QUEUEING MODELS FOR ASSEMBLY-LIKE SYSTEMS

BY

JOHN M. HARRISON

TECHNICAL REPORT NO. 133
NOVEMBER 30, 1970

SUPPORTED BY THE ARMY, NAVY AND AIR FORCE
UNDER CONTRACT NO0014-67-A-0112-0052 (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government
Distribution of this Document is Unlimited

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
QUEUEING MODELS FOR ASSEMBLY-LIKE SYSTEMS

by

JOHN M. HARRISON

TECHNICAL REPORT NO. 133
NOVEMBER 30, 1970

Supported by the Army, Navy and Air Force
with the Office of Naval Research

Gerald J. Lieberman, Project Director

This research was supported in part by the Office of Naval Research under Contract N00014-67-A-0112-0058 (NR-047-061) and National Science Foundation Grant GK-2925.

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government

Distribution of this document is unlimited

DEPARTMENT OF OPERATIONS RESEARCH
and
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>INTRODUCTION AND SUMMARY</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1 The Real-World Systems Being Modeled</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 The Queueing Models Employed</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.3 Type of Results Obtained</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1.4 Implications for Real-World Systems</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>1.5 Overview</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>PRELIMINARIES</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2.1 General Notation and Terminology</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2.2 Weak Convergence of Random Functions</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2.3 Convergence of Expectations</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>2.4 Some Miscellaneous Lemmas</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>2.5 Single-Server Queues</td>
<td>59</td>
</tr>
</tbody>
</table>

**PART I: LIMIT THEOREMS FOR THE SINGLE-STATION MODEL**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>3.1 Formulation and Notation</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>3.2 Basic Representations and Preliminary Results</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>THE CASE $\rho &lt; 1$</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>4.1 The Process ${w_n^*}$</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>4.2 The Processes ${T_n}$, ${w_n}$ and ${w_n}$</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>4.3 Special Cases</td>
<td>96</td>
</tr>
<tr>
<td>5</td>
<td>THE CASE $\rho &gt; 1$</td>
<td>102</td>
</tr>
<tr>
<td>6</td>
<td>THE CASE $\rho = 1$</td>
<td>108</td>
</tr>
<tr>
<td>7</td>
<td>ADDITIONAL REMARKS; EXAMPLES AND APPLICATIONS</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>7.1 The Processes $D(t)$ and $Q(t)$</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>7.2 Initial Conditions</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>7.3 Deterministic Flows</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>7.4 Examples and Applications</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>7.5 Additional Remarks</td>
<td>133</td>
</tr>
<tr>
<td>Chapter</td>
<td>Chapter Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>8</td>
<td>SINGLE-SERVER QUEUES IN SERIES</td>
<td>136</td>
</tr>
<tr>
<td>9</td>
<td>ASSEMBLY-LIKE NETWORKS</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>9.1 Notation and Terminology</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>9.2 Basic Representations and Preliminary Results</td>
<td>168</td>
</tr>
<tr>
<td>10</td>
<td>LIMIT THEOREMS FOR A GENERAL ASSEMBLY-LIKE NETWORK</td>
<td>177</td>
</tr>
<tr>
<td>11</td>
<td>SPECIAL CASES AND ADDITIONAL REMARKS</td>
<td>193</td>
</tr>
<tr>
<td></td>
<td>11.1 Totally Unbalanced Systems</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td>11.2 Externally Balanced Systems</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>11.3 Additional Remarks</td>
<td>199</td>
</tr>
<tr>
<td><strong>PART III: RELATED TOPICS AND DIRECTIONS FOR FUTURE RESEARCH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>DESIGN AND CONTROL OF ASSEMBLY-LIKE SYSTEMS</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>12.1 Some Equilibrium Distributions for Modified Markovian Systems</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>12.2 Dynamic Control of Markovian Systems</td>
<td>220</td>
</tr>
<tr>
<td>13</td>
<td>OTHER DIRECTIONS FOR FUTURE RESEARCH</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>APPENDIX</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>REFERENCES</td>
<td>242</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction and Summary

In this dissertation an apparently new class of queueing models is identified and studied. The basic model is a multiple input generalization of a GI/G/1 queueing system, but much more complex network models are also considered. The principle results obtained are "functional" limit theorems for the various stochastic processes associated with the models.

1.1 The Real-World Systems Being Modeled

An assembly operation, examples of which abound in the industrial world, can be schematically represented as in Figure 1 below.

![Diagram]

Figure 1. Schematic representation of an assembly operation.

An assembler is supplied with several different types of input items. His job is to assemble these various items into finished products. It may be that more than one input item of a given type is required to make a finished product, but the number is constant from one finished
product to the next (i.e., the finished products are identical). However, for ease of discussion, let us assume throughout the remainder of this section that exactly one input item of each type is required for each assembly.

One can imagine situations in which the assembler would always have an ample supply of input items. Typically, however, input items of each type are supplied intermittently through time, and each of the input processes involves some stochastic variability. Similarly, assembly times are usually random variables. Thus inventories of the various types of input items accumulate at the assembly station, and the sizes of these inventories fluctuate randomly as time passes. If any of the inventory levels fall to zero, then the server becomes idle until the shortage is rectified. A system of this type will be referred to hereafter as an "assembly-like" operation, since there exist physical systems which do not literally involve assembly and yet have the same essential nature. For example, it might be that two different types of customers arrive separately at a service facility but can be served only in pairs, each pair consisting of one customer of each type.

Consider now an industrial complex in which input items are converted into finished products by passing through several stages of sub-assembly plus a final assembly. An example of such an assembly network, consisting of several input processes plus several single-server assembly stations, is schematically represented in Figure 2 below. This particular system has six input processes, two sub-assembly stations and one final assembly station. Input items of types 1 and 2 are converted into type 1 sub-assemblies by the server at station 1, and
input items of types 4, 5 and 6 are converted into type 2 sub-assemblies by the server at station 2. Input items of type 3 and sub-assemblies of types 1 and 2 are then converted into finished products by the server at station 3. Assembly networks occurring in the real world are typically much larger than the system portrayed in Figure 2. The essential feature of gradual consolidation through sub-assembly is well demonstrated by this simple example, however.

The input processes to an assembly network normally involve some stochastic variability, and assembly times at the various stations of the system are usually random variables. Thus the phenomenon of randomly fluctuating inventory levels again arises. If any of the inventory levels at a given station within the network falls to zero, then the assembler there becomes idle until the shortage is rectified. Changes in inventory levels at the various stations of the system are of course closely interconnected.

Henceforth, we shall speak of "assembly-like" networks in order
to emphasize the fact that there exist complex systems which may not literally involve assembly of tangible items and yet are essentially equivalent to an assembly network like that in Figure 2. In fact, there are many important examples of such systems. We shall pursue this point further in the next section after some additional terminology has been introduced.

1.2 The Queueing Models Employed

In Part I of this dissertation a queueing theoretic model of an assembly-like operation will be formulated and analyzed, our specific assumptions being as follows.

(i) Input items of \( K \) different types arrive individually from outside the system. The times between consecutive arrivals of type \( k \) items \( (k = 1,...,K) \) are independent and identically distributed (i.i.d) random variables. Thus the input process for each type of item is a renewal process, and we further assume that these \( K \) renewal processes are mutually independent of one another.

(ii) Without loss of generality it is assumed that exactly one input item of each type is required to make a finished product. The apparently more general situation where \( n_k \) items of type \( k \) are required for a finished product can be reduced to this standard case by taking the "time between consecutive arrivals of type \( k \) items" to be an \( n_k \)-fold convolution.

