. 15
November 20, 1969

Prepared under U. S. Public Health
Service Grant USPHS GM-14554-03

Reproduction in whole or in part is permitted
for any purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Abstract.

Consider a particle moving in two dimensional space along the path \( y = \alpha + \beta x \) according to some location function \( r(t) \). Suppose that \( M \) observations \( (x, y) \) of the particle's location are available over a fixed segment of the path, each subject to circular normal error, and that the observations are received sequentially as the particle moves along the path. This is a regression problem when both variables are subject to error, with the additional information that the observation means are in known order.

In this dissertation, it is proved that this additional information makes possible a class of estimators of the path, to be known as the grouped orthogonal least squares estimates, which are asymptotically efficient, in the sense that the asymptotic variance of the angular error \( \sqrt{M}(\hat{\beta}_M - \beta) \) where \( \beta = \tan^{-1}\beta \) achieves the smallest attainable value.

The full statement of the problem is given in the introduction and summary, and alternative solutions are proposed. The antecedents of the problem are given in Chapter I which also contains asymptotic distribution theory for the various solutions, and the results on optimality of the grouped orthogonal least squares estimators. Chapter II is a discussion in terms of applications, including moderate sample size considerations, some Monte Carlo results, and recommendations for choosing an estimator.

In Chapter III, we consider estimation of the other parameters, deviations from the assumptions, and a simple multivariate extension.
In Chapter IV, we consider the generalization of the tracking problem in which the ranks of the observation means are not known exactly but only partially through some independent stochastic information.
Acknowledgment

I am deeply indebted to Associate Professor Bradley Efron for suggesting the problem treated in this dissertation and for his guidance during its preparation. I am especially thankful for his sharp insight and remarkable patience.

Professor Ingram Olkin and Associate Professor Paul Switzer gave the manuscript a careful and conscientious reading and their comments were invaluable.

Finally, Margaret Lof remained a good humored and accurate typist during preparation and endless revision of the final version of the dissertation.
Table of Contents

Introduction

Chapter I
A. A Historical Digression
B. The Problem is Solved. Asymptotic Distribution Theory for Alternative Estimates of the Path
C. Asymptotic Optimality of Estimation Based Upon Grouping

Chapter II: Small Sample Considerations. Choosing an Estimate
A. Introductory Remarks
B. Some Monte Carlo Work
C. Distribution Theory for Finite Sample Sizes
D. Choosing an Estimate of θ.

Chapter III: Some Additional Results
A. Estimating Other Parameters
B. Estimating θ When the Error is not Circular Normal
C. Extending the Tracking Problem to 3 Dimensions
D. Partial Information on Ranks of Means Through Multiple Observations.

Chapter IV: Partial Ranking Problem
A. Introduction
B. The Tracking Problem When the Information on Ranks has Normally Distributed Error
C. The Tracking Problem When the Information on Ranks has Exponentially Distributed Error
D. Comparison of the Estimates
E. Efficient Estimation and the Cramér Rao Lower Bound
F. Estimating θ
Introduction and Summary: The Problem

This dissertation concerns a special version of the regression problem when both variables are subject to error, to be referred to as the tracking problem. In the tracking problem, the true centers of the various observations will be assumed to have known order along the path. It is this problem which is formulated in this introduction and solved in Chapters I and II and III. In Chapter IV, we consider a modified problem in which the order of the true centers along the path must itself be estimated by some independent stochastic information.

Suppose a particle is moving along a linear path $y = \alpha + \beta x$ in two dimensional space according to some unspecified location function $r(t)$ with positive velocity $(r'(t) > 0)$ where $t$ can be taken to represent time. Suppose further that during a time interval $[T_0, T_1]$ observations are taken on the particle at equidistant points in time, each observation subject to circular normal error, with the errors independent. If the observations are received sequentially as the particle moves along the path, the tracking problem involves regression when both variables are subject to error, and the statistician has available in addition the order of the observation centers along the path.

For the univariate problem, we assume without loss of generality that the time period in question is $[0, 1]$. Let $M$ observations be taken, assuming again without loss of generality that they are taken at times
\[ t_i = i/M \quad i = 1, 2, \ldots, M \]

\[ r_i = r(t_i) \quad i = 1, 2, \ldots, M. \]

If the path in question is \( y = \alpha + \beta x \) with \( \theta = \tan^{-1} \beta \), we have observations

\[ (x_i, y_i) \quad i = 1, 2, \ldots, M \]

where

\[ x_i = r_i \cos \theta + \delta_i \quad i = 1, 2, \ldots, M \]

\[ y_i = \alpha + r_i \sin \theta + \epsilon_i \]

and

\[ \begin{pmatrix} \delta_i \\ \epsilon_i \end{pmatrix} \sim N \begin{pmatrix} 0 \\ \sigma^2 I \end{pmatrix} \quad i = 1, 2, \ldots, M \]

\[ I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]

and the errors are mutually independent.

The \( r_i \) are not assumed to be equally spaced along the line and, as \( M \) increases, the observation centers become dense along a fixed segment of the line.

Although regression problems are customarily discussed in terms of estimating \( \beta \), we shall be seeking an estimate of \( \theta = \tan^{-1} \beta \) primarily to facilitate discussion of various optimality properties. This will become clear during the analysis.
Let \( \hat{\theta}_M \) be a sequence of estimates of \( \theta \), for \( M = 1, 2, \ldots \).

Definition IS.1: If \( \hat{\theta}_M \) has the property that the distribution of 
\( (\hat{\theta}_M - \theta) \) is independent of \( \theta \) for all \( M \), we shall say that the 
sequence of estimates is coordinate free.

In such cases, in order to simplify the analysis, a special 
coordinate system will be chosen so that the path is parallel to \( x \) 
axis. This can be done without loss of generality, provided that 
\( \hat{\theta}_M - \theta \) is the quantity of interest. Then we have observations

\[
(x_i, y_i) \quad i = 1, 2, \ldots, M
\]

where

\[
x_i = r_i + \delta_i \quad \text{for all } i
\]

\[
y_i = \epsilon_i
\]

and \( (\delta_i, \epsilon_i) \) are as before.

The central result of this dissertation is that the information 
on the order of the centers available in this version of the tracking 
problem makes possible a class of coordinate free estimates of \( \theta \),
to be known as the grouped orthogonal least squares estimates, which 
achieve minimum asymptotic variance. We will also consider five
alternative natural solutions for the tracking problem. These five
solutions are presented in 5 subsections which follow this paragraph,
along with their principal properties. The detailed discussion is
left for the body of the dissertation. Let
\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]

\[ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \]

In all cases, we estimate the path by the line

\[(y - \bar{y}) = (x - \bar{x}) \tan \hat{\theta}\]

the line through \((\bar{x}, \bar{y})\) with slope \(\tan \hat{\theta}\) where \(\hat{\theta}\) is some estimate of \(\theta\).

(a). The Orthogonal Least Squares Line.

It is well known that when information on the order of the centers is not available, the maximum likelihood estimator of \(\theta\) (call it \(\hat{\theta}_M^{(1)}\)), or equivalently the orthogonal least squares line, is found by estimating \(\theta\) from the equation

\[
\tan 2\hat{\theta}_M^{(1)} = \frac{\sum_{i=1}^{M} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{M} (x_i - \bar{x})^2 - \sum_{i=1}^{M} (y_i - \bar{y})^2}
\]

we assume without loss of generality that \(\theta = 0\), as this procedure is coordinate free. [See [5] pp. 275.6 for a detailed discussion.]

This procedure minimizes residual orthogonal distance from the points \((x_i, y_i)\) to the line.

Definition IS.2: The sequence of estimates \((y - \bar{y}) = (x - \bar{x}) \tan \hat{\theta}_M^{(1)}\) of the line \(y = \alpha + \beta x\) will be known as the orthogonal least squares line.
(b). The Grouped Orthogonal Least Squares Line.

To define the class of grouped orthogonal least squares estimates, let \( g(t) \) be a (grouping) function defined on \([0, 1]\) such that

\[
g(t) \in c[0, 1] \\
g(t) > 0 \text{ for all } t \\
\int_0^1 g(t)dt = 1
\]

and define

\[
G(t) = \int_0^t g(u)du.
\]

Now define the \( i^{th} \) group to be those observations received during the time interval

\[
[G^{-1}(i-1/k), G^{-1}(i/k)] = d_i
\]

where we assume that \( k \), the number of groups, is given. Let

\[
n_i = \text{the number of observations received during } d_i.
\]

Define

\[
\bar{x}_i = \frac{\sum_{t \in d_i} x_j}{n_i} \\
\bar{y}_i = \frac{\sum_{t \in d_i} y_j}{n_i} \\
\bar{r}_i = \frac{\sum_{t \in d_i} r_j}{n_i}.
\]
Note that \( k \) is chosen for each \( M \) so that \( k/M \to 0 \) as \( M \to \infty \), then

\[
\min_{i=1,2,\ldots,k} \left[ n_i \right] \to \infty \text{ as } M \to \infty.
\]

Although the numbers

\[
\bar{x}_i, \bar{y}_i, \bar{r}_i \quad i = 1, 2, \ldots, k
\]

depend upon \( kM \), and the grouping function \( g(t) \), we shall keep this in mind without explicit reference in the notation. Finally, we estimate \( \theta \) by \( \hat{\theta}_M^{(2)} \) from the equation

\[
\frac{\tan 2\hat{\theta}_M^{(2)}}{2} = \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2}.
\]

This line results when we minimize residual orthogonal distance to the points

\[
(\bar{x}_i, \bar{y}_i) \quad i = 1, 2, \ldots, k
\]

with weights proportional to the number of observations in the group.

**Definition IS.3:** The sequence of estimates

\[
(y - \bar{y}) = \hat{\theta}_M^{(2)}(x - \bar{x}) \quad M = 1, 2, \ldots
\]

of the line \( y = \alpha + \beta x \) will be known as the **grouped orthogonal least squares line.**

Now, the line \( (y - \bar{y}) = \hat{\theta}_M^{(2)}(x - \bar{x}) \) depends upon the choice of \( g(t) \), the grouping function. The estimator resulting when \( g(t) = 1 \) was originally proposed by Efron in [6]. The procedure is coordinate free and we can assume \( \theta = 0 \) without loss of generality.
(c). The Linear Trend Estimate.

If we assume that as the particle moves along the path \( y = \alpha + \beta x \), the path function \( r(t) \) will be approximately linear during the period of observation. In particular, suppose

\[
\begin{align*}
  x_i &= a + b_i + \delta_i, \\
  y_i &= c + d_i + \epsilon_i, \quad i = 1, 2, \ldots, M
\end{align*}
\]

where

\[
\begin{pmatrix}
  \delta_i \\
  \epsilon_i
\end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 I \right).
\]

In this case, an alternative estimate of \( \theta \) (call it \( \hat{\theta}^{(3)}_M \)) would be constructed by

\[
\hat{b}_M = \frac{\sum_{i=1}^{M} (x_i - \bar{x})(1 - i)}{\sum_{i=1}^{M} (1 - i)^2}
\]

\[
\hat{d}_M = \frac{\sum_{i=1}^{M} (y_i - \bar{y})(1 - i)}{\sum_{i=1}^{M} (1 - i)^2}
\]

where

\[
\bar{I} = \frac{\sum_{i=1}^{M} i}{M} = \frac{M+1}{2}
\]

and

\[
\hat{\beta}^{(3)}_M = \frac{\hat{d}_M}{\hat{b}_M}
\]
\[
\sum_{i=1}^{M} (y_i - \bar{y})(i - \bar{i}) = \frac{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{i})}{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{i})}
\]

and

\[
\hat{\theta}_M^{(3)} = \tan^{-1}\hat{\theta}_M^{(3)}.
\]

Definition IS.4: The sequence of estimates

\[
(y - \bar{y}) = (x - \bar{x})\tan \hat{\theta}_M^{(3)}
\]

of the line \(y = \alpha + \beta x\) will be known as the \textit{linear trend estimate}.

The linear trend estimate is a sequence of estimates whose angular error \((\hat{\theta}_M^{(3)} - \theta)\) is independent of \(\theta\) in distribution, since \(\hat{a}, \hat{b}, \hat{c}, \hat{d}\) are chosen to minimize a coordinate free quantity

\[
\sum_{i=1}^{M} [(x_i - \alpha - \beta i)^2 + (y_i - \gamma - \delta i)^2].
\]

Thus, in discussing the linear trend estimate we shall assume \(\theta = 0\).


Suppose the statistician groups the observations as for the grouped orthogonal least squares estimate, but then proceeds as though the \(x\) observations are not subject to error. The resulting line, which minimizes residual vertical distance to the group means, uses the estimate

\[
\hat{\beta}_M^{(4)} = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - x)(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}
\]
and

\[ \hat{\beta}^{(4)}_M = \tan^{-1} \hat{\beta}^{(4)}_M . \]

**Definition IS.5:** The sequence of estimates

\[ (y - \bar{y}) = \hat{\beta}^{(4)}_M (x - \bar{x}) \]

of the line \( y = \alpha + \beta x \) will, for any \( g(t) \), be known as the **grouped naive least squares line**.

\( (e) \). The **Naive Least Squares Estimate**.

If the statistician chose not to group and behaved as though the \( x \) coordinate were observed without error, he would use the estimate

\[ \hat{\beta}^{(5)}_M = \frac{\sum_{i=1}^{M} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{M} (x_i - \bar{x})^2} \]

and

\[ \hat{\beta}^{(5)}_M = \tan^{-1} \hat{\beta}^{(5)}_M . \]

**Definition IS.6:** The sequence of estimates

\[ (y - \bar{y}) = (x - \bar{x})\tan \hat{\beta}^{(5)}_M \]

of the line \( y = \alpha + \beta x \) will be known as the **naive least squares line**.

This line minimizes residual vertical distance.

These, then, are the five alternative strategies for estimating the path, and the idea of real interest is that of grouping. Chapter I
contains the asymptotic properties of the various estimates of \( \theta \). In particular, \( \hat{\theta}_M^{(2)} \) and \( \hat{\theta}_M^{(4)} \), the two estimates based upon grouping, are shown to be asymptotically efficient in the sense that both are asymptotically normal and achieve asymptotic variance equal to a lower bound of Cramér-Rao type. \( \hat{\theta}_M^{(3)} \) is also asymptotically efficient in this sense if \( r(t) \) is linear. \( \hat{\theta}_M^{(1)} \) is not asymptotically efficient, and \( \hat{\theta}_M^{(5)} \) is not consistent.

In Chapter II, we turn to the small sample properties of the various estimates of \( \theta \), and include some Monte Carlo work. The sequence \( \hat{\theta}_M^{(4)} \) is shown to have undesirable small sample properties and the sequence \( \hat{\theta}_M^{(3)} \) is shown to have asymptotic variance which is stable against nonlinearity of \( r(t) \). It is shown that there exist special cases in which the estimate \( \hat{\theta}_M^{(3)} \) is superior to \( \hat{\theta}_M^{(2)} \), in the sense that \( \hat{\theta}_M^{(3)} \) approaches its asymptotic distribution "more rapidly" than does \( \hat{\theta}_M^{(2)} \). Guidelines are set for choosing between these estimates in practice. In Chapter III, we consider estimation of the other parameters, deviations from the assumptions and a simple multivariate extension.

