Statistical PERT: Improvements in the
Determination of the Project Completion Time Distribution

by

Thomas C. Baker, Jr. and Robert L. Sielken Jr.
Statistical PERT: Improvements in the Determination of the Project Completion Time Distribution

by

Thomas C. Baker, Jr. and Robert L. Sielken Jr.

THEMIS OPTIMIZATION RESEARCH PROGRAM
Technical Report No. 58
August 1978

INSTITUTE OF STATISTICS
Texas A&M University

Research conducted through the Texas A&M Research Foundation and sponsored by the Office of Naval Research Contract N00014-78-C-0426 Project NR047-179

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

This document has been approved for public release and sale; its distribution is unlimited.

ATTACHMENT II
Statistical PERT: Improvements in the Determination of the Project Completion Time Distribution

Thomas C. Baker, Jr. and Robert L. Sielken Jr.
ABSTRACT

This report develops improvements to a new project scheduling procedure, Statistical PERT, being developed at the Institute of Statistics, Texas A&M University. The project scheduling algorithm is a five step iterative procedure capable of determining a minimum cost project schedule when the activities making up the project have durations which are random variables. The cost of an activity is assumed to be a convex piecewise linear function of the activity's mean duration. The problem is to determine the activity mean durations which both minimize the total project cost and insure that the mean (or some specified percentile) of the corresponding project completion time distribution is less than or equal to a specified project deadline. The entire distribution of the project's completion time under the minimum cost schedule is a valuable by-product.

A critical step, Subnetwork Analysis, in the proposed procedure is improved and extended. Subnetwork Analysis determines an estimate of the duration distribution, $F(t)$, for each subnetwork identified in the previous steps. This estimate is extended to include an extrapolation of upper and lower bounds on $F(t)$. This report also develops a new sampling procedure which results in improved estimators for the bounds on $F(t)$. 

TABLE OF CONTENTS

Section                                                                 Page

1. A STATISTICAL APPROACH TO PROJECT SCHEDULING .......................... 1
   1.1 Introduction ............................................................................. 1
   1.2 The Project Scheduling Problem ............................................. 1
   1.3 Outline of the New Approach to Project Scheduling ............... 4
   1.4 An Integrated System of Computer Programs ........................... 6
   1.5 The Determination of the Subnetwork Duration Distribution 7

2. ANALYSIS OF A SUBNETWORK ..................................................... 9
   2.1 Introduction ............................................................................. 9
   2.2 Formation of Clusters ............................................................. 11
   2.3 Bounding the Discrete Subnetwork Duration Distribution F 19
       2.3.1 Upper Bounds on F ......................................................... 19
       2.3.2 Lower Bounds on F ......................................................... 22
       2.3.3 The Tightness of the Bounds on F ................................. 23
   2.4 Using Sampling to Estimate the Upper and Lower Bounds 4
       on F .......................................................................................... 24
   2.5 Estimating F by Extrapolating Between the Upper and 2
       Lower Bounds on F .................................................................. 25
   2.6 A Summary of the Subnetwork Analysis Procedure .................. 26

3. SAMPLE-BASED ESTIMATORS FOR A DISCRETE DISTRIBUTION FUNCTION 29
   3.1 Introduction ............................................................................. 29
   3.2 Some Proposed Estimators ....................................................... 30
       3.2.1 The Empirical Distribution Function, G_1(t) .............. 33
       3.2.2 The Modified Empirical Distribution Function, 33
            G_2(t) ............................................................................... 33
# TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.3 The Continuous Estimator, ( G_3(t) )</td>
<td>33</td>
</tr>
<tr>
<td>3.2.4 The Mixed Estimator, ( G_4(t) )</td>
<td>34</td>
</tr>
<tr>
<td>3.2.5 The Mixed Estimator, ( G_5(t) )</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Criteria for a Good Estimator</td>
<td>34</td>
</tr>
<tr>
<td>3.4 Choosing Between Simple Random or Systematic Sampling</td>
<td>36</td>
</tr>
<tr>
<td>3.4.1 An Ordering Scheme</td>
<td>36</td>
</tr>
<tr>
<td>3.4.2 Implementing the Ordering Scheme</td>
<td>41</td>
</tr>
<tr>
<td>3.5 The Simulation Study</td>
<td>42</td>
</tr>
<tr>
<td>3.5.1 A Comparison of ( G_1(t), \ldots, G_5(t) ) Under Systematic Sampling</td>
<td>45</td>
</tr>
<tr>
<td>3.5.2 The Performance of ( G_2(t) ) Under Systematic and Random Sampling</td>
<td>51</td>
</tr>
<tr>
<td>4. ESTIMATION OF A DISCRETE DISTRIBUTION FUNCTION BY EXTRAPOLATING UPPER AND LOWER BOUNDS</td>
<td>53</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>53</td>
</tr>
<tr>
<td>4.2 The Extrapolation Problem</td>
<td>53</td>
</tr>
<tr>
<td>4.3 A Linear Programming Solution to the Extrapolation Problem</td>
<td>55</td>
</tr>
<tr>
<td>4.4 An Example of the Linear Programming Solution</td>
<td>58</td>
</tr>
<tr>
<td>5. POTENTIAL MODIFICATIONS OF THE SUBNETWORK ANALYSIS PROCEDURE</td>
<td>61</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>61</td>
</tr>
<tr>
<td>5.2 Explicit Evaluation of the Subnetwork Duration Distribution</td>
<td>63</td>
</tr>
<tr>
<td>5.3 Approximating the Subnetwork Duration Distribution ( F(t) )</td>
<td>66</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.3.1 Robillard and Trahan's Lower Bound on $F(t)$</td>
<td>67</td>
</tr>
<tr>
<td>5.3.2 Kleindorfer's Upper and Lower Bounds on $F(t)$</td>
<td>69</td>
</tr>
<tr>
<td>5.3.3 Incorporating Different Methods of Approximating $F(t)$ into Subnetwork Analysis</td>
<td>72</td>
</tr>
<tr>
<td>5.4 Additional Probability Inequalities as Bases for Upper and Lower Bounds on $F(t)$</td>
<td>74</td>
</tr>
<tr>
<td>6. CONCLUDING REMARKS</td>
<td>76</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>77</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Activity Durations for the Subnetwork in Figure 2</td>
</tr>
<tr>
<td>2</td>
<td>The Activity Durations for the Subnetwork in Figure 8</td>
</tr>
<tr>
<td>3</td>
<td>The Approximate Ordering of the x-values for the Subnetwork in Figure 8</td>
</tr>
<tr>
<td>4</td>
<td>Simulation Results for the Highly Skewed Left $F_{23,2}(t)$</td>
</tr>
<tr>
<td>5</td>
<td>Simulation Results for the Skewed Left $F_{8',2}(t)$</td>
</tr>
<tr>
<td>6</td>
<td>Simulation Results for the Symmetric $F_{5,5}(t)$</td>
</tr>
<tr>
<td>7</td>
<td>Simulation Results for the Skewed Right $F_{2,8}(t)$</td>
</tr>
<tr>
<td>8</td>
<td>Simulation Results for the Highly Skewed Right $F_{2,23}(t)$</td>
</tr>
<tr>
<td>9</td>
<td>Ratios of the Empirical Behavior of $G_{2}(t)$ Using Systematic Sampling to that Using Random Sampling</td>
</tr>
<tr>
<td>10</td>
<td>Extrapolation Data</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure                                  Page

1  A small project represented as a directed network .......... 3
2  Subnetwork with activities labeled (activity number; mean duration) ........................................ 12
3  Subnetwork for determining the associates of activity 2 when $\lambda = 1$ ........................................ 14
4  Subnetwork for determining the associates of activity 7 when $\lambda = 1$ ........................................ 14
5  Subnetwork for determining the associates of activity 9 when $\lambda = 1$ ........................................ 15
6  Subnetwork for determining the eliminants of activity 3 when $\theta = 2$ ........................................ 15
7  $G_1(t), G_2(t), G_3(t), G_4(t),$ and $G_5(t)$ for the sample data in (3.7) ........................................ 32
8  A small subnetwork ........................................ 39
9  Duration distributions used in the simulation study .......... 43
10 Extrapolation results for the data in Table 10 .............. 60
11 Subnetworks considered by Hartley and Wortham (1966) .... 64
12 Subnetworks considered by Ringer (1969) ..................... 65
1. A STATISTICAL APPROACH TO PROJECT SCHEDULING

1.1 Introduction

The many technological advances of the last century have resulted in a drastic increase in the magnitude and complexity of man's enterprises. This, in turn, has brought about an acute need for detailed and effective project planning. Thus, in recent years, a search for a general technique which can be employed to simplify the task of cost-effective project scheduling has been undertaken. A host of promising strategies have been proposed, and a few have even enjoyed widespread use. However, the methods currently in use have possibly serious shortcomings (see, for example, Sielken and Hartley (1977)). Therefore, under the sponsorship of the Office of Naval Research, the Institute of Statistics has undertaken the development, implementation, and evaluation of a new project scheduling system that yields reliable results and can be economically applied to very large scheduling problems. This report is a part of that undertaking.

1.2 The Project Scheduling Problem

Project scheduling problems arise in a wide variety of contexts. Consequently, a number of varying formulations of the problem are currently in use. Since these formulations are not all exactly equivalent, this subsection gives the specific formulation considered in this work.
A project is, in general, made up of a series of "tasks" or "activities" which consume time. These activities are represented graphically by directed arcs. The origin point and terminal point of an arc are both called "nodes". The graphical representation of a project, showing the precedence relationships among the various activities, is called a "network"; the first node in a network is usually referred to as the "source" while the last node is usually called the "sink". In addition, the following basic rules are adhered to:

1) Before a particular activity may begin, every other activity whose terminal node is that activity's origin node must be completed.

2) Arcs imply logical precedence only; the length of the arc has no significance.

3) The network cannot contain any loops or cycles.

For example, a small project might consist of activities A, B, C, D, and E with the following precedence relationships:

   i) A must be completed before either C or D can be started;
   ii) B must be completed before D can be started; and
   iii) C and D must both be completed before E can be started.

The corresponding network representation is shown in Figure 1. The arc labeled F does not correspond to any "real" activity but is a "dummy" activity merely representing the precedence relation that A must be completed before D can be started. The circles numbered 1, 2, ..., 5 represent the activities' origin and terminal nodes.
The time required to complete an activity is a random variable. The cost of an activity is a convex piecewise linear function of the activity's mean duration. Thus, a "project schedule" is a specification of each activity's mean duration. The "total project cost" is simply the sum of the corresponding activity costs. The time to complete the entire project is a random variable whose distribution depends upon the activity duration distributions. The objective is to determine a minimum cost project schedule such that the mean (or some percentile) of the corresponding project completion time distribution is less than or equal to a specified project deadline.
1.3 Outline of the New Approach to Project Scheduling

In 1974 the development of a new approach to project scheduling was begun with the support of the Office of Naval Research. The new project scheduling procedure that has resulted is an iterative algorithm involving the following five general steps:

Step 1. Deterministic Scheduling: Find a minimum cost project schedule which completes the project by TARGET TIME when each activity's duration is exactly its mean duration and hence deterministic instead of random. (The initial value of TARGET TIME is usually the specified project deadline.)

Step 2. Simplification: Let each activity's duration be a random variable with distribution corresponding to that activity's mean duration chosen during Deterministic Scheduling. Replace various configurations of activities by single activities. The duration distribution for a replacement activity is the distribution of the time to complete all of the activities in the configuration it is replacing. The result of this step is a simplified project network with fewer activities.

Step 3. Decomposition: Partition the simplified project network into several subnetworks in such a way that the resultant subnetworks can be linked together in
either series or parallel to form the simplified project network.

Step 4. Subnetwork Analysis: Analyze separately each of the subnetworks determined during Decomposition. Within a subnetwork each activity's duration distribution is approximated by a two-point discrete distribution with matching mean, variance, and third moment. Determine the subnetwork duration distribution corresponding to these discrete activity duration distributions.

Step 5. Synthesis: Combine the approximate subnetwork duration distributions to obtain an approximate completion time distribution for the entire project. If the mean (or some specified percentile), $\hat{T}$, of this project completion time distribution is sufficiently close to the specified project deadline, the "optimal" project schedule has been found. Otherwise, reset TARGET TIME to $\text{New TARGET TIME} = \text{Old TARGET TIME} \times (\text{Project Deadline}/\hat{T})$ and return to Step 1.

A general discussion and relatively nonmathematical overview of this project scheduling procedure is contained in Baker and Sielken (1978)(see also Sielken and Hartley (1977)). The detailed documentation of the development thus far of each step is as follows:

Step 1. Dunn and Sielken (1977);
Step 2. Hartley and Wortham (1966) and Ringer (1969);
Step 3. Sielken and Fisher (1976);
Step 4. Sielken, Ringer, Hartley, and Arseven (1974) and Sielken, Hartley, and Spoeri (1976);

1.4 An Integrated System of Computer Programs

Prior to this research, separate computer programs had been written to perform each of the five steps. (These programs are fully documented in the references cited for each step.) However, from a user's viewpoint, this arrangement was awkward because the programs had to be executed one at a time and the output from each step had to be manually modified for use by the next program in the sequence. Thus, one of the objectives of this research has been to fashion the individual programs into an integrated package that is more practicable from a user's viewpoint.

The ability to schedule large projects with as many as 1000 activities in the project network and as many as 500 activities in any simplified subnetwork was one of the desired characteristics of the computer implementation. This objective prohibited the combination of the five original programs into a single large program since doing so would drastically limit the size of the project that could be analyzed because of the computer core storage restrictions. Thus, the computer implementation of the new project scheduling procedure has been in the form of several individual programs internally linked
together. The resulting integrated system of computer programs requires that the user supply the project description and algorithm parameters only to the main (first) program. From then on each program automatically prepares the proper input for the remaining programs in the iterative procedure and stores this information on either disk or tape from which it is retrieved as needed. The job control language automatically calls the individual programs as it cycles through the five step iterative algorithm.

Since the new project scheduling procedure is an iterative procedure, it may repeat Steps 1-5. However, as pointed out in Sielken and Hartley (1977), after the initial performance of Steps 1-3 their subsequent performance is greatly simplified. Thus, special simplified versions of the programs for these steps are called when these steps are repeated. Needless to say, the preparation of these simplified versions has greatly improved the efficiency of the computer implementation.

The new project scheduling software package that has resulted is fully documented in the User's Guide found in Baker and Sielken (1978). Included in Baker and Sielken (1978) is an example with a complete listing of the system's input and output.

1.5 The Determination of the Subnetwork Duration Distribution

Two new theoretical contributions to the project scheduling procedure are documented in this report. Both are improvements to the statistical methodologies used in determining the subnetwork
duration distributions. Section 2 of this report contains a
detailed description of the procedure (including the improvements)
used to determine the subnetwork duration distributions. Sections 3
and 4 present detailed documentations of the improvements.

From a statistical viewpoint a subnetwork's duration is defined
easily enough as the maximum of the paths through the subnetwork.
In Figure 1 (p. 3), for example, there are really three paths;
namely,

\[ P_1 = A + C + E \]
\[ P_2 = A + F + D + E \]
\[ P_3 = B + D + E . \]  \hspace{1cm} (1.1)

The project duration is simply the maximum of \( P_1, P_2, \) and \( P_3 \). However,
the difficulty is that the paths are usually dependent since the paths
often have activities in common. For example, the paths \( P_1 \) and \( P_2 \)
have activities \( A \) and \( E \) in common. Section 5 contains a review of the
few known general results concerning the distribution of the maximum
of dependent random variables. Also in Section 5 is an indication of
how these general results could be used to modify the Subnetwork
Analysis procedure described in Section 2.
2. ANALYSIS OF A SUBNETWORK

2.1 Introduction

The objective of Subnetwork Analysis is to determine each subnetwork's duration distribution.

At the end of Step 2 each activity in the subnetwork has a specified duration distribution. This distribution is now approximated by a two-point discrete distribution. In particular, an activity, say A, is now conceptualized as having two possible duration times, say $l_A$ for a lower duration and $u_A$ for an upper duration. The probability that the activity duration is $l_A$ is assumed to be $P_A$, and correspondingly the probability that the activity duration is $u_A$ is assumed to be $Q_A = 1 - P_A$. The values of $l_A$, $u_A$, and $P_A$ are chosen so that the mean, variance, and third moment of the discrete distribution are the same as the mean, variance, and third moment of activity A's specified duration distribution.

Let $n$ be the number of activities in the subnetwork. Let $v = 1, 2, \ldots, 2^n$ index the $2^n$ possible configurations of activity durations when each activity is either at its upper duration or at its lower duration. Let

$$p_v = \text{probability of the } v\text{-th activity duration configuration}$$

$$= \prod_{i=1}^{n} [P_i (1 - \delta_{v,i}) + Q_i \delta_{v,i}]$$

(2.1)

where
$\delta_{v,i} = 1$ if the duration for the $i$-th activity is $u_i$

in the $v$-th activity duration configuration

= 0 if the duration for the $i$-th activity is $l_i$

in the $v$-th activity duration configuration. \hspace{1cm} (2.2)

Then the subnetwork duration distribution when each activity has its
two-point discrete distribution is

$$ F(t) = \sum_{v=1}^{2^n} P_v I_t(t_v) \hspace{1cm} (2.3) $$

where

$t_v =$ the subnetwork duration when the activity durations

are in the $v$-th configuration \hspace{1cm} (2.4)

and

$$ I_t(t_v) = 1 \text{ if } t_v \leq t, $$

$$ = 0 \text{ if } t_v > t. \hspace{1cm} (2.5) $$

The discrete distribution function $F$ is an approximation to the subnetwork's exact duration distribution.