(iii) Having completed one assembly, the server (assembler) immediately begins another if one input item of each type is available. Otherwise he stops work completely (becomes idle) and waits until the required
inputs become available.

(iv) Service times (assembly times) are i.i.d. random variables and also independent of the input processes.

(v) The server uses input items of each type in the order of their arrival, and no bound is placed upon the number of input items of any type which can accumulate over time.

Readers familiar with queueing theory will recognize that this model is a multiple input generalization of the much studied GI/G/1 queueing system. Its salient feature is a very special kind of batch servicing, each batch containing exactly one customer of each type \( k = 1, \ldots, K \). Since assembly operations abound in the real world, the model seems a natural one, and yet it apparently has not been studied before.

Since the great majority of physical systems described by our model do not involve human customers, the terminology of "items" will be maintained throughout. In conformity with standard queueing theoretic usage, however, we shall speak of the "waiting time" of a type \( k \) item rather than its time in inventory, of the "queue" of type \( k \) items rather than the inventory or stock of them, and of a "server" rather than an assembler.

In Part II a more general queueing model of an assembly-like model is used for a single-server queueing system with a renewal input process, i.i.d. service times and first-come-first-served queue discipline.
network is formulated and analyzed. In modeling network systems we shall not speak explicitly of input processes. It is convenient (and clearly equivalent) to assume that the system is composed entirely of \( J \) stations, some of which serve only to generate input items and the remainder of which are assembly-like operations. Those stations which serve only to generate input items will be called external stations, and it is convenient to assume that they are numbered \( 1, \ldots, J_E \) while the remaining stations are numbered \( J_E + 1, \ldots, J \). With these conventions the network of Figure 2 can be alternately represented as in Figure 3 below.

![Diagram of network]

Finished Products

Figure 3. The assembly-like network of Figure 2 with external stations added.

In discussing the network model it is important to give precise meaning to the term "item". For each \( j = 1, \ldots, J \) we use the term "type \( j \) item" to mean an entity created by server \( j \) at the completion of a service. The assumptions of our network model are then as follows.
(i) Service times at each station in the network are i.i.d. and service times at the various stations are mutually independent.

(ii) The network contains a single terminal station (or sink) such that the outputs from all other stations eventually funnel into it as input. It is convenient to assume that the terminal station is numbered \( J \). Type \( J \) items represent finished products and depart the system immediately upon their creation.

(iii) For each station \( j = 1, \ldots, J-1 \) there is a unique successor \( S(j) \). Type \( j \) items proceed immediately to station \( S(j) \) upon their creation. For each \( i = J_k + 1, \ldots, J \) we define the set of predecessors of station \( i \), \( P(i) = \{ j : j = S(j) \} \). The service discipline at each such station \( i \) is that of an assembly-like operation. In particular server \( i \) needs one item of each type \( j \in P(i) \) for each of his services, and he uses items of each type in the order of their arrival.

(iv) Servers at external stations are never idle.

(v) The network is acyclic in the sense that the output from a station cannot eventually return to act as input to the same station.

It was mentioned above that our network model is applicable to many important types of systems which may not literally involve assembly. To demonstrate this fact, note that an "item" of any given type need not actually be tangible. Rather it can be viewed as a completed prerequisite or a clearance. With this interpretation, an assembly-like

\[ */ \text{Note that for network systems this assumption does entail a loss of generality.} \]
network can be more broadly described as follows. The system is composed of J "workers", each of whom has an assigned task which he is required to perform repeatedly. Some of these workers, however, are constrained in the performance of their assigned tasks by a set of precedence requirements. In particular, if \( J_E < i \leq J \) then each worker \( j \in P(i) \) must complete the \( n^{th} \) performance of his task before worker \( i \) can begin the \( n^{th} \) performance of his task. The famous Program Evaluation and Review Technique (PERT) has been devised as a means of analyzing precedence networks of this type. (The reader unfamiliar with PERT is referred to Hillier and Lieberman (1967), pp. 208-235, for an introductory description and further references.) PERT is concerned with the performance of a precedence network type of system when each worker is required to perform his task just once. As has already been indicated, we shall be concerned with the problems associated with repeated passage through a precedence network.

The relationship between our network model and that studied by PERT will be discussed further in Chapter 11. For the time being it will be easier to think in terms of applications involving actual assembly. Hereafter we shall use the term "single-station system" to mean an assembly-like operation, and in a corresponding fashion we shall speak of "the single-station model". Thus a single-station system is actually an assembly-like network which has only one non-external station.

1.3 Type of Results Obtained

The basic stochastic processes associated with the single-station model are as follows. Suppose for the sake of concreteness that the
input processes are turned on at time zero, the server being initially idle and having no stocks of any items. Let \( T_n \) be the time at which the server completes his \( n^{th} \) service \( (n = 1, 2, \ldots) \), and let \( w_n^k \) be the waiting time (exclusive of service time) of the \( n^{th} \) arriving item of type \( k \) \( (k = 1, \ldots, K; \ n = 1, 2, \ldots) \). Although these are not the only processes that we shall study (the others being introduced shortly), they do constitute the center of attention. The completion time process \( \{T_n\} \) is of fundamental interest because of the information it contains about the output of the system. The waiting time processes \( \{w_n^k\} \) are basic indicators of the amount of "congestion" arising within the system due to the stochastic variability of interarrival and service times.

Our primary goal in Part I is to develop limit theorems for \( T_n \) and \( w_n^k \) as \( n \to \infty \) in the single-station model. Actually, rather than treating the individual processes \( \{w_n^k\} \), we shall more generally investigate the vector process \( \{w_n\} \), where

\[
\begin{pmatrix}
  w_n^1 \\
  \vdots \\
  w_n^k \\
  \vdots \\
  w_n^K
\end{pmatrix}, \quad n = 1, 2, \ldots
\]

Let \( u_k^k \) be a generic notation for a time between consecutive arrivals of type \( k \) items \( (k = 1, \ldots, K) \), and let \( v \) be a generic notation for a service time. It is obvious that the relative magnitude of the parameters

\[
a_k = E(u_k^k), \quad k = 1, \ldots, K
\]
and

\[ b = E(v) \]

is critical to the limiting behavior of the waiting time process \( \{w_n\} \).

In order to facilitate a discussion of asymptotic behavior, let us digress for a moment to agree upon some terminology. We shall speak of \( X \) as being a random variable (r.v.) only if \( X \) is finite almost everywhere, and we shall speak of \( F(\cdot) \) as being a distribution function (d.f.) only if \( F(-\infty) = 0 \) and \( F(+\infty) = 1 \). Thus we do not speak of "defective" or "extended" random variables. We reserve the terminology of convergence in distribution to mean convergence in distribution to another random variable, making it redundant to say that \( X_n \) converges in distribution to a non-defective limit. Let \( \{X_n\} \) be r.v.'s with distribution functions \( \{F_n\} \). If \( F_n(x) \to F(x) \) as \( n \to \infty \) at all continuity points \( x \) of \( F(\cdot) \), and if \( 0 < F(+\infty) < 1 \), then we shall say that \( X_n \) is asymptotically defective, meaning that some but not all of the probability mass associated with \( X_n \) escapes to \( +\infty \) as \( n \to \infty \). If \( F(+\infty) = 0 \) or equivalently \( X_n \quad \text{P} \to +\infty \), meaning that all of the probability mass escapes to \( +\infty \), then we shall simply say that \( X_n \) blows up. Also, we take random vectors to have all components finite almost everywhere and reserve the terminology of convergence in distribution for such vectors to mean convergence in distribution to another random vector.