Chapter IV is devoted to a modification of the tracking problem in which the ranks of the true centers are not known exactly. The problem is formulated in the introduction to that chapter.
Chapter I

A. A Historical Digression.

We begin this chapter with some brief comments on the historical setting for this work. Consider M pairs of random variables

$$(x_i, y_i) \quad i = 1, 2, \ldots, M.$$  

Let the expected value of $x_i$ be denoted by

$$E(x_i) = X_i \quad i = 1, 2, \ldots, M$$

and similarly denote

$$E(y_i) = Y_i \quad i = 1, 2, \ldots, M.$$  

Let

$$\epsilon_i = y_i - Y_i$$

$$\delta_i = x_i - X_i \quad i = 1, 2, \ldots, M.$$  

$$\text{Var}(\epsilon_i) = \sigma_\epsilon$$

$$\text{Var}(\delta_i) = \sigma_\delta$$

Further suppose

I. The random variables $\epsilon_1, \ldots, \epsilon_M$ each have the same distribution and are uncorrelated.

II. The random variables $\delta_1, \ldots, \delta_M$ each have the same distribution and are uncorrelated.
III. The random variables $\epsilon_i$ and $\delta_j$ ($i, j = 1, 2, \ldots, M$) are uncorrelated.

IV. A single linear relation holds between the true values $X$ and $Y$, i.e.

$$Y_i = \alpha + \beta X_i \quad i = 1, 2, \ldots, M.$$ 

This is the regression problem when both variables are subject to error, as formulated by Abraham Wald [1]. Wald sought consistent estimators of $\alpha$, $\beta$, $\sigma_e$, $\sigma_\delta$ through the following procedure. Let

$$X_{[i]} \quad i = 1, 2, \ldots, M$$

denote the ordered $x$ values so that

$$x_{[i]} \leq x_{[j]} \iff i \leq j \quad i, j = 1, 2, \ldots, M$$

and let

$$Y_{[i]} \quad i = 1, 2, \ldots, M$$

be the $y$ values ordered so that $X_{[i]}$ and $Y_{[i]}$ belong to the same pair. Suppose

$$M/2 = n$$

an integer. If

$$\gamma_1 = \frac{(x_{[1]} + \ldots + x_{[n]}) - (x_{[n+1]} + \ldots + x_{[M]})}{M}$$

and

$$\gamma_2 = \frac{(y_{[1]} + \ldots + y_{[n]}) - (y_{[n+1]} + \ldots + y_{[M]})}{M}$$

12
then Wald proposed that $\beta$ be estimated by the estimator

$$\hat{\beta} = \gamma_2 / \gamma_1.$$  

Wald proved that $\hat{\beta}$ is a consistent estimator of $\beta$ if

V. The limit inferior of

$$\frac{(X_{[1]} + \ldots + X_{[n]}) - (X_{[n+1]} + \ldots + X_{[M]})}{M}$$

is positive as $M \to \infty$ through the even integers.

Of course, the mean values $X_{[1]}$ are not observable. Hence the procedure of Wald hinges upon the property that condition V is satisfied when the $X_{[1]}$ are not observable.

A paper of J. Neyman and Elizabeth L. Scott [2] focused on the difficulties inherent in assigning the observations $(x_1, y_1)$ to the two groups on the basis of the magnitude of $x_1$. In the formulation of Neyman and Scott, the $x_1$ were taken to be $M$ independent and identically distributed observations from some distribution, while groups 1 and 2 are defined as a proportion $p_1$ so that the group 1 $x$ values consist of

$$X_{[1]}, \ldots, X_{[p_1 M]}$$

and a proportion $p_2$ such that the group 2 $x$ values consist of

$$X_{[(1-p_2)M+1]} + \ldots + X_{[M]}$$
that is, the $p_1 M$ smallest and the $p_2 M$ largest values. Now let

$$(\bar{x}_1, \bar{y}_1), (\bar{x}_2, \bar{y}_2)$$

be the group averages for groups 1 and 2. Define

$$\gamma_3 = \frac{\bar{x}_2 - \bar{x}_1}{(p_1 + p_2)M}$$

$$\gamma_4 = \frac{\bar{y}_2 - \bar{y}_1}{(p_1 + p_2)M}$$

and

$$\hat{\beta}_{p_1, p_2} = \gamma_4 / \gamma_3.$$
In another formulation of the regression problem, with

\[ X_i \sim N(0, \sigma_x^2) \]

\[ \delta_i \sim N(0, \sigma_\delta^2) \quad i = 1, 2, \ldots, M \]

\[ \epsilon_i \sim N(0, \sigma_\epsilon^2) \]

Reiersol [3] pointed out that the parameter \( \beta \) is not identifiable.

Thus, investigation of the regression problem when both variables are subject to error has involved the addition of various structural assumptions about the observations. The many directions in which research has proceeded are summarized extensively in [4].

This paper focuses upon the idea implicit in Wald's work that information about the ranks of the variables \( X_i \) could be used to find a reasonable estimate of the line. The idea of using the rank information to form groups, as briefly mentioned in the introduction and developed extensively in this chapter, was originally proposed by Efron in an unpublished paper [6].


In the introduction, we presented a special version of the regression problem with both variables subject to error, and five estimates of the path \( y = \alpha + \beta x \) were proposed for this problem, to be known as the tracking problem. The estimation problem is to be discussed in terms of estimating \( \theta = \tan^{-1} \beta \) and we are considering five sequences of estimates, which we have denoted
\[ \hat{\theta}_M^{(i)} \quad i = 1, 2, 3, 4, 5. \]

For \( i = 2, 4 \), the sequence also depends upon other parameters not specified in the notation, namely \( g(t) \) and \( k \) (see the introduction).

In this section we find limiting distributions for all five sequences of estimates. We shall make the

**Assumption:** \( r(t) \) is a Riemann integrable function on \([0, 1]\).

Although the discussion could proceed without this assumption, we shall see that this assumption allows us to avoid some uninteresting complications.

**Definition IB.1:**

\[ \tilde{r}(t) = \int_0^1 r(t) dt. \]

**Definition IB.2:**

\[ P = \int_0^1 [r(t) - \tilde{r}(t)]^2 dt. \]

**Definition IB.3:**

\[ Q = \sigma^2 / P. \]

\( Q \), to be termed the "essential error" will be shown in section IC to be the minimum asymptotic variance attainable for unbiased estimates of \( \theta \).

Recalling from the introduction that we have observations

\[ (x_i, y_i) \quad i = 1, 2, \ldots, M \]
with
\[
\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left( \begin{pmatrix} r_i \\ 0 \end{pmatrix}, \sigma^2 I \right)
\]
for all \(i\)

and with \(\hat{\theta}_M^{(2)}\) as defined in the introduction, the central result of this section is the following.

Theorem IB.1: If \(\hat{\theta}_M^{(2)}\) is the sequence of estimates of \(\theta\) for some grouping function \(g(t)\), and if \(k \to \infty\) and \(M \to \infty\) in such a fashion that
\[\frac{k}{M} \to c \leq 1\]
then
\[
\sqrt{M}(\hat{\theta}_M^{(2)} - \theta) \overset{d}{\to} N(0, Q + cQ^2)
\]
and this result is independent of the choice of \(g(t)\).

In particular, if \(c = 0\)
\[
\sqrt{M}(\hat{\theta}_M^{(2)} - \theta) \overset{d}{\to} N(0, Q)
\]

The proof of this result will be presented through a number of lemmas.

If
\[
\gamma_M^{(2)} = \frac{\tan 2\hat{\theta}_M^{(2)}}{2} = \frac{\sum n_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum n_i (\bar{x}_i - \bar{x})^2 - \sum n_i (\bar{y}_i - \bar{y})^2}
\]
then by the Taylor expansion of \(\frac{\tan 2\theta}{2}\),
\[
\gamma_M^{(2)} = \hat{\theta}_M^{(2)} + o_p([\hat{\theta}_M^{(2)}]^3)
\]
We will show that \( \hat{\gamma}_M^{(2)} \) is \( O_p(1/\sqrt{M}) \). But then \( \hat{\beta}_M^{(2)} \) is \( O_p(1/\sqrt{M}) \), \( [\hat{\beta}_M^{(2)}]^{-1} \) is \( O_p(1/M^{3/2}) \) and \( \sqrt{M}(\hat{\beta}_M^{(2)} - \theta) \) and \( \sqrt{M}(\hat{\gamma}_M^{(2)} - \theta) \) will have the same limiting distribution, when \( \theta = 0 \). Since the procedure is coordinate free, we assume \( \theta = 0 \) without loss of generality.

Definition IB.4:

\[
N_M = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})
\]

\[
D_M = U_M - V_M
\]

where

\[
U_M = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2
\]

\[
V_M = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2
\]

and both \( N_M \) and \( D_M \) depend implicitly upon \( g(t) \) and \( k \).

Lemma IB.1:

\( D_M \) is distributed as the difference \( U_M - V_M \) where

\[
U_M \sim \sigma^2 x^2_{k-1} \left( \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2 \right)
\]

that is, \( \sigma^2 \) times a noncentral \( x^2 \) with \( k-1 \) degrees of freedom and noncentrality parameter \( \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2 \), and

\[
V_M \sim \sigma^2 x^2_{k-1} \text{ independent of } U_M.
\]

Proof: This is a standard "analysis of variance" result, see for example [8] pp. 32-38.

QED
Lemma IB.2:

\[
\frac{D_M}{M} \xrightarrow{P} P \quad \text{as} \quad M \to \infty, \ k \to \infty.
\]

Proof: We have

\[
E\left(\frac{D_M}{M}\right) = \frac{\sum_{i=1}^{k} p_i (\bar{r}_i - \bar{r})^2}{M}
\]

and from elementary properties of calculus

\[
\lim_{m \to \infty} \frac{\sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2}{M} = P
\]

since this is the definition of Riemann integration. Further

\[
\text{Var}\left(\frac{D_M}{M}\right) = \frac{\sigma^4 k^4 (k-1)}{M^2}
\]

so that

\[
\lim_{M \to \infty} \text{Var}\left(\frac{D_M}{M}\right) = 0 \quad \text{See [8], p. 412.}
\]

The result then follows immediately from Chebycheff's inequality. QED

Corollary IB.1:

\[
\frac{M}{D_M} \xrightarrow{P} \frac{1}{P} \quad \text{as} \quad M \to \infty, \ k \to \infty.
\]

Proof: \(\frac{M}{D_M}\) is a continuous function of \(\frac{D_M}{M}\). QED

Lemma IB.3:

\[
\frac{N_M}{\sqrt{M}} \xrightarrow{d} N(0, \sigma^2 P + \sigma^2 k^4)
\]
as \( k \to \infty, M \to \infty \) such that

\[
k/M \to c.
\]

Proof: Consider the independent \( k \)-dimensional vectors

\[
\mathbf{z} = \left( \begin{array}{c} \sqrt{n_1} (\bar{x}_1 - \bar{x}) \\ \vdots \\ \sqrt{n_1} (\bar{x}_k - \bar{x}) \end{array} \right)
\]

\[
\mathbf{y} = \left( \begin{array}{c} \sqrt{n_1} (\bar{y}_1 - \bar{y}) \\ \vdots \\ \sqrt{n_1} (\bar{y}_k - \bar{y}) \end{array} \right)
\]

\[
\mathbf{r} = \left( \begin{array}{c} \sqrt{n_1} (\bar{r}_1 - \bar{r}) \\ \vdots \\ \sqrt{n_1} (\bar{r}_k - \bar{r}) \end{array} \right).
\]

These three \( k \)-vectors all lie in the \((k - 1)\) dimensional vector space orthogonal to the vector

\[
\mathbf{t}_0 = \left( \begin{array}{c} \sqrt{n_1} \\ \vdots \\ \sqrt{n_1} \end{array} \right).
\]

To see this, simply note that

\[
\mathbf{z}'\mathbf{t}_0 = \mathbf{y}'\mathbf{t}_0 = \mathbf{r}'\mathbf{t}_0 = 0
\]

and a vector \( \mathbf{\bar{t}} \) orthogonal to \( \mathbf{t}_0 \) yields

\[
\mathbf{\bar{t}}'\mathbf{x} \sim N(\mathbf{\bar{t}}'\mathbf{\bar{r}}, \sigma^2)
\]

\[
\mathbf{\bar{t}}'\mathbf{y} \sim N(0, \sigma^2).
\]

Hence, if \( H_{k-1} \) is the \( k - 1 \) dimensional vector space orthogonal to
$\vec{y}_0$, let $\vec{x}_P$, $\vec{y}_P$ and $\vec{y}_P$ be the projections of $\vec{x}$, $\vec{y}$ and $\vec{y}$ into $H_{k-1}$. Then $\vec{x}_P$ is a $k - 1$ dimensional normal vector with covariance matrix $\sigma^2 I_{k-1}$ in this space, and so also is $\vec{y}_P$, with respect to any orthonormal basis. To illustrate this, let

$$Z = \begin{pmatrix} \sqrt{n} \vec{y}_1 \\ \vdots \\ \sqrt{n} \vec{y}_k \end{pmatrix} \sim N(0, \sigma^2 I_k)$$

then

$$\vec{y} = Z - \left( Z \cdot \frac{\vec{y}_0}{\sqrt{n}} \right) \frac{\vec{y}_0}{\sqrt{n}}$$

that is, $\vec{y}$ is just a spherical normal random vector, with one component removed.

Now, $\vec{x}_P$ and $\vec{y}_P$ are independent. Thus

$$\frac{N}{\sqrt{M}} = \frac{\vec{x}_P \cdot \vec{y}_P}{\sqrt{M}}$$

$$= \frac{1}{\sqrt{M}} \|\vec{x}_P\| \|\vec{y}_P\|$$

where $\vec{y}_P$ is the projection of $\vec{y}_P$ onto the vector $\vec{x}_P$. But then

$$\vec{y}_P \sim N(0, \sigma^2)$$

since $\vec{y}_P$ is spherical normal. In addition, we know from the previous argument that

$$\|\vec{x}\|^2 \sim \sigma^2 x^2_{k-1}(\|\vec{x}\|^2).$$
Now \( \|x\|^2 \) can be written

\[
\|x\|^2 = (U + \left[ \sum_{i=1}^{k} n_i \left( \bar{r}_i - \bar{r} \right)^2 \right]^{\frac{1}{2}})^2 + \sigma^2 x_{k-1}^2
\]

where

\[ U \sim N(0, \sigma^2) \]

independently of \( x_{k-1}^2 \). Also

\[
\frac{\|x\|^2}{M} \xrightarrow{p} P + c\sigma^2 \quad \text{as} \quad M, \ k \to \infty
\]

since

\[
\frac{\sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2}{M} \to P \quad \text{as} \quad M, \ k \to \infty
\]

and

\[
\frac{x_{k-1}^2}{M} \xrightarrow{p} c \quad \text{as} \quad M, \ k \to \infty.
\]

Thus, finally, calling upon Slutsky's theorem ([9] p. 236)

\[
\frac{N_M}{\sqrt{M}} \xrightarrow{p} N(0, \sigma^2 P + c\sigma^2).
\]

QED

Now we are prepared for

Proof of Theorem IB.1:

The result follows from our lemmas. For

\[
\sqrt{M} \tan \frac{2\alpha(2)}{M} = \frac{N_M}{\sqrt{M}} \cdot \frac{M}{D_M}
\]

and calling again upon Slutsky's limit theorems for rational functions of random variables, we have
\[
\tan \frac{2^{(2)}}{M^2} \xrightarrow{\sqrt{M}} \mathcal{L}_{1/\nu'} N(0, \sigma^2 \nu' + c\sigma^2)
\]
as
\[
M, k \to \infty, k/M \to c
\]
or
\[
\sqrt{M} \gamma^{(2)}_M \xrightarrow{\mathcal{L}} N(0, Q + cQ^2).
\]
We note further that, by our previous remarks
\[
\sqrt{M}(\hat{\theta}^{(2)}_M - \theta) \xrightarrow{\mathcal{L}} N(0, Q + cQ^2)
\]
as
\[
M, k \to \infty, k/M \to c
\]
for all \( \theta \). To minimize the asymptotic variance, we let \( k \to \infty \) sufficiently slowly that \( k/M \to 0 \). If \( k/M \to 0 \) but \( k, M \to \infty \), then
\[
\sqrt{M}(\hat{\theta}^{(2)}_M - \theta) \xrightarrow{\mathcal{L}} N(0, Q).
\]
\[\text{QED}\]
Corollary IB.2:

As \( M \to \infty \)
\[
\sqrt{M}(\hat{\theta}^{(1)}_M - \theta) \xrightarrow{\mathcal{L}} N(0, Q + Q^2).
\]

Proof: The result follows immediately, for \( \hat{\theta}^{(1)}_M \) is a particularization of \( \hat{\theta}^{(2)}_M \) with \( k = M \). \[\text{QED}\]

Of course, there is a good deal to be said about these results.
But, for now, we continue with the discussion of asymptotic results,
and leave the discussion for the chapter on applications. To complete
this section, we derive the asymptotic distribution theory for the
estimation sequences $\hat{\beta}_M^{(3)}$, $\hat{\beta}_M^{(4)}$, and $\hat{\beta}_M^{(5)}$.