The goal of Subnetwork Analysis is to determine $F$.

Since the number, $n$, of activities in the subnetwork may be fairly large, the complete enumeration of the $2^n$ discrete subnetwork durations may sometimes be impractical. When this happens, the discrete subnetwork duration distribution $F$ must be approximated. The approximation of $F$ will be based on the activities which are mostly likely to influence the subnetwork duration. The identification of these important activities and their interrelationships is discussed in the next subsection which is a review of the procedures originating in Sielken,

Each subnetwork is assumed to be an acyclic network with one source, one sink, and no cut vertices.

2.2 Formation of Clusters

The mean duration for activity A is

\[ m_A = P_A \cdot \mu_A + Q_A \cdot \mu_A' \]  \hspace{1cm} (2.6)

and the standard deviation of activity A's duration is

\[ s_A = [P_A \cdot \sigma_A^2 + Q_A \cdot \mu_A'^2 - m_A^2]^{1/2} \]  \hspace{1cm} (2.7)

When each activity duration takes on a fixed (nonrandom) value, the subnetwork's duration is the duration of the longest path through the subnetwork where the "length" of an activity is its duration.

For example, consider the subnetwork described in Table 1 and displayed in Figure 2. When each activity duration is its mean duration, then the subnetwork's duration is 32, corresponding to the path consisting of activities 2, 7, and 9.

**Definition 1**: A critical activity is an activity on the longest path when all the subnetwork's activity durations are set to their means.

Thus in the example the critical activities are 2, 7, and 9.

The search for the activities which are most likely to influence the subnetwork duration begins with the critical activities. Each critical activity initiates a separate set of activities called a
TABLE 1
Activity Durations for the Subnetwork in Figure 2

<table>
<thead>
<tr>
<th>Activity</th>
<th>$\xi_A$</th>
<th>$u_A$</th>
<th>$P_A$</th>
<th>$m_A$</th>
<th>$s_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>2.00</td>
<td>.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>8.00</td>
<td>10.50</td>
<td>.2</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>9.55</td>
<td>15.67</td>
<td>.6</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1.50</td>
<td>4.00</td>
<td>.8</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3.35</td>
<td>5.52</td>
<td>.7</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>4.00</td>
<td>6.00</td>
<td>.5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>8.73</td>
<td>16.90</td>
<td>.6</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>12.00</td>
<td>14.50</td>
<td>.2</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>5.00</td>
<td>15.00</td>
<td>.5</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 2
Subnetwork with activities labeled (activity number; mean duration).
"cluster". Initially there are several clusters. In the example the initial clusters are

\[ C_1 = \{2\}, \quad C_2 = \{7\}, \quad \text{and} \quad C_3 = \{9\}. \quad (2.8) \]

Some of the non-critical activities may influence the subnetwork's duration when not all of the activity durations are at their mean values.

**Definition 2**: An associate of a critical activity \( A \) is a non-critical activity which is on the longest path when all activity durations are set to their mean except for the critical activity \( A \) which has its duration reduced from \( m_A \) to \( \max(m_A - \lambda s_A, 0) \) where \( \lambda \) is a nonnegative parameter.

Thus the associates of a critical activity \( A \) are those activities whose effect on the subnetwork's duration are related to activity \( A \)'s duration. In the example, for \( \lambda = 1 \) the associates of the critical activities 2, 7, and 9 can be determined by considering Figures 3, 4, and 5 respectively. In Figure 3 the longest path is still the critical path 2, 7, and 9, so that activity 2 has no associates. In Figure 4 the longest path is 2, 5, 6, and 9, so that activities 5 and 6 are the associates of activity 7. In Figure 5, the longest path is 2, 5, and 8, so that activities 5 and 8 are associates of activity 9.

The associates of each critical activity are determined and added to the cluster containing that critical activity. Thus, in the example the clusters are expanded to

\[ C_1 = \{2\}, \quad C_2 = \{7, 5, 6\}, \quad \text{and} \quad C_3 = \{9, 5, 8\}. \quad (2.9) \]
Figure 3
Subnetwork for determining the associates of Activity 2 when $\lambda = 1$.

Figure 4
Subnetwork for determining the associates of Activity 7 when $\lambda = 1$. 
Figure 5
Subnetwork for determining the associates of Activity 9 when $\lambda = 1$.

Figure 6
Subnetwork for determining the eliminants of Activity 3 when $\theta = 2$. 
The idea underlying the clusters is that they should be sets of activities whose effects on the subnetwork's duration are interrelated. Thus, if two clusters contain any activities in common, the activities in these two clusters all have an interrelated effect on the subnetwork's duration, so the two clusters are combined into one cluster. In the example clusters $C_2$ and $C_3$ both contain activity 5, so they are combined. The resulting clusters are

$$C_1 = \{2\} \text{ and } C_2 = \{5, 6, 7, 8, 9\}.$$  \hspace{1cm} (2.10)

A non-critical activity may also influence the subnetwork's duration if its duration exceeds its mean.

**Definition 3**: An eliminant of a non-critical activity $A$ is a critical activity which is not on the longest path when all activity durations are set to their means except for activity $A$ which has its duration increased from $m_A$ to $m_A + \theta s_A$ where $\theta$ is a nonnegative parameter.

For instance, if $\theta = 2$, the eliminants of the non-critical activity 3 in the example can be determined from Figure 6. There the longest path is 1, 3, 6, and 9, so that the eliminants of activity 3 are the critical activities 2 and 7. In the example, when $\theta = 2$, none of the other non-critical activities (1, 4, 5, 6, and 8) have any eliminants. For a specified value of $\theta$ the eliminants of every non-critical activity are determined. If a non-critical activity $A$ has eliminants, then the effect of $A$'s eliminants on the subnetwork duration is related to $A$'s duration, so $A$ is added to every cluster containing at least one
of its eliminants. Thus in the example the clusters become

\[ C_1 = \{2, 3\} \quad \text{and} \quad C_2 = \{3, 5, 6, 7, 8, 9\} \]  
(2.11)

After the clusters have been expanded on the basis of eliminants, any two clusters containing common elements are combined. Therefore in the example, \( C_1 \) and \( C_2 \) are combined to form a single cluster

\[ C_1 = \{2, 3, 5, 6, 7, 8, 9\} \]  
(2.12)

In general, after the determination of associates and eliminants for specified values of \( \lambda \) and \( \theta \) and the subsequent combining of clusters, there may still be more than one cluster and some of the non-critical activities may not be in any cluster. Usually the larger the values of \( \lambda \) and \( \theta \) the greater the number of activities in the clusters and the smaller the number of clusters. The clusters that remain represent sets of activities such that the effects on the subnetwork's duration of the activity durations for the activities within a set are all interrelated. Activities in different clusters have roughly independent effects on the subnetwork's duration. Activities not in any cluster have essentially no effect on the subnetwork's duration.

The consideration of critical activities, associates, eliminants, and the formation of clusters of related activities is obviously only one way of identifying the activities which have an important effect on the subnetwork's duration and their interrelationships. However, this particular procedure does have the following desirable properties:
Property 1: If $\lambda_2 > \lambda_1$, then any activity which would be an associate of a critical activity $A$ when $\lambda = \lambda_1$ would also be an associate of $A$ when $\lambda = \lambda_2$.

Property 2: If $\theta_2 > \theta_1$, then any critical activity which would be an eliminant of a non-critical activity $A$ when $\theta = \theta_1$ would also be an eliminant of $A$ when $\theta = \theta_2$.

Property 3: For any fixed value of $\lambda$, the set of activities in the union of the clusters is monotonically nondecreasing as $\theta \rightarrow \infty$.

Property 4: The number of clusters is nonincreasing as $\theta \rightarrow \infty$.

Property 5: If $s_A > 0$ for a non-critical activity $A$, then there exists $\theta_A < \infty$ such that $A$ will have some eliminants for any $\theta \geq \theta_A$.

Property 6: If $s_A > 0$ for every non-critical activity $A$ and

$$\theta^* = \max\{\theta_A; A \text{ non-critical}\},$$

then for $\theta \geq \theta^*$ all activities will be in one cluster.

Most of these properties are fairly straightforward; however, Property 6 requires some special justification. This justification is based on the following definition and theorem which is proven in Sielken, Ringer, Hartley, and Arseven (1974).

**Definition 4:** In any acyclic network a **bridge** over any two consecutive arcs $A_1$ and $A_2$ is any arc $A_3$ such that all paths from the source to the sink passing through $A_3$ do not pass through either $A_1$ or $A_2$. 
Theorem 1: In any acyclic network with no cut vertices there is at least one bridge for any pair of consecutive arcs.

Property 5 implies that all activities will belong to some cluster if \( \theta \geq \theta^* \). Now consider any two consecutive activities \( A_1 \) and \( A_2 \) on the critical path. Theorem 1 implies that there is a bridge over \( A_1 \) and \( A_2 \), say \( A_3 \). Since the critical path passes through \( A_1 \) and \( A_2 \), \( A_3 \) cannot be on the critical path. Therefore, if \( \theta \geq \theta^* \geq \theta_{A_3} \), \( A_1 \) and \( A_2 \) will be eliminants of \( A_3 \) and hence will be in the same cluster as \( A_3 \). Thus, since each cluster contains at least one original critical activity and any two consecutive critical path activities belong to the same cluster when \( \theta \geq \theta^* \), there is only one cluster when \( \theta \geq \theta^* \) and Property 6 is established.

2.3 Bounding the Discrete Subnetwork Duration Distribution \( F \)

2.3.1 Upper Bounds on \( F \)

Suppose that the cluster formation procedure described in subsection 2.2 has been carried out on a subnetwork for some specified values of \( \theta \) and \( \lambda \) and yielded \( K \) clusters. For each cluster \( C \) so determined, let \( n_c \) be the number of activities in the cluster and let \( v = 1, \ldots, 2^{n_c} \) index the \( 2^{n_c} \) configurations of activity durations corresponding to

(a) the duration for each activity \( A \) not in \( C \) being equal to its lower point \( \lambda_A \), and

(b) the durations for the activities in \( C \) being at each of the \( 2^{n_c} \) possible combinations of their upper and lower points.

Then define
\[ F^+(C; t) = \sum_{v=1}^{n_c} p_v I_t(t_v) \]  

where \( p_v \), \( t_v \), and \( I_t(t_v) \) are defined in (2.1), (2.4), and (2.5) respectively. The distribution function \( F^+(C; t) \) is an upper bound on \( F \). This can be shown by considering the following:

**Theorem 2:** For any cluster \( C \), any \( t \), and any activity \( A \) not in \( C \),

\[ F^+(C \cup \{A\}; t) \leq F^+(C; t) \]  

(For the proof of this theorem, see Sielken, Hartley, and Spoerl (1975).) A straightforward application of Theorem 2 yields

**Theorem 3:** For any two clusters \( C_1 \) and \( C_2 \) and any \( t \),

\[ F^+(C_1 \cup C_2; t) \leq \min\{F^+(C_1; t), F^+(C_2; t)\} \]  

If \( C^* \) represents the set (cluster) of all activities in the subnetwork, then

\[ F(t) = F^+(C^*; t) \]  

(2.14)

Since \( C \) is a subset of \( C^* \), either Theorem 2 or Theorem 3 implies

\[ F^+(C; t) \geq F(t) \]  

(2.15)

for any cluster \( C \).

Theorems 2 and 3 can also be used to define some tighter upper bounds on the subnetwork's duration distribution than \( F^+(C; t) \). Historically, two different improved bounds have been employed, and both have been incorporated into the current subnetwork analysis
procedure. They are

\[ F_1^+(t; \theta, \lambda) = \prod_{i=1}^{K} \left( U C_i ; t \right) \]  \hspace{1cm} (2.16)

and

\[ F_2^+(t; \theta, \lambda) = \min_{1 \leq i \leq K} F^+(C_i; t) . \]  \hspace{1cm} (2.17)

Let \( F^+(t; \theta, \lambda) \) denote either \( F_1^+(t; \theta, \lambda) \) or \( F_2^+(t; \theta, \lambda) \). Then, since Property 2 of the cluster formation procedure implies that as \( \theta \) increases the clusters expand or are combined, Theorems 2 and 3 imply that \( F^+(t; \theta, \lambda) \) is a nonincreasing function of \( \theta \) for every \( t \) and any \( \lambda \). Property 6 and (2.14) imply that for \( \theta > \theta^* \)

\[ F^+(t; \theta, \lambda) = F(t) \]  \hspace{1cm} (2.18)

for every \( t \) and any \( \lambda \). Also (2.14) along with the definitions (2.16) and (2.17) imply

\[ F^+(t; \theta, \lambda) \geq F(t) \]  \hspace{1cm} (2.19)

for all \( t, \theta, \) and \( \lambda \). These results are summarized in Theorem 4.

**Theorem 4:** (a) \( F^+(t; \theta, \lambda) \) is a nonincreasing function of \( \theta \)

for every \( t \) and any \( \lambda \);

(b) there exists a finite value \( \theta^* \) such that \( \theta > \theta^* \)

implies \( F^+(t; \theta, \lambda) = F(t) \) for every \( t \) and \( \lambda \);

and

(c) for any \( \theta, \lambda, \) and \( t \)

\[ F^+(t; \theta, \lambda) \geq F(t) . \]
2.3.2 Lower Bounds on $F$

Let $n_c$ denote the number of activities in cluster C, and let $v = 1, \ldots, 2^n_c$ index the $2^n_c$ configuration of activity durations corresponding to

(a) the duration for each activity $A$ not in the cluster being equal to its upper point $u_A$, and

(b) the durations for activities in the cluster being at each of the $2^n_c$ possible combinations of the upper and lower points.

Then define

$$F^-(C; t) = \sum_{v=1}^{2^n_c} p_v I_t(t_v)$$

where $p_v$, $t_v$, and $I_t(t_v)$ are as previously defined. Also define

$$F^-(t; \theta, \lambda) = \max_{1 \leq i \leq K} F^-(C_i; t)$$

and

$$F^-(t; \theta, \lambda) = \max_{1 \leq i \leq K} F^-(C_i; t).$$

Using an argument completely analogous to that used to prove Theorem 4, Sielken, Hartley, and Spoeri (1975) also proved

**Theorem 5:** (a) $F^-(t; \theta, \lambda)$ is a nondecreasing function of $\theta$ for any fixed value of $\lambda$;

(b) there exists a finite value $\theta^*$ such that $\theta \geq \theta^*$ implies

$$F^-(t; \theta, \lambda) = F(t)$$

for every $t$ and any $\lambda$; and
(c) for any $\theta$, $\lambda$, and $t$

$$F^-(t; \theta, \lambda) \leq F(t).$$

(Again, $F^-(t; \theta, \lambda)$ is a generic term used to denote either $F^-_1(t; \theta, \lambda)$ or $F^-_2(t; \theta, \lambda)$.) Thus, $F^-(t; \theta, \lambda)$ is a valid lower bound on $F$.

2.3.3 The Tightness of the Bounds on $F$

That the $F^-_1$-bounds are tighter than the $F^-_2$-bounds can be seen as follows. The evaluation of $F^-_2(t; \theta, \lambda)$ involves the determination of $F^-_i(C_i; t)$ for each $i$ whereas $F^-_1(t; \theta, \lambda) = F^-(_K \cup C_i; t)$. Let $L_i$ be the length of the longest path when

1) the activities in $C_i$ are at a particular configuration of their upper and lower durations and
2) all activities not in $C_i$ have their upper durations.

Let $L_U$ be the length of the longest path when

1) the configuration of upper and lower durations for the activities in $C_i$ is the same as in the determination of $L_i$,
2) the activities in $\cup_{j=1}^K C_j - C_i$ are at any combination of their upper and lower durations, and
3) all activities not in $\cup_{j=1}^K C_j$ have their upper durations.

Then $L_i \geq L_U$ since every activity duration in the determination of $L_i$ is greater than or equal to its corresponding duration in the determination of $L_U$. Since $L_i \geq L_U$ for any configuration of upper and lower durations for the activities in $C_i$,

$$F^-(_K \cup C_j; t) \geq F^-(_K C_i; t)$$

and

$$F^-(_K C_j; t) \geq F^-(_K C_i; t)$$  \hspace{1cm} (2.23)
\[ F_1^-(t; \theta, \lambda) = F_2^-(U C_j; t) \geq \max_{1 \leq i \leq K} F_2^-(C_i; t) = F_2^-(t; \theta, \lambda). \quad (2.24) \]

A similar argument can be used to show
\[ F_1^+(t; \theta, \lambda) = F_2^+(U C_j; t) \leq \min_{1 \leq i \leq K} F_2^+(C_i; t) = F_2^+(t; \theta, \lambda). \quad (2.25) \]

The extent of the differences between the two upper bounds and two lower bounds depends heavily on the structure of the particular sub-network being analyzed and is a topic that should be considered in future empirical studies.