To motivate a basic qualitative result for assembly-like systems, note that if \( K = 1 \) then our single-station model reduces to an ordinary GI/G/1 queue. For such a system it is well known that \( w_n \)
converges in distribution if \( \rho \equiv \frac{b}{a_1} < 1 \), and \( w_n \) blows up if \( \rho \geq 1 \).

The case of \( w_n \) asymptotically defective cannot occur. For our single-station model the critical system parameter is found to be

\[
\rho = \frac{b}{\max_{1 \leq k \leq K} a_k} = \min_{1 \leq k \leq K} \left( \frac{b}{a_k} \right),
\]

which will be called the system traffic intensity (this being the name used for the corresponding quantity in a GI/G/1 queue). If \( \rho \geq 1 \), we find that \( w_n^k \) blows up for all \( k = 1, \ldots, K \) as \( n \to \infty \), although the asymptotic behavior of the system is much more subtle when \( \rho = 1 \) than when \( \rho > 1 \). Moreover, even when \( \rho < 1 \), it is apparent that if \( a_1 > a_k \) then \( w_n^k \) must blow up as \( n \to \infty \), since input items of type \( k \) arrive at a mean rate which is strictly greater than that for type 1 items. Thus, assuming \( K > 1 \), there is only one case in which there exists any hope that the system will stabilize in the sense of \( w_n \) converging in distribution. That is the situation where \( b < a_1 = \cdots = a_K \).

In this case, however, we find that \( w_n^k \) is asymptotically defective for all \( k \). To see why, suppose that \( K = 2 \), and \( b < a_1 = a_2 \). Let \( \{u_1^n\} \) and \( \{u_2^n\} \) denote the sequences of interarrival times for type 1 items and type 2 items respectively. Then the \( n^{th} \) item of type \( k \) arrives at time \( \sum_{i=1}^{n} u_i^k (k = 1, 2) \). Letting \( x^+ \) denote the positive part of \( x \) as usual, it is then clear that

\[ (1.1) \quad w_n^1 \geq \left[ \sum_{i=1}^{n} u_i^2 - u_i^1 \right]^+. \]

That is, the time that the \( n^{th} \) item of type 1 waits is at least as
large as the time (if any) that elapses between its arrival and that of the \( n \)th type 2 item. Since \( \{w_n^1 - w_n^2\} \) is a sequence of i.i.d. random variables, it follows from the Central Limit Theorem that the right-hand side of (1.1) is asymptotically defective, exactly half of the associated probability mass escaping to \( +\infty \) as \( n \to \infty \). Thus \( w_n^1 \) is at best asymptotically defective, and the same statement holds for \( w_n^2 \) by symmetry. Another qualitative result is that no more than one of the \( K \) waiting time sequences \( \{w_n^k\} \) can converge in distribution when \( \rho < 1 \), and if one of them does, then the rest blow up.

Having determined that the waiting time vectors \( w_n \) do not converge in distribution, we seek to characterize the way in which this important process blows up. The characterizations obtained are comparable to those developed by various different authors for the GI/G/1 queueing system in the "heavy traffic" situation where \( \rho \geq 1 \). In recent years considerable study has been devoted to queues in heavy traffic, a central problem being to find a sequence \( \{c_n\} \) of translation constants, a sequence \( \{d_n\} \) of scaling constants, and a non-degenerate r.v. \( X \) such that

\[
\frac{w_n - c_n}{d_n} \xrightarrow{D} X \quad \text{as} \quad n \to \infty.
\]

The normalizing constants \( \{c_n\} \) and \( \{d_n\} \) show the rate at which \( w_n \) blows up. In all cases studied thus far (involving either GI/G/1 queues or more complex types of systems) it has developed that the appropriate normalization is of the form \( c_n = cn, \ d_n = d\sqrt{n} \), \( c \) and \( d \) being fixed constants with \( d > 0 \). Thus one obtains
(1.2) \[ \frac{1}{d \sqrt{n}} (w_n - cn) \overset{D}{\rightarrow} X \quad \text{as} \quad n \to \infty. \]

For all values of \( \rho \) (not just \( \rho > 1 \)), the limit theorems developed in Part I for the waiting time process \( \{w_n\} \) are also of the form (1.2). In our model, however, \( c \) must be a constant vector, since \( \{w_n\} \) is a vector process. Actually, we shall obtain much more general "functional" limit theorems for the vector process \( \{w_n\} \), from which results like (1.2) follow as corollaries. Functional limit theorems for various stochastic processes associated with GI/G/1 queues in heavy traffic were first obtained by Whitt (1968) and later extended to much more complex systems in heavy traffic by Iglehart and Whitt (1969 a,c). The reader is also referred to Whitt (1968) for a comprehensive survey of heavy traffic results for GI/G/1 queueing systems. The general subject of functional limit theorems is discussed in Section 2.2 below. For the time being, we simply remark that they contain considerably more information than ordinary limit theorems like (1.2). In particular, as immediate corollaries to the functional limit theorems developed below for the process \( \{w_n\} \), we obtain limit theorems for the accumulated waiting time vectors

\[ W_n = \sum_{i=1}^{n} w_i, \quad n = 1, 2, \ldots \]

associated with the single-station model. Since the \( k \text{th} \) component of \( W_n \) is the total waiting time for the first \( n \) items of type \( k \), it is more or less obvious that \( W_n \) cannot converge in distribution as \( n \to \infty \) without normalization. By the very nature of the completion
time process \( \{T_n\} \), it is clear that this remark applies to it as well. Thus all of our basic limit theorems are for normalized processes. We emphasize that this is true for the \( \{w_n\} \) process not because attention is restricted to any special "heavy traffic" cases but rather because in an assembly-like system any traffic amounts to heavy traffic.

To summarize, the most important results obtained in Part I are as follows. For each possible combination of the system parameters \( s_1, \ldots, s_K \) and \( b \), functional limit theorems are presented for both the completion time process \( \{T_n\} \) and the vector waiting time process \( \{w_n\} \). As corollaries to these basic results we are able to show convergence in distribution for properly normalized versions of \( T_n \) and \( w_n \). As a corollary to the latter result we are also able to show convergence in distribution for a properly normalized version of \( W_n \). All of these limit theorems are distribution free, requiring only the assumption of independent renewal input processes and i.i.d. service times.

Despite the rather discouraging behavior of the processes \( \{w_n\} \), there is associated with the single-station model another waiting time process which exhibits a more conventional but very intriguing asymptotic behavior when \( \rho < 1 \). Let the \( K \)-tuple consisting of the \( n^{th} \) arriving item of each type \( k = 1, \ldots, K \) be called the \( n^{th} \) "unit". Assuming that no items of any type are initially present, it is this group of items on which the server operates in his \( n^{th} \) service. Let \( w^*_n \) denote the waiting time for the \( n^{th} \) unit. Equivalently stated, \( w^*_n \) is the time that expires between the instant at which the last of those items comprising the \( n^{th} \) unit arrives and the instant at which the \( n^{th} \)
service begins. Clearly then

\[ w^*_n = \min_{1 \leq k \leq K} w^n_k, \quad n = 1, 2, \ldots \]

The sequence \( \{w^*_n\} \) is of some interest in itself, since it represents
the waiting time process in a single-server queuing system with i.i.d.
service times and an input process which is the minimum of \( K \) renewal
processes. It will be shown in Part I that \( w^*_n \) converges in distribu-
tion without normalization if and only if \( \rho < 1 \). Moreover, the
limiting distribution can be expressed in terms of the limiting waiting
time distributions in certain GI/G/1 queuing systems.