Recall the discussion of the introduction, pointing that $\hat{\beta}_M^{(3)}$,
the linear trend estimator of $\theta$, is the maximum likelihood estimator
when $r(t)$ is linear. Of course we do not wish to assume $r(t)$ is
linear, so for the general case we have

Theorem IB.2:

If $r(t)$ is a Riemann integrable function as assumed earlier and

$$
\hat{\beta}_M^{(3)} = \frac{\sum_{i=1}^{M} (y_i - \bar{y})(i - \bar{t})}{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{t})}
$$

where

$$
\bar{t} = \frac{1}{M} \sum_{i=1}^{M} i
$$

and

$$
\hat{\beta}_M^{(3)} = \tan^{-1}\hat{\beta}_M^{(3)}
$$

then

$$
\sqrt{M}(\hat{\beta}_M^{(3)} - \theta) \leq \mathcal{N}(0, Q(c(r, t)))
$$

as $M \to \infty$ where

$$
c(r, t) = \frac{\int_0^1 [r(t) - \bar{r}(t)]^2 dt \int_0^1 (t - \bar{t})^2 dt}{\int_0^1 [r(t) - \bar{r}(t)][t - \bar{t}] dt^2}.
$$

24
In comment, note \( c^{-1}(r, t) \) is simply the square of the covari-
ation between \( r(t) \) and a linear function, and that \( r(t) \) being linear
will imply \( c(r, t) = 1 \), so that \( \sqrt{M}(\hat{\beta}_{M}^{(3)} - \theta) \overset{d}{\to} N(0, Q) \), the limiting
distribution found for \( \sqrt{M}(\hat{\beta}_{M}^{(2)} - \theta) \).

Proof: Since the procedure is coordinate free, the distribution of
angular error will again be independent of \( \theta \), hence we assume \( \theta = 0 \).
But as \( \hat{\theta} \to 0 \)

\[
\hat{\beta} = \hat{\theta} + O(\theta^3)
\]

from the Taylor expansion of \( \tan \hat{\theta} \). Hence the limiting distribution
of \( \sqrt{M} \hat{\beta}_{M}^{(3)} \) will be the same as that of \( \sqrt{M} \hat{\beta}_{M}^{(3)} \) when \( \theta = 0 \). Write

\[
\sqrt{M} \hat{\beta}_{M}^{(3)} = \frac{-\frac{3}{2} \sum_{i=1}^{M} (y_i - \bar{y})(i - \bar{I})}{m^{-2} \sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{I})}
\]

where

\[
\bar{I} = \frac{M+1}{2}
\]

The argument will follow that of Theorem IB.1.

\[
\frac{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{I})}{M^2} \sim N \left( \frac{\sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{I})}{M^2}, \frac{\sum_{i=1}^{M} (i - \bar{I})^2}{M^2} \right)
\]

and

\[
\frac{\sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{I})}{M^2} \to \int_{0}^{1} [r(t) - \bar{r}(t)][t - \frac{1}{2}] dt
\]
as \( M \to \infty \), as a consequence of the definition of Riemann integration.

Also

\[
\sigma^2 \sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{x}) / M^4 \to 0 \text{ as } M \to \infty .
\]

Consequently, as in Lemma IB.2 and Corollary IB.1

\[
\frac{M}{M^2} \sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{x}) \xrightarrow{P} \int_0^1 [r(t) - \bar{r}(t)] [t - \frac{1}{2}] dt
\]

as \( M \to \infty \), and

\[
\frac{M^2}{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{x})} \xrightarrow{P} \frac{1}{1} \int_0^1 [r(t) - \bar{r}(t)] [t - \frac{1}{2}] dt .
\]

Furthermore,

\[
\frac{M}{M^{3/2}} \sum_{i=1}^{M} (y_i - \bar{y})(i - \bar{y}) / \sim N \left( 0, \sigma^2 \frac{\sum_{i=1}^{M} (i - \bar{x})^2}{M^3} \right)
\]

and as \( M \to \infty \)

\[
\frac{M}{M^3} \sum_{i=1}^{M} (i - \bar{x})^2 \to \int_0^1 [t - \frac{1}{2}]^2 dt
\]

again from Riemann integration definitions. Collecting these results and calling once again upon Slutsky's theorem, we have

\[
\sqrt{M} s_M^{(3)} \xrightarrow{D} N \left( 0, \frac{\sigma^2 \int_0^1 [t - \frac{1}{2}]^2 dt}{\left[ \int_0^1 [r(t) - \bar{r}(t)] [t - \frac{1}{2}] dt \right]^2} \right)
\]

and recalling that
\[ q = \int_0^1 \frac{\sigma^2}{[r(t) - \bar{r}(t)]^2} dt \]

we have,

\[ \sqrt{M} \hat{\beta}_M^{(3)} \overset{d}{\to} N(0, Qc(r, t)) \]

as \( M \to \infty \), and consequently for all \( \theta \)

\[ \sqrt{M} (\hat{\theta}_M^{(3)} - \theta) \overset{d}{\to} N(0, Qc(r, t)). \]

QED

To complete this section, we shall discuss the sequences \( \hat{\beta}_M^{(4)} \)

and \( \hat{\beta}_M^{(5)} \), where

\[ \hat{\beta}_M^{(4)} = \frac{\sum_{i=1}^k n_i (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y})}{\sum_{i=1}^k n_i (\tilde{x}_i - \bar{x})^2} \]

and

\[ \hat{\beta}_M^{(5)} = \frac{\sum_{i=1}^M (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^M (x_i - \bar{x})^2} \]

and

\[ \hat{\beta}_M^{(1)} = \tan^{-1} \hat{\beta}_M^{(1)} \quad i = 4, 5. \]

\( \hat{\beta}_M^{(5)} \) is termed the "naive least squares" estimator because it minimizes residual vertical distance and would be used by an observer under the impression that the \( x_i \) were observed without error. \( \hat{\beta}_M^{(4)} \) minimizes residual vertical distance to the \( k \) group averages. We shall show that the sequence \( \sqrt{M}(\hat{\theta}_M^{(4)} - \theta) \) agrees in asymptotic distribution with \( \sqrt{M}(\hat{\theta}_M^{(2)} - \theta) \). Thus we have
Theorem IB.3: If

\[ \hat{\beta}_M^{(4)} = \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2} \]

that is, the grouped naive least squares estimation sequence based upon some choice of \( g(t) \) and a choice of \( k \) for each \( M \), and if

\[ \hat{\theta}_M^{(4)} = \tan^{-1} \hat{\beta}_M^{(4)} \]

then for all \( \theta \neq \pi/2 \)

\[ \sqrt{M}(\hat{\theta}_M^{(4)} - \theta) \leq N(0, Q) \]

as \( M \to \infty, k \to \infty \), provided that

\[ k^2/M \to 0. \]

If, on the other hand

\[ \lim_{M \to \infty} k^2/M > 0 \]

but

\[ k/M \to 0 \]

then

\[ \lim_{m \to \infty} \hat{\theta}_M^{(4)} = \theta \]

that is, \( \hat{\theta}_M^{(4)} \) is still consistent. \( \square \)
The reader should take note of the condition $k^2/M \to 0$ which was not required for previous results. This condition is necessary to make the bias asymptotically negligible.

Proof: As before, write

$$\sqrt{M} \beta_M^{(4)} = \frac{M^{\frac{1}{2}} \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x}) (\bar{y}_i - \bar{y})}{M^{-1} \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2}.$$ 

Since $\beta_M^{(4)}$ is not based upon a coordinate free procedure, we cannot assume $\beta = 0$. But

$$E(\bar{x}_i - x) = (\bar{x}_i - \bar{x}) \cos \theta$$
$$E(\bar{y}_i - y) = (\bar{x}_i - \bar{x}) \sin \theta$$

for all $i$.

Consequently

$$E \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2}{M} = \frac{\cos^2 \theta \cdot \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2 + (k-1) \sigma^2}{M}$$

$$\text{Var} \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2}{M} = \frac{4 \cos^2 \theta \cdot \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2 + 2(k-1) \sigma^4}{M^2}$$

and we have

$$\lim_{M \to \infty} \text{Var} \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2}{M} = 0.$$ 

Calling upon Lemma IB.2 and Corollary IB.1, we have that
\[
\left[ \frac{\sum_{i=1}^{k} n_i (\tilde{x}_i - \bar{x})^2}{M} - \frac{\cos^2 \theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2}{M} - \frac{(k-1)\sigma^2}{M} \right] \rightarrow P_0
\]
as \(M \rightarrow \infty\), and consequently by Slutsky's theorem
\[
\sqrt{M}(\hat{\beta}_M^{(4)} - \beta)
\]
has the same asymptotic distribution as
\[
\sqrt{M}\left( \frac{\sum_{i=1}^{k} n_i (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y})}{\cos^2 \theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2 + (k-1)\sigma^2} - \beta \right)
\]
which we shall denote \(T_M\).

We see that
\[
\frac{\cos^2 \theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2 + (k-1)\sigma^2}{M} \rightarrow P \cos^2 \theta
\]
as \(M \rightarrow \infty\), so that the condition \(\theta \neq \pi/2\) insures that the denominator of \(\hat{\beta}_M^{(4)}\) does not converge to zero.

Now, we note that
\[
E \frac{\sum_{i=1}^{k} n_i (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y})}{M} = \cos \theta \sin \theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2
\]
so that
\[
ET_M = \sqrt{M}\left( \frac{\cos \theta \sin \theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2}{\cos^2 \theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2 + (k-1)\sigma^2} - \beta \right)
\]
\[ -\sqrt{M} \left( -\frac{(k-1)\sigma^2}{\cos^2\theta \sum n_i (\bar{r}_i - \bar{r})^2} + \frac{1}{2} \left( \frac{(k-1)\sigma^2}{\cos^2\theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2} \right)^2 + \ldots \right) \tan \theta \]

by the Taylor expansion of \( 1/(1+x) \) where we must assume \( M \) sufficiently large that

\[ \left| \frac{(k-1)\sigma^2}{\cos^2\theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2} \right| < 1. \]

Now, by the integrability of \( r(t) \)

\[ \frac{(k-1)\sigma^2}{\cos^2\theta \sum_{i=1}^{k} n_i (\bar{r}_i - \bar{r})^2} = O\left( \frac{k}{M} \right) \]

so that

\[ E_{T_M} = \sqrt{M} \tan \theta \cdot \left( -O\left( \frac{k}{M} \right) + O\left( \frac{k^2}{M^2} \right) \ldots \right) \]

\[ = \tan \theta \cdot \left( -O\left( \frac{k}{\sqrt{M}} \right) + O\left( \frac{k^2}{M^{3/2}} \right) \ldots \right) \]

and since \( k^2/M \to 0 \) as \( M \to \infty \)

\[ \lim_{M \to \infty} E_{T_M} = 0, \]

that is

\[ \lim_{M \to \infty} B_M(\hat{\beta}_M^{(h)} - \beta) = 0. \]

To complete the proof, write
\[
\begin{align*}
\tilde{x} &= \left( \sqrt{n_i} (\tilde{x}_i - \bar{x}) \right) \\
\tilde{y} &= \left( \sqrt{n_i} (\tilde{y}_i - \bar{y}) \right) \\
\tilde{z}_1 &= \left( \sqrt{n_i} (\tilde{z}_1 - \bar{z}) \cos \theta \right) \\
\tilde{z}_2 &= \left( \sqrt{n_i} (\tilde{z}_1 - \bar{z}) \sin \theta \right).
\end{align*}
\]

As before, these \( k \)-vectors lie in the \( k-1 \) dimensional subspace orthogonal to \( \tilde{z}_0 = \left( \sqrt{n_i} \right) \).

Write

\[
\sum_{i=1}^{k} n_i (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y}) = \tilde{z}_1^T \tilde{z}_2 + (\tilde{x} - \tilde{z}_1)' \tilde{z}_2 + \tilde{z}_1'(\tilde{y} - \tilde{z}_2)
\]

so that

\[
\text{Var} \sum_{i=1}^{k} n_i (\tilde{x}_i - \bar{x})(\tilde{y}_i - \bar{y}) = \text{Var}[(\tilde{x} - \tilde{z}_1)' \tilde{z}_2 + \tilde{z}_1'(\tilde{y} - \tilde{z}_2)].
\]

Now, both \( \tilde{x} \) and \( \tilde{y} \) are normally distributed with covariance matrix \( \sigma^2 I_{k-1} \) in the subspace orthogonal to \( \tilde{z}_0 \) (any orthonormal basis) and they are independent. Thus, as proved in Lemma IB.3

\[
(\tilde{x} - \tilde{z}_1)' \tilde{z}_2 \sim N(0, \sigma^2 \sin^2 \theta \sigma^2)
\]

and

\[
\tilde{z}_1'(\tilde{y} - \tilde{z}_2) \sim N(0, \sigma^2 \cos^2 \theta \sigma^2)
\]

as \( M \to \infty \). Further, since these two random variables are uncorrelated,
they are asymptotically independent, which implies

\[
\frac{(\vec{x} - \vec{r}_1)'(\vec{y} + (\vec{r}_2)'(\vec{x} - \vec{r}_1)'}{\sqrt{M}} \overset{p}{\rightarrow} N(0, \sigma^2 P)
\]

as \( M \to \infty \). Finally

\[
\cos^2 \theta \sum_{i=1}^{k} n_i (\vec{r}_i - \vec{r})^2 + (k-1) \sigma^2 \over M = \cos^2 \theta \cdot P + O(1/\sqrt{M})
\]

and

\[
\frac{1}{\cos^2 \theta \cdot P + O(1/\sqrt{M})} = \frac{1}{\cos^2 \theta \cdot P} \left[ 1 - O(1/\sqrt{M}) + \ldots \right].
\]

Collecting this sequence of results, we have

\[
\sqrt{M}(\hat{\beta}_M^{(4)} - \beta) \overset{p}{\rightarrow} N(0, Q/\cos^2 \theta)
\]

as \( M \to \infty \).