2.4 Using Sampling to Estimate the Upper and Lower Bounds on F

The only instance in which upper and lower bounds on F are computed rather than F itself is when it is computationally impractical to determine the longest path for each of the \(2^n\) activity duration configurations.

For given \(\theta\) and \(\lambda\), the evaluation of \(F_1^+(t; \theta, \lambda)\) only requires the determination of the longest path for each of \(2^{n_U}\) activity configurations where \(n_U\) is the number of activities in the union of the \(K\) clusters \(C_o = U C_j; i.e.,\)
\[ n_U = \sum_{j=1}^{K} n_j. \quad (2.26) \]

The evaluation of \(F_1^-(t; \theta, \lambda)\) also entails only \(2^{n_U}\) longest path determinations. Likewise, the evaluation of \(F_2^+(t; \theta, \lambda)\) or \(F_2^-(t; \theta, \lambda)\) only requires the determination of the longest path for each of
\[ n_S = \sum_{i=1}^{n_U} 2^{n_i} \quad (2.27) \]
activity configurations. Since \(2^n\) is always greater than or equal to
$n_s, F^+_2(t; \theta, \lambda)$ and $F^-_2(t; \theta, \lambda)$ are the most economical bounds to compute in terms of the number of longest path determinations required. However, for any given $\theta$ and $\lambda$, $F^+_1(t; \theta, \lambda)$ and $F^-_1(t; \theta, \lambda)$ are tighter bounds than $F^+_2(t; \theta, \lambda)$ and $F^-_2(t; \theta, \lambda)$, respectively. Thus, in making the choice of which one of the two sets of bounds to compute, there is a trade-off between the accuracy of the bounds and the effort required to compute them.

Since the cluster formation procedure is such that the clusters expand or are pooled as $\theta$ increases, it may happen that for particular $\theta$ and $\lambda$, $2^{n_U}$ and $2^{n_i}$ for some $i$ are both quite large even though $\theta$ is only moderately large. In this case it again becomes impractical to examine all the required activity configurations involved in determining either the $F_1$-bounds or the $F_2$-bounds. Consequently, if for the specified values of $\theta$ and $\lambda$, $2^{n_U}$ (or $2^{n_i}$ for some $i$, as the case may be) is excessively large, Subnetwork Analysis will compute estimates of the corresponding upper and lower bounds based on only a sample of the total number of possible configurations. The actual estimators used in this situation are described and developed in Section 3.

2.5 Estimating $F$ by Extrapolating Between the Upper and Lower Bounds on $F$

Theorems 4 and 5 of subsection 2.3 imply that if $\theta_{i+1} > \theta_i$ and $\lambda_{i+1} > \lambda_i$ for all $i = 1, \ldots, I$ then

$$F^+(t; \theta_1, \lambda_1) \geq F^+(t; \theta_2, \lambda_2) \geq \ldots \geq F^+(t; \theta_I, \lambda_I) \geq F(t) \geq \ldots \geq F^-(t; \theta_{I-1}, \lambda_{I-1}) \geq \ldots \geq \geq F^-(t; \theta_1, \lambda_1)$$

(2.28)
for all t. Thus, if \( F^+(t; \theta, \lambda) \) and \( F^-(t; \theta, \lambda) \) have been calculated for I pairs \( (\theta_i, \lambda_i) \), then \( F(t) \) may be estimated by extrapolating between \( F^+(t; \theta_i, \lambda_i) \) and \( F^-(t; \theta_i, \lambda_i) \). As currently written, Subnetwork Analysis calculates upper and lower bounds on the subnetwork's approximate duration distribution for a sequence of three \((\theta, \lambda)\) pairs, \((\theta, \lambda) = (1, 1), (2, 2), (3, 2)\). An extrapolation procedure is then used to obtain an estimate of \( F \). The procedure that has been developed for this purpose is documented in Section 4 of this report.

2.6 A Summary of the Subnetwork Analysis Procedure

The following is a step-by-step description of the subnetwork analysis procedure in summary form. Recall that the objective of Subnetwork Analysis is to determine an "approximation", say \( \hat{F} \), to the subnetwork's duration distribution.

(a) If \( n = 1 \), let \( \hat{F} \) be the actual activity duration distribution for the one activity comprising the subnetwork, and stop.
Otherwise, go to Step b.

(b) Identify the two-point discrete distribution \((\xi_A, u_A, p_A, q_A)\) for every activity A in the subnetwork.

(c) Ascertain the user's choice of

1. \( NMAX \), the maximum value of \( m \) for which all \( 2^m \) activity duration configurations are to be explicitly considered,
2. the \((\theta, \lambda)\) pairs to be considered if not the standard pairs \((1, 1), (2, 2), \) and \((3, 2)\),
(3) whether the bounds on F are to be \((F_1^-, F_1^+)\) or \((F_2^-, F_2^+)\) if \(n > N_{\text{MAX}}\), and

(4) SAMSIZ, the sample size to be taken if, in the determination of bounds on F for some \((\theta, \lambda)\) pair, the number of activity configurations in the cluster being considered exceeds \(2^{N_{\text{MAX}}}\).

(d) If the number of activities in the subnetwork doesn't exceed \(N_{\text{MAX}}\), compute the subnetwork's discrete duration distribution, \(F\), explicitly, let \(\hat{F} = F\), and stop. Otherwise, go to Step e.

(e) Do Steps f - i for every \((\theta, \lambda)\) pair. Then go to Step j.

(f) Form the clusters corresponding to \((\theta, \lambda)\). If the bounds are to be \((F_1^-, F_1^+)\), go to Step g. If the bounds are to be \((F_2^-, F_2^+)\), go to Step h.

(g) Form the union of the clusters and determine \(n_{\cup}\). If \(n_{\cup} \leq N_{\text{MAX}}\), evaluate the bounds \((F_1^-, F_1^+)\) on the basis of all \(2^{n_{\cup}}\) activity duration configurations. If \(n_{\cup} > N_{\text{MAX}}\), take a sample of size SAMSIZ from the \(2^{n_{\cup}}\) activity duration configurations and form both \(F_1^-\) and \(F_1^+\) on the basis of this single sample. Go to Step e.

(h) Do the following for each cluster, \(C_1\). Let \(n_1\) denote the number of activities in the cluster. If \(n_1 \leq N_{\text{MAX}}\), evaluate \(F^-(C_1; t)\) and \(F^+(C_1; t)\) on the basis of all \(2^{n_1}\) activity duration configurations. If \(n_1 > N_{\text{MAX}}\), take a sample of size SAMSIZ from the \(2^{n_1}\) activity duration configurations and form both \(F^-(C_1; t)\) and \(F^+(C_1; t)\) on the basis of this single sample.
(i) Form $F^-_2$ and $F^+_2$ from the $F^-_{C_i(t)}$'s and $F^+_{C_i(t)}$'s respectively. Go to Step e.

(j) Form $\hat{F}$ by extrapolating the $(F^-, F^+)$ bounds determined for the $(\theta, \lambda)$ pairs. Stop.

This process is repeated for every subnetwork in the simplified project network.
3. SAMPLE-BASED ESTIMATORS FOR A DISCRETE
DISTRIBUTION FUNCTION

3.1 Introduction

In the subnetwork analysis procedure the calculation of \( F^-(C; t) \) or \( F^+(C; t) \), say \( F(C; t) \), for a cluster \( C \) comprised of \( n_c \) activities requires \( 2^n_c \) longest path determinations. If \( 2^n_c \) is too large from a practical standpoint, \( F(C; t) \) must be approximated on the basis of a sample of the possible \( 2^n_c \) activity duration configurations.

The estimation of \( F(C; t) \) involves two aspects

1) the identification of an acceptable method of sample selection, and

2) the determination of the form of the estimator.

Because of the practical difficulties (computer storage requirements, etc.) involved in implementing other sampling schemes, only simple random sampling (with replacement) and systematic sampling were considered in this research. Of these two, systematic sampling is the preferred technique. The reasons for this preference are presented in subsection 3.4.

Now \( F(C; t) \) is the distribution function of a discrete random variable \( X \) (the length of the subnetwork's longest path) which has a known number,

\[
M = 2^n_c
\]  

(3.1)

of possible values which are not necessarily distinct. Since the probability that a particular activity, \( A \), attains its lower duration is known to be \( P_A \) and the probability that it attains its upper
duration is known to be $Q_A$, the random variable $X$ is such that when an activity duration configuration is realized not only is the numerical value of $X$ observed but also the probability, $p$, of that activity duration configuration is available. This departure from the usual estimation situation has been exploited in the formation of the estimators for $F(C; t)$.

3.2 Some Proposed Estimators

The five estimators of $F(C; t)$ that were considered in this research are

$$G_1(t) = \frac{\sum_{i=1}^{m} I_t(x_i)}{m};$$  \hspace{1cm} (3.2)

$$G_2(t) = \frac{\sum_{i=1}^{m} p_i I_t(x_i)}{m};$$  \hspace{1cm} (3.3)

$$G_3(t) = \begin{cases} 0, & t < x_1 \\ \frac{\sum_{i=1}^{m} p_i I_t(x_i) + (t-x_i)p_{i+1}}{m} - \frac{(x_{j+1} - x_j)}{m}, & x_1 \leq t < x_m \\ 1, & t \geq x_m \end{cases};$$  \hspace{1cm} (3.4)
\begin{equation}
G_4(t) = \begin{cases} 
0, & t < x_1 \\
\frac{\sum_{i=1}^{m-1} p_i I_t(x_i) + (t-x_{j+1})(1-\sum_{i=1}^{m} p_i)}{(x_{j+1}-x_j)} , & x_1 \leq t < x_m \\
1, & t \geq x_m 
\end{cases}
\end{equation}

(3.5)

and

\begin{equation}
G_5(t) = \begin{cases} 
0, & t < x_1 \\
\frac{\sum_{i=1}^{m-1} p_i I_t(x_i) + (t-x_{j})(1-\sum_{i=1}^{m} p_i)}{(x_{j+1}-x_j)} , & x_1 \leq t < x_m \\
1, & t \geq x_m 
\end{cases}
\end{equation}

(3.6)

where in each case, the $x_i$'s represent an ordered sample of size $m$ from the population of subnetwork durations corresponding to the $M$ activity duration configurations, $p_i$ is the probability of the activity duration configuration corresponding to $x_i$, $j$ is the largest integer such that $x_j \leq t$ and $I_t(\cdot)$ is as defined in (2.5).

Even though sampling will normally only be employed if $M$ is very large, for illustration purposes consider $M = 20$ and that the following ordered sample of size $m = 5$ has been obtained:

\begin{align*}
x_1 &= 2, \quad x_2 = 2.5, \quad x_3 = 4.0, \quad x_4 = 4.5, \quad x_5 = 5 \\
p_1 &= .02, \quad p_2 = .03, \quad p_3 = .05, \quad p_4 = .07, \quad p_5 = .03.
\end{align*}

(3.7)

The $G_1(t), G_2(t), \ldots, G_5(t)$ for this sample data are displayed in Figures 7a - 7e, respectively.
Figure 7

$G_1(t)$, $G_2(t)$, $G_3(t)$, $G_4(t)$, and $G_5(t)$ for the sample data in (3.7).
The discussion in subsections 3.2.1 - 3.2.5 assumes that the m sample points have been selected without replacement.

3.2.1 The Empirical Distribution Function, $G_1(t)$

In many situations requiring the estimation of a distribution function, the empirical distribution function, $G_1(t)$, has very desirable properties. However, these properties are derived from the assumption that the sample values have been observed with roughly the same relative frequency as they occur in the population. For the situation under consideration, though, every $x_i$ occurs in the sample with relative frequency $m_i/M$ regardless of the true value of $p_i$. Consequently, $G_1(t)$ was of interest in this study only as a basis for comparison and generalization.

3.2.2 The Modified Empirical Distribution Function, $G_2(t)$

The major disadvantage of $G_1(t)$ is that it ignores the information contained in the $p_i$'s. Instead of assigning weight $1/m$ to each sampled point, $G_2(t)$ assigns to $x_i$ the weight

$$p_i / \sum_{i=1}^{m} p_i.$$  \hspace{1cm} (3.8)

3.2.3 The Continuous Estimator, $G_3(t)$

Although $F$ is discrete, the subnetwork's actual duration distribution is continuous. Therefore, it was anticipated that a continuous estimator might be in order. The estimator $G_3(t)$ is
continuous and equals \( G_2(t) \) at every sampled point but interpolates linearly between sample points.

3.2.4 The Mixed Estimator, \( G_4(t) \)

The estimator \( G_4(t) \) has the advantages of a continuous estimator between sampled values but preserves the discrete nature at the sampled values. Like \( G_3(t) \), \( G_4(t) \) also equals \( G_2(t) \) at every sampled point. However, at each sampled point \( x_i \), \( G_4(t) \) has a jump of size \( p_i \). For \( t \) between \( x_i \) and \( x_{i+1} \), \( G_4(t) \) interpolates linearly between \( G_2(x_i) \) and \( G_2(x_{i+1}) - p_{i+1} \).

3.2.5 The Mixed Estimator, \( G_5(t) \)

Like \( G_4(t) \), the estimator \( G_5(t) \) also assigns the discrete jump sizes to the sampled points. However, this estimator spreads the probability that is unaccounted for by the sampled values evenly over the range \( x_1 \) to \( x_m \).

3.3 Criteria for a Good Estimator

The quantity being estimated is a distribution function. Thus, a "good" estimator, say \( G(t) \), should have the properties of a distribution function; namely,

\[
(1) \ 0 < G(t) < 1, \ \ -\infty < t < \infty \quad \text{and}
\]

\[
(2) \ G(t_i) < G(t_j) \quad \text{for} \ t_i < t_j .
\]

In addition, it is desirable (but not requisite) that the estimator be "consistent" in the sense that the estimate of the distribution
function based on a sample containing all possible $x_i$ is the true distribution function of $X$.

All of the estimators exhibit the properties of a distribution function if the sampling is performed *without replacement*. However, the estimators $G_4(t)$ and $G_5(t)$ may not be between zero and one for all $t$ if sampling is with replacement.

Only estimators $G_2(t)$, $G_4(t)$, and $G_5(t)$ are always consistent. This follows immediately since if $m = M$ and the sample corresponds to all $M$ activity duration configurations, then

$$
\sum_{i=1}^{m} p_i = 1.
$$

Furthermore only $G_2(t)$ satisfies

$$
\lim_{m \to \infty} G(t) = F(t)
$$

if the sampling is done with replacement.

Since the mean and upper percentiles of the subnetwork's duration distribution are the quantities of primary interest, this work also sought an estimator whose estimates of a distribution function's mean, $\mu$, 90-th percentile, $P_{90}$, and 95th percentile, $P_{95}$, exhibit a high degree of precision. The simulation study described in subsection 3.5 was designed to determine the suitability of the proposed estimators in this regard.
3.4 Choosing Between Simple Random or Systematic Sampling

On the basis of computational ease alone, sampling with replacement is the preferred method. However, only \( G_2(t) \) satisfies (3.10) if the sampling is done with replacement. Also \( G_4(t) \) and \( G_5(t) \) are not necessarily distribution functions if the sampling is done with replacement. On the other hand, if systematic sampling is employed these difficulties do not arise. In addition, the simulation study described in subsection 3.5 indicates that estimates derived from systematic samples contain more information than estimates based on simple random sampling. This was anticipated since Cochran (1946) showed that for at least partially ordered populations the variance of \( \bar{x} = \frac{\sum x_i}{m} \) under systematic sampling is always less than it is under simple random sampling. Hence, the algorithm does its sampling via the systematic technique if possible.

Unfortunately, the way a computer represents integers in its memory makes systematic sampling impractical of \( M > M_o = 2^{a-1} - 1 \) where \( a \) is the number of binary bits in an integer word for the particular machine being used. For most modern IBM computers, \( a = 32 \), and thus \( M_o = 2,147,483,647 \). When \( M > M_o \), the algorithm uses random sampling with replacement.

3.4.1 An Ordering Scheme

Unfortunately, the relative magnitude of the subnetwork duration corresponding to a particular activity duration configuration cannot be determined in general unless all configurations are considered
explicitly. Hence, the following approximate ordering scheme was devised.

Let \( v = 1, \ldots, n_c \) index the \( 2^n_c \) configurations of activity durations corresponding to

(1) the duration of each activity not in the cluster being equal to either its upper duration or its lower duration depending on whether a lower bound or an upper bound, respectively, is being determined and

(2) the durations for the activities in the cluster being at each of the \( 2^n_c \) possible combinations of their upper and lower points.