In Part II we identify the following stochastic processes associated
with the general network model. Let \( T^j_n \) denote the time at which server
\( j \) (i.e., the server at station \( j \)) completes his \( n \)th service
\( (j = 1, \ldots, J; n = 1, 2, \ldots) \). Let \( w^j_n \) denote the waiting time (exclusive
of service time) at station \( S(j) \) of the \( n \)th arriving type \( j \) item
\( (j = 1, \ldots, J-1; n = 1, 2, \ldots) \). (Actually the phrase "at station \( S(j) \"
is superfluous, since type \( j \) items are no longer considered to exist
after they have undergone service at station \( S(j) \).) It is not meaning-
ful to speak of waiting times for type \( J \) items, since they are assumed
to depart the system immediately upon their creation. The significance
of the completion time processes \( \{T^j_n\} \) is apparent for both precedence
network applications and actual assembly systems. In the former case,
however, the waiting times \( \{w^j_n\} \) will generally have a somewhat
unconventional interpretation. We define the vector processes
\[
T_n = \begin{pmatrix}
\eta_n^1 \\
\vdots \\
\eta_n^J
\end{pmatrix}, \quad w_n = \begin{pmatrix}
w_n^1 \\
\vdots \\
w_n^{J-1}
\end{pmatrix}, \quad n = 1, 2, \ldots .
\]

Assuming that the assembly-like network does not simply reduce to a set of stations arranged in series, it is shown that the waiting time vectors \( w_n \) do not converge in distribution, regardless of the values of the mean service times \( b_1, \ldots, b_J \). At least one of the component random variables \( w_n^j \) is asymptotically defective.

The principle results obtained in Part II are functional limit theorems for the vector processes \( \{T_n\} \) and \( \{w_n\} \). Such results are developed for all possible values of the mean service times \( b_1, \ldots, b_K \) without making any specific distributional assumptions. From the functional limit theorem for \( \{T_n\} \) we are able to show convergence in distribution for \( T_n \) properly normalized. As corollaries to the functional limit theorem for \( \{w_n\} \) we are able to show convergence in distribution both for a properly normalized version of \( w_n \) and for a properly normalized version of the accumulated waiting time vector

\[
W_n = \sum_{i=1}^n w_i.
\]

Thus far, in discussing both the single-station model and its general network counterpart, we have mentioned only discrete-parameter stochastic processes. For the single-station model, the following continuous-parameter processes are also of obvious interest. Assuming the system to be initially empty, let \( D(t) \) denote the number of services that are completed by time \( t \) \((0 \leq t < \infty)\). Alternatively stated, \( D(t) \) is the number of finished products that depart the system in the interval
Let \( Q^k(t) \) denote the number of type \( k \) items present in the system at time \( t \) \( (1 \leq k \leq K; \ 0 \leq t < \infty) \). Letting

\[
Q(t) = \begin{bmatrix}
Q^1(t) \\
\vdots \\
Q^K(t)
\end{bmatrix}, \quad 0 \leq t < \infty,
\]

we shall refer to \( \{D(t), \ 0 \leq t < \infty\} \) as the departure process and to \( \{Q(t), \ 0 \leq t < \infty\} \) as the queue length process. Similarly, for the general network model we define \( D^j(t) \) to be the number of services completed by server \( j \) up to time \( t \) \( (j = 1, \ldots, J; \ 0 \leq t < \infty) \) and \( Q^j(t) \) to be the number of type \( j \) items present in the system at time \( t \) \( (j = 1, \ldots, J-1; \ 0 \leq t < \infty) \). We then define the vector processes

\[
D(t) = \begin{bmatrix}
D^1(t) \\
\vdots \\
D^J(t)
\end{bmatrix}, \quad Q(t) = \begin{bmatrix}
Q^1(t) \\
\vdots \\
Q^{J-1}(t)
\end{bmatrix}, \quad 0 \leq t < \infty.
\]

Before discussing the type of results obtained for \( D(t) \) and \( Q(t) \), it is appropriate to make some general remarks concerning the overall content and organization of Parts I and II.

In Part II an elaborate notational system is developed which allows a very compact treatment of assembly-like networks. In fact, the treatment is so compact that Part II actually contains only one theorem, a functional result for the vector process \( \{T_n\} \) which is valid for any given set of mean service times \( b_1, \ldots, b_J \). From it, all of the other results are obtained essentially as corollaries. Using this method of attack, it is easily shown that the asymptotic behavior of the
continuous-parameter processes $D(t)$ and $Q(t)$ is essentially
equivalent to that of their discrete-parameter counterparts $\{T_n\}$ and
$\{w_n\}$. By demonstrating the nature of their equivalence, we are able
to give an equally extensive treatment to the continuous-parameter
processes with very little additional effort.

The greatest disadvantage of the analytical approach taken in Part
II is reduced clarity and readability. Essentially simple ideas tend
to be obscured by the imposing notation, and results are stated in so
general a form that it becomes difficult to grasp their basic qualitative
implications. It is for this reason that we have chosen to first give
a complete and almost totally independent treatment of the single-
station model. Throughout Part I we have attempted to state and prove
results in the most direct way possible. The separate cases $\rho < 1$,$\rho > 1$ and $\rho = 1$ are treated in separate chapters in order to emphasize
both the qualitative and the quantitative differences in the behavior
of the system under various load conditions. Also, a number of special
cases are discussed in which particular results reduce to a signif-
ically simpler form. Despite its expository advantages, the approach
taken in Part I has some notable drawbacks, the most serious of which
concerns the processes $D(t)$ and $Q(t)$. Because the cases $\rho < 1$,$\rho > 1$ and $\rho = 1$ are treated separately, one must show three separate
times that limit theorems for the continuous-parameter processes are
essentially equivalent to those for their discrete-parameter counterparts.
In order to avoid such needless repetition, attention has been restricted
to discrete-parameter processes in the main chapters of Part I. In a
concluding chapter, the corresponding results for $D(t)$ and $Q(t)$ are
stated, and the reader is referred to Part II for a rigorous proof.

Finally, it should be mentioned that very few of the limit theorems
presented either in Part I or in Part II are stated in a completely
explicit form. For purposes of discussion, consider a statement of
ordinary convergence in distribution like (1.2). In most cases we shall
specify the limiting distribution only indirectly, by indicating, for
example, that $X$ is distributed as some function of $K$ independent
normal random variables.*/* Unless it is otherwise indicated, the
reader should assume that this distribution is not known explicitly (at
least by the author). Of course the fact that such distributions have
not appeared as yet in the literature of probability theory does not
mean that they cannot be found, as will be discussed under "suggestions
for future research".

1.4 Implications for Real-World Systems

In this section we shall briefly indicate how the limit theorems
presented in Parts I and II can serve as a basis for decision making
in the design and operation of real-world systems. We use the term
"assembly-like system" to refer both to single-station systems and to
more general network structures.