Furthermore \( \theta = \tan^{-1} \beta \), and \( \hat{\beta}_M^{(4)} = \tan^{-1} \hat{\beta}_M^{(4)} \) an infinitely differentiable function except at \( \pi/2 \). Therefore

\[
\sqrt{M}(\hat{\beta}_M^{(4)} - \beta) = \frac{\partial}{\partial \beta} \left| \sqrt{M}(\hat{\beta}_M^{(4)} - \beta) + O(1/\sqrt{M}) + \ldots \right|
\]

\[
= \cos \theta \cdot \sqrt{M}(\hat{\beta}_M^{(4)} - \beta) + O(1/\sqrt{M})
\]

by Taylor's expansion, so that

\[
\sqrt{M}(\hat{\beta}_M^{(4)} - \beta) \overset{p}{\rightarrow} N(0, Q)
\]

as \( M \to \infty \).
Thus, as promised, we have shown that $\hat{\beta}_M^{(4)}$ and $\hat{\beta}_M^{(2)}$, when normalized, have the same limiting distribution. In Chapter II, we shall argue for the superiority of $\hat{\beta}_M^{(2)}$ as an estimate from small sample considerations, but to complete this section, we have

Corollary IB.3:

\[
\hat{\beta}_M^{(5)} \text{ is not consistent}. \quad \Box
\]

Proof:

\[
\frac{\sum_{i=1}^{M} (x_i - \bar{x})(y_i - \bar{y})}{M} \xrightarrow{P} \cos \theta \sin \theta \cdot P
\]

\[
\frac{\sum_{i=1}^{M} (x_i - \bar{x})^2}{M} \xrightarrow{P} \cos^2 \theta \cdot P + \sigma^2
\]

as $M \to \infty$. Thus

\[
\hat{\beta}_M^{(5)} \xrightarrow{P} \frac{\cos \theta \sin \theta \cdot P}{\cos^2 \theta \cdot P + \sigma^2} \neq \beta
\]

as $M \to \infty$. \quad \text{QED}

This completes the presentation of the asymptotic distribution theory for the five alternative sequences of estimates of $\theta$. In the next section, we shall show that $Q$, the asymptotic variance achieved by the sequences $\hat{\beta}_M^{(2)}$ and $\hat{\beta}_M^{(4)}$ is in fact the minimum asymptotic variance achievable for unbiased estimates of $\theta$.
C. The Asymptotic Optimality of Estimation Based upon Grouping.

We introduce this section with a summary of our asymptotic results. By assuming \( r(t) \) integrable, we have

\[
P = \int_0^1 [r(t) - \overline{r(t)}]^2 \, dt
\]

and

\[
Q = \sigma^2 / P.
\]

For the five sequences of estimates under consideration, we have shown that \( \hat{\theta}_M^{(5)} \), the naive least squares estimate, is not consistent. For \( \hat{\theta}_M^{(1)} \), the maximum likelihood estimate when rank information is ignored

\[
\sqrt{M}(\hat{\theta}_M^{(1)} - \theta) \overset{d}{\rightarrow} N(0, Q + Q^2)
\]

as \( M \to \infty \). For \( \hat{\theta}_M^{(2)} \), the grouped orthogonal least squares estimate,

\[
\sqrt{M}(\hat{\theta}_M^{(2)} - \theta) \overset{d}{\rightarrow} N(0, Q)
\]

as \( M \to \infty \), provided that \( k/M \to 0 \). For \( \hat{\theta}_M^{(3)} \), the linear trend estimate

\[
\sqrt{M}(\hat{\theta}_M^{(3)} - \theta) \overset{d}{\rightarrow} N(0, Qc(r, t))
\]

as \( M \to \infty \), where

\[
c(r, t) = \frac{\int_0^1 [r(t)-\overline{r(t)}]^2 \, dt \int_0^1 (t-\frac{1}{2})^2 \, dt}{\left[\int_0^1 [r(t)-\overline{r(t)}][t-\frac{1}{2}] \, dt \right]^2}.
\]
For $\hat{\theta}_M^{(4)}$, the grouped naive least squares estimate

$$\sqrt{M}(\hat{\theta}_M^{(4)} - \theta) \overset{d}{\to} N(0, Q)$$

as $M \to \infty$, provided $k^2/M \to 0$.

In this section, we prove that $Q$ is the minimum variance attainable for this problem.

Theorem IC.1:

The minimum asymptotic variance attainable for $\sqrt{M}(\hat{\theta}_M - \theta)$ as $M \to \infty$, where $\hat{\theta}_M$ is any unbiased estimate of $\theta$, is $Q$.

Proof: Suppose our observations are distributed as in the original statement of the problem, with

$$E(x_i - \bar{x}) = (r_i - \bar{r})\cos \theta$$

$$E(y_i - \bar{y}) = (r_i - \bar{r})\sin \theta$$

except that $r_i$ are now known. We can estimate the actual line at least as accurately in this case, for we have the option of randomizing the $r_i$ and thereby putting ourselves back into the original situation. Thus if we estimate $\theta$ (or $\tan \theta$) through the method of maximum likelihood, we must find

$$\max \left\{ \sum_{i=1}^{M} ((x_i - \bar{x}) - (r_i - \bar{r})\cos \theta)^2 + ((y_i - \bar{y}) - (r_i - \bar{r})\sin \theta)^2 \right\}$$

or

$$\min \left\{ \sum_{i=1}^{M} (x_i - \bar{x})(r_i - \bar{r})\cos \theta + \sum_{i=1}^{M} (y_i - \bar{y})(r_i - \bar{r})\sin \theta \right\}$$
minimizing w.r.t. \( \theta \), we find

\[
\hat{\theta}_M = \tan \hat{\theta}_M = \frac{\sum (y_i - \bar{y})(r_i - \bar{r})}{\sum (x_i - \bar{x})(r_i - \bar{r})}.
\]

Thus, from the Taylor development of \( \tan \hat{\theta}_M \) to second order,

\[
(\hat{\theta}_M - \theta)_M = O(\theta^2)\text{.}
\]


Consequently, as discussed more fully in section IB,

\[
\lim_{M \to \infty} L \sqrt{M} \hat{\theta}_M = \lim_{M \to \infty} L \sqrt{M} \hat{\theta}_M
\]

when \( \theta = 0 \).

The limiting distribution for \( \hat{\theta}_M \) is derived exactly as was that of the linear trend estimate in Section IB. Briefly

\[
\frac{\sum_{i=1}^{M} (x_i - \bar{x})(r_i - \bar{r})}{\sum_{i=1}^{M} (y_i - \bar{y})(r_i - \bar{r})} \overset{P}{\to} \frac{1}{P}.
\]

Thus, for the problem posed here the normalized sequence of unbiased estimates of \( \theta \) has limiting variance no smaller than \( Q \). Since we can do no better in the tracking problem as formulated in the introduction, \( Q \) is a lower bound on asymptotic variance for that problem as well.

QED
In this sense, both $\hat{\theta}_M^{(2)}$ and $\hat{\theta}_M^{(4)}$ can be said to be asymptotically optimal. In addition, if $r(t)$ is linear, the sequence $\hat{\theta}_M^{(3)}$ is also asymptotically optimal.

Thus, from the large sample point of view, $\hat{\theta}_M^{(2)}$ and $\hat{\theta}_M^{(4)}$ are competitive estimators, while in special cases $\hat{\theta}_M^{(3)}$ may be a competitor. In the next chapter, we shall use small sample arguments to justify the exclusion of $\hat{\theta}_M^{(4)}$ as an inappropriate estimate, and to give guidelines for choosing between $\hat{\theta}_M^{(2)}$ and $\hat{\theta}_M^{(3)}$.  

38
Chapter II. Small Sample Considerations, Choosing an Estimate

A. Introductory Remarks.

In the Introduction and Summary we posed a special version of the regression problem with both variables subject to error, which we called the tracking problem. We proposed in the Introduction five alternative procedures for estimating the line \( y = \alpha + \beta x \). Discussing the problem in terms of estimating \( \theta = \tan^{-1} \beta \), we found in Chapter I that \( \hat{\theta}_M^{(2)} \), the grouped orthogonal least squares estimate, minimizing orthogonal distance and \( \hat{\theta}_M^{(4)} \), the grouped naive least squares estimate (minimizing vertical distance) were asymptotically efficient in the sense that \( \sqrt{M}(\hat{\theta}_M^{(i)} - \theta) i = 2, 4 \) is asymptotically normal as \( M \to \infty \) and the limiting variance achieves a minimum of the Cramèr Rao type. A third alternative, \( \hat{\theta}_M^{(3)} \), the linear trend estimate was shown to be asymptotically efficient in this sense if and only if \( r(t) \), the path function, is linear.

In this chapter, we continue to discuss the problem of choosing an estimate of \( \theta \), giving emphasis now to the distribution theory for finite sample sizes. Section B contains some Monte Carlo results, Section C includes finite sample size distribution theory for the estimates \( \hat{\theta}_M^{(2)} \), \( \hat{\theta}_M^{(3)} \), and \( \hat{\theta}_M^{(4)} \), and Section D contains a formulation of guidelines for choosing an estimate.

In this chapter, we drop the assumption that \( r(t) \) is integrable, and handle the discussion in terms of summands instead of integrals. In discussing estimation by grouping, we assume \( g(t) = 1 \), i.e. equal sized groups, for this is the case of usual interest and the discussion
is thereby greatly simplified. Nevertheless, the reader can easily verify that the discussion can be modified to apply to any choice of \( g(t) \). Thus, we assume

\[
k = \text{number of groups} \\
n = k/M = \text{number of observations per group}.
\]

Thus \( \bar{x}_i, \bar{y}_i, \bar{r}_i \ i = 1, 2, \ldots, k \) now apply to \( k \) observation averages where each group has size \( n \).

Definition IIA.1:

\[
Q_k = \frac{\sigma^2}{n \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2}.
\]

Now we found \( \hat{\beta}^{(2)}_M \) from the equation

\[
\tan \left( \frac{2\hat{\beta}^{(2)}_M}{2} \right) = \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2}
\]

so that

\[
\text{Var} \left( \frac{\tan \frac{2\hat{\beta}^{(2)}_M}{2}}{2} \right) \approx \frac{\text{Var} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\left[ \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 \right]^2}
\]

\[
\approx Q_k + (k - 1)Q_k^2.
\]

Thus, to consistently achieve accuracy on the order of \( 5^\circ \) we must have
\[ Q_k + (k - 1)Q_k^2 < 1/100 \]

since \( \text{Var}(\hat{\theta}_M^{(2)}) \leq \text{Var} \left( \frac{\tan 2\hat{\theta}_M^{(2)}}{2} \right) \) when \( \theta = 0 \). Throughout this chapter, this inequality will be assumed to hold and we will continue to assume \( \theta = 0 \) where appropriate. A brief digression is necessary at this point.

Let \( x \) be a normal random variable such that

\[ x \sim N(\mu, \Delta^2) \]

where \( \mu > 0 \) and

\[ \frac{\Delta}{\mu} \leq 1/5. \]

Now \( E(1/x) \) and \( E(1/x^2) \) are undefined. To circumvent this technical difficulty, define

\[ x_T^\mu = \begin{cases} x, & |x - \mu| \leq \mu/2 \\ \mu, & \text{otherwise}. \end{cases} \]

\( x_T^\mu \) is the truncate of \( x \), agreeing with \( x \) everywhere except at the tails having probability less than .01.

Lemma IIA.1:

\[ \frac{1}{\mu} \leq E\left( \frac{1}{x_T^\mu} \right) \leq \frac{1.1}{\mu}. \]

Proof:

\[
E \frac{1}{x_T^\mu} = \int_{-\mu/2}^{\mu/2} \frac{1}{\mu+x} \frac{1}{\Delta} \phi\left(\frac{x}{\Delta}\right) dx
\]

\[
= \int_{-\mu/2}^{\mu/2} \frac{1}{\mu} \left(1 - \frac{x}{\mu} - \frac{x^2}{\mu^2} \ldots\right) \frac{1}{\Delta} \phi\left(\frac{x}{\Delta}\right) dx.
\]
The odd order terms vanish by symmetry, while the even order terms past \( \frac{X^2}{\mu^2} \) are negligible, e.g.

\[
E \left( \frac{X^4}{\mu^5} \right) = \frac{3\Delta^4}{\mu^5} \leq \frac{3}{\mu^5}.
\]

Therefore

\[
E \left( \frac{1}{X_T} \right) \leq \frac{1}{\mu} \left( 1 + \frac{\Delta^2}{2} + \frac{3\Delta^4}{4} + \ldots \right)
\]

\[
\leq \frac{1}{c} \left( 1 + \frac{1}{25} + \frac{3}{(25)^2} + \ldots \right)
\]

\[
\leq \frac{1.1}{\mu}
\]

and by Jensen's inequality

\[
\frac{1}{\mu} \leq E \left( \frac{1}{X_T} \right).
\]

QED

Lemma IIA.2:

\[
\frac{1}{\mu} \leq E \left( \frac{1}{X_T^2} \right) \leq \frac{1.1}{\mu}.
\]

Proof:

\[
E \left( \frac{1}{X_T^2} \right) = \int_{-\mu/2}^{\mu/2} \frac{1}{(\mu + x)^2} \frac{1}{\Delta} \varphi \left( \frac{X}{\Delta} \right) dx
\]

\[
\int_{-\mu/2}^{\mu/2} \frac{1}{\mu^2} \left( 1 - 2 \frac{X}{\mu} + 3 \frac{X^2}{\mu^2} + \ldots \right) \frac{1}{\Delta} \varphi \left( \frac{X}{\Delta} \right) dx
\]

and the argument proceeds as in Lemma IIA.1.

QED
With these preliminaries at hand, we turn to the Monte Carlo result.

B. Some Monte Carlo Work.

The problem of estimating the track of a particle moving along a straight line in the plane according to a path function \( r(t) \) was investigated by a series of Monte Carlo trials. In each trial, we took

\[
\begin{pmatrix}
  x_i \\
  y_i
\end{pmatrix} \sim_N \begin{pmatrix}
  r(i) \\
  0
\end{pmatrix}, \sigma^2_i \quad i = 1, 2, \ldots, 500
\]

for some choice of \( r(i) \) such that \( r(500) - r(0) = 10 \), and with \( \sigma^2 = 37.5 \) for each choice of \( r(i) \) and for each of 200 iterations of the experiment, \( \theta \) was estimated by the estimates \( \hat{\theta}_M^{(1)} \), \( \hat{\theta}_M^{(2)} \) and \( \hat{\theta}_M^{(4)} \). The objective was to gain information about improvements in efficiency through grouping, comparative performance of \( \hat{\theta}_M^{(2)} \) and \( \hat{\theta}_M^{(3)} \) and stability of \( \text{Var} \hat{\theta}_M^{(k)} \) against nonlinearity of \( r(i) \). For each choice of \( r(i) \), a sample mean and sample standard deviation was computed for each estimate over the 200 iterations. A summary of these results is as follows. The mean of \( \hat{\theta}_M^{(i)} \) refers to the average outcome for the 200 trials, and the variance refers to the sample variance over the 200 trials. \( k \) was taken to be 25, close to the optimal value as discussed in Section C.