The activity duration configuration whose corresponding subnetwork duration, say \( x_v \), is approximately the \( v \)-th smallest subnetwork duration can be determined from \( g_{n_c}(v) \) defined by

\[
g_{n_c}(v) = \text{the } k\text{-th smallest binary integer containing exactly } i \text{ "1" s}
\]

where \( i \) is the smallest integer such that

\[
v \leq \sum_{j=1}^{i} \binom{n_c}{j}
\]

and

\[
k = v - \sum_{j=0}^{i-1} \binom{n_c}{j}
\]

with

\[
\binom{n_c}{j} = \frac{n_c!}{(n_c - j)!j!}
\]
and
\[ \sum_{j=0}^{n_c - 1} j \equiv 0. \tag{3.14} \]

In particular the activity duration configuration corresponding to approximately the \( v \)-th smallest subnetwork duration has the \( j \)-th activity in the cluster equal to its lower duration if the \( j \)-th digit (counting from the least significant digit) in \( g_{n_c}^v \) is 0 and equal to its upper duration if the \( j \)-th digit in \( g_{n_c}^v \) is 1.

For \( v = 1 \), \( g_{n_c}^1 = 0_2 \) (base 2), so under the approximate ordering \( x_1 \) equals the subnetwork's duration when every activity has its lower duration which in fact is the smallest possible \( x \)-value. Similarly, \( x_{2^n_c} \) is the subnetwork's duration when every activity has its upper duration and is the largest possible \( x \)-value. For \( 1 \leq v_s < v_t \leq 2^n_c \), \( x_{v_s} \) is not necessarily less than or equal to \( x_{v_t} \). However, for \( v_s \) very much smaller than \( v_t \), the activity configuration corresponding to \( g_{n_c}^v \) has more activities at their upper duration than the one corresponding to \( g_{n_c}^{v_s} \). Hence, \( x_{v_t} \) is likely to be larger than \( x_{v_s} \).

For example, Table 3 gives the approximate ordering of the \( x \)-values for the small subnetwork pictured in Figure 8 and described in Table 2 for the case when all five of the subnetwork's activities are in the cluster C.
Figure 8

A small subnetwork.

TABLE 2

<table>
<thead>
<tr>
<th>Activity</th>
<th>( \ell )</th>
<th>( u )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>
TABLE 3

The Approximate Ordering of the x-values for the Subnetwork in Figure 8

<table>
<thead>
<tr>
<th>v</th>
<th>g_5(v)</th>
<th>Activity durations corresponding to g_5(v)</th>
<th>x_v</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>00000_2</td>
<td>4 5 3 8 5</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>00001_2</td>
<td>4 5 3 8 7</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>00010_2</td>
<td>4 5 3 10 5</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>00100_2</td>
<td>4 6 3 8 5</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>01000_2</td>
<td>8 5 3 8 5</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>10000_2</td>
<td>8 5 3 10 5</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>00011_2</td>
<td>4 5 3 10 7</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>00110_2</td>
<td>4 5 3 8 7</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>01001_2</td>
<td>4 6 3 8 7</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>01101_2</td>
<td>4 6 3 10 5</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>01100_2</td>
<td>4 6 3 8 5</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>10010_2</td>
<td>8 5 3 8 5</td>
<td>18</td>
</tr>
<tr>
<td>12</td>
<td>10100_2</td>
<td>8 5 3 10 5</td>
<td>18</td>
</tr>
<tr>
<td>13</td>
<td>10100_2</td>
<td>8 5 6 8 5</td>
<td>19</td>
</tr>
<tr>
<td>14</td>
<td>11000_2</td>
<td>8 6 3 8 5</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>11000_2</td>
<td>8 6 3 10 7</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td>00111_2</td>
<td>4 5 6 10 7</td>
<td>14</td>
</tr>
<tr>
<td>17</td>
<td>01011_2</td>
<td>4 6 3 10 7</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>01101_2</td>
<td>4 6 6 8 7</td>
<td>17</td>
</tr>
<tr>
<td>19</td>
<td>01110_2</td>
<td>4 6 6 10 5</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>10011_2</td>
<td>8 5 3 10 7</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>10101_2</td>
<td>8 5 6 8 7</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>10110_2</td>
<td>8 5 6 10 5</td>
<td>19</td>
</tr>
<tr>
<td>23</td>
<td>10110_2</td>
<td>8 5 6 10 7</td>
<td>19</td>
</tr>
<tr>
<td>24</td>
<td>11001_2</td>
<td>8 6 3 8 7</td>
<td>18</td>
</tr>
<tr>
<td>25</td>
<td>11010_2</td>
<td>8 6 3 10 5</td>
<td>18</td>
</tr>
<tr>
<td>26</td>
<td>11100_2</td>
<td>8 6 6 8 5</td>
<td>19</td>
</tr>
<tr>
<td>27</td>
<td>01111_2</td>
<td>4 6 6 10 7</td>
<td>17</td>
</tr>
<tr>
<td>28</td>
<td>10111_2</td>
<td>8 5 6 10 7</td>
<td>21</td>
</tr>
<tr>
<td>29</td>
<td>11011_2</td>
<td>8 6 3 10 7</td>
<td>18</td>
</tr>
<tr>
<td>30</td>
<td>11101_2</td>
<td>8 6 6 8 7</td>
<td>21</td>
</tr>
<tr>
<td>31</td>
<td>11110_2</td>
<td>8 6 6 10 5</td>
<td>19</td>
</tr>
<tr>
<td>32</td>
<td>11111_2</td>
<td>8 6 6 10 7</td>
<td>21</td>
</tr>
</tbody>
</table>
3.4.2 Implementing the Ordering Scheme

An efficient algorithm for finding the k-th smallest binary integer containing exactly i "1"s was developed. The algorithm is as follows:

1. Let
   \[ NP = \text{the number of binary digits whose values are as yet undetermined}, \]
   \[ NI = \text{the number of the NP remaining digits that are to be assigned the value "1"}, \]
   \[ J = \text{the location of the digit whose value is currently being determined as counted from the right}. \]

2. Set \( NP = n_c \), \( NI = i \), \( J = n_c \), \( B = \binom{n_c}{i} \), \( R = B - k \).

3. If \( NI < 1 \), assign the value "0" to all remaining digits and stop. Otherwise, set
   \[ B = B \times NI/NP \]
   \[ NI = NI - 1 \]
   \[ NP = NP - 1 \]
   \[ RR = R - B. \]
   If \( RR \leq 0 \), go to 4. Otherwise, go to 5.

4. Assign the J-th right-most digit the value "1". Set \( J = J - 1 \). Go to 3.

5. Assign the J-th right-most digit the value "0". Set
J = J - 1

R = RR

B = B (NP - NI)/NP

NP = NP - 1

RR = R - B.

If RR £ 0, go to 4. Otherwise, do 5 again.

3.5 The Simulation Study

Since a subnetwork's duration distribution is the distribution of the maximum path length, most subnetwork duration distributions are skewed left. Nevertheless, the behavior of the proposed estimators \(G_1(t), \ldots, G_6(t)\) was determined for samples drawn from populations exhibiting a variety of distributional shapes. Since the beta distribution with probability density function (p.d.f.)

\[
B_{\alpha,\beta}(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1 - t)^{\beta-1}, \quad \alpha, \beta > 0, \quad 0 \leq t \leq 1
\]

is a finite range distribution which can assume a wide variety of shapes, the subnetwork duration distributions corresponded to

\(B_{23,2}(t), B_{8,2}(t), B_{5,5}(t), B_{2,8}(t), \text{ and } B_{2,23}(t)\)

in the simulation study. (These represent the shapes highly skewed left, skewed left, symmetrical, skewed right, and highly skewed right, respectively, as indicated in Figure 9.)
Figure 9

Duration distributions used in the simulation study.
The discrete subnetwork duration distribution, say $F_{\alpha,\beta}$, corresponding to $B_{\alpha,\beta}$ was constructed by randomly selecting

$$t_1 \leq t_2 \leq \ldots \leq t_999 \leq t_{1000} = 1.0,$$

and defining

$$F_{\alpha,\beta}(t) = 0, \quad t < t_1$$

$$= \int_{0}^{t_1} B_{\alpha,\beta}(x)dx, \quad t_1 \leq t < t_{i+1}$$

$$= 1, \quad t \geq 1.0. \quad (3.16)$$

Therefore in the simulation study $M = 1000$. Also, since $\int_{0}^{1} B_{\alpha,\beta}(x)dx$ does not exist in closed form unless $t = 1$, it was evaluated at each $t_1 < 1$ using an approximation due to Peizer and Pratt (1963).

Samples of size $m$ were then taken from each of the $F_{\alpha,\beta}(t)$ and the estimators $G_1(t), \ldots, G_5(t)$ were calculated. In the subnetwork analysis procedure, sampling is only employed when practical considerations dictate the explicit consideration of only a relatively small proportion of the subnetwork's activity duration configurations. Hence, the only sample sizes considered by this work were $m = 10, 20, 50,$ and $100$ which represent sampling proportions of $1\%, 2\%, 5\%,$ and $10\%$, respectively.

The performance of the five proposed estimators with respect to estimation of the parameters $\mu, p_{95},$ and $p_{90}$ was evaluated for each of the $F_{\alpha,\beta}(t)$ on the basis of the following three criteria
(i) mean deviation,

$$MD(\gamma) = \frac{200}{200} \sum_{i=1}^{200} \frac{|\hat{\gamma}_i - \gamma|}{},$$

(3.17)

(ii) mean square error,

$$MSE(\gamma) = \frac{200}{200} \sum_{i=1}^{200} \frac{(\hat{\gamma}_i - \gamma)^2}{},$$

and

(3.18)

(iii) bias,

$$BIAS(\gamma) = \frac{200}{200} \sum_{i=1}^{200} \frac{\hat{\gamma}_i}{200} - \gamma$$

(3.19)

where in each case

\( \gamma = \) the parameter being estimated (\( \gamma = \mu, \ P_{95}, \) or \( P_{90} \));

\( \hat{\gamma}_i = \) an estimate of \( \gamma \) based on the \( i \)-th sample of size \( m \) drawn;

and

\( 200 = \) the number of samples of size \( m \) drawn.

### 3.5.1 A Comparison of \( G_1(t), \ldots, G_5(t) \) under Systematic Sampling

Tables 4-8 indicate the simulation results for the proposed estimators under systematic sampling. The following observations may be made:

(1) For every case considered, \( G_1(t) \) and \( G_5(t) \) yielded highly biased estimates of at least one of the 3 parameters \( \mu, \ P_{95}, \) and \( P_{90} \) while \( G_2(t), G_3(t), \) and \( G_4(t) \) are only moderately biased.
TABLE 4

Simulation Results for the Highly Skewed Left \( F_{23, 2}(t) \) Having \( \mu = 92122, P_{95} = 98763, P_{90} = 97805^* \)

<table>
<thead>
<tr>
<th>Percent Sampling</th>
<th>Estimator</th>
<th>( t )</th>
<th>MD(( \hat{\mu} ))</th>
<th>MSE(( \hat{\mu} ))</th>
<th>BIAS(( \hat{\mu} ))</th>
<th>( \hat{\mu}_{95} )</th>
<th>MSE(( \hat{\mu}_{95} ))</th>
<th>BIAS(( \hat{\mu}_{95} ))</th>
<th>( \hat{\mu}_{90} )</th>
<th>MSE(( \hat{\mu}_{90} ))</th>
<th>BIAS(( \hat{\mu}_{90} ))</th>
<th>( \hat{\mu}_{90} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>( G_1 )</td>
<td>50149</td>
<td>41972</td>
<td>17700</td>
<td>-41972</td>
<td>100000</td>
<td>1237</td>
<td>15</td>
<td>-1237</td>
<td>100000</td>
<td>2195</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>91275</td>
<td>2592</td>
<td>99</td>
<td>-846</td>
<td>97518</td>
<td>2586</td>
<td>126</td>
<td>-1246</td>
<td>95593</td>
<td>3538</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>85677</td>
<td>6454</td>
<td>508</td>
<td>-6445</td>
<td>95508</td>
<td>3471</td>
<td>201</td>
<td>-3255</td>
<td>93806</td>
<td>4241</td>
<td>269</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>85724</td>
<td>6408</td>
<td>503</td>
<td>-6398</td>
<td>95575</td>
<td>3406</td>
<td>196</td>
<td>-3188</td>
<td>93888</td>
<td>4166</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>50482</td>
<td>41640</td>
<td>17340</td>
<td>-41640</td>
<td>95200</td>
<td>3562</td>
<td>128</td>
<td>-3563</td>
<td>90523</td>
<td>7282</td>
<td>536</td>
</tr>
<tr>
<td>2%</td>
<td>( G_1 )</td>
<td>49986</td>
<td>42135</td>
<td>17774</td>
<td>-42135</td>
<td>100000</td>
<td>1237</td>
<td>15</td>
<td>-1237</td>
<td>97642</td>
<td>1343</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>91480</td>
<td>1971</td>
<td>60</td>
<td>-641</td>
<td>97424</td>
<td>1801</td>
<td>56</td>
<td>-1339</td>
<td>96218</td>
<td>2189</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>88844</td>
<td>3320</td>
<td>158</td>
<td>-3278</td>
<td>96381</td>
<td>2437</td>
<td>85</td>
<td>-2382</td>
<td>94870</td>
<td>2987</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>88894</td>
<td>3286</td>
<td>155</td>
<td>-3227</td>
<td>96470</td>
<td>2350</td>
<td>81</td>
<td>-2293</td>
<td>94958</td>
<td>2903</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>50928</td>
<td>41193</td>
<td>16972</td>
<td>-41193</td>
<td>95430</td>
<td>3334</td>
<td>114</td>
<td>-3334</td>
<td>91132</td>
<td>6674</td>
<td>458</td>
</tr>
<tr>
<td>5%</td>
<td>( G_1 )</td>
<td>49884</td>
<td>42238</td>
<td>17843</td>
<td>-42238</td>
<td>97469</td>
<td>1294</td>
<td>21</td>
<td>-1294</td>
<td>92400</td>
<td>5405</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>91902</td>
<td>1101</td>
<td>19</td>
<td>-220</td>
<td>98225</td>
<td>854</td>
<td>14</td>
<td>-539</td>
<td>97481</td>
<td>1182</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>90844</td>
<td>1524</td>
<td>33</td>
<td>-1277</td>
<td>97466</td>
<td>1335</td>
<td>29</td>
<td>-1297</td>
<td>96660</td>
<td>1375</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>90895</td>
<td>1496</td>
<td>33</td>
<td>-1227</td>
<td>97531</td>
<td>1283</td>
<td>28</td>
<td>-1232</td>
<td>96735</td>
<td>1322</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>52108</td>
<td>40014</td>
<td>16019</td>
<td>-40014</td>
<td>95897</td>
<td>2867</td>
<td>85</td>
<td>-2867</td>
<td>92388</td>
<td>5418</td>
<td>310</td>
</tr>
<tr>
<td>10%</td>
<td>( G_1 )</td>
<td>49851</td>
<td>42271</td>
<td>17869</td>
<td>-42271</td>
<td>97010</td>
<td>1754</td>
<td>33</td>
<td>-1754</td>
<td>90963</td>
<td>6843</td>
<td>469</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>91922</td>
<td>904</td>
<td>13</td>
<td>-200</td>
<td>98598</td>
<td>561</td>
<td>4</td>
<td>-165</td>
<td>97616</td>
<td>764</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>91375</td>
<td>1048</td>
<td>17</td>
<td>-746</td>
<td>98107</td>
<td>695</td>
<td>8</td>
<td>-656</td>
<td>97226</td>
<td>882</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>91429</td>
<td>1030</td>
<td>17</td>
<td>-692</td>
<td>98153</td>
<td>657</td>
<td>7</td>
<td>-609</td>
<td>97307</td>
<td>822</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>54153</td>
<td>37968</td>
<td>14442</td>
<td>-37968</td>
<td>96490</td>
<td>2274</td>
<td>55</td>
<td>-2274</td>
<td>93792</td>
<td>4012</td>
<td>179</td>
</tr>
</tbody>
</table>

* All entries have been multiplied by \( 10^5 \).
TABLE 5

Simulation Results for the Skewed Left $F_{8,2}(t)$ Having $\mu = 80127$, $P_{95} = 96008$, $P_{90} = 94100^*$