Suppose that an assembly-like system is constructed which essentially
satisfies the assumptions of our model, and suppose further that the

*/* Similarly, the diffusion processes obtained as weak limits in our
functional results are typically shown only to have the same distribution
as some function of multi-dimensional Brownian Motion.
designers of the system intend for it to operate continuously over an indefinite period of time. From our results it follows that the in-process inventories (or queues) of at least some types of items generated within the system will grow without bound as time passes. From a purely technical viewpoint, the explanation for this behavior can be found in the various assumptions of stochastic independence which are so critical to the model. In real-world systems which are operated on a continuing basis, a feedback phenomenon occurs, even if unintentionally, causing workers to speed up, slow down or simply stop their activities in response to certain conditions. Thus the development of excessive in-process inventories is avoided by introducing dependence among the activities of the system's various "servers". Rather than allowing an informal "control policy" of this type to arise through consensus, the decision-maker must recognize the inherent instability of an assembly-like configuration and build into the system some efficient mechanism for checking the development of excessive congestion. This can be accomplished either through an alteration in the original design (e.g., restricting some or all queue lengths) or through the adoption of some dynamic control policy. The limit theorems developed in Parts I and II are of little use in determining how a continuing system can be optimally designed or controlled. Part III, however, contains an introductory discussion of such problems.

As the preceding paragraph indicates, one must conclude that our basic models are not realistic for continuing systems. A great deal of manufacturing, however, is done on the basis of production runs, and in this situation our assumptions of independence are likely to be
quite realistic. If a collection of manufacturing and assembly operations are joined together for the production of some relatively large but fixed number $n$ of finished products, then there exists no possibility of in-process inventory levels actually growing without bound. It is unlikely then that the various operations comprising the system will be closely coordinated. Imagine now that for some given run size $n$ a decision-maker can choose among several possible system designs. These various designs may have different numbers of stations, different mean service times at corresponding stations, different service time distributions, etc. Assuming $n$ to be relatively large, a primary application of our limit theorems is to provide characterizations of the system's behavior under each possible design, thus providing a basis for decision making. These characterizations are of several different types. First is the revelation of which important stochastic processes tend to blow up as $n$ becomes large and the rate at which they do so (as revealed by normalization constants). Also revealed are the system parameters (mean service times, variances, etc.) to which critical processes are sensitive. It will be shown, for example, that the development of congestion at a given station within a complex network is typically unaffected by the removal of all but a few of the other stations which precede it. Finally, of course, limit theorems like (1.2) can be used as approximation theorems for large but finite values of $n$. In a production run situation, our limit theorems give approximations for the probability of completing the run in a specified length of time, the expected total cost of carrying in-process inventory under a linear cost structure, etc. A specific problem of choosing
among several hypothetical system designs is discussed for purposes of illustration in Section 7.4.

A closely related problem is that of choosing among several possible system designs when production continues indefinitely through time but is done in a series of independent production runs of identical size. In this case one must determine an optimal run size under each possible design and then choose among the various designs on the basis of minimal expected cost per finished product. The usefulness of our limit theorems in attacking problems of this type is also illustrated in Section 7.4.

Our results are also useful in the analysis of problems which are not as precisely structured as those discussed above. Suppose, for example, that finished products are manufactured in a sequence of essentially independent production runs of approximately the same size, with all inventory levels falling nearly to zero before the beginning of the next run. After the obvious simplifying assumptions are made, our limit theorems can still be used to compute approximate measures of system performance. These approximations of course become increasingly crude as more and stronger simplifying assumptions are introduced.

1.5 Overview

Very briefly, the sequencing and content of succeeding chapters is as follows. Chapter 2 collects some well known results from probability theory and the theory of queues which will be needed later. Of particular importance is the discussion of weak convergence and functional limit theorems in Section 2.2.
Part I (Chapters 3-7) deals with the single-station model. In Chapter 3 the assumptions of the model are formally stated, notation is fixed, and some important preparatory lemmas are proved. Limit theorems for \( \{w_n^*\}, \{w_n\}, \{T_n\} \) and \( \{W_n\} \) are then developed for the three separate cases \( \rho < 1, \rho > 1 \) and \( \rho = 1 \) in Chapters 4, 5 and 6 respectively. In addition to some examples and miscellaneous remarks, Chapter 7 contains a complete treatment of the processes \( \{D(t)\} \) and \( \{Q(t)\} \).

Part II (Chapters 8-11) deals with the general network model. In order to lay some necessary groundwork for the study of assembly-like networks, Chapter 8 develops some limit theorems for single-server queues in series. The notation and terminology for assembly-like networks are then presented in Chapter 9 along with some important preparatory lemmas. In Chapter 10, limit theorems for \( \{T_n\}, \{v_n\}, \{W_n\}, \{D(t)\}, \) and \( \{Q(t)\} \) in a general assembly-like network are stated and proved. Since these theorem statements are quite complex, the first two sections of Chapter 11 are devoted to showing how they simplify in some special cases of interest. The final section of Chapter 11 contains some additional remarks and discussion regarding the general network model.

Part III (Chapters 12 and 13) is concerned with other types of problems associated with assembly-like systems which might prove fruitful areas for future research. Chapter 12 contains some fragmentary results, but Chapter 13 is purely expository.
CHAPTER 2

Preliminaries

The basic purpose of this chapter is to collect some well known results from probability theory and the theory of queues which will be needed later. We begin by compiling some general notation and terminology which will be used throughout, including one important piece of non-standard notation.

2.1 General Notation and Terminology

The symbols 0 and 1 will be used to denote a vector all of whose components are zero and a vector all of whose components are one respectively. Otherwise, we shall rely upon context rather than any special notation to differentiate between vector and scalar quantities. Components of a vector will generally be indexed by superscripts. Exponents are used so infrequently that no confusion should result from this practice. We write \( b = (b^k) \) to mean that \( b \) is a vector with components \( b^k \). Similarly, \( A = (a_{ij}) \) means that \( A \) is a matrix whose elements are denoted by \( a_{ij} \). All vectors encountered should be interpreted as column vectors.

In the usual way, we define the "positive part" operator \((\cdot)^+\) by

\[
x^+ = \begin{cases} 
  x & \text{if } x > 0 \\
  0 & \text{if } x \leq 0 
\end{cases}
\]
Also as usual, we write \( f(t) = o(g(t)) \) to mean that \( \frac{f(t)}{g(t)} \to 0 \) as \( t \to \infty \), and we write \( f(t) \sim g(t) \) to mean that \( \frac{f(t)}{g(t)} \to 1 \) as \( t \to \infty \).

The symbol \( I \) is used only to denote the identity matrix. The transpose of a matrix \( A \) is denoted by \( A^T \). If a square matrix \( \Sigma \) is positive semi-definite, then we write \( \Sigma^{1/2} \) to denote the square root of \( \Sigma \), i.e., the square matrix satisfying \( (\Sigma^{1/2})^T \Sigma^{1/2} = \Sigma \).

We write \( X \overset{D}{=} Y \) to mean that the random variables \( X \) and \( Y \) have the same distribution (are equivalent in law).

The real line is denoted by \( \mathbb{R} \), and \( K \)-dimensional Euclidean space is denoted by \( \mathbb{R}^K \). The non-negative half-line \([0, \infty)\) is denoted by \( \mathbb{R}^+ \).