43
<table>
<thead>
<tr>
<th>$r(i)$</th>
<th>Mean $\hat{\theta}_M^{(1)}$</th>
<th>Var $\hat{\theta}_M^{(1)}$</th>
<th>Mean $\hat{\theta}_M^{(2)}$</th>
<th>Var $\hat{\theta}_M^{(2)}$</th>
<th>Mean $\hat{\theta}_M^{(4)}$</th>
<th>Var $\hat{\theta}_M^{(4)}$</th>
<th>Run No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{50}$</td>
<td>$-1.3 \times 10^{-4}$</td>
<td>$6.8 \times 10^{-2}$</td>
<td>$8.2 \times 10^{-3}$</td>
<td>$1.2 \times 10^{-2}$</td>
<td>$8.2 \times 10^{-3}$</td>
<td>$9.2 \times 10^{-3}$</td>
<td>1</td>
</tr>
<tr>
<td>$\frac{1}{50} + \sin \frac{2\pi i}{50}$</td>
<td>$-9.9 \times 10^{-5}$</td>
<td>$6.8 \times 10^{-2}$</td>
<td>$8.2 \times 10^{-3}$</td>
<td>$1.2 \times 10^{-2}$</td>
<td>$8.2 \times 10^{-3}$</td>
<td>$9.2 \times 10^{-3}$</td>
<td>2</td>
</tr>
<tr>
<td>$\frac{1}{50} + \sin \frac{2\pi i}{50} \cdot \frac{1}{50}$</td>
<td>$8.7 \times 10^{-3}$</td>
<td>$5.3 \times 10^{-2}$</td>
<td>$-1.7 \times 10^{-3}$</td>
<td>$1.1 \times 10^{-2}$</td>
<td>$-4.1 \times 10^{-5}$</td>
<td>$8.6 \times 10^{-3}$</td>
<td>3</td>
</tr>
<tr>
<td>$\frac{1}{50} + \text{N}(0, .005)$</td>
<td>$8.3 \times 10^{-3}$</td>
<td>$6.3 \times 10^{-2}$</td>
<td>$2.2 \times 10^{-2}$</td>
<td>$1.1 \times 10^{-2}$</td>
<td>$1.6 \times 10^{-2}$</td>
<td>$9.4 \times 10^{-3}$</td>
<td>4</td>
</tr>
<tr>
<td>$\frac{1}{50} + \text{N}(0, .02)$</td>
<td>$-1.3 \times 10^{-2}$</td>
<td>$5.9 \times 10^{-2}$</td>
<td>$-1.9 \times 10^{-3}$</td>
<td>$1.2 \times 10^{-2}$</td>
<td>$-3.1 \times 10^{-3}$</td>
<td>$9.4 \times 10^{-3}$</td>
<td>5</td>
</tr>
<tr>
<td>$\frac{1}{500} \cdot 10$</td>
<td>$2.1 \times 10^{-2}$</td>
<td>$5.5 \times 10^{-2}$</td>
<td>$5.9 \times 10^{-4}$</td>
<td>$1.1 \times 10^{-2}$</td>
<td>$-9.5 \times 10^{-5}$</td>
<td>$1.1 \times 10^{-2}$</td>
<td>6</td>
</tr>
<tr>
<td>$\frac{1}{500} \cdot 10$</td>
<td>$4.1 \times 10^{-3}$</td>
<td>$5.8 \times 10^{-2}$</td>
<td>$8.2 \times 10^{-3}$</td>
<td>$1.1 \times 10^{-2}$</td>
<td>$8.2 \times 10^{-3}$</td>
<td>$9.3 \times 10^{-3}$</td>
<td>7</td>
</tr>
</tbody>
</table>
The seven choices of \( r(i) \) were comprised of a linear function, a linear function plus a sinusoidal component of two different magnitudes, a linear function plus normal variation of two different variances, a pure quadratic and a pure cubic. The reader may note that both grouped orthogonal least squares estimation and linear trend estimation reduce variance by a factor of 5 to 6 in comparison with the ungrouped orthogonal least squares estimate. In addition, the linear trend estimate is superior in mean square error to the grouped orthogonal least squares estimate in all seven cases. This is of course purely a function of the choice of parameters but does illustrate the stability of the mean square error of the linear trend estimate in a variety of situations. The linear trend estimate has variance within 10% of \( Q_M \) in all cases excepting the case of the cubic polynomial.

C. Distribution Theory for Finite Sample Size.

In this section, we show that when

\[
Q_k + (k - 1)q_k^2 \leq 1/100
\]

\( \hat{\beta}_M^{(2)} \) has a distribution closely approximating its asymptotic distribution, that \( \hat{\beta}_M^{(3)} \) approaches its limiting distribution even more rapidly, with variance which is stable against nonlinearity of \( r(t) \), and that \( \hat{\beta}_M^{(4)} \) has highly significant bias.

The random variables for which finite sample size distribution theory is available are the random variables \( \tan \frac{\theta_M^{(2)}}{2}, \hat{\beta}_M^{(3)} \) and \( \hat{\beta}_M^{(4)} \), each of whose expectation is undefined. However, this is a
technical difficulty associated with small probabilities in the tails of the distributions, and we are actually interested in the bounded random variables \( \hat{\delta}_M^{(1)} \) \( i = 2, 3, 4 \). Consequently, we circumvent this difficulty by replacing the denominators of \( \tan \frac{2\hat{\delta}_M^{(2)}}{2} \), \( \hat{\delta}_M^{(3)} \) and \( \hat{\delta}_M^{(4)} \) by their truncates as defined in IIA. It will not be possible to be as rigorous here as in Chapter I.

Let

\[
\tan \frac{2\hat{\delta}_M^{(2)}}{2} = \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2}
\]

with each group having \( n \) members. As before, we assume \( \theta = 0 \). We will denote

\[
D_M = \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2.
\]

Assertion 1: When \( Q_k + (k - 1)Q_k^2 \leq 1/100 \), we may replace the denominator, \( D_M' \) of \( \hat{\delta}_M^{(2)} \) by its expected value.

As shown in Chapter I, \( D_M \) is distributed as the difference of two independent \( x^2 \) random variables with \( k - 1 \) degrees of freedom, one of which is noncentral. Consequently

\[
D_M \rightarrow N\left( \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2, \frac{4\sigma^2}{n} \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + \frac{4\sigma^4}{n^2} (k - 1) \right)
\]

as \( M, k \rightarrow \infty \). Furthermore
\[
\frac{4\sigma^2}{n} \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + \frac{4\sigma}{n^2} (k-1) \left[ \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 \right]^2 = 4(Q_k + (k - 1)Q_k^2) \leq 1/25.
\]

Thus, using the approximate normality of \( D_n \) and Lemma IIA.1, we have as an approximation discounting the tails with probability < .01

\[
\frac{1}{k} \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 \leq \mathbb{E} \left( \frac{1}{n} \right) \leq \frac{1.1}{\sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2}.
\]

Note that a rigorous justification of this approximation would require a sophisticated argument, using for example the Berry-Esseen theorem. Because of the many approximations made here, such rigor seems inappropriate. For empirical support, refer to Section IIB.

Using Assertion 1, we may say that for

\[
Q_k + (k - 1)Q_k^2 \leq 1/100
\]

we can approximate the distribution of \( \tan 2\tilde{\Theta}^{(2)}/M \) by that of

\[
\frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2}.
\]

But, from Lemma IB.3

\[
\frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2} \sim N(0, Q_k + (k - 1)Q_k^2)
\]

47
approximately and has exactly that mean and variance.

To get any idea of the efficiency of \( \hat{\theta}_M^{(2)} \), we need to investigate how closely \( \hat{Q}_k + (k - 1)Q_k^2 \) can be made to approximate \( Q_M \), the Cramér Rao lower bound. First, we assume that \( Q_k \) is approximately equal to \( Q_M \), an assumption which is reasonable if \( k \) is large relative to track length \((r(1) - r(0))\), for this assumption is equivalent to assuming

\[
n \sum_{i=1}^{k} (\tilde{r}_i - \bar{r})^2 = \sum_{i=1}^{M} (r_i - \bar{r})^2.
\]

Secondly, if \( Q_k \) is relatively unaffected by the choice of \( k \), we may decide to choose \( k \) so that

\[
(k - 1)Q_k^2 \leq \frac{1}{10} Q_k
\]

or

\[
(k - 1) \leq \frac{1}{10} Q_k.
\]

When this holds, \( \hat{Q}_k + (k - 1)Q_k^2 \) would be brought within 5% of \( Q_k \). Thus the actual choice of \( k \) may in general involve preliminary estimation of the track and of \( 4\sigma^2 \).

Collecting all of this discussion, we assert that when \( k \) can be chosen such that

\[
Q_k^2 + (k - 1)Q_k \leq \frac{1}{100}
\]

it is possible to choose \( k \) such that, approximately

\[
\frac{\tan 2\hat{\theta}_M^{(2)}}{2} \sim N(0, Q_k)
\]

48
for $\beta = 0$, and consequently
\[
\hat{\beta}_M^{(2)} \sim N(\theta, \varrho_k)
\]
for all $\theta$.

We now turn to a discussion of the estimator
\[
\hat{\beta}_M^{(3)} = \frac{\sum_{i=1}^{M} (y_i - \bar{y})(i - \bar{i})}{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{i})}.
\]

Assertion 2: If $r(t)$ is linear, for $Q_k + (k - 1)Q_k^2 \leq 1/25$, $\hat{\beta}_M^{(3)}$ can be approximated in distribution by
\[
\frac{\sum_{i=1}^{M} (y_i - \bar{y})(i - \bar{i})}{\sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{i})}.
\]

To see this note that
\[
\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{i}) \sim N\left(M \sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{i}), \sigma^2 M \sum_{i=1}^{M} (i - \bar{i})^2\right)
\]
and
\[
\sigma^2 M \sum_{i=1}^{M} (i - \bar{i})^2
\]
\[
\left[\frac{\sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{i})}{\sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{i})}\right]^2 = Q_M
\]

when $r(t)$ is linear. But
\[
Q_M < Q_k < Q_k + (k - 1)Q_k^2 \leq 1/25
\]

49
thus by Lemma IIA.1, and Lemma IIA.2

\[
E \left( \frac{1}{\sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{I})} \right) = \frac{1}{\sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{I})}
\]

\[
E \left( \frac{1}{\left[ \sum_{i=1}^{M} (x_i - \bar{x})(i - \bar{I}) \right]^2} \right) = \frac{1}{\left[ \sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{I}) \right]^2}.
\]

Thus, if \( r(t) \) is linear, \( \hat{\beta}_M^{(3)} \) can be approximated by its asymptotic distribution with \( M \) one fourth as large as that necessary for \( \hat{\beta}_M^{(2)} \).

Assertion 3: \( c(r, t) \) is near one for lower degree polynomials.

Case 1: \( r(t) = t^2 \). Then by straightforward integration we find

\[
c(r, t) = \frac{45}{46}.
\]

Case 2:

\( r(t) = t^3 \).

Then we find

\[
c(r, t) = \frac{84}{100}.
\]

Now if we look back to Assertion 2, we see that for general \( r(t) \)

\[
\sigma^2 \sum_{i=1}^{M} (i - \bar{I})^2 \geq \left[ \sum_{i=1}^{M} (r_i - \bar{r})(i - \bar{I}) \right]^2 
\]

\[
= Q_M c(r, t).
\]
Collecting these results, we see that for \( r(t) \) a lower order polynomial, the linear trend estimator may for certain sample sizes have smaller variances than the grouped orthogonal least squares estimate. This point will be discussed in more detail in the next section.

To complete this section, consider the random variable

\[
\frac{\sum_{i=1}^{k} (\bar{y}_i - \bar{y})(\bar{x}_i - \bar{x})}{\cos^2 \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + (k-1)\frac{\sigma^2}{n}}
\]

which we know to be approximately equivalent to \( \hat{\beta}_M^{(4)} \) in distribution.

Assertion 4: The random variable

\[
\frac{\sum_{i=1}^{k} (\bar{y}_i - \bar{y})(\bar{x}_i - \bar{x})}{\cos^2 \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + (k-1)\frac{\sigma^2}{n}}
\]

is a biased estimate of \( \beta \), and the bias is unbounded as a function of \( \beta \). 

To see this, we note

\[
\left| \frac{\text{E} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\cos^2 \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + (k-1)\frac{\sigma^2}{n}} - \beta \right| = \left| \frac{\cos \theta \sin \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2}{\cos^2 \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + (k-1)\frac{\sigma^2}{n}} - \beta \right|
\]

51
\[ \tan \theta \left( -\frac{(k-1)\sigma_n^2}{\cos^2 \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2} + \left[ \frac{(k-1)\sigma_n^2}{\cos^2 \theta \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2} \right]^2 \right). \]

These four assertions form the basis for a discussion of the choice of an estimate, which we take up in the next section.

D. Choosing an Estimate of \( \theta \).

In the last section, we argued that the statistician will choose as an estimate of \( \theta \) either \( \hat{\theta}_M^{(2)} \) for some choice of \( k \), or \( \hat{\theta}_M^{(3)} \), or \( \hat{\theta}_M^{(4)} \). In this section, some guidelines for that decision are offered.

The estimate \( \hat{\theta}_M^{(4)} \) is biased as an estimate of \( \theta \), and the amount and direction of the bias is lost in the complexities of the distribution theory. Since \( \hat{\theta}_M^{(4)} \) is asymptotically no more efficient than \( \hat{\theta}_M^{(2)} \), there seems obvious advantage in choosing \( \hat{\theta}_M^{(2)} \) or \( \hat{\theta}_M^{(3)} \) rather than \( \hat{\theta}_M^{(4)} \). Thus, since \( \hat{\theta}_M^{(4)} \) is in no way superior to \( \hat{\theta}_M^{(2)} \), and the question of bias does not arise with \( \hat{\theta}_M^{(2)} \), it is recommended that \( \hat{\theta}_M^{(4)} \) never be used to estimate \( \theta \).

Thus, the statistician will choose either \( \hat{\theta}_M^{(2)} \) or \( \hat{\theta}_M^{(3)} \) to estimate \( \theta \). Since \( \hat{\theta}_M^{(2)} \) is always asymptotically efficient, one would expect in general to choose \( \hat{\theta}_M^{(2)} \) as the appropriate estimate. However, there is one situation in which \( \hat{\theta}_M^{(3)} \) may be a superior estimate.

In deriving the distribution for the random variables \( \hat{\beta}_M^{(1)} \) in Section IIIC, we replaced the denominator random variable by its expectation. We shall use the term "denominator error" to refer to the
inaccuracy in describing the distribution of $\hat{\delta}^{(1)}_M$ introduced by this approximation.

We saw in Section B that the denominator error for the estimate $\hat{\delta}^{(2)}_M$ is insignificant if

$$Q_k + (k - 1)Q_k^2 \leq 1/100$$

(Assertion 1) and we argued that appropriate choice of $k$ could make

$$(k - 1)Q_k^2 \ll Q_k.$$  

We also saw (Assertion 4) that the denominator error for the estimate $\hat{\delta}^{(3)}_M$ is insignificant if

$$c(r, t)Q_M \leq 1/25$$

and $c(r, t)$ is near one for $r(t)$ a polynomial of low degree.

Our Monte Carlo work shows that denominator error can increase the variance of the estimate $\hat{\delta}^{(1)}_M$ by a factor as large as 5 over what it would be if the denominator were truly nonrandom. Thus, in choosing an estimate, we must make denominator error insignificant whenever possible.

From the arguments above, if

$$Q_k > 1/100$$

but

$$c(r, t)Q_M < 1/25$$
the denominator error for $\hat{\theta}_M^{(3)}$ may be assumed insignificant, while the same is not true for $\hat{\theta}_M^{(2)}$. The identification of a situation in which these conditions hold must rely upon preliminary estimation of $\sigma^2$, the variance parameter in the circular normal error, and $r(t)$, questions which are discussed in Chapter III. It seems likely that the existence and identification of this unusual situation would be a rare event. Consequently, for the general case, the statistician must rely upon the estimate $\hat{\theta}_M^{(2)}$.