<table>
<thead>
<tr>
<th>Percent</th>
<th>Estimator</th>
<th>n</th>
<th>$n(t^*)$</th>
<th>MSE($n(t^*)$)</th>
<th>BLAS($\hat{P}$)</th>
<th>$\hat{P}$</th>
<th>MSE($\hat{P}$)</th>
<th>BLAS($\hat{P}$)</th>
<th>$\hat{P}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>G$_1$</td>
<td>50149</td>
<td>29978</td>
<td>9071</td>
<td>-29978</td>
<td>100000</td>
<td>3992</td>
<td>159</td>
<td>-1992</td>
</tr>
<tr>
<td></td>
<td>G$_2$</td>
<td>79420</td>
<td>3674</td>
<td>203</td>
<td>-707</td>
<td>92007</td>
<td>5005</td>
<td>384</td>
<td>-4001</td>
</tr>
<tr>
<td></td>
<td>G$_3$</td>
<td>73134</td>
<td>7247</td>
<td>711</td>
<td>-6993</td>
<td>89697</td>
<td>6356</td>
<td>563</td>
<td>-6310</td>
</tr>
<tr>
<td></td>
<td>G$_4$</td>
<td>73184</td>
<td>7206</td>
<td>704</td>
<td>-6944</td>
<td>89770</td>
<td>6292</td>
<td>554</td>
<td>-6238</td>
</tr>
<tr>
<td></td>
<td>G$_5$</td>
<td>50374</td>
<td>29754</td>
<td>8853</td>
<td>-29754</td>
<td>95034</td>
<td>981</td>
<td>10</td>
<td>-973</td>
</tr>
<tr>
<td>2%</td>
<td>G$_1$</td>
<td>49986</td>
<td>30141</td>
<td>9105</td>
<td>-30141</td>
<td>100000</td>
<td>3992</td>
<td>159</td>
<td>-3992</td>
</tr>
<tr>
<td></td>
<td>G$_2$</td>
<td>79937</td>
<td>2305</td>
<td>101</td>
<td>-191</td>
<td>93913</td>
<td>3182</td>
<td>179</td>
<td>-2095</td>
</tr>
<tr>
<td></td>
<td>G$_3$</td>
<td>77170</td>
<td>3617</td>
<td>190</td>
<td>-2957</td>
<td>92033</td>
<td>4031</td>
<td>255</td>
<td>-3975</td>
</tr>
<tr>
<td></td>
<td>G$_4$</td>
<td>77220</td>
<td>3584</td>
<td>187</td>
<td>-2907</td>
<td>92097</td>
<td>3984</td>
<td>251</td>
<td>-3911</td>
</tr>
<tr>
<td></td>
<td>G$_5$</td>
<td>50679</td>
<td>29449</td>
<td>8672</td>
<td>-29449</td>
<td>95066</td>
<td>959</td>
<td>9</td>
<td>-941</td>
</tr>
<tr>
<td>5%</td>
<td>G$_1$</td>
<td>49884</td>
<td>30243</td>
<td>9150</td>
<td>-30243</td>
<td>97469</td>
<td>1467</td>
<td>26</td>
<td>1461</td>
</tr>
<tr>
<td></td>
<td>G$_2$</td>
<td>79795</td>
<td>1626</td>
<td>43</td>
<td>-332</td>
<td>95200</td>
<td>1762</td>
<td>47</td>
<td>-808</td>
</tr>
<tr>
<td></td>
<td>G$_4$</td>
<td>78777</td>
<td>1974</td>
<td>61</td>
<td>-1350</td>
<td>94267</td>
<td>1971</td>
<td>61</td>
<td>-1741</td>
</tr>
<tr>
<td></td>
<td>G$_5$</td>
<td>51543</td>
<td>28584</td>
<td>8172</td>
<td>-29594</td>
<td>05134</td>
<td>900</td>
<td>9</td>
<td>-874</td>
</tr>
<tr>
<td>10%</td>
<td>G$_1$</td>
<td>49851</td>
<td>30277</td>
<td>9168</td>
<td>-30277</td>
<td>97010</td>
<td>1002</td>
<td>12</td>
<td>1002</td>
</tr>
<tr>
<td></td>
<td>G$_2$</td>
<td>79708</td>
<td>1987</td>
<td>53</td>
<td>-420</td>
<td>95166</td>
<td>1140</td>
<td>21</td>
<td>-842</td>
</tr>
<tr>
<td></td>
<td>G$_3$</td>
<td>79167</td>
<td>2157</td>
<td>62</td>
<td>-960</td>
<td>94745</td>
<td>1382</td>
<td>29</td>
<td>-1263</td>
</tr>
<tr>
<td></td>
<td>G$_4$</td>
<td>79220</td>
<td>2143</td>
<td>61</td>
<td>-907</td>
<td>94822</td>
<td>1344</td>
<td>28</td>
<td>-1186</td>
</tr>
<tr>
<td></td>
<td>G$_5$</td>
<td>53046</td>
<td>27081</td>
<td>7340</td>
<td>-27081</td>
<td>95224</td>
<td>820</td>
<td>8</td>
<td>-783</td>
</tr>
</tbody>
</table>

* All entries have been multiplied by $10^5$. 
<table>
<thead>
<tr>
<th>Percent Sampling</th>
<th>Estimator</th>
<th>m(t)</th>
<th>MD</th>
<th>MSE</th>
<th>BLAS(p_0</th>
<th>( \hat{\theta}_9</th>
<th>MSE(\hat{\theta}_9</th>
<th>BLAS(\hat{\theta}_9</th>
<th>( \hat{\theta}_{90}</th>
<th>MSE(\hat{\theta}_{90}</th>
<th>BLAS(\hat{\theta}_{90}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>( G_1 )</td>
<td>50149</td>
<td>2502</td>
<td>84</td>
<td>-42</td>
<td>100000</td>
<td>25108</td>
<td>6304</td>
<td>25108</td>
<td>100000</td>
<td>30082</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>49798</td>
<td>3664</td>
<td>205</td>
<td>-309</td>
<td>71385</td>
<td>5230</td>
<td>645</td>
<td>-3507</td>
<td>67119</td>
<td>7398</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>43308</td>
<td>6985</td>
<td>658</td>
<td>-6800</td>
<td>67175</td>
<td>8260</td>
<td>1023</td>
<td>-7718</td>
<td>62090</td>
<td>8498</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>43360</td>
<td>6939</td>
<td>651</td>
<td>-6748</td>
<td>67251</td>
<td>8192</td>
<td>1011</td>
<td>-7642</td>
<td>62160</td>
<td>8437</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>50133</td>
<td>34</td>
<td>0</td>
<td>25</td>
<td>94973</td>
<td>20080</td>
<td>4032</td>
<td>20080</td>
<td>89947</td>
<td>20029</td>
</tr>
<tr>
<td>2%</td>
<td>( G_1 )</td>
<td>49986</td>
<td>1224</td>
<td>21</td>
<td>-121</td>
<td>100000</td>
<td>25108</td>
<td>6304</td>
<td>25108</td>
<td>97642</td>
<td>27724</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>49808</td>
<td>3334</td>
<td>164</td>
<td>-299</td>
<td>73208</td>
<td>4345</td>
<td>259</td>
<td>-1684</td>
<td>68825</td>
<td>4068</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>46864</td>
<td>4075</td>
<td>264</td>
<td>-3243</td>
<td>70871</td>
<td>4926</td>
<td>357</td>
<td>-4022</td>
<td>66021</td>
<td>4974</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>46918</td>
<td>4050</td>
<td>261</td>
<td>-3189</td>
<td>70938</td>
<td>4876</td>
<td>351</td>
<td>-3955</td>
<td>66085</td>
<td>4934</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>50130</td>
<td>34</td>
<td>0</td>
<td>22</td>
<td>94920</td>
<td>20028</td>
<td>4011</td>
<td>20027</td>
<td>89850</td>
<td>19932</td>
</tr>
<tr>
<td>5%</td>
<td>( G_1 )</td>
<td>49884</td>
<td>514</td>
<td>3</td>
<td>-223</td>
<td>97469</td>
<td>22576</td>
<td>5101</td>
<td>22576</td>
<td>92400</td>
<td>22482</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>49743</td>
<td>2328</td>
<td>76</td>
<td>-364</td>
<td>74020</td>
<td>2527</td>
<td>97</td>
<td>-872</td>
<td>69763</td>
<td>2683</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>48610</td>
<td>2530</td>
<td>97</td>
<td>-1497</td>
<td>72953</td>
<td>2923</td>
<td>133</td>
<td>-1939</td>
<td>68610</td>
<td>2665</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>48665</td>
<td>2515</td>
<td>95</td>
<td>-1442</td>
<td>73019</td>
<td>2873</td>
<td>118</td>
<td>-1873</td>
<td>68673</td>
<td>2632</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>50122</td>
<td>117</td>
<td>0</td>
<td>14</td>
<td>94761</td>
<td>19869</td>
<td>3948</td>
<td>19869</td>
<td>89600</td>
<td>19682</td>
</tr>
<tr>
<td>10%</td>
<td>( G_1 )</td>
<td>49851</td>
<td>307</td>
<td>1</td>
<td>257</td>
<td>97010</td>
<td>22117</td>
<td>4894</td>
<td>22117</td>
<td>90963</td>
<td>21044</td>
</tr>
<tr>
<td></td>
<td>( G_2 )</td>
<td>50026</td>
<td>1251</td>
<td>21</td>
<td>-81</td>
<td>74652</td>
<td>1564</td>
<td>41</td>
<td>-240</td>
<td>69884</td>
<td>2006</td>
</tr>
<tr>
<td></td>
<td>( G_3 )</td>
<td>49446</td>
<td>1286</td>
<td>25</td>
<td>-661</td>
<td>74138</td>
<td>1782</td>
<td>48</td>
<td>-754</td>
<td>69281</td>
<td>1999</td>
</tr>
<tr>
<td></td>
<td>( G_4 )</td>
<td>49503</td>
<td>1278</td>
<td>24</td>
<td>-603</td>
<td>74194</td>
<td>1760</td>
<td>47</td>
<td>-699</td>
<td>69343</td>
<td>1975</td>
</tr>
<tr>
<td></td>
<td>( G_5 )</td>
<td>50136</td>
<td>123</td>
<td>0</td>
<td>28</td>
<td>94570</td>
<td>19677</td>
<td>3872</td>
<td>19677</td>
<td>89288</td>
<td>19370</td>
</tr>
</tbody>
</table>

* All entries have been multiplied by \( 10^5 \).
TABLE 7

Simulation Results for the Skewed Right $F_{2,8}(t)$ Having $\mu = 20067$, $P_{95} = 42929$, $P_{90} = 36912^*$

<table>
<thead>
<tr>
<th>Percent</th>
<th>Sampling</th>
<th>Estimator</th>
<th>$n$</th>
<th>$\delta(n)$</th>
<th>Mean</th>
<th>MSE</th>
<th>BIAS</th>
<th>$\delta(95)$</th>
<th>MSE</th>
<th>BIAS(95)</th>
<th>$\delta(90)$</th>
<th>MSE</th>
<th>BIAS(90)</th>
<th>BLAS</th>
<th>$\delta(90)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td></td>
<td>G1</td>
<td>50149</td>
<td>30082</td>
<td>1103</td>
<td>3082</td>
<td>100000</td>
<td>57071</td>
<td>32571</td>
<td>57071</td>
<td>100000</td>
<td>63088</td>
<td>39801</td>
<td></td>
<td>63088</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G2</td>
<td>20092</td>
<td>3583</td>
<td>181</td>
<td>141</td>
<td>41767</td>
<td>5436</td>
<td>490</td>
<td>1162</td>
<td>35497</td>
<td>6699</td>
<td>718</td>
<td>-1414</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G3</td>
<td>15120</td>
<td>5244</td>
<td>389</td>
<td>-4946</td>
<td>37129</td>
<td>6688</td>
<td>671</td>
<td>-5800</td>
<td>30775</td>
<td>7338</td>
<td>871</td>
<td>-6136</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G4</td>
<td>15148</td>
<td>5220</td>
<td>385</td>
<td>-4919</td>
<td>37188</td>
<td>6640</td>
<td>663</td>
<td>-5741</td>
<td>30839</td>
<td>7285</td>
<td>861</td>
<td>-6073</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G5</td>
<td>49894</td>
<td>29827</td>
<td>8897</td>
<td>29827</td>
<td>94794</td>
<td>52044</td>
<td>27086</td>
<td>52044</td>
<td>89948</td>
<td>53036</td>
<td>28128</td>
<td>53036</td>
<td></td>
</tr>
<tr>
<td>2%</td>
<td></td>
<td>G1</td>
<td>49986</td>
<td>29919</td>
<td>8972</td>
<td>29919</td>
<td>100000</td>
<td>57071</td>
<td>32571</td>
<td>57071</td>
<td>97641</td>
<td>60730</td>
<td>36908</td>
<td>60730</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G2</td>
<td>19867</td>
<td>2283</td>
<td>84</td>
<td>-200</td>
<td>41938</td>
<td>3809</td>
<td>221</td>
<td>-991</td>
<td>36214</td>
<td>4003</td>
<td>268</td>
<td>-668</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G3</td>
<td>17288</td>
<td>3233</td>
<td>156</td>
<td>-2779</td>
<td>39630</td>
<td>4479</td>
<td>304</td>
<td>-3299</td>
<td>33979</td>
<td>4576</td>
<td>314</td>
<td>-2933</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G4</td>
<td>17332</td>
<td>3200</td>
<td>153</td>
<td>-2735</td>
<td>39683</td>
<td>4446</td>
<td>301</td>
<td>-3247</td>
<td>34041</td>
<td>4534</td>
<td>309</td>
<td>-2870</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G5</td>
<td>49562</td>
<td>29495</td>
<td>8700</td>
<td>29495</td>
<td>94918</td>
<td>51988</td>
<td>27028</td>
<td>51988</td>
<td>89841</td>
<td>52929</td>
<td>28015</td>
<td>52929</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td>G1</td>
<td>49884</td>
<td>29817</td>
<td>8894</td>
<td>29817</td>
<td>97469</td>
<td>54540</td>
<td>29750</td>
<td>54540</td>
<td>92400</td>
<td>55489</td>
<td>30793</td>
<td>55489</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G2</td>
<td>20034</td>
<td>1375</td>
<td>32</td>
<td>-33</td>
<td>42323</td>
<td>2114</td>
<td>72</td>
<td>-606</td>
<td>36311</td>
<td>2596</td>
<td>102</td>
<td>-600</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G3</td>
<td>19008</td>
<td>1793</td>
<td>42</td>
<td>-1059</td>
<td>41319</td>
<td>2390</td>
<td>87</td>
<td>-1610</td>
<td>35474</td>
<td>2739</td>
<td>113</td>
<td>-1437</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G4</td>
<td>19058</td>
<td>1764</td>
<td>41</td>
<td>-1008</td>
<td>41368</td>
<td>2371</td>
<td>84</td>
<td>-1561</td>
<td>35530</td>
<td>2707</td>
<td>111</td>
<td>-1382</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G5</td>
<td>48635</td>
<td>28568</td>
<td>8162</td>
<td>28568</td>
<td>94753</td>
<td>51823</td>
<td>26857</td>
<td>51823</td>
<td>89594</td>
<td>52682</td>
<td>27755</td>
<td>52682</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td>G1</td>
<td>49851</td>
<td>29784</td>
<td>8871</td>
<td>29784</td>
<td>97010</td>
<td>54080</td>
<td>29249</td>
<td>54080</td>
<td>90963</td>
<td>54051</td>
<td>29216</td>
<td>54051</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G2</td>
<td>20093</td>
<td>960</td>
<td>13</td>
<td>26</td>
<td>42990</td>
<td>1788</td>
<td>48</td>
<td>61</td>
<td>36832</td>
<td>1778</td>
<td>48</td>
<td>-380</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G3</td>
<td>19563</td>
<td>1062</td>
<td>15</td>
<td>-504</td>
<td>42488</td>
<td>1780</td>
<td>47</td>
<td>-441</td>
<td>36122</td>
<td>1902</td>
<td>54</td>
<td>-790</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G4</td>
<td>19615</td>
<td>1040</td>
<td>15</td>
<td>-452</td>
<td>42538</td>
<td>1781</td>
<td>47</td>
<td>-391</td>
<td>36183</td>
<td>1867</td>
<td>52</td>
<td>-729</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G5</td>
<td>47141</td>
<td>27075</td>
<td>7332</td>
<td>27075</td>
<td>94547</td>
<td>51617</td>
<td>26643</td>
<td>51617</td>
<td>89182</td>
<td>52270</td>
<td>27322</td>
<td>52270</td>
<td></td>
</tr>
</tbody>
</table>