Following the notation of Anderson (1958), p. 26, we use the symbol \( \mathcal{N}(\mu, \Sigma) \) to denote the multivariate normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \). In this notation \( \Sigma \) is restricted to be square, symmetric and positive semi-definite, but it need not be positive definite. The case of \( \Sigma \) not positive definite includes the possibility that one or more components of the random vector have zero variance and hence are equal almost everywhere to a constant. We also use the symbol \( \mathcal{N}(\mu, \Sigma) \) to denote a random vector with the specified distribution. We shall use the fact that if \( A \) is a matrix of appropriate dimension then (cf. Anderson (1958), p. 26)

\[
\mathcal{N}(\mu, \Sigma) \overset{D}{=} \mathcal{N}(A\mu, A\Sigma A^T).
\]

Thus in particular
\[ N(0, I)^{1/2} \] 

The notation \( \{\xi(t), 0 \leq t \leq 1\} \), or more simply \( \xi \) or \( \xi(\cdot) \), will be used to denote standard Brownian Motion over the unit interval. More generally, \( \xi^K \) denotes standard K-dimensional Brownian Motion over the unit interval. By "standard" K-dimensional Brownian Motion we mean that each component process has zero drift and unit variance, the component processes being mutually independent. Readers unfamiliar with Brownian Motion (or the Wiener process) are referred to Breiman (1968), Chapter 12.

In each part of this dissertation we shall identify a number of different stochastic processes. Each process, however, is generated from a basic countable collection of mutually independent random variables, and thus it is immediate that there exists a probability space \((\Omega, \mathcal{F}, P)\) on which all the processes are defined. Henceforth the underlying probability space will not be mentioned explicitly. Despite this fact, the symbolism \( P[\cdot] \) will be used to denote the probability of the event in braces, and we trust that the meaning of this notation will be clear from context.

Except possibly for the symbol \( D \), all of the notation discussed thus far conforms with standard usage. We shall use one other general notation, however, which is definitely not standard. This is the notion of a sequence of random variables which are "delta of \( n^p \)."

**Definition:** If \( \{X_n\} \) is a sequence of r.v.'s and \( p > 0 \), we shall write
\[ X_n = \Delta(n^p) \]

to mean that

\[ \frac{1}{n^p} \left( \max_{1 \leq i \leq n} |X_i| \right)_p \to 0 \quad \text{as} \quad n \to \infty. \]

The notation \( Y_n = Z_n + \Delta(n^p) \) is understood to mean that \( (Y_n - Z_n) = \Delta(n^p) \).

This concept is roughly similar to the "little o" notation for sequences of real numbers, its intuitive content being that if \( X_n = \Delta(n^p) \), then the entire vector \((X_1, \ldots, X_n)\) is insignificant in comparison with \((n^p, \ldots, n^p)\) for large values of \( n \). Just why this notation is so convenient will become clear in the next section. We shall use without comment the obvious facts that if \( \{X_n\}, \{X_n'\}, \{Y_n\} \) and \( \{Y_n'\} \) are r.v.'s such that \( X_n = X_n' + \Delta(n^p) \) and \( Y_n = Y_n' + \Delta(n^p) \) then

\[ X_n + Y_n = X_n' + Y_n' + \Delta(n^p) \]

\[ \max(X_n, Y_n) = \max(X_n', Y_n') + \Delta(n^p) \]

\[ \max_{1 \leq i \leq n} (X_i) = \max_{1 \leq i \leq n} (X_i') + \Delta(n^p) \]

2.2 Weak Convergence of Random Functions

The classical Central Limit Theorem (C.L.T.) is concerned with the asymptotic distribution of the partial sums \( S_n = X_1 + \cdots + X_n \) \((S \equiv 0)\), where \( \{X_1\} \) is a sequence of random variables. In particular, if the
$X_i$ are i.i.d. with mean zero and variance $\sigma^2$, then the Lindeberg-Lévy form of the C.L.T. states that

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{D} N(0, \sigma^2) \quad \text{as } n \to \infty .$$

A much more general result can be obtained, however, by considering the following sequence of random functions generated by $\{S_n\}$. Let

$$(2.1) \quad \nu_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots ,$$

where square brackets denote the integer part operator. */ For each $n = 1, 2, \ldots$ $\nu_n(t)$ is a piecewise constant, right-continuous function on [0,1] with jumps only at values of $t$ which are integer multiples of $\frac{1}{n}$. A typical sample path for a function $\nu_n(\cdot)$ is shown in Figure 3 below for $n = 8$.

$$\nu_8(t) = \frac{1}{\sqrt{8}} S_{[8t]}$$

Figure 3. A typical sample path of $\nu_8(t)$.

*/ That is, $[nt]$ is the largest integer less than or equal to $nt$. Hereafter, square brackets have this meaning if and only if they appear in a subscript.
It is useful to think of \( S_i \) as the position of a random walk after \( i \) steps. The graph of \( \nu_n(\cdot) \) is obtained by plotting the first \( n \) steps of the random walk at integer multiples of \( 1/n \), extending the set of \( n+1 \) points (including the origin) to a function on \([0,1]\) in a piecewise constant, right-continuous fashion, and then adjusting the vertical scale by a factor of \( \frac{1}{\sqrt{n}} \). Donsker's Theorem, the key result presented in this section establishes that the sample path of the random walk, as encapsulated in the random functions \( \nu_n(\cdot) \), converges "in distribution" to that of Brownian Motion as the number of steps becomes large.

From this basic theorem the Lindeberg-Lévy C.L.T. and a multitude of other useful results follow as easy corollaries. In order to make a rigorous statement of Donsker's theorem, of course, one needs a generalized definition of convergence in distribution. Toward that end, we shall compile a few basic results from the theory of weak convergence of probability measures on a metric space. Billingsley (1968) has given an excellent account of this general theory, and the reader is referred to his book for a thorough discussion of the definitions and results outlined here.

Let \((M, \rho)\) be a metric space and let \( \mathcal{F} \) be the \( \sigma \)-field generated by the open sets of \( M \) (i.e., \( \mathcal{F} \) is the class of Borel subsets of \( M \)).

**Definition:** If \( \{P_n\} \) and \( P \) are probability measures on \( \mathcal{F} \) which satisfy

\[
\lim_{n \to \infty} \int_M f \, dP_n = \int_M f \, dP
\]

for all bounded, continuous functions \( f : M \to \mathbb{R} \), then we say that \( P_n \)
converges weakly to \( P \) as \( n \to \infty \) and write \( P_n \Rightarrow P \).

Rather than dealing explicitly with sequences of measures on \( \mathcal{F} \), it is usually more convenient to speak of sequences of random elements of \( M \). Thus, for ease of exposition, we establish the following definition.

**Definition:** If \( \{X_n\} \) and \( X \) are random elements of \( M \) (i.e., measurable mappings from some probability space into \( M \)), then we say that \( X_n \) converges weakly to \( X \) as \( n \to \infty \) and write \( X_n \Rightarrow X \) if and only if \( P_n \Rightarrow P \), where \( \{P_n\} \) and \( P \) are the measures on \( \mathcal{F} \) induced by \( \{X_n\} \) and \( X \) respectively.

In the particular case where \( M \) is \( \mathbb{R}^K \) and \( \rho \) is the usual Euclidean metric, weak convergence is equivalent to the familiar notion of convergence in distribution. That is, if \( \{X_n\} \) and \( X \) are random \( K \)-vectors (random elements of \( \mathbb{R}^K \)), then \( X_n \Rightarrow X \) if and only if \( X_n \overset{D}{\to} X \).

In applying the theory of weak convergence, the following result is of fundamental importance. It is easily proved and appears in Section 5 of Billingsley (1968) along with other more general propositions of a similar nature. This result will be used many times in the following chapters, and we shall refer to it as the "Continuous Mapping Theorem" rather than by number.
Theorem 1* (Continuous Mapping Theorem): Suppose that \((M, \rho)\) and 
\((M', \rho')\) are two metric spaces and that \(\varphi : M \to M'\) is continuous (and 
therefore also measurable). If \(\{X_n\}\) and \(X\) are random elements of 
\(M\) with \(X_n \Rightarrow X\) in \((M, \rho)\), then \(\varphi(X_n) \Rightarrow \varphi(X)\) in 
\((M', \rho')\).