Because of the effect of denominator error, the efficiency of $\hat{\theta}_M^{(2)}$ is not a linear function of $M$. For $M, k$ such that

$$Q_k > 1/100$$

denominator error causes $\hat{\theta}_M^{(2)}$ to be a poor estimate of $\theta$, a situation which is intuitively clear and is given some empirical support in the Monte Carlo result. Thus, in a problem of this type, it is imperative that $M$ be chosen sufficiently large and $k$ chosen such that

$$Q_k + (k - 1)Q_k^2 \leq 1/100.$$
Chapter III: Some Additional Results.

A. Estimating Other Parameters.

In this section estimates of $\sigma^2$ and $r(t)$ are proposed, and the asymptotic distribution theory derived. We assume here that $\beta$ is estimated by $\hat{\beta}_M^{(2)}$.

It is natural to estimate $\sigma^2$ by the residual orthogonal distance, namely

$$(M - 1)\hat{\sigma}_M^2 = \frac{\sum_{i=1}^{M} (y_i - \bar{\hat{y}} - \hat{\beta}_M^{(2)}(x_i - \bar{x}))^2}{1 + \hat{\beta}_M^{(2)}^2}$$

assuming without loss of generality that $\beta = 0$. Now

$$\hat{\beta}_M^{(2)} = O_p(1/\sqrt{M})$$

$$\sum_{i=1}^{M} (y_i - \bar{y})(x_i - \bar{x}) = O_p(\sqrt{M})$$

$$\sum_{i=1}^{M} (x_i - \bar{x})^2 = O_p(M)$$

Consequently

$$(M - 1)\hat{\sigma}_M^2 = \frac{\sum(y_i - \bar{y})^2 + O_p(1)}{1 + O_p(1/M)}$$

Furthermore

$$\sum_{i=1}^{M} (y_i - \bar{y})^2 = \sigma^2 \chi^2_{M-1}$$
Collecting these results

\[(M - 1)\hat{\sigma}_M^2 \Rightarrow \frac{\sigma^2}{M-1}\]

as \(M \rightarrow \infty\). Thus \(\hat{\sigma}_M^2\) is asymptotically unbiased and has known asymptotic distribution.

To estimate \(r(t)\), suppose the statistician is prepared to assume that \(r(t)\) is a polynomial of degree less than or equal to \(d\). Let

\[f = M/d\]

Then, for any \(M\), we find the group averages as for the grouped orthogonal least squares procedure, where now the number of groups is \(d\). We have

\[(\bar{x}_1, \bar{y}_1) \quad i = 1, 2, \ldots, d\]

Define

\[\bar{s}_i = \frac{(\bar{x}_1 - \bar{x}) + \hat{\beta}_M^{(2)}(\bar{y}_1 - \bar{y})}{\sqrt{1 + \hat{\beta}_M^{(2)}}}\]

so that \(\bar{s}_i\) is simply the distance between \((\bar{x}, \bar{y})\) and \((\bar{x}_1, \bar{y}_1)\) onto the line \(y - \bar{y} = \hat{\beta}_M^{(2)}(x - \bar{x})\). Thus \(\bar{s}_i\) is an estimate of \(\bar{r}_i - \bar{r}\). Now

\[(\bar{x}_1 - \bar{x}) = (\bar{r}_1 - \bar{r}) + O_p(1/\sqrt{T})\]

\[(\bar{y}_1 - \bar{y}) = O_p(1/\sqrt{T})\]

and

\[\hat{\beta}_M^{(2)} = O_p(1/\sqrt{M})\]

56
Consequently

\[ \sqrt{f}[\bar{s}_i - (\bar{r}_i - \bar{r})] \overset{d}{\rightarrow} N(0, \sigma^2 (1 - 1/d)) \]

as \( M \to \infty \) and

\[ \text{Cov}(\bar{s}_i, \bar{s}_j) \to -\sigma^2/d \]

as \( M \to \infty \) for all \( i, j = 1, 2, \ldots, d \).

Thus the random variables \( \bar{s}_i, i = 1, 2, \ldots, d \) are converging jointly in distribution to the multivariate normal distribution with covariance matrix known up to a constant multiple. Since the observations were assumed to be equally spaced in time, or spaced in some other way known to the statistician, we have asymptotically the problem of estimating the polynomial \( r(t) - \bar{r}(t) \) when \( r(t) \) has degree \( \leq d \) and we have \( d \) observations at \( d \) different points, each being the polynomial at that point plus normally distributed error. Since the covariance matrix of the errors is known up to a constant multiple, we have asymptotically the standard problem in polynomial regression.

It is unfortunate that nothing can be said for moderate sample sizes for these two problems. The distribution theory for moderate sample sizes is hopelessly complex, hence if it is necessary in practice to rely upon these large sample results.

B. Estimating \( \theta \) when the Error is not Circular Normal.

It is important to consider the robustness of the grouped orthogonal least squares procedure against deviations from the assumptions. We note that the grouped orthogonal least squares estimate is clearly
robust against nonnormality, since the observations are immediately averaged. It is necessary only that the distribution of the errors obey one of the classical criteria for normal convergence of sums, e.g. the Lindeberg Feller condition. Then $\widehat{\theta}_M^{(2)}$ will be asymptotically normally distributed with $\sigma^2$ in $Q_k$ being the variance in the asymptotic distribution of the error sample mean.

Suppose, however, that

$$x_i = r_i + \delta_i \quad i = 1, 2, \ldots, M$$

$$y_i = \epsilon_i$$

where $\theta$ is assumed to be equal to zero without loss of generality and

$$\begin{pmatrix} \delta_i \\ \epsilon_i \end{pmatrix} \sim N \left( 0, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \right)$$

that is the error distribution is normal but not circular normal.

Theorem IIIB.1: If

$$\hat{\beta}_M^{(\theta)} = \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2}$$

and $\hat{\beta}_M^{(2)} = \tan^{-1}\hat{\beta}_M^{(2)}$, as before, then

$$\sqrt{M}(\hat{\theta}_M^{(2)} - \theta) \overset{d}{\rightarrow} N(0, Q)$$

as $M \to \infty$. 

58
In other words, the estimate $\hat{\theta}_M^{(2)}$ is still asymptotically efficient in this more general case. Note that the Cramer-Rao lower bound does not change.

Proof: It is easy to prove that

$$\mathbb{E} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 + \frac{k-1}{n} \left( \sigma_x^2 - \sigma_y^2 \right)$$

$$\text{Var} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k} (\bar{r}_i - \bar{r})^2 \left[ \frac{6\sigma_x^2}{n} - \frac{2\sigma_y^2}{n} \right] + (k^2 - 1) \left[ \frac{\sigma_x^4}{n^2} + \frac{\sigma_y^4}{n^2} \right]$$

$$- 2 \left[ \frac{\sigma_x^2 \sigma_y^2}{n^2} (k-1)^2 + 2 \frac{\sigma_{xy}^2}{n^2} (k-1) \right]$$

so that

$$\frac{\text{Var} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2}{\mathbb{E} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2} \to 0$$

as $M \to 0$. Thus by Slutsky's theorem the denominator of $\hat{\theta}_M^{(2)}$ can be replaced by its expected value in the limit. Now $\hat{a}_i$ can be given the representation

$$\hat{a}_i = \frac{\sigma_{xy}}{\sigma_y^2} \epsilon_i + \sqrt{\frac{\sigma_{xy}^2}{\sigma_x^2} \frac{\sigma_{xy}^2}{\sigma_y^2}} \ u_i$$

where $u_i$ and $\epsilon_i$ are independent and $u_i \sim N(0, 1)$. Using this representation and the argument followed in Lemma IB.2 it is easily
shown that

$$\frac{\sqrt{n}}{\sqrt{k}} \sum_{i=1}^{k} (x_i - \bar{x})(y_i - \bar{y}) \leq_{\text{N}(0, \sigma^2)} \frac{\sigma_y^2}{k} \sum_{i=1}^{k} (r_i^2 - \bar{r})^2 + \frac{k-1}{M} \left[ \sigma_x^2 \sigma_y^2 + 2 \sigma_{xy}^2 \right]$$

as $M \to \infty$. Collecting these results, and using the fact that $k/M \to 0$, we have

$$\sqrt{M} \hat{\beta}_M^{(2)} \to_{\text{N}} \left( 0, \frac{M \sigma_y^2}{n} \sum_{i=1}^{k} (r_i - \bar{r})^2 + \frac{(k-1)M}{n} \frac{\sigma_x^2 \sigma_y^2 + 2 \sigma_{xy}^2}{\sum_{i=1}^{k} (r_i - \bar{r})^2} \right)$$

as $M \to \infty$, and thus since the second term in the variance is asymptotically negligible

$$\sqrt{M} \hat{\beta}_M^{(2)} \leq_{\text{N}(0, Q)}$$

as $M \to \infty$.

Thus, for all $\theta$,

$$\sqrt{M}(\hat{\beta}_M^{(2)} - \theta) \to_{\text{N}(0, Q)}$$

C. Extending the Tracking Problem to 3 Dimensions.

One of the simplest and most natural extensions of the two dimensional tracking problem is to the problem of estimating a linear track in Euclidean 3 space. Suppose that we have a straight line in 3 space and a particle moving along it according to a path function $r(t)$ and observations
\[
\begin{pmatrix}
  x_1 \\
  y_1 \\
  z_1
\end{pmatrix}
\sim N
\begin{pmatrix}
  r_x(t_i) \\
  r_y(t_i) \\
  r_z(t_i)
\end{pmatrix}
\begin{pmatrix}
  \sigma_i
\end{pmatrix}
\quad i = 1, 2, \ldots, M
\]

where the setting is otherwise equivalent to that in Section II A. To solve this problem, we need only note that this line can be described by the two linear equations

\[
\begin{align*}
  z &= \alpha + \beta x \\
  z &= \gamma + \delta y
\end{align*}
\]

so that, since the two planes described by these equations are orthogonal, this problem reduces immediately to a pair of problems of the type previously considered. Note that \((\hat{\alpha}, \hat{\beta})\) will not be independent of \((\hat{\gamma}, \hat{\delta})\).

D. Partial Information on Ranks of Means through Multiple Observations.

We have seen in the foregoing work that for the problem posed originally by Efron, knowledge of the order of the means leads to efficient estimation of the angle of the path with respect to the \(x\) axis. In biological problems in which the independent variable is subject to error, it is rarely the case that the ranks of the means are known. However, it is often possible to take multiple observations on the \(x\) or \(y\) coordinate. In this section, we illustrate with a simple example how such a possibility might lead to improved estimation procedures.

Consider, for example, two variables \(x\) and \(y\) which are known to be linearly related
\[ y = \alpha + \beta x. \]

We assume without loss of generality that \( \beta = 0 \). Suppose that the statistician can take an observation either of the form

\[
\begin{pmatrix}
    x_i \\
    y_i
\end{pmatrix} \sim N\left( \begin{pmatrix}
    \mu_i \\
    \alpha
\end{pmatrix}, \begin{pmatrix}
    \sigma^2 I
\end{pmatrix} \right)
\]

where \( \mu_i \) is uniformly distributed on \( [-\frac{1}{2}, \frac{1}{2}] \) or of the form

\[
\begin{pmatrix}
    x_{i1} \\
    x_{i2} \\
    y_i
\end{pmatrix} \sim N\left( \begin{pmatrix}
    \mu_i \\
    \alpha
\end{pmatrix}, \begin{pmatrix}
    \sigma^2 I
    & 0 \\
    0 & I
\end{pmatrix} \right)
\]

which simply represents taking two independent observations on the \( x \) coordinate. Statistician one chooses to take \( M \) observations of the first type and to estimate \( \theta \) by

\[
\tan 2\hat{\theta}_M = \frac{\sum_{i=1}^{M} (x_i - \bar{x})(y_i - \bar{y})}{2 \left( \sum_{i=1}^{M} (x_i - \bar{x})^2 - \sum_{i=1}^{M} (y_i - \bar{y})^2 \right)}
\]

where \( \theta = \tan^{-1} \beta \). From Chapter II, we know that

\[
\sqrt{M} \hat{\theta}_M \sim N\left( 0, \frac{2}{P} + \frac{\lambda^2}{P^2} \right)
\]

as \( M \to \infty \), and since

\[
P = \int_{-\frac{1}{2}}^{\frac{1}{2}} t^2 dt = 1/12
\]
\[ \sqrt{M} \hat{\theta}_M \sim \mathcal{N}(0, 12\sigma^2 + 144\sigma^4) \]

Now \( \hat{\theta}_M \) is not efficient. Furthermore, since statistician one cannot group the data, he is likely to be plagued by large "denominator error", that is an increase in variance due to the variance of the denominator in \( \tan 2\hat{\theta}_M/2 \). Consequently, for certain sample sizes statistician one will do very poorly indeed.

Statistician two, on the other hand, chooses to take \( 2/3 M \) observations of the form

\[
\begin{pmatrix}
  x_{i1} \\
  x_{i2} \\
  y_i
\end{pmatrix}
\quad i = 1, 2, \ldots, 2/3 M
\]

so that \( x_1, x_{12} \) are independent and identically distributed around the same \( \mu_1 \). He then defines a two group estimate by the following rule. Define Groups \( G_1 \) and \( G_2 \) by

\[
(x_{i2}, y_i) \in \begin{cases}
  G_1, & x_{i1} < 0 \\
  G_2, & x_{i2} > 0
\end{cases}
\]

and let \( (\bar{x}_1, \bar{y}_1), (\bar{x}_2, \bar{y}_2) \) be the group averages for groups \( G_1, G_2 \). The critical feature of this procedure is that the information on ranks, namely \( x_{11} \), is independent of \( (x_{12}, y_i) \). Now, let

\[
\hat{\theta}'_M = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}
\]

and \( \hat{\theta}'_M = \tan^{-1}\hat{\theta}'_M \). Now
\[ \bar{x}_2 - \bar{x}_1 \xrightarrow{P} E(\bar{x}_2) - E(\bar{x}_1) \quad \text{as} \quad M \to \infty. \]

Writing

\[ E(\bar{x}_j) = \bar{\mu}_j \quad j = 1, 2 \]

we have

\[ E(\bar{\mu}_1) = \frac{3}{2M} \sum_{i=1}^{2/3M} [1 - \Phi(\mu_i)]\mu_i \]

\[ = \int_{-\frac{1}{2}}^{\frac{1}{2}} [1 - \Phi(x)]dx \]

and integrating by parts

\[ E(\bar{\mu}_1) \approx -.11 \]

and similarly

\[ E(\bar{\mu}_2) = .11. \]

Now \( E(\hat{\beta}_M') = 0 \), and

\[ \sqrt{M}\left(\hat{\beta} - \frac{\bar{y}_2 - \bar{y}_1}{\bar{\mu}_2 - \bar{\mu}_1}\right) \xrightarrow{P} 0 \]

as \( M \to \infty \). Consequently

\[ M \text{Var} \frac{\hat{\gamma}_M'}{M} \to \frac{3\sigma^2}{M} + \frac{3\sigma^2}{M} \]

\[ \left(\frac{.22}{2}\right)^2 \]

as \( M \to \infty \), or
\[
M \text{ Var } \hat{B}_M \rightarrow \frac{6}{M(.22)^2} \approx \frac{124}{M}.
\]

Thus, even this simple procedure leads to an increase in asymptotic efficiency, relative to the naive procedure. Furthermore, grouping handles the problem of denominator error for moderate sample sizes.

It is clear that this procedure can be generalized. Increasing the number of groups will further improve efficiency. And the reduction of denominator error can often be a critical factor. But, as is clear from the simplicity of this discussion, the question of independent information on ranks of the means remains essentially unexplored.
Chapter IV. Tracking Problem When the Information on Ranks is Stochastic.