* All entries have been multiplied by $10^5$. 
<table>
<thead>
<tr>
<th>Sampling</th>
<th>Percent</th>
<th>Estimator</th>
<th>$\tau$</th>
<th>$(t)_{\text{ok}}$</th>
<th>$(t)_{\text{BSR}}$</th>
<th>BIAS($\hat{\tau}$)</th>
<th>$p_{0.05}$</th>
<th>$p_{0.01}$</th>
<th>$p_{0.001}$</th>
<th>BIAS($\hat{p}$)</th>
<th>$p_{0.05}$</th>
<th>$p_{0.01}$</th>
<th>$p_{0.001}$</th>
<th>BIAS($\hat{p}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G_1</td>
<td>1%</td>
<td>G_1</td>
<td>50149</td>
<td>42077</td>
<td>17788</td>
<td>42077</td>
<td>100000</td>
<td>81691</td>
<td>66734</td>
<td>81691</td>
<td>100000</td>
<td>84556</td>
<td>71497</td>
<td>84556</td>
</tr>
<tr>
<td>G_2</td>
<td></td>
<td>3950</td>
<td>2188</td>
<td>74</td>
<td>-1529</td>
<td>15888</td>
<td>4864</td>
<td>369</td>
<td>-242</td>
<td>13125</td>
<td>5164</td>
<td>399</td>
<td>-2319</td>
<td>399</td>
</tr>
<tr>
<td>G_3</td>
<td></td>
<td>3894</td>
<td>4192</td>
<td>197</td>
<td>-4178</td>
<td>12648</td>
<td>5909</td>
<td>511</td>
<td>-5662</td>
<td>9795</td>
<td>5934</td>
<td>493</td>
<td>-5649</td>
<td>493</td>
</tr>
<tr>
<td>G_4</td>
<td></td>
<td>49724</td>
<td>41652</td>
<td>197</td>
<td>94966</td>
<td>76657</td>
<td>58762</td>
<td>76657</td>
<td>89932</td>
<td>74488</td>
<td>55484</td>
<td>74488</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2%</td>
<td>G_1</td>
<td>49986</td>
<td>41914</td>
<td>17588</td>
<td>41914</td>
<td>100000</td>
<td>81691</td>
<td>66734</td>
<td>81691</td>
<td>100000</td>
<td>82197</td>
<td>67591</td>
<td>82197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_2</td>
<td>7664</td>
<td>1217</td>
<td>23</td>
<td>-408</td>
<td>17050</td>
<td>3421</td>
<td>173</td>
<td>-126</td>
<td>2867</td>
<td>126</td>
<td>-1233</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_3</td>
<td>5749</td>
<td>2404</td>
<td>73</td>
<td>-2322</td>
<td>14976</td>
<td>3706</td>
<td>206</td>
<td>-2224</td>
<td>12394</td>
<td>3532</td>
<td>173</td>
<td>-3050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_4</td>
<td>5757</td>
<td>2394</td>
<td>72</td>
<td>-2315</td>
<td>15037</td>
<td>3656</td>
<td>201</td>
<td>-3272</td>
<td>12453</td>
<td>3489</td>
<td>168</td>
<td>-2991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_5</td>
<td>49257</td>
<td>41184</td>
<td>16963</td>
<td>41184</td>
<td>94908</td>
<td>76599</td>
<td>89825</td>
<td>74381</td>
<td>55325</td>
<td>74381</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>G_1</td>
<td>49884</td>
<td>41812</td>
<td>17468</td>
<td>41812</td>
<td>97469</td>
<td>79160</td>
<td>62667</td>
<td>79160</td>
<td>92400</td>
<td>76956</td>
<td>59226</td>
<td>6956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_2</td>
<td>7964</td>
<td>858</td>
<td>12</td>
<td>-108</td>
<td>17740</td>
<td>2256</td>
<td>68</td>
<td>-569</td>
<td>15147</td>
<td>1821</td>
<td>47</td>
<td>-297</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_3</td>
<td>7090</td>
<td>1232</td>
<td>22</td>
<td>-983</td>
<td>16952</td>
<td>2238</td>
<td>67</td>
<td>-1357</td>
<td>14276</td>
<td>1934</td>
<td>52</td>
<td>-1169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_4</td>
<td>7123</td>
<td>1203</td>
<td>21</td>
<td>-949</td>
<td>17011</td>
<td>2203</td>
<td>65</td>
<td>-1298</td>
<td>14330</td>
<td>1906</td>
<td>50</td>
<td>-1114</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_5</td>
<td>47970</td>
<td>39897</td>
<td>15922</td>
<td>39897</td>
<td>94746</td>
<td>76437</td>
<td>58426</td>
<td>76437</td>
<td>89587</td>
<td>74141</td>
<td>54970</td>
<td>74141</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>G_1</td>
<td>49851</td>
<td>41779</td>
<td>17544</td>
<td>41779</td>
<td>97010</td>
<td>78700</td>
<td>61940</td>
<td>78700</td>
<td>90963</td>
<td>75518</td>
<td>57031</td>
<td>75518</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_2</td>
<td>7953</td>
<td>579</td>
<td>6</td>
<td>-119</td>
<td>17991</td>
<td>1054</td>
<td>18</td>
<td>-318</td>
<td>15410</td>
<td>1085</td>
<td>16</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_3</td>
<td>7469</td>
<td>762</td>
<td>10</td>
<td>-603</td>
<td>17476</td>
<td>1221</td>
<td>22</td>
<td>-833</td>
<td>14854</td>
<td>1059</td>
<td>16</td>
<td>-590</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_4</td>
<td>7513</td>
<td>738</td>
<td>9</td>
<td>-559</td>
<td>17527</td>
<td>1190</td>
<td>21</td>
<td>-782</td>
<td>14896</td>
<td>1038</td>
<td>16</td>
<td>-548</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G_5</td>
<td>45895</td>
<td>37823</td>
<td>14311</td>
<td>37823</td>
<td>94548</td>
<td>76238</td>
<td>58123</td>
<td>76238</td>
<td>89198</td>
<td>73754</td>
<td>54398</td>
<td>73754</td>
</tr>
</tbody>
</table>

* All entries have been multiplied by $10^5$. 

(2) In the vast majority of the cases considered, \( G_3(t) \) and \( G_4(t) \) performed virtually the same in all respects while \( G_2(t) \) performed approximately twice as well as either \( G_3(t) \) or \( G_4(t) \).

(3) The relative performances of \( G_1(t), \ldots, G_5(t) \) remained virtually unchanged as \( m \) increased.

(4) As would be expected, all five estimators increased in precision as sample size increased.

On the basis of these observations, \( G_2(t) \) appears to be the "best" estimator and is the one implemented by the subnetwork analysis procedure.

3.5.2 The Performance of \( G_2(t) \) under Systematic and Random Sampling

Table 9 presents a comparison of the performance of the estimator \( G_2(t) \) for both systematic sampling and random sampling under a variety of sampling conditions. In almost every case, systematic sampling was superior to random sampling and hence is the preferred technique.
<table>
<thead>
<tr>
<th>Percent</th>
<th>Distribution</th>
<th>sampled</th>
<th>$N(\hat{\theta})$</th>
<th>$\text{MSE}(\hat{\theta})$</th>
<th>$\text{VAR}(\hat{\theta})$</th>
<th>$\text{BIAS}(\hat{\theta})$</th>
<th>$\text{MD} (\hat{\theta})$</th>
<th>$\text{MSE}(\hat{\theta})$</th>
<th>$\text{VAR}(\hat{\theta})$</th>
<th>$\text{BIAS}(\hat{\theta})$</th>
<th>$\text{MD} (\hat{\theta})$</th>
<th>$\text{MSE}(\hat{\theta})$</th>
<th>$\text{VAR}(\hat{\theta})$</th>
<th>$\text{BIAS}(\hat{\theta})$</th>
<th>$\text{MD} (\hat{\theta})$</th>
<th>$\text{MSE}(\hat{\theta})$</th>
<th>$\text{VAR}(\hat{\theta})$</th>
<th>$\text{BIAS}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>$F_{8,2}$</td>
<td>.5065</td>
<td>.2416</td>
<td>.2581</td>
<td>.2613</td>
<td>.5860</td>
<td>.2748</td>
<td>.2660</td>
<td>.5370</td>
<td>.5646</td>
<td>.2591</td>
<td>.2819</td>
<td>.4677</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{5,5}$</td>
<td>.5472</td>
<td>.2804</td>
<td>.2813</td>
<td>.4256</td>
<td>.5506</td>
<td>.2986</td>
<td>.3425</td>
<td>.4397</td>
<td>.6902</td>
<td>.4827</td>
<td>.5582</td>
<td>.4472</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{2,8}$</td>
<td>.6163</td>
<td>.3514</td>
<td>.3535</td>
<td>.2655</td>
<td>.5379</td>
<td>.3037</td>
<td>.3210</td>
<td>.3265</td>
<td>.6721</td>
<td>.4680</td>
<td>.4830</td>
<td>.4722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2%</td>
<td>$F_{8,2}$</td>
<td>.4958</td>
<td>.2664</td>
<td>.2724</td>
<td>.1744</td>
<td>.6764</td>
<td>.3800</td>
<td>.4096</td>
<td>.5607</td>
<td>.6690</td>
<td>.3602</td>
<td>.3772</td>
<td>.5559</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{5,5}$</td>
<td>.7842</td>
<td>.5655</td>
<td>.5620</td>
<td>4.4626</td>
<td>.6850</td>
<td>.3984</td>
<td>.4408</td>
<td>.4737</td>
<td>.6317</td>
<td>.3405</td>
<td>.3668</td>
<td>.3863</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{2,8}$</td>
<td>.5858</td>
<td>.3373</td>
<td>.3374</td>
<td>-.4000</td>
<td>.5599</td>
<td>.3031</td>
<td>.3063</td>
<td>.5132</td>
<td>.5955</td>
<td>.3681</td>
<td>.3651</td>
<td>.9369</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>$F_{8,2}$</td>
<td>.5932</td>
<td>.3440</td>
<td>.3378</td>
<td>.8058</td>
<td>.6447</td>
<td>.3730</td>
<td>.4166</td>
<td>.4635</td>
<td>.6010</td>
<td>.3187</td>
<td>.3387</td>
<td>.5020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{5,5}$</td>
<td>.9260</td>
<td>.7835</td>
<td>.7732</td>
<td>15.1666</td>
<td>.7586</td>
<td>.4663</td>
<td>.4545</td>
<td>.8635</td>
<td>.7185</td>
<td>.5324</td>
<td>.5476</td>
<td>.2069</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{2,8}$</td>
<td>.6255</td>
<td>.4102</td>
<td>.4102</td>
<td>1.2692</td>
<td>.5039</td>
<td>.2696</td>
<td>.2635</td>
<td>.6615</td>
<td>.6204</td>
<td>.3541</td>
<td>.3438</td>
<td>.1047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>$F_{8,2}$</td>
<td>1.1572</td>
<td>1.1521</td>
<td>1.1555</td>
<td>1.6627</td>
<td>1.5619</td>
<td>2.2203</td>
<td>2.8222</td>
<td>5.8276</td>
<td>5.7114</td>
<td>.3281</td>
<td>.2500</td>
<td>.9418</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{5,5}$</td>
<td>.7099</td>
<td>.4468</td>
<td>.4468</td>
<td>1.3728</td>
<td>.5955</td>
<td>.3628</td>
<td>.3703</td>
<td>.3283</td>
<td>.6487</td>
<td>.4571</td>
<td>.4812</td>
<td>.4057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F_{2,8}$</td>
<td>.6387</td>
<td>.3513</td>
<td>.3611</td>
<td>.8929</td>
<td>.6441</td>
<td>.4247</td>
<td>.4247</td>
<td>.9838</td>
<td>.6515</td>
<td>.4000</td>
<td>.3916</td>
<td>-6.7857</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. ESTIMATION OF A DISCRETE DISTRIBUTION FUNCTION
BY EXTRAPOLATING UPPER AND LOWER BOUNDS

4.1 Introduction

Since it is sometimes impractical to completely enumerate a
subnetwork's discrete duration distribution, F, the bounds \( F^+(t; \theta, \lambda) \)
and \( F^-(t; \theta, \lambda) \) are calculated as a first step in the determination
of an estimate, \( \hat{F} \), of F. Theorems 4 and 5 of section 2 imply that
for \( \theta \) very large both \( F^+(t; \theta, \lambda) \) and \( F^-(t; \theta, \lambda) \) may serve as
adequate estimates of F. Unfortunately, it becomes increasingly
laborious to calculate these quantities as \( \theta \to \infty \). Hence the extrapo-
lation procedure described in this section was devised as a practical
alternative to evaluating the upper and lower bounds for large \( \theta \).

4.2 The Extrapolation Problem

Suppose that for a particular subnetwork, the numerical values
of \( F^+(t; \theta, \lambda) \) and \( F^-(t; \theta, \lambda) \) are available for each of the combina-
tions of \( t = t_1, \ldots, t_I \) and \( (\theta, \lambda) = (\theta_1, \lambda_1), \ldots, (\theta_J, \lambda_J) \) where
\begin{align*}
(1) & \quad t_i \leq t_{i+1} \text{ for all } i \\
(2) & \quad \theta_j \leq \theta_{j+1} \text{ and } \lambda_j \leq \lambda_{j+1} \text{ for all } j.
\end{align*}
The specific goal of the extrapolation procedure is to estimate F at
the points \( t_1, t_2, \ldots, t_I \).

Let
\[ \omega = 1/(1+\theta) \]
\[ \omega_j = 1/(1+\theta_j), \quad j = 1, \ldots, J, \quad (4.1) \]
and define

\[(a) \ H^+(t, \omega) = F^+(t; 0, \lambda) \] and
\[(b) \ H^-(t, \omega) = F^-(t; 0, \lambda). \]  \hspace{1cm} (4.2)

Then the results of Theorems 4 and 5 can be restated as follows

\[(a) \ H^+(t, \omega) \] is a nondecreasing function of \( \omega \) for every 
\( t; \ H^-(t, \omega) \) is a nonincreasing function of \( \omega \) for 
\( \) every \( t; \) 
\[(b) \) for any \( \omega \) and \( t \)
\[ H^+(t, \omega) \geq F(t) \geq H^-(t, \omega) ; \]
and
\[(c) \) there exists a finite value \( \omega^* \) such that \( \omega \leq \omega^* \)
\[ implies \ H^+(t, \omega) = H^-(t, \omega) = F(t) \) for every \( t. \]

Thus, estimating \( H^+(t, 0) \) and \( H^-(t, 0) \) is the same as estimating \( F(t). \) Although viable estimates of \( F(t) \) can be obtained in a variety of ways, the proposed procedure uses the known quantities \( H^+(t_i, \omega_j) \) and \( H^-(t_i, \omega_j) \) \( (i = 1, \ldots, I; \ j = 1, \ldots, J) \) to estimate functions \( \hat{H}^+(t, \omega) \) and \( \hat{H}^-(t, \omega) \) satisfying

\[ (1) \ \hat{H}^+(t_i, \omega_j) \geq \hat{H}^+(t_i, \omega_{j+1}) \) for each \( i \) and \( j; \)
\[ (2) \ \hat{H}^-(t_i, \omega_j) \leq \hat{H}^-(t_i, \omega_{j+1}) \) for each \( i \) and \( j; \) and
\[ (3) \ \hat{H}^+(t_i, 0) = \hat{H}^-(t_i, 0) \leq \hat{H}^+(t_{i+1}, 0) = \hat{H}^-(t_{i+1}, 0) \) for 
\[ each \ i. \]  \hspace{1cm} (4.3)
and then estimates \( F(t_i) \) by

\[
\hat{F}(t_i) = \hat{H}^+(t_i, 0) = \hat{H}^-(t_i, 0) \quad i = 1, \ldots, I. \tag{4.4}
\]

The basic idea is simply for each \( t_i \) to fit a function \( \hat{H}^+(t_i, \omega) \) (as a function of \( \omega \)) to the sequence of upper bounds on \( F(t_i) \) namely \( \hat{H}^+(t_i, \omega_1), \ldots, \hat{H}^+(t_i, \omega_J) \) and also fit a function \( \hat{H}^-(t_i, \omega) \) to the lower bounds on \( F(t_i) \) under the restriction that

\[
\lim_{\omega \to 0} \hat{H}^+(t_i, \omega) = \lim_{\omega \to 0} \hat{H}^-(t_i, \omega). \tag{4.5}
\]

Since \( F(t_1) \leq \ldots \leq F(t_I) \), the additional restriction that

\[
\hat{H}^+(t_1, 0) \leq \ldots \leq \hat{H}^+(t_I, 0) \tag{4.6}
\]

is imposed so that

\[
\hat{F}(t_1) \leq \ldots \leq \hat{F}(t_I). \tag{4.7}
\]

### 4.3 A Linear Programming Solution to the Extrapolation Problem

The determination of \( \hat{H}^+ \) and \( \hat{H}^- \) is as follows. For each \( i \) let

\[
\hat{H}^+(t_i, \omega) = \alpha_{0i} + \alpha_{1i} \omega + \alpha_{2i} \omega^2 \tag{4.8}
\]

and

\[
\hat{H}^-(t_i, \omega) = \beta_{0i} + \beta_{1i} \omega + \beta_{2i} \omega^2 \tag{4.9}
\]

where \( \alpha_{0i}, \alpha_{1i}, \alpha_{2i}, \beta_{0i}, \beta_{1i}, \) and \( \beta_{2i} \) \( i = 1, \ldots, I \) are all constants determined so that (4.3) holds. Since (4.4) is a quadratic function in \( \omega \), requirement (1) of (4.3) is met by requiring
Similarly, requirement (2) of (4.3) is met by restricting
\[ \beta_{ii} \leq 0 \quad i = 1, \ldots, I. \] (4.11)

Finally, requirement (3) is met by requiring
\[ \alpha_{0i} = \beta_{0i} \leq \alpha_{0,i+1} = \beta_{0,i+1} \quad i = 1, \ldots, I. \] (4.12)

Of course, when the restrictions (4.10), (4.11), and (4.12) are enforced, it is not always possible to have
\[ (1) \; \hat{H}^+(t_i, \omega_j) = H^+(t_i, \omega_j) \quad \text{and} \]
\[ (2) \; \hat{H}^-(t_i, \omega_j) = H^-(t_i, \omega_j) \] (4.13)
for all \( i \) and \( j \). Hence, the constants \( \alpha_{0i}, \alpha_{li}, \alpha_{2i}, \beta_{0i}, \beta_{li}, \) and \( \beta_{2i} \) (\( i = 1, \ldots, I \)) are determined by minimizing
\[
\sum_{j=1}^{J} a(\omega_j) \left[ \sum_{i=1}^{I} \left| \hat{H}^+(t_i, \omega_j) - H^+(t_i, \omega_j) \right| + \left| \hat{H}^-(t_i, \omega_j) - H^-(t_i, \omega_j) \right| \right] \] (4.14)
under the restrictions (4.10) - (4.12) where the \( a(\omega_j) \) is a specified nonnegative weighting constant.