Of course continuity of the mapping \(\varphi\) in the theorem statement means 
continuity under the metrics \(\rho\) and \(\rho'\). In many of the applications 
that we shall encounter, \(M'\) will be \(\mathbb{R}^K\) and \(\rho'\) the Euclidean 
metric. In this case the theorem states that if \(X_n \Rightarrow X\) and 
\(\varphi : M \to \mathbb{R}^K\) is continuous, then \(\varphi(X_n) \Rightarrow \varphi(X)\).

Only one other result from the general theory of weak convergence 
will be needed. Before stating it, note that if \(\{X_n\}\) and \(\{Y_n\}\) are 
measurable mappings from some probability space into \((M, \rho)\), then 
\(\{\rho(X_n, Y_n)\}\) is just a sequence of random variables defined on the 
original space. Thus the statement that \(\rho(X_n, Y_n) \xrightarrow{P} 0\) is meaningful 
without any generalized definition of convergence in probability. For 
a proof of the following result see Billingsley (1968), p. 25.

Theorem 2: If \(\{X_n\}, \{Y_n\}\) and \(X\) are random elements of \((M, \rho)\) such 
that \(X_n \Rightarrow X\) and \(\rho(X_n, Y_n) \xrightarrow{P} 0\), then \(Y_n \Rightarrow X\).

We introduce now the function space \(D[0, 1]\) which plays the role

* A single sequential numbering system is used for lemmas, theorems 
and corollaries within each chapter. Thus the first three results of 
a given chapter may be Lemma 1, Theorem 2, and Corollary 3. References 
to results appearing in another chapter are prefaced by the referenced 
chapter's number. Thus Theorem 3.2 is Theorem 2 of Chapter 3.
of $M$ in most of the weak convergence applications encountered here.

**Definition:** Let $D[0,1]$ denote the space of all functions $x: [0,1] \to \mathbb{R}$ which are right-continuous at each point $t \in [0,1]$ and have left-hand limits at each point $t \in (0,1)$.

We define a metric $d$ on $D[0,1]$ as follows. Let $\Lambda$ denote the set of all strictly increasing continuous functions from $[0,1]$ onto itself. (If $\lambda \in \Lambda$ then necessarily $\lambda(0) = 0$ and $\lambda(1) = 1$.) If $\lambda \in \Lambda$, then we write $\lambda t$ to mean $\lambda(t)$. For $x,y \in D[0,1]$, the metric $d(x,y)$ is defined to be the infimum of those positive $\varepsilon$ such that

$$
\sup_{0 \leq t \leq 1} |\lambda t - t| \leq \varepsilon \quad \text{and} \quad \sup_{0 \leq t \leq 1} |x(t) - y(\lambda t)| \leq \varepsilon.
$$

This metric generates the so-called Skorohod topology on $D[0,1]$, and in Chapter 3 of Billingsley (1968) it is shown that $D[0,1]$ is a separable metric space under $d$. A sequence of elements $\{x_n\} \subset D[0,1]$ converges in the Skorohod topology to $x \in D[0,1]$ if and only if there exists $\{\lambda_n\} \subset \Lambda$ such that

$$(2.2) \quad \lim_{n \to \infty} x_n(\lambda_n t) = x(t), \quad \lim_{n \to \infty} \lambda_n t = t,$$

*/ In the same chapter it is shown that $D[0,1]$ is not complete under $d$ but that there exists another metric $d_0$ which also generates the Skorohod topology and under which $D[0,1]$ is complete. This fact is very useful in the theoretical development, but since both metrics generate the same topology, it is not necessary for our purposes to introduce $d_0$.  

32
both limits being uniform in \( t \).

It is immediate that the random functions \( \{v_n\} \) defined by (2.1) are random elements of \( D[0,1] \), so the basic concepts required for a statement of Donsker's Theorem are now available. Recall that the symbol \( \xi \) is used to denote standard Brownian Motion over the unit interval. Since the sample paths of Brownian Motion are almost surely continuous, \( \xi \) is a random element of \( D[0,1] \).

**Theorem 3 (Donsker's Theorem):** If \( \{X_i\} \) is an i.i.d. sequence of random variables with \( E(X_1) = 0 \) and \( 0 < E(X_1^2) = \sigma^2 < \infty \), then

\[
v_n \Rightarrow \sigma \xi \quad \text{as} \quad n \to \infty,
\]

where the random functions \( v_n(\cdot) \) are defined by (2.1). */

For a proof of this important result see Billingsley (1968), Section 16. The power of Donsker's Theorem is due largely to the wealth of results that can be obtained from it through application of the Continuous Mapping Theorem. In order to demonstrate the continuity of some mappings which will be of particular interest here, we state first a useful property of \( D[0,1] \). See Billingsley (1968), p. 110, for a proof.

**Lemma 4:** For each \( x \in D[0,1] \) and each \( \epsilon > 0 \) there exist points \( t_0, t_1, \ldots, t_n \) such that \( 0 = t_0 < \cdots < t_n = 1 \) and

*/ Henceforth all statements of weak convergence are understood to mean weak convergence in the metric space \( (D[0,1], d) \) or the corresponding product space \( (D[0,1]^K, d^K) \) to be introduced shortly.
\[
\sup_{t_{i-1} \leq s, t \leq t_i} |x(s)-x(t)| \leq \varepsilon, \quad i = 1, \ldots, n.
\]

Theorem 5:

(a) Let \( \varphi: \mathbb{I} \rightarrow \mathbb{R} \) be continuous. Then the mapping \( F: D[0,1] \rightarrow D[0,1] \)
defined by
\[
[F(x)](t) = (\varphi \circ x)(t), \quad 0 \leq t \leq 1, \quad x \in D[0,1]
\]
is continuous.

(b) If \( x \in D[0,1] \) then \( \sup_{0 \leq t \leq 1} |x(t)| < \infty \). Moreover, the mapping \( G: D[0,1] \rightarrow D[0,1] \)
defined by
\[
[G(x)](t) = \sup_{0 \leq s \leq t} x(s), \quad 0 \leq t \leq 1, \quad x \in D[0,1]
\]
is continuous.

(c) The projection \( \pi: D[0,1] \rightarrow \mathbb{I} \) defined by
\[
\pi(x) = x(1), \quad x \in D[0,1]
\]
is continuous.

(d) If \( x \in D[0,1] \) then \( x(\cdot) \) is Riemann integrable. Moreover, the
(Riemann) integration functional \( \psi: D[0,1] \rightarrow \mathbb{R} \) given by
\[
\psi(x) = \int_0^1 x(t)dt, \quad x \in D[0,1]
\]
is continuous.

Proof: (a) Suppose that \( \{X_n\} \subseteq D[0,1], \ x \in D[0,1] \), and \( x_n \overset{Sk}{\rightarrow} x \) (i.e.,
\( x_n \) converges to \( x \) in the Skorohod topology as \( n \rightarrow \infty \)). We need only
show that $F(x_n) \xrightarrow{Sk} F(x)$. Choose \( \{\lambda_n\}_{n=1}^{\infty} \) to satisfy (2.2). Since \( \phi \) is continuous and \( x_n \) is uniformly bounded (see part (d) below), (2.2) continues to hold when \( x_n \) is replaced by \( F(x_n) = \phi \circ x_n \) and \( x \) is replaced by \( F(x) = \phi \circ x \). Thus \( F(x_n) \xrightarrow{Sk} F(x) \).