A. Introduction.

In previous chapters, we considered the tracking problem as a problem in which observations are received sequentially and the ranks of the observation means along the track are known exactly. There are some situations in which the information on the ranks of the observation means is itself stochastic. In this chapter, we consider tracking problems having this additional stochastic aspect.

Suppose a Biologist is studying two aspects, \( x \) and \( y \), of captured young birds and that he wishes to estimate a relationship of the form

\[
y = \alpha + \beta x.
\]

Suppose also that

\[
x_i \sim N(r_i \cos \theta, \sigma^2)
\]

\[
y_i \sim N(r_i \sin \theta, \sigma^2)
\]

where \( r_i \) is the age of the \( i \)th bird. Further \( r_i \) is not known exactly, but is estimated through \( t_i \), the wing spread, where for young birds

\[
t_i \sim \text{Exp}(r_i, \delta)
\]

an exponential distribution to the right of \( r_i \) with parameter \( \delta \). We see that this is a kind of tracking problem where rank \( \{r_i\} \) is no longer known but must be estimated in some fashion from \( t_i \).
In another example, a Marine Biologist studying three types of ocean bacteria knows that the population density \( y \) of type I and the population density \( x \) of type II obey the relationship

\[
y = \alpha + \beta x
\]

and that at the \( i^{th} \) measurement

\[
x_i \sim N(r_i \cos \theta, \sigma^2)
\]

\[
y_i \sim N(r_i \sin \theta, \sigma^2)
\]

where \( x_i = r_i \cos \theta \) and \( y_i = r_i \sin \theta \) are the true population densities of types I and II and \( r_i \) is the population density of type III. However \( r_i \) itself must be estimated by a measurement

\[
t_i \sim N(r_i, \sigma^2).
\]

In both of these examples, the parameter rank \( (r_i) \) is not observable.

In this chapter, we show that the grouped orthogonal least squares procedure and the linear trend procedure as defined in the introduction to this dissertation can be modified to give a solution of the generalized problem. In the development of the theory, we assume that \( r(t) \), the track function, is linear.

We show that the grouped orthogonal least squares estimate of \( \theta \), \( \hat{\theta}^{(2)} \), and the linear trend estimate of \( \theta \), \( \hat{\theta}^{(3)}_M \), have essentially equal asymptotic variance for a wide range of values of \( \delta \) in both
the normal error and exponential error problems. We infer that \( \hat{s}_M^{(2)} \) is the preferred estimate whenever \( r(t) \) has a nonlinear component. In Section B, we show that both estimates are better than 95% efficient in the sense of Cramér-Rao for selected values of \( \delta \), when \( r(t) \) is linear.

B. The Tracking Problem When the Information on Ranks has Normally Distributed Error.

Suppose we have \( M \) observations

\[
\begin{pmatrix}
X_i \\
Y_i
\end{pmatrix} \sim N\left( \begin{pmatrix}
\alpha + r_i \cos \theta \\
\alpha + r_i \sin \theta
\end{pmatrix}, \sigma_i^2 \right)
\quad i = 1, 2, \ldots, M
\]

where \( r_i = a + bi \) for all \( i \). Suppose further that \( r_i \) is not observed, but instead we observe

\[
t_i \sim N(r_i, \sigma^2).
\]

This is the general tracking problem with track

\[
y = \alpha + \beta x
\]

where \( \beta = \tan \theta \), except that now the ranks of the \( r_i \) are not known and we must use the information in the \( t_i \) observations.

Since we are considering estimation of \( \theta \) by \( \hat{s}_M^{(2)} \), the grouped orthogonal least squares estimate, and \( \hat{s}_M^{(3)} \), the linear trend estimate, both of which are rotation invariant procedures, we may assume \( \theta = 0 \) without loss of generality. To simplify analysis, we assume

68
\[ r_i \sim U[0, 1] \text{ for all } i \]

i.e. the \( r_i \) are i.i.d. and uniform on \([0, 1]\), for this is an equivalent problem asymptotically. Then we have

\[
(x_i, y_i, r_i, t_i) \quad i = 1, 2, \ldots, M
\]

where

\[
\begin{pmatrix}
  x_i \\
  y_i
\end{pmatrix} \sim N\left( \begin{pmatrix} r_i \\
  0 
\end{pmatrix}, \sigma^2 I \right) \text{ for all } i, \text{ given } r_i
\]

\[ r_i \sim U[0, 1] \]

and

\[ t_i \sim N(r_i, \sigma^2) \text{ for all } i, \text{ given } r_i. \]

Thus our problem is to estimate \( \theta = \tan^{-1} \beta \) where we have assumed \( \theta = 0 \) without loss of generality. We begin by redefining the estimates \( \hat{\beta}_M^{(2)} \) and \( \hat{\beta}_M^{(3)} \).

Let the observations be ranked according to the ranks of the \( t_i \) so that

\[
(x_{[j]}, y_{[j]})
\]

is such that the associated \( t \) value, \( t_{[j]} \) has rank \( j \) among the \( M \) observations. Now we redefine \( \hat{\beta}_M^{(2)} \) as follows:
Let
\[
\hat{\beta}_M^{(2)} = \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2}
\]

where \(k\) is the number of groups, \(n = M/k\) and
\[
\bar{x}_i = \frac{\sum_{j=ni+1}^{ni+l} x[j]}{n}
\]
\[
\bar{y}_i = \frac{\sum_{j=ni+1}^{ni+l} y[j]}{n}
\]

Then \(\hat{\gamma}_M^{(2)} = \tan^{-1} \hat{\beta}_M^{(2)}\). Further we define \(\hat{\gamma}_M^{(3)}\) by
\[
\hat{\beta}_M^{(3)} = \frac{\sum_{j=1}^{M} (y[j] - \bar{y})(j - \bar{j})}{\sum_{j=1}^{M} (x[j] - \bar{x})(j - \bar{j})}
\]

and \(\hat{\gamma}_M^{(3)} = \tan^{-1} \hat{\beta}_M^{(3)}\). Thus, these estimates are equivalent to those of previous chapters except that we now proceed as though the ranks of the \(t\) values were actually the true ranks of the observation means. The efficiencies of \(\hat{\beta}_M^{(2)}\) and \(\hat{\beta}_M^{(3)}\) will be a function of \(\delta\), the parameter of variance of the \(t_i\) values as estimates of \(r_i\). Let us look at the asymptotic situation as \(M \to \infty\). First we consider the estimate \(\hat{\beta}_M^{(2)}\) for the case when \(k\) remains fixed.

Note that, unconditionally, \(t_i\) are i.i.d. for all \(M\) according to the distribution
\[ f(t) = \phi(t/\delta) - \phi(t-1/\delta) \]

and that the conditional density of \( r \) given \( t \) is

\[ f(r|t = t_0) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2}\left(\frac{r-t_0}{\delta}\right)^2} \]

\[ \phi(t_0/\delta) - \phi(t_0-1/\delta) \]

Theorem IVB.1: When \( k \) is fixed

\[ \sqrt{M} \hat{g}^{(2)} \mathop{\Rightarrow}^{P} N \left( 0, \frac{\sigma^2}{1/k} \sum_{i=1}^{k} (c_i - \bar{c})^2 \right) \]

as \( M \to \infty \), where

\[ F^{-1}(t'_i) = i/k \quad i = 0, 1, \ldots, k \]

and

\[ c_i = E(r|t \in (t'_{i-1}, t'_i)) \]

Proof: Refer to Theorem IB.1. Since \( k \) is fixed and \( M \to \infty \), we know that \( \hat{x}_i \) and \( \tilde{y}_i \) converge to constants by the law of large numbers. Now \( \tilde{y}_i \) and \( \tilde{y} \) converge in probability to zero so that
\[
\sum_{i=1}^{k} (\tilde{y}_i - \tilde{y})^2 \to 0.
\]

Let

\[ n_i = \text{the number of } t \text{ values less than } t'_i. \]

Then \( n_i \) is Binomial with parameters \( i/k \) and \( M \). As \( M \to \infty \)

\[ n_i = n \cdot i + o(\sqrt{M}). \]

But then, let

\[ x^*_i = \text{the mean value of all } x_i \text{ such that } t_1 \in \{t'_{i-1}, t'_i\}. \]

Because \( n_i = n \cdot i + o(\sqrt{M}) \), both \( x^*_i \) and \( \bar{x}_i \) are comprised of a number of values of the order of \( M \), and the number of values which they do not hold in common is of the order of \( \sqrt{M} \), so that

\[ |\tilde{x}^*_i - \tilde{x}_i| \overset{P}{\to} 0 \text{ as } M \to \infty. \]

Further

\[ x^*_i \overset{P}{\to} c_i \text{ as } M \to \infty \]

by the law of large numbers, hence

\[ \bar{x}_i \overset{P}{\to} c_i \text{ as } M \to \infty. \]

Then

\[ \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 - \sum_{i=1}^{k} (\tilde{y}_i - \tilde{y})^2 \to \sum_{i=1}^{k} (c_i - \bar{c})^2 \]

as \( M \to \infty. \)
Furthermore, as a direct consequence of Lemma IV.3,

\[ \sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \overset{d}{\to} N(0, \frac{k^2}{M} \sum_{i=1}^{k} (\bar{c}_i - \bar{c})^2) \]

Finally, when \( \theta = 0 \), we know that \( \sqrt{M} \hat{\beta}^{(2)}_M \) and \( \sqrt{M} \hat{\theta}^{(2)}_M \) where

\[ \hat{\theta}^{(2)}_M = \tan^{-1} \hat{\beta}^{(2)}_M \]

converge in law to the same distribution as \( M \to \infty \).

This completes the proof. \( \Box \)

In Section D, we give numerical results for various values of \( \theta \).

Suppose now that \( k \to \infty \), but sufficiently slowly that \( k^2/M \to 0 \) as \( M \to \infty \) and \( \max_{i=1, \ldots, M} |\bar{x}_i - E(\bar{x}_1)| \overset{P}{\to} 0 \) as \( M \to \infty \). Although such a rate is not easily characterized, one would in this way obtain the best possible asymptotic performance for the grouped orthogonal least squares estimate, thus

**Theorem IVB.2:** If \( k \to \infty \) at such a rate

\[ \sqrt{M} \hat{\theta}^{(2)}_M \to N(0, r^2/\text{Var}(r|t)) \]

where

\[ \text{Var}(r|t) = E[(r|t) - E(r)]^2 \]

i.e. the variance of the conditional distribution.

**Proof:** The proof proceeds as in Theorem IVB.1. However, since \( k \to \infty \) we have also that

\[ \sum_{i=1}^{K} (c_i - \bar{c})^2 \to E((E(r|t) - E(r))^2 \]

since
\[ c_1 = E(x| t \in \{ t_{1-1}, t_{1} \} ) \quad \text{QED} \]

Again see Section D for numerical results as a function of \( \theta \).

We turn now to consideration of the estimate \( \hat{\theta}_M^{(3)} = \tan^{-1} \hat{\alpha}_M^{(3)} \).

Consider the random variable rank \( [t_1] \) which is i.i.d. for all \( M \).

Then rewrite

\[
\hat{\beta}_M^{(3)} = \frac{\sum_{i=1}^{M} (y_i - \bar{y})(\text{rank}(t_i) - M/2)}{\sum_{i=1}^{M} (x_i - \bar{x})(\text{rank}(t_i) - M/2)}.
\]

Dividing both numerator and denominator by \( M \), we see that both numerator and denominator are covariance coefficients, where \( y \) and \( t \) are independent and \( x \) and \( t \) are dependent.

Writing

\[
\sqrt{M} \hat{\beta}_M^{(3)} = \frac{\frac{1}{M^{3/2}} \left[ \sum_{i=1}^{M} y_i \cdot \text{rank}(t_i) - \frac{M^2 \bar{y}}{2} \right]}{\frac{1}{M^2} \left[ \sum_{i=1}^{M} x_i \cdot \text{rank}(t_i) - \frac{M^2 \bar{x}}{2} \right]}
\]

we have

Theorem IVB.3:

\[
\sqrt{M} \hat{\beta}_M^{(3)} \xrightarrow{d} N\left( 0, \frac{\sigma^2}{12 \text{ Cov}^2(x, \frac{\text{rank}(t)}{M})} \right)
\]

as \( M \to \infty \).

Proof: Write

\[
\sqrt{M} \hat{\beta}_M^{(3)} = \frac{N_M}{D_M}.
\]
\( D_M \) is a sample covariance coefficient so

\[
D_M \overset{p}{\to} \text{Cov}\left(r, \frac{\text{rank}(t)}{M}\right) \quad \text{as} \quad M \to \infty
\]

by the law of large numbers. Further

\[
N_M \overset{\mathcal{L}}{\to} \frac{1}{\sqrt{M}} \left[ \frac{1}{M} \sum_{i=1}^{M} \frac{\text{rank}(t_i)}{M} - \frac{MV^2}{2} \right]
\]

is simply \( \sqrt{M} \) times a covariance coefficient between independent random variables. By the central limit theorem

\[
N_M \overset{\mathcal{L}}{\to} N(0, \text{Var}(y)\text{Var}\left(\frac{\text{rank}(t)}{M}\right))
\]

or

\[
N_M \overset{\mathcal{L}}{\to} N(0, \sigma^2/12)
\]

since \( \frac{\text{rank}(t)}{M} \) converges in law to a uniform distribution on \([0, 1]\).

This completes the proof.

QED

We will evaluate the term \( \text{Cov}\left(r, \frac{\text{rank}(t)}{M}\right) \) in the section on numerical results. Before giving the numerical results, we turn to the theoretical discussion of the case in which \( t \) has an exponential distribution.

C. The Tracking Problem When the Information on Ranks has Exponentially Distributed Error.

Suppose we have \( M \) observations

\[
(x_i, y_i) \quad i = 1, 2, \ldots, M
\]
whose distribution conditioned on the parameter $r_i$ is

\[
\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left(\begin{pmatrix} r_i \\ 0 \end{pmatrix}, \sigma^2 I \right) \quad i = 1, 2, \ldots, M
\]

where $r_i$ is itself uniformly distributed on $[0, 1]$. Suppose further that $r_i$ is not observed but instead $t_i$ where

\[
f(t_i/r_i) = 1/\delta \ e^{-\delta t_i} e^{5r_i} \quad t_i > r_i.
\]

The situation is comparable to that in Section B and our analysis proceeds as before. The c.d.f. of $t$ is now

\[
F(t) = \begin{cases} 
0 & t < 0 \\
-t - 5^{-1} e^{-5t} (e^{5t} - 1) & 0 < t < 1 \\
1 - 5^{-1} e^{-5t} (e^5 - 1) & 1 < t
\end{cases}
\]

and

\[
f(r|t) = \begin{cases} 
0 & r > t \\
5e^{5r}/(e^{5t} - 1) & 0 < r < t < 1 \\
5e^{5r}/(e^5 - 1) & 0 < r < t, t > 1.
\end{cases}
\]

But our analysis of Section B was independent of the exact form of $F(t)$. Thus we have

**Theorem IVC.1:** If $k$ is fixed as $M \to \infty$

\[
\sqrt{M} \frac{\hat{q}_M^{(2)}}{\sigma^2} \to N\left(0, \frac{\sigma^2}{1/k \sum_{i=1}^k (c_i - \bar{c})^2} \right)
\]

as $M \to \infty$ where
\[ F(t'_i) = \frac{i}{k} \]

and

\[ c_i = E(r|t \in (t'_{i-1}, t'_i)) \]

Theorem IVC.2: If \( k \to \infty \) at such a rate that

\[ \frac{k^2}{M} \to 0 \quad \text{and} \quad \max_{i=1,2,\ldots,k} |\bar{x}_i - E(\tilde{x}_i)|_p \to 0 \quad \text{as} \quad M \to \infty \]

then

\[ \sqrt{M} \hat{\theta}_M^{(2)} \xrightarrow{p} N(0, \frac{\sigma^2}{\text{Var}(r|t)}) \]

where

\[ \text{Var}(r|t) = E[(E(r|t) - E(r))^2] \]

Theorem IVC.3:

\[ \sqrt{M} \hat{\theta}_M^{(3)} \xrightarrow{p} N\left(0, \frac{\sigma^2}{12 \text{Cov}^2\left(r, \frac{\text{rank}(t)}{M}\right)}\right) \]

as \( M \to \infty \).