The weights, \( a(\omega) \), in (4.14) should reflect the increase in information about \( F(t) \) as \( \omega \to 0 \) (i.e., \( \theta \to \infty \)). In the algorithm the weight \( a(\omega) \) has been defined to be
\[ a(\omega) = 1 - 2\omega^2 + 3\omega^3. \]
The coefficients in the cubic function were selected so

\[ a(0) = 1, \quad a(1) = 0, \quad \text{and} \quad \frac{da(\omega)}{d\omega} \bigg|_{\omega=0} = \frac{da(\omega)}{d\omega} \bigg|_{\omega=1} = 0 \]  

(4.15)

Hence, the points \( \omega = .25 \) and \( \omega = .5 \) which correspond to \( \theta = 3 \) and \( \theta = 1 \), respectively, have weights .84375 and .50000, respectively.

The minimization of (4.14) subject to (4.10) - (4.12) can be restated as

\[
\text{minimize } \sum_{j=1}^{J} \sum_{i=1}^{I} a(\omega_j) u_{ij} + v_{ij} \\
\text{subject to } \sum_{k=1}^{I} y_k \leq 1 \\
-u_{ij} \leq \sum_{k=1}^{I} y_k + \alpha_{1i} \omega_j + (\alpha_{21i} - \alpha_{212}) \omega_j^2 - H^+(t_{ij}, \omega_j) \leq u_{ij} \text{ for all } i, j \\
-v_{ij} \leq \sum_{k=1}^{I} y_k - \beta_{1i} \omega_j + (\beta_{21i} - \beta_{212}) \omega_j^2 - H^-(t_{ij}, \omega_j) \leq v_{ij} \text{ for all } i, j \\
(4.16)
\]

\[ u_{ij}, v_{ij}, \alpha_{1i}, \beta_{1i}, \alpha_{21i}, \beta_{21i}, \beta_{212} \geq 0 \text{ for all } i, j \]

where

\[ \alpha_{2i} = \alpha_{21i} - \alpha_{212} , \]

(4.17)

\[ \beta_{2i} = \beta_{21i} - \beta_{212} , \text{ and} \]

(4.18)

\[ \alpha_{0i} = \beta_{0i} = \sum_{k=1}^{I} y_k . \]

(4.19)

This is a linear programming problem which may be solved using any standard method. In the computer implementation of the subnetwork
analysis procedure, a streamlined version of the revised simplex algorithm was especially prepared and implemented to solve this problem. Once this linear programming problem has been solved, $\hat{F}$ is retrieved through the relation

$$\hat{F}(t_i) = \sum_{k=1}^{i} \gamma_k$$

for all $i$.

(4.20)

If the upper bounds $F^+(t, \theta, \lambda)$ and lower bounds $F^-(t; \theta, \lambda)$ were determined without sampling then

$$H^+(t_i, w_1) > \ldots > H^+(t_i, w_j) \geq F(t_i) \geq H^-(t_i, w_1) > \ldots > H^-(t_i, w_j).$$

However this relationship does not necessarily hold if sampling is used in the determination of the bounds. Hence the determination of $\hat{H}^+(t, w)$ and $\hat{H}^-(t, w)$ does not include the restriction

$$\hat{H}^-(t_i, w_j) \leq \hat{F}(t_i) \leq \hat{H}^+(t_i, w_j) \quad i = 1, \ldots, I.$$

It should also be noted that, if a weighted least squares criterion had been used instead of minimization of a weighted sum of absolute residuals, then the determination of $\hat{F}(t)$ would have been a quadratic programming problem instead of a somewhat simpler linear programming problem.

4.4 An Example of the Linear Programming Solution

Using the simplex algorithm referred to in subsection 4.3, the linear programming problem (4.16) was solved for the data in Table 10. Figure 10 indicates the fits obtained.
### TABLE 10

Extrapolation Data

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$F^+(H^+)$</th>
<th>$F^-(H^-)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1 = 1.5$ ($\omega_1 = .4$)</td>
<td>$t_1$</td>
<td>.7</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>$t_2$</td>
<td>.8</td>
<td>.3</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>1.0</td>
<td>.5</td>
</tr>
<tr>
<td>$\theta_2 = 2.5$ ($\omega_2 = .2851$)</td>
<td>$t_1$</td>
<td>.4</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>$t_2$</td>
<td>.6</td>
<td>.3</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>.75</td>
<td>.6</td>
</tr>
<tr>
<td>$\theta_3 = 3.5$ ($\omega_3 = .2222$)</td>
<td>$t_1$</td>
<td>.35</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>$t_2$</td>
<td>.55</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>.75</td>
<td>.65</td>
</tr>
</tbody>
</table>
As required, $\hat{F}(t_1) = .31 \leq \hat{F}(t_2) = .43 \leq \hat{F}(t_3) = .71$.

Figure 10

Extrapolation results for the data in Table 10.
5. POTENTIAL MODIFICATIONS OF THE SUBNETWORK ANALYSIS PROCEDURE

5.1 Introduction

The objective of Subnetwork Analysis is to determine each subnetwork's duration distribution, say \( F(t) \). When this step is begun, each activity has a specified duration distribution. Let

- \( n \) = number of activities in the subnetwork,
- \( X_i \) = the duration of activity \( i \), and
- \( F_{X_i}(t) \) = the c.d.f. for activity \( i \).

Also, let

- \( m \) = the number of paths through the subnetwork, and
- \( Y_j \) = the length of the \( j \)-th path through the subnetwork

\[
Y_j = \sum_{i=1}^{n} \delta_{ij} X_i
\]  

(5.1)

where

\[
\delta_{ij} = \begin{cases} 1 & \text{if activity } i \text{ is on the } j \text{-th path} \\ 0 & \text{otherwise.} \end{cases}
\]

Let the maximum path length be

\[
Y^* = \max_{1 \leq j \leq m} Y_j.
\]  

(5.2)

Then

\[
F(t) = P(Y^* \leq t) = \int_{-\infty}^{t} \cdots \int_{-\infty}^{t} dF_{Y_1}, \ldots, dF_{Y_m \left( t_1, \ldots, t_m \right)}
\]  

(5.3)

where

\[
F_{Y_1, \ldots, Y_m \left( t_1, \ldots, t_m \right)} = \text{the joint distribution of the } m \text{ paths.}
\]
The activity distributions are assumed to be independent. Thus, the marginal distribution of $Y_j$, say $F_{Y_j}(t)$, is the convolution of the path's activity duration distributions; that is,

$$F_{Y_j}(t) = \int_\ldots \int F_{X_{j_1}}(t - \sum \delta_{i,j} X_i) \prod_{i \neq j_1} dF_{X_i}(X_i) \quad (5.4)$$

where $j_1$ is the index of an activity on the path. Furthermore, if

$$F_{Y_j|X}(t)$$

denotes the conditional distribution of $Y_j$ given a set, $X$, of activity duration values, then

$$F_{Y_j|X}(t) = \int_\ldots \int F_{X_{j_1}}(t - \sum \delta_{i,j} X_i) \prod_{i \neq j_1} dF_{X_i}(X_i) \quad (5.5)$$

which is the convolution of the path's activity durations not in $X$.

If there is no activity that is in two or more of $Y_{j_1}, \ldots, Y_{j_k}$ (that is, these paths have no activities in common), then $Y_{j_1}, \ldots, Y_{j_k}$ are independent, and

$$F_{Y_{j_1}, \ldots, Y_{j_k}}(t_{j_1}, \ldots, t_{j_k}) = \prod_{i=1}^k F_{Y_i}(t_{j_i}) \quad (5.6)$$

However, if

$$X = \{X_i | \text{activity } i \text{ is in more than one of } Y_1, \ldots, Y_m\}$$

is a nonempty set, then $Y_1, \ldots, Y_m$ are dependent, and

$$F(t) = \int_\ldots \int \left[ \prod_{j=1}^m F_{Y_j|X}(t - \sum \delta_{i,j} X_i) \right] \prod_{X_i \in X} dF_{X_i}(X_i) \quad (5.7)$$

5.2 Explicit Evaluation of the Subnetwork Duration Distribution

For simple subnetworks it is relatively easy to identify all of the paths and give an explicit expression for $F(t)$ via (5.7). In particular Hartley and Wortham (1966) considered series, parallel, and Wheatstone Bridge subnetworks (see Figure 11). Ringer (1969) extended this work to include Double Wheatstone Bridge and Criss-Cross subnetworks (see Figure 12). These exact expressions for the subnetwork duration distribution form the basis of Step 2, Simplification, in the project scheduling procedure. Interestingly, this implies that the subnetworks actually considered in Step 4, Subnetwork Analysis, do not have any of these simple activity configurations in them, and hence are generally fairly complex.

To utilize (5.7) to determine $F(t)$, all the paths through the subnetwork must be identified and then the numerical evaluation of (5.7) performed. Martin (1964) presented a clever method for performing the numerical evaluation of (5.7) when the activity duration distributions were all piecewise polynomial functions with finite ranges. Martin's technique is most readily suited to subnetworks primarily composed of activities in series or parallel. Unfortunately, the subnetworks generally encountered in the Subnetwork Analysis step are not of this form. Furthermore, Martin's technique becomes computationally impractical for large subnetworks.
Activities in Parallel

Two Activities in Series

Wheatstone Bridge

Figure 11

Subnetworks considered by Hartley and Wortham (1966).
Double Wheatstone Bridge

Criss-Cross

Figure 12

Subnetworks considered by Ringer (1969).
5.3 Approximating the Subnetwork Duration Distribution $F(t)$

Since the explicit evaluation of the exact expression for the subnetwork duration distribution $F(t)$ given in (5.7) is generally impractical for other than simple subnetworks, several authors have considered approximating $F(t)$. A review of the classical approximation procedures is given in Moder and Phillips (1974). The more recent approximation procedures are essentially based on either sophisticated Monte Carlo simulation or the determination of upper and lower bounds for $F(t)$. The Subnetwork Analysis procedure developed in Sections 2 - 4 is one of these approximation procedures. That Subnetwork Analysis procedure basically estimates $F(t)$ by extrapolating a sequence of upper and lower bounds on the subnetwork's discrete duration distribution $F$ - with specialized Monte Carlo techniques sometimes employed in the determination of the upper and lower bounds.

Noteworthy papers on the Monte Carlo simulation of $F(t)$ include Van Slyke (1963), Gaver and Burt (1968), and Burt and Garman (1971).

The two outstanding published techniques for determining upper and lower bounds on $F(t)$ are due to Robillard and Trahan (1977) and Kleindorfer (1971). These techniques are briefly discussed in subsections 5.3.1 and 5.3.2, respectively.

Subsection 5.3.3 indicates several ways that the Monte Carlo techniques and the upper and lower bounds of Kleindorfer (1971) and Robillard and Trahan (1977) can be incorporated into the general Subnetwork Analysis procedure.
5.3.1 Robillard and Trahan's Lower Bound on \( F(t) \)

Robillard and Trahan (1977) proposed a lower bound, \( F^-(t) \), for \( F(t) \) based on a Bonferroni inequality. Specifically,

\[
P( \max_{1 \leq j \leq m} Y_j > t) = P( \bigcup_{j=1}^{m} \{ Y_j > t \}) \leq \sum_{j=1}^{m} P(Y_j > t),
\]

so

\[
F(t) = P( \max_{1 \leq j \leq m} Y_j \leq t) = 1 - P( \max_{1 \leq j \leq m} Y_j > t) \geq 1 - \sum_{j=1}^{m} P(Y_j > t)
\]

\[
= 1 - \sum_{j=1}^{m} [1 - F_{Y_j}(t)] = 1 - m + \sum_{j=1}^{m} F_{Y_j}(t) \equiv F^-(t).
\]

Robillard and Trahan (1977) evaluate the term \( \sum_{j=1}^{m} F_{Y_j}(t) \) using the characteristic functions, say \( \psi_{Y_1}(\tau) \), \( \ldots \), \( \psi_{Y_m}(\tau) \), of \( Y_1 \), \( \ldots \), \( Y_m \).

Let \( I_\tau \) denote the integration corresponding to the inversion of a characteristic function. Then

\[
\sum_{j=1}^{m} F_{Y_j}(t) = \sum_{j=1}^{m} I_\tau[\psi_{Y_j}] = I_\tau[\sum_{j=1}^{m} \psi_{Y_j}]
\]

(5.10)

where the last equality follows from the linearity of integration.

For example, if \( Y_1 \), \( \ldots \), \( Y_m \) are all continuous random variables, then

\[
\sum_{j=1}^{m} I_\tau[\psi_{Y_j}(\tau)] = \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-i\tau x} \psi_{Y_j}(\tau) dx \, dt
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-i\tau x} \sum_{j=1}^{m} \psi_{Y_j}(\tau) dx \, dt
\]

(5.11)

\[
= I_\tau[\sum_{j=1}^{m} \psi_{Y_j}(\tau)].
\]
Although $\psi_j^x (t)$ can be written as the product of the characteristic functions for the individual activities in $Y_j$, this approach to evaluating $\Sigma_{j=1}^m \psi_j^x (t)$ would require the explicit enumeration of all the subnetwork's paths which is computationally impractical for large complex subnetworks. Therefore Robillard and Trahan (1977) developed the following recursive scheme for evaluating $\Sigma_{j=1}^m \psi_j^x (t)$. Let $\psi_{x_1}^x (t), \ldots, \psi_{x_n}^x (t)$ denote the characteristic functions of the individual activity durations $x_1, \ldots, x_n$. Let the $k$-th activity originate at node $\text{Orig}_k$ and terminate at node $\text{Term}_k$. If

\begin{align*}
B_i &= \{ k \mid \text{Term}_k = i \}, \\
\phi(t, 1) &= 1, \text{ and} \\
\phi(t, i) &= \Sigma_{k \in B_i} \phi(t, \text{Orig}_k) \psi_{x_k}^x (t), \quad i = 2, \ldots, N, \tag{5.14}
\end{align*}

where $N$ is the number of nodes, then

\begin{equation}
\Sigma_{j=1}^m \psi_j^x (t) = \phi(t, N). \tag{5.15}
\end{equation}

Although it is not explicitly noted by Robillard and Trahan (1977), the number of paths, $m$, can also be recursively generated. If

\begin{equation}
m_1 = 1 \tag{5.16}
\end{equation}

and

\begin{equation}
m_i = \Sigma_{k \in B_i} m_k \quad i = 2, \ldots, N, \tag{5.17}
\end{equation}

then

\begin{equation}
m = m_N. \tag{5.18}
\end{equation}

Apart from any numerical inaccuracies in the computation (5.14), the tightness of the lower bound $F^-(t)$ is the same as the tightness of the Bonferroni inequality (5.8).
Robillard and Trahan (1977) also note that another Bonferroni inequality implies that

\[
P(\max_{1 \leq j \leq m} Y_j > t) = P\left( \bigcup_{j=1}^{m} \{Y_j > t\}\right) \geq \sum_{j=1}^{m} P(Y_j > t) - \sum_{i<j} P(Y_i > t, Y_j > t)
\]

(5.19)

Unfortunately, to use (5.19) as the basis for an upper bound on \( F(t) \) seems to require the explicit enumeration of the paths because of the joint nature of \( P(Y_i > t, Y_j > t) \) and the lack of a convenient upper bound for \( P(Y_i > t, Y_j > t) \).

5.3.2 Kleindorfer's Upper and Lower Bounds on \( F(t) \)

Let the subnetwork's activities be numbered \( i = 1, \ldots, n \) in such a way that, if \( i < j \) and both activities \( i \) and \( j \) are on a path, then activity \( i \) precedes activity \( j \). Let \( A_i \) denote the set of activities which immediately precede activity \( i \) on some path. As before, let

\[
X_i = \text{the duration of activity } i
\]

and also define

\[
P_i(t) = P(U_i \leq t) \quad (5.20)
\]

\[
V_i = U_i + X_i = \text{the completion time for activity } i, \text{ and}
\]

\[
Q_i = P(V_i \leq t).
\]

Kleindorfer (1971) proposed upper and lower bounds for \( F(t) \) by recursively defining upper bounds, \( P'_i(t) \) and \( Q'_i(t) \), and lower bounds, \( P''_i(t) \) and \( Q''_i(t) \) on \( P_i(t) \) and \( Q_i(t) \), respectively. The upper bounds \( P'_i(t) \) are based upon the inequality
\[
\min_{j \in A_i} P(V_j \leq t) \geq P(\max_{j \in A_i} V_j \leq t) = P_i(t), \quad (5.21)
\]

and \(Q'_i(t)\) is simply the convolution of \(F_{X_i}(t)\) and \(P'_i(t)\). The lower bounds \(P''_i(t)\) are based upon the inequality

\[
P_i(t) = P(\max_{j \in A_i} V_j \leq t) \geq \prod_{j \in A_i} P(V_j \leq t), \quad (5.22)
\]

and \(Q''_i(t)\) is the convolution of \(F_{X_i}(t)\) and \(P''_i(t)\). (Although Kleindorfer proves a version of (5.22), the inequality as stated follows from the more general results of Esary, Proschan, and Walkup (1967).)

The recursive relations for \(P'_i(t), Q'_i(t), P''_i(t),\) and \(Q''_i(t)\) are as follows: For notational convenience assume that activity 1 is an activity with zero duration which precedes the rest of the subnetwork and that activity \(n\) is an activity with zero duration which follows the completion of the rest of the subnetwork. Furthermore, assume that \(X_i\) is a discrete, nonnegative random variable taking on values in \(S\) for all \(i\). Then, for \(t \geq 0,\)

\[
P'_1(t) = P_1(t) = Q_1(t) = Q'(t) = 1, \quad (5.23)
\]

\[
P_i(t) = \min_{j \in A_i} Q'_j(t), \quad i = 2, \ldots, n, \quad (5.24)
\]

\[
Q'_i(t) = \sum_{s \in S} P(X_i = s)P'_i(t - s), \quad i = 2, \ldots, n, \quad (5.25)
\]

\[
P''_1(t) = P_1(t) = Q_1(t) = Q''(t) = 1, \quad (5.26)
\]

\[
P'_i(t) = \prod_{j \in A_i} Q'_j(t), \quad i = 2, \ldots, n, \quad (5.27)
\]

\[
Q''_i(t) = \sum_{s \in S} P(X_i = s)P''_1(t - s), \quad i = 2, \ldots, n. \quad (5.28)
\]
Finally,

\[ P_n(t) \leq F(t) \leq P'_n(t). \] (5.29)

The computational beauty of these bounds on \( F(t) \) is that they do not necessitate the enumeration of all of the subnetwork's paths.

The tightness of these bounds on \( F(t) \) depends on the structure of the subnetwork. Since the recursive relations (5.23) - (5.28) sequentially bound the \( P_i(t) \) in terms of the bounds for the completion time distributions of the activities immediately preceding activity \( i \), the differences \( P'_i(t) - P_i(t) \) and \( P_i(t) - P''_i(t) \) essentially cumulate as \( i \) increases. Therefore, the bounds on \( F(t) \) will generally tend to be tighter the shorter the subnetwork's paths. Furthermore, the difference

\[ \min_{j \in A_1} P(V_j \leq t) - P(\max_{j \in A_1} V_j \leq t) \] (5.30)

tends to decrease as the \( V_j \)'s have more and more activities in common; whereas, the difference

\[ P(\max_{j \in A_1} V_j \leq t) - \Pi_{j \in A_1} P(V_j \leq t) \] (5.31)

tends to increase as the \( V_j \)'s have more and more activities in common. Thus subnetwork structures that lead to tight upper bounds on \( F(t) \), lead to loose lower bounds on \( F(t) \), and vice versa. Of course, the tightness of both the upper and lower bounds tends to decrease as the number of paths increases.
5.3.3 Incorporating Different Methods of Approximating $F(t)$ into Subnetwork Analysis

The Monte Carlo simulation techniques and the bounding procedures of Kleindorfer (1971) and Robillard and Trahan (1977) referred to thus far in subsection 5.3 could be used to modify the current Subnetwork Analysis procedure discussed in Sections 2–4. Since the empirical experience with the modifications to be briefly described in the remainder of this subsection is generally extremely limited, these potential modifications are really subjects for future research.

A Monte Carlo simulation of the subnetwork duration distribution $F(t)$ could, of course, essentially replace the current Subnetwork Analysis procedure. A less radical revision would be to carry out the cluster formation procedure described in subsection 2.2 for a fixed (presumably large) value of $(\theta, \lambda)$; let IMPORTANT be the set of all activities in the union of the clusters; and then estimate $F(t)$ by fixing the durations of the activities not in IMPORTANT at their mean values and doing a Monte Carlo simulation of the durations for the activities in IMPORTANT. The durations of the activities in IMPORTANT could be simulated from either their actual distributions or their approximate two-point discrete distributions. Another potential modification would be to perform the current Subnetwork Analysis procedure as is except that the upper and lower bounds $F^+(C; t)$ and $F^-(C; t)$ used in determining $F^-(t; \theta, \lambda)$ and $F^+(t; \theta, \lambda)$ could be determined with the durations for activities in $C$ determined by a Monte Carlo simulation of their actual duration distributions or to their two-point discrete distributions.
Another possible replacement for the current Subnetwork Analysis procedure would be to determine Kleindorfer's upper bound on $F(t)$ and either Kleindorfer's or Robillard and Trahan's lower bound on $F(t)$ and then use the average of these two bounding distributions as the estimate of $F(t)$. (Of course, the maximum of Kleindorfer's and Robillard and Trahan's lower bounds is also a valid lower bound.)

Again a less radical revision would be to carry out the cluster formation procedure described in subsection 2.2 for a fixed (presumably large) value of $(\theta, \lambda)$; let IMPORTANT be the set of all activities in the union of the clusters; and then estimate $F(t)$ by fixing the durations of the activities not in IMPORTANT at their mean values and averaging the upper and lower bounds for the subnetwork duration distribution when the durations for the activities in IMPORTANT have either their actual distributions or their two-point discrete distributions. The durations for the activities not in IMPORTANT could, alternatively, be fixed at their lower values when the upper bound is being determined and be fixed at their upper values when the lower bound is being determined. Finally, another potential modification would be to perform the current Subnetwork Analysis procedure as is except that the upper and lower bounds $F^+(C;t)$ and $F^-(C;t)$ could be either Kleindorfer's or Robillard and Trahan's bounds determined with the durations for the activities in $C$ having either their actual distributions or their two-point discrete distributions.

Presumably, a project scheduler might settle for a project schedule which has the probability of the project's completion by the specified deadline bounded from below by a specified amount.
In such instances lower bounds on the subnetwork duration distributions suffice. Then the Subnetwork Analysis procedure could be replaced by a procedure which simply determines either Kleindorfer's or Robillard and Trahan's lower bound. Alternatively, the cluster formation procedure could be carried out for a specified value of \((\theta, \lambda)\), the set IMPORTANT of all activities in the union of the clusters formed, and then either Kleindorfer's or Robillard and Trahan's lower bound computed with the durations for the activities outside IMPORTANT fixed and the durations for the activities in IMPORTANT having either their actual distributions or their two-point discrete distributions.

5.4 Additional Probability Inequalities as Bases for Upper and Lower Bounds on \(F(t)\)

In addition to the ones cited in subsections 5.3.1 and 5.3.2, there are other known probability inequalities which imply upper and lower bounds on \(F(t) = P(\max_{1 \leq j \leq m} Y_j \leq t)\). Three upper bounds on \(P(\max_{1 \leq j \leq m} Y_j \leq t)\) and the authors who proposed them are:

(i) Chung and Erdos (1952),

\[
P(\max_{1 \leq j \leq m} Y_j \leq t) \leq 1 - \left\{ \left[ \sum_{j=1}^{m} P(Y_j > t) \right]^2 / \left[ \sum_{j=1}^{m} P(Y_j > t) \right] \right\} + \sum_{i \neq j} P(Y_i > t, Y_j > t)\}
\]

(ii) Dawson and Sankoff (1967),

\[
p(\max_{1 \leq j \leq m} Y_j \leq t) \leq 1 - \frac{2}{r} \left[ \sum_{j=1}^{m} P(Y_j > t) \right] - \frac{1}{r-1} \sum_{i < j} P(Y_i > t, Y_j > t)
\]

(5.32)  

(5.33)
where $r$ is the greatest integer less than or equal to

$$
\sum_{i \neq j}^{m} P(Y_i > t, Y_j > t) / \sum_{j=1}^{m} P(Y_j > t); \tag{5.34}
$$

and (iii) Kounias (1968),

$$
P(\max_{1 \leq j \leq m} Y_j \leq t) \leq 1 - \{ \sum_{j \in L} P(Y_j > t) - \sum_{i < j, i, j \in L} P(Y_i > t, Y_j > t) \} \tag{5.35}
$$

where $L$ is any subset of \{1, 2, ..., $m$\} with two or more elements.

A lower bound, proposed by Hunter (1976), is

$$
P(\max_{1 \leq j \leq m} Y_j \leq t) \geq 1 - \sum_{j=1}^{m} P(Y_j > t) - \sum_{(i,j) \in T} P(Y_i > t, Y_j > t) \tag{5.36}
$$

where $T$ is any connected set of $m - 1$ pairs $(i,j)$ such that either $(.,k)$ or $(k,.)$ is in the set for each $k = 1, ..., m$.

The primary difficulty in evaluating these bounds is that the subnetwork's paths must be explicitly enumerated in order to compute the $P(Y_i > t, Y_j > t)$. Should this computational difficulty be overcome, however, the bounds could be incorporated into the Subnetwork Analysis procedure as per the discussion in subsection 5.3.3.
6. CONCLUDING REMARKS

This report had as its goal the improvement and implementation of a new project scheduling procedure currently being developed at the Institute of Statistics, Texas A&M University. The project scheduling procedure has been improved by significantly extending the very critical Subnetwork Analysis procedure. In particular, a suitable sampling procedure and estimator for bounds on the subnetwork's duration distribution, \( F(t) \), has been developed and incorporated. In addition, a procedure for extrapolating upper and lower bounds on \( F(t) \) to obtain an estimate of \( F(t) \) has also been determined and implemented.

A computer system implementing the project scheduling procedure (including the improvements in Subnetwork Analysis) has been prepared and is documented in Baker and Sielken (1978).

In addition, some possible alternatives to the current Subnetwork Analysis procedure have been suggested. These alternatives are interesting topics for future research.

The authors wish to acknowledge their gratitude to the Office of Naval Research for the support of this research under contracts N00014-68-A-0140 and N00014-76-C-0038. Several present and past members of the Institute of Statistics at Texas A&M University have also contributed to the development of this new project scheduling procedure and its computer implementation: E. Arseven (Lederle Laboratories), P. P. Biemer, C. S. Dunn, N. E. Fisher (Compucon Inc.), L. J. Ringer, and R. K. Spoeri (Bureau of the Census).

The authors also want to particularly acknowledge the considerable contributions of H. O. Hartley to Statistical PERT.
REFERENCES


This report develops improvements to a new project scheduling procedure, Statistical PERT, being developed at the Institute of Statistics, Texas A&M University. The project scheduling algorithm is a five step iterative procedure capable of determining a minimum cost project schedule when the activities making up the project have durations which are random variables. The cost of an activity is assumed to be a convex piecewise linear function of the activity's mean duration. The problem is to determine the activity mean durations which both minimize the total project cost and insure that the mean (or some specified percentile) of the corresponding project completion time distribution is less than or equal to a specified project deadline. The entire distribution of the project's completion time under the minimum cost schedule is a valuable by-product.

A critical step, Subnetwork Analysis, in the proposed procedure is improved and extended. Subnetwork Analysis determines an estimate of the duration distribution, $F(t)$, for each subnetwork identified in the previous steps. This estimate is extended to include an extrapolation of upper and lower bounds on $F(t)$. This report also develops a new sampling procedure which results in improved estimators for the bounds on $F(t)$. 

ATTACHMENT III
<table>
<thead>
<tr>
<th>KEY WORDS</th>
<th>LINK A</th>
<th></th>
<th>LINK B</th>
<th></th>
<th>LINK C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
<td>WT</td>
</tr>
<tr>
<td>Statistical PERT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Scheduling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum Cost Scheduling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Activity Durations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Completion Time Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Duration Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bounding Network Duration Distributions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ATTACHMENT III (Continued)
BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

Operations Research
Office of Naval Research
(Code 434)
Arlington, Virginia 22217

Copies

Director
Office of Naval Research
Branch Office

Copies

495 Summer Street
Boston, Massachusetts 02210
Attn: Dr. A.L. Powell

3

1

Director, Naval Research Labora- tory
Attn: Library, Code 2029 (ONRL)
Washington, D.C. 20390

2

Prof. Emanuel Parzen
Institute of Statistics
Texas A&M University
College Station, TX 77843

2

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

12

Office of Naval Research
San Francisco Area Office
760 Market St. - Room 447
San Francisco, California 94103

1

Defense Logistics Studies
Information Exchange
Army Logistics Management Center
Fort Lee, Virginia 23801

1

Technical Library
Naval Ordnance Station
Indian Head, Maryland 20640

1

Naval Research Laboratory
Washington, D.C. 20390

6

Bureau of Naval Personnel
Navy Department
Technical Library
Washington, D.C. 20370
STOP 82

2

Office of Naval Research
New York Area Office
207 West 24th Street
New York, New York 10011
Attn: Dr. J. Laderman

1

Library, Code 0212
Naval Postgraduate School
Monterey, California 93940

1

Director, Office of Naval Research
Branch Office
1030 East Green Street
Pasadena, California 91101
Attn: Dr. A.R. Laufer

1

Naval Ordnance Station
Louisville, Kentucky 40214

1

Naval Research Laboratory
Washington, D.C. 20390

Copies

Yale University
Department of Statistics
Box 2179 - Yale Station
New Haven, CT 06520
Attn: Prof. I.R. Savage

1

Library
Naval Electronics Laboratory Center
San Diego, California 92132

1

Mr. William Dejka
Code 360
Naval Electronics Laboratory Center
San Diego, California 92132

1

Professor Lucien A. Schmit, Jr.
Room 6731, Boelter Hall
School of Engng. & Applied Sci.
University of California
Los Angeles, California 90024

1

Dr. A.L. Slafkosky
Scientific Advisor
Commandant Marine Corps (Code AX)
Washington, D.C. 20380

1
<table>
<thead>
<tr>
<th>Copies</th>
<th>Copies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Harvard University</td>
</tr>
<tr>
<td></td>
<td>Department of Statistics</td>
</tr>
<tr>
<td></td>
<td>Cambridge, Massachusetts 02139</td>
</tr>
<tr>
<td></td>
<td>Attn: Prof. W.G. Cochran</td>
</tr>
<tr>
<td>1</td>
<td>Columbia University</td>
</tr>
<tr>
<td></td>
<td>Department of Industrial Engineering</td>
</tr>
<tr>
<td></td>
<td>New York, New York 10027</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. C. Derman</td>
</tr>
<tr>
<td></td>
<td>New York University</td>
</tr>
<tr>
<td></td>
<td>Institute of Mathematical Sciences</td>
</tr>
<tr>
<td></td>
<td>New York, New York 10453</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. W.M. Hirsch</td>
</tr>
<tr>
<td></td>
<td>University of North Carolina</td>
</tr>
<tr>
<td></td>
<td>Statistics Department</td>
</tr>
<tr>
<td></td>
<td>Chapel Hill, North Carolina 27515</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. W.L. Smith and Prof. M.R. Leadbetter</td>
</tr>
<tr>
<td></td>
<td>Purdue University</td>
</tr>
<tr>
<td></td>
<td>Division of Mathematical Sciences</td>
</tr>
<tr>
<td></td>
<td>Lafayette, Indiana 47079</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. H. K. Rubin</td>
</tr>
<tr>
<td></td>
<td>University of California, San Diego</td>
</tr>
<tr>
<td></td>
<td>Department of Mathematics</td>
</tr>
<tr>
<td></td>
<td>P.O. Box 109</td>
</tr>
<tr>
<td></td>
<td>La Jolla, California 92038</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. M. Rosenblatt</td>
</tr>
<tr>
<td></td>
<td>Florida State University</td>
</tr>
<tr>
<td></td>
<td>Department of Statistics</td>
</tr>
<tr>
<td></td>
<td>Tallahassee, Florida 32306</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. R.A. Bradley</td>
</tr>
<tr>
<td></td>
<td>New York University</td>
</tr>
<tr>
<td></td>
<td>Department of Industrial Engineering &amp; Operations Research</td>
</tr>
<tr>
<td></td>
<td>Bronx, New York 10453</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. J.H.K. Kao</td>
</tr>
<tr>
<td></td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td></td>
<td>Department of Statistics</td>
</tr>
<tr>
<td></td>
<td>Madison, Wisconsin 53706</td>
</tr>
<tr>
<td>1</td>
<td>Attn: Prof. G.E.P. Box</td>
</tr>
</tbody>
</table>

- University of Chicago
- Statistical Research Center
- Chicago, Illinois 60637
- Attn: Prof. W. Kruskal

- Princeton University
- Department of Mathematics
- Princeton, New Jersey 08540
- Attn: Prof. J.W. Tukey

- Stanford University
- Department of Statistics
- Stanford, California 94305
- Attn: Prof. T.W. Anderson

- University of California
- Department of Statistics
- Berkeley, California 94720
- Attn: Prof. P.J. Bickel

- University of Washington
- Department of Mathematics
- Seattle, Washington 98105
- Attn: Prof. Z.W. Birnbaum

- Cornell University
- Department of Mathematics
- Ithaca, New York 14850
- Attn: Prof. J. Kiefer
Office of Gifts & Exchanges  
Library  
Texas A&M University  
College Station, Texas 77843  

Copies  
2

Archives  
Texas A&M University  
College Station, TX 77843  

Copies  
1