In Section D, results of numerical integrations are given so that \( \hat{\theta}_M^{(2)} \) and \( \hat{\theta}_M^{(3)} \) are compared in asymptotic variance for various values of 8.

D. Comparison of the Estimates.

In this section, we report the results of numerical integrations which give asymptotic variances for the estimates \( \hat{\theta}_M^{(2)} \) and \( \hat{\theta}_M^{(3)} \) for various values of 8.
1. The Normal Case.

We considered the case \( k = 10 \). To find the asymptotic variance of \( \hat{\sigma}_M^{(2)} \) for \( k = 10 \), we must find \( t'_i \) where

\[
f(t'_i) = i/10 \quad i = 1, \ldots, 9
\]

and

\[
c_i = E(r \mid t \in \{t'_{i-1}, t'_i\}) \quad i = 1, 2, \ldots, 10.
\]

Since \( F(t) \) is monotone, any simple approximation procedure will find the values \( t'_i \). Then

\[
c_i = 10 \int_{t'_{i-1}}^{t'_i} E(r \mid t) df(t) dt
\]

where

\[
E(r \mid t) = \frac{1}{\Phi(t/\delta) - \Phi(t-1/\delta)} \int_0^1 r \varphi(t - r/\delta) dr
\]

\[
= t + \delta \left\{ \frac{\varphi(t/\delta) - \varphi(t - 1/\delta)}{\Phi(t/\delta) - \Phi(t - 1/\delta)} \right\}
\]

and

\[
df(t) = \Phi(t/\delta) - \Phi(t - 1/\delta).
\]

This function is directly integrable and the value of \( \frac{1}{10} \sum_{i=1}^{10} (c_i - \bar{c})^2 \) for various values of \( \delta \) is reported in Table I. To find the asymptotic variance of \( \hat{\sigma}_M^{(2)} \) when \( k \to \infty \) we must integrate.
\[
\int_{-\infty}^{\infty} \left[ t + \delta \left( \frac{\phi(t/\delta) - \phi(t-1/\delta)}{\phi(t/\delta) - \phi(t-1/\delta)} \right) \right]^2 dt
\]

and the results of numerical integration for various values of \( \delta \) are also given in Table I.

To find the asymptotic variance of \( \hat{\theta}_M^{(3)} \), we must numerically evaluate the integral

\[
E[\text{rank}(t)] = M \int_{-\infty}^{\infty} \int_{0}^{1} r F(t) \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2}(\frac{t-r}{\delta})^2} dr dt
\]

\[
= M \int_{-\infty}^{\infty} F(t) \int_{0}^{1} r \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2}(\frac{t-r}{\delta})^2} dr dt
\]

\[
= M \int_{-\infty}^{\infty} F(t) [ F(t) - \phi(\frac{t-1}{\delta}) ] dt.
\]

The results appear in Table I.

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( \sqrt{\mathbb{M}} \text{Var}(\hat{\theta}_M^{(2)}) ), ( k=10 )</th>
<th>( \sqrt{\mathbb{M}} \text{Var}(\hat{\theta}_M^{(2)}) ), ( k=\infty )</th>
<th>( \sqrt{\mathbb{M}} \text{Var}(\hat{\theta}_M^{(3)}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.125</td>
<td>( \sigma^2 / .070 )</td>
<td>( \sigma^2 / .071 )</td>
<td>( \sigma^2 / .069 )</td>
</tr>
<tr>
<td>.25</td>
<td>( \sigma^2 / .048 )</td>
<td>( \sigma^2 / .049 )</td>
<td>( \sigma^2 / .048 )</td>
</tr>
<tr>
<td>.5</td>
<td>( \sigma^2 / .020 )</td>
<td>( \sigma^2 / .021 )</td>
<td>( \sigma^2 / .021 )</td>
</tr>
<tr>
<td>1</td>
<td>( \sigma^2 / .006 )</td>
<td>( \sigma^2 / .006 )</td>
<td>( \sigma^2 / .006 )</td>
</tr>
<tr>
<td>2</td>
<td>( \sigma^2 / .002 )</td>
<td>( \sigma^2 / .002 )</td>
<td>( \sigma^2 / .001 )</td>
</tr>
<tr>
<td>4</td>
<td>( \sigma^2 / .0004 )</td>
<td>( \sigma^2 / .0004 )</td>
<td>( \sigma^2 / 1 \times 10^{-5} )</td>
</tr>
</tbody>
</table>

79
We see that all three procedures lead to approximately the same asymptotic distribution. We will discuss these data in more detail after we look at the exponential case.

2. The Exponential Case.

To find the asymptotic distribution of \( \hat{\theta}_M^{(2)} \) for this case when \( k = 10 \), we first approximated

\[
t_i' \quad i = 1, 2, \ldots, 9
\]

where \( F(t_i') = i/10 \), and then evaluated the integral

\[
\int_{t_i'}^{t_i} E(r|t)dF(t) = c_i
\]

where \( F(t) \) is defined in Section C and

\[
E(r|t) = \begin{cases} 
\frac{te^{st} - 1}{e^{st} - 1} & t < 1 \\
\frac{e^{s-1}(e^{st} - 1)}{e^{s} - 1} & t > 1
\end{cases}
\]

Results of the numerical integration are reported in Table II for various values of \( \delta \). To find the asymptotic variance of \( \hat{\theta}_M^{(2)} \) when \( k = \infty \), one evaluates the integral

\[
\int_{0}^{\infty} E^2(r|t)dF(t) .
\]

Again see Table II for results of the numerical integration. Finally, to find the asymptotic variance of \( \hat{\theta}_M^{(3)} \), one must find
$$E\{r \cdot \frac{\text{rank}(t)}{M}\}.$$ 

Since 

$$E\{r \cdot \frac{\text{rank}(t)}{M}\mid r, t\} = rF(t)$$ 

we have 

$$E\{r \cdot \frac{\text{rank}(t)}{M}\} = \int_0^1 \int_r^\infty r \cdot F(t) f_r(t) dr dt$$ 

where 

$$f_r(t) = 8e^r e^{-8t} \quad t > r.$$ 

Again see Table II.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\sqrt{M \text{ Var}(\hat{\theta}^{(2)}_M), k=10}$</th>
<th>$\sqrt{M \text{ Var}(\hat{\theta}^{(2)}_M), k=\infty}$</th>
<th>$\sqrt{M \text{ Var}(\hat{\theta}^{(3)}_M)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>$\sigma^2/0.066$</td>
<td>$\sigma^2/0.068$</td>
<td>$\sigma^2/0.068$</td>
</tr>
<tr>
<td>4</td>
<td>$\sigma^2/0.053$</td>
<td>$\sigma^2/0.054$</td>
<td>$\sigma^2/0.054$</td>
</tr>
<tr>
<td>3</td>
<td>$\sigma^2/0.046$</td>
<td>$\sigma^2/0.047$</td>
<td>$\sigma^2/0.043$</td>
</tr>
<tr>
<td>2.5</td>
<td>$\sigma^2/0.040$</td>
<td>$\sigma^2/0.041$</td>
<td>$\sigma^2/0.058$</td>
</tr>
<tr>
<td>2</td>
<td>$\sigma^2/0.034$</td>
<td>$\sigma^2/0.035$</td>
<td>$\sigma^2/0.030$</td>
</tr>
<tr>
<td>1.5</td>
<td>$\sigma^2/0.027$</td>
<td>$\sigma^2/0.027$</td>
<td>$\sigma^2/0.021$</td>
</tr>
<tr>
<td>1</td>
<td>$\sigma^2/0.018$</td>
<td>$\sigma^2/0.018$</td>
<td>$\sigma^2/0.012$</td>
</tr>
<tr>
<td>0.75</td>
<td>$\sigma^2/0.014$</td>
<td>$\sigma^2/0.014$</td>
<td>$\sigma^2/0.008$</td>
</tr>
<tr>
<td>0.5</td>
<td>$\sigma^2/0.009$</td>
<td>$\sigma^2/0.009$</td>
<td>$\sigma^2/0.004$</td>
</tr>
</tbody>
</table>
From Tables I and II, we see that the three estimation procedures are almost equally efficient and that efficiency is sensitive to the parameter $s$.

Since our analysis proceeded on the assumption that $r(t)$ is linear, one would in general prefer $\hat{\theta}_M^{(2)}$ the grouped orthogonal least squares estimate, since $\hat{\theta}_M^{(3)}$ was shown in Chapter II to lose efficiency when $r(t)$ is not linear.

In Section E, we discuss efficient estimation and find the Cramér-Rao lower bound for this problem, and in Section F we discuss estimation of $s$.

E. Efficient Estimation and the Cramér-Rao Lower Bound.

In this section we show that, both the grouped orthogonal least squares procedure and the linear trend procedure are better than 95% efficient for estimating $\theta$ when $t$ has small variance, in the sense of Cramér-Rao.

The method is as follows. Once we have the density function $f_\theta(x, y, t)$, we know that the Cramér-Rao lower bound for estimating $\theta$ is of the form

$$\sigma_\theta^2(\theta) \geq \frac{1}{nE \left[ \frac{\partial}{\partial \theta} \log f_\theta(x, y, t) \right]^2}$$

for unbiased estimates. Further, we know that

$$E \left[ \frac{\partial}{\partial \theta} \log f_\theta(x, y, t) \right]^2 = \int_x \int_y \int_t \frac{\partial}{\partial \theta} \left[ f_\theta(x, y, t) \right]^2 f_\theta(x, y, t) \, dx \, dy \, dt$$

82
an integration which can be performed numerically. It remains only to
find \( f_\theta(x, y, t) \).

For the normal case, we have

\[
f_\theta(x, y, t) = \int_0^1 \frac{1}{(2\pi)^{3/2}\delta} e^{-\frac{1}{2}(x-r \cos \theta)^2 - \frac{1}{2}(y-r \sin \theta)^2 - \frac{1}{2}(t-r)^2} dr
\]

when we assume \( \sigma = 1 \) and \( \delta \) known. Thus \( \theta \) is the only parameter
and it is possible to integrate with respect to \( r \), differentiate with
respect to \( \theta \) and find the Cramér-Rao lower bound by numerical integra-
tion. It happens that

\[
\left[ \frac{\partial}{\partial \theta} f_\theta(x, y, t) \right]^2 \frac{1}{f_\theta(x, y, t)} = y^2 e^{-y^2/2} R(x, t)
\]

and \( y \) can be integrated directly.

For the exponential case, we have

\[
f_\theta(x, y, t) = \int_0^1 \frac{\delta}{2\pi} e^{-\frac{1}{2}(x-r \cos \theta)^2 - \frac{1}{2}(y-r \sin \theta)^2 + 8r - 8t} dr
\]

where again \( \sigma = 1 \) and \( \delta \) is known. Once more it is true that we
can integrate with respect to \( r \), that

\[
\left[ \frac{\partial}{\partial \theta} f_\theta(x, y, t) \right]^2 \frac{1}{f_\theta(x, y, t)} = y^2 e^{-y^2/2} R(x, t)
\]

and calculation of the Cramér Rao-lower bound involves direct integra-
tion with respect to \( y \) and a double numerical integration. The
Cramér Rao lower bound and the efficiencies of the estimates are
tabled below for two cases of normal error and two cases of exponential error.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\operatorname{Var} \hat{\delta}(2)_M$, $k=10$</th>
<th>$\operatorname{Var} \hat{\delta}(2)_M$, $k=\infty$</th>
<th>$\operatorname{Var} \hat{\delta}(3)_M$</th>
<th>Cramér-Rao Lower Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125</td>
<td>1/0.048</td>
<td>1/0.049</td>
<td>1/0.048</td>
<td>1/0.050</td>
</tr>
<tr>
<td>0.25</td>
<td>1/0.070</td>
<td>1/0.071</td>
<td>1/0.070</td>
<td>1/0.072</td>
</tr>
</tbody>
</table>

Exponential Case

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\operatorname{Var} \hat{\delta}(2)_M$, $k=10$</th>
<th>$\operatorname{Var} \hat{\delta}(2)_M$, $k=\infty$</th>
<th>$\operatorname{Var} \hat{\delta}(3)_M$</th>
<th>Cramér-Rao Lower Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>1/0.053</td>
<td>1/0.054</td>
<td>1/0.054</td>
<td>1/0.055</td>
</tr>
<tr>
<td>0.4</td>
<td>1/0.066</td>
<td>1/0.068</td>
<td>1/0.068</td>
<td>1/0.068</td>
</tr>
</tbody>
</table>

From Table IV, we see that for $\delta$ in this range, all these estimates are 95-100% efficient. We infer thereby that there exists no substantially superior estimate, that the three estimations are substantially equivalent when $r(t)$ is linear, and that the choice of an estimate must depend upon questions such as the linearity of $t$, deviations from the assumptions, etc.

F. Estimating $\delta$.

If the statistician is going to use the $t$ values as rank information, he will want to know how closely the $t$ values approximate the $r$ values, i.e. how large is $\operatorname{Var}(t/r)$ in comparison with $\operatorname{Var}(t)$ or the track length.
Suppose the statistician has a good idea about the length of the track, which he estimates to be $L$. Then

$$\text{Var}(t) = E \text{Var}(t/r) + \text{Var} E[t/r].$$

In the normal case, this simplifies to

$$\text{Var}(t) = \sigma^2 + \text{Var}(r)$$

while in the exponential case we have

$$\text{Var}(t) = 1/\sigma^2 + E(r + 1/\sigma)$$

$$= 1/\sigma^2 + 1/\sigma + E(r).$$

If the statistician has estimated the total track length, he can now estimate $\text{Var}(t)$ directly from the $t$ observations and solve the appropriate equation above for an estimate of $\sigma$. This would be the usual procedure.

In other cases, the statistician may be able to estimate independently the rank correlation between the $r$ ranks and the $t$ ranks, say from data of a previous study. Some sample values of the rank correlation between $r$ and $t$ when $r \sim U[0, 1]$ are given in Table III.

<table>
<thead>
<tr>
<th>Normal Case</th>
<th>Exp. Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>Rank Cor.</td>
</tr>
<tr>
<td>.125</td>
<td>.91</td>
</tr>
<tr>
<td>.25</td>
<td>.75</td>
</tr>
<tr>
<td>.5</td>
<td>.50</td>
</tr>
<tr>
<td>1</td>
<td>.26</td>
</tr>
<tr>
<td>2</td>
<td>.13</td>
</tr>
<tr>
<td>4</td>
<td>.01</td>
</tr>
<tr>
<td>85</td>
<td>.75</td>
</tr>
<tr>
<td>.25</td>
<td>.30</td>
</tr>
<tr>
<td>.25</td>
<td>.21</td>
</tr>
<tr>
<td>.25</td>
<td>.12</td>
</tr>
</tbody>
</table>
In any case, it is essential to get some estimate of $S$, since the procedure is highly sensitive to $S$. One of the above methods will usually be satisfactory.
Bibliography