(b) The boundedness of \( x(\cdot) \) is immediate from Lemma 4, so \( G \) is well defined. The continuity of \( G \) follows from an argument by Whitt (1968), p. 63, after minor modifications.

(c) Immediate from the characterization (2.2) of Skorohod convergence and the observation that \( \lambda_n(1) = 1 \) for all \( n \).

(d) The Riemann integrability of \( x(\cdot) \) follows easily from Lemma 4, the boundedness of \( x(\cdot) \), and the Cauchy criterion for integrability (cf. Bartle (1964), p. 321). In order to prove that continuity of \( \psi,^{*}/ \) suppose that \( \{x_n\} \subset D[0,1], x \in D[0,1], \) and \( x_n \xrightarrow{Sk} x \). Then \( x_n(t) \rightarrow x(t) \) at all continuity points \( t \) of \( x \) (cf. Billingsley (1968), p. 112). Moreover, \( x \) can have at most countably many discontinuities (cf. Billingsley (1968), p. 110), so \( x_n(t) \rightarrow x(t) \) for almost all \( t \in [0,1] \). If it can be shown that the functions \( x_n(\cdot) \) are uniformly bounded for sufficiently large \( n \), then it follows from the Bounded Convergence Theorem that \( \int x_n \rightarrow \int x \), which completes the proof.

To show uniform boundedness for sufficiently large \( n \), choose \( \{\lambda_n\} \subset \Lambda \) to satisfy (2.2). Since the convergence in (2.2) is uniform, there exists for each \( \epsilon > 0 \) an integer \( N_\epsilon \) such that

\[ \sum_{n=1}^{\infty} \int \left| x_n(t) - x(t) \right| \, dt < \epsilon \]

\( \xrightarrow{\epsilon} \) The author is indebted to Thomas Magnanti for suggesting the following proof.
\[ \sup_{0 \leq t \leq 1} |x_n(\lambda_n t) - x(t)| \leq \varepsilon \text{ for all } n \geq N_\varepsilon. \]

Letting \( B = \sup_{0 \leq t \leq 1} |x(t)| < \infty, \) it follows
\[ \sup_{0 \leq t \leq 1} |x_n(\lambda_n t)| \leq B + \varepsilon \text{ for all } n \geq N_\varepsilon. \] (2.3)

But for each \( n = 1, 2, \ldots \), \( \lambda_n \) maps \([0,1]\) onto itself, so
\[ \sup_{0 \leq t \leq 1} |x_n(\lambda_n t)| = \sup_{0 \leq t \leq 1} |x_n(t)| \text{ for all } n. \] (2.4)

Combining (2.3) and (2.4), we have that \( |x_n(t)| \leq B + \varepsilon \) for all \( t \in [0,1] \) and all \( n \geq N_\varepsilon \), which completes the proof.

Q.E.D.

Given the continuity of the projection \( \pi \), one obtains the Lindeberg-Lévy C.L.T. as an immediate corollary of Donsker's Theorem. Defining \( \nu_n(\cdot) \) as in (2.1) and assuming that \( \{X_i\} \) satisfies the hypotheses of Donsker's Theorem, we have from the Continuous Mapping Theorem that
\[ \pi(\nu_n) = \nu_n(1) = \frac{1}{\sqrt{n}} S_n \xrightarrow{d} \pi(\sigma_{\xi}) = \sigma_{\xi}(1). \]

Since \( \sigma_{\xi}(1) \) is normally distributed with mean zero and variance \( \sigma^2 \), the Central Limit Theorem has been obtained. To see in a concrete way that Donsker's Theorem does in fact contain more information than the ordinary C.L.T., consider the problem of finding a limit theorem for \( \max_{0 \leq i \leq n} S_i \). Define \( g:D[0,1] \to \mathbb{R} \) by letting

\[ g(S) = (S - \mathbb{E}S)^2. \]
\[ g(x) = (\pi G)(x) = \sup_{0 \leq t \leq 1} x(t), \quad x \in D[0,1] \]

Defining \( v_n(\cdot) \) as in (2.1), we have that

\[ \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} S_i = g(v_n), \quad n = 1, 2, \ldots \]

Since \( g \) is the composition of two continuous mappings, it is continuous. Assuming that \( \{X_i\} \) satisfies the hypothesis of Donsker's Theorem, we have that \( v_n \Rightarrow \sigma \xi \), and thus from the Continuous Mapping Theorem

\[ g(v_n) \overset{D}{=} g(\sigma \xi) = \sup_{0 \leq t \leq 1} [\sigma \xi(t)] = \sigma \sup_{0 \leq t \leq 1} \xi(t), \]

or

\[ (2.5) \quad \frac{1}{\sigma \sqrt{n}} \max_{0 \leq i \leq n} S_i \overset{D}{\Rightarrow} \sup_{0 \leq t \leq 1} \xi(t). \]

Since the supremum of Brownian Motion is known to have a half-normal distribution, (2.5) can be rewritten as

\[
\lim_{n \to \infty} P\left\{ \frac{1}{\sigma \sqrt{n}} \max_{0 \leq i \leq n} S_i \leq x \right\} = \begin{cases} 
0 & \text{if } x \leq 0 \\
\sqrt{\frac{2}{\pi}} \int_0^x e^{-y^2/2} dy & \text{if } x > 0
\end{cases}
\]

Billingsley has very descriptively renamed Donsker's Theorem "the functional central limit theorem", meaning that it generalizes the ordinary C.L.T. by treating the random functions generated by a sequence
of partial sums rather than just the normalized partial sums themselves. In the spirit of this usage, one might describe Donsker's Theorem as a "functional generalization" or "functional version" of the ordinary C.L.T. For all of the sequences of random variables \( \{Z_n\} \) encountered in this dissertation, we shall call

\[
\xi_n(t) = \frac{1}{\sqrt{n}} Z_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots
\]

the sequence of random functions generated by \( \{Z_n\} \), and a result of the form

\[(2.6) \quad \xi_n \Rightarrow \xi \text{ as } n \to \infty \]

will be called a functional limit theorem for the r.v.'s \( \{Z_n\} \). If

\[(2.7) \quad \frac{1}{\sqrt{n}} Z_n \overset{D}{\to} Z \text{ as } n \to \infty , \]

then necessarily \( \xi(1) \overset{D}{=} Z \), and (2.6) represents a functional generalization of (2.7).

One use of Donsker's Theorem is as an aid in computing the distribution of certain functionals of Brownian Motion (cf. Billingsley (1968), pp. 70-72). As an example, consider the problem of computing the distribution of \( \psi(\xi) = \int_0^1 \xi(t)dt \). Let \( \{X_i\} \) be a sequence of i.i.d. random variables distributed \( \text{N}(0,1) \), and define the random functions \( v_n \) as in (2.1). It is clear from the piecewise constant nature of \( v_n(\cdot) \) that
\[
\psi(n) = \int_0^1 \frac{1}{n} v_n(t) dt = \sum_{i=0}^{n-1} \frac{1}{n} \left( \frac{1}{\sqrt{n}} \right) S_i = \frac{1}{\sqrt{n^3}} \sum_{i=1}^{n-1} (n-i)X_i.
\]

Since the \( X_i \) are distributed \( N(0,1) \), this means that
\[
\psi(n) \overset{D}{\rightarrow} \frac{1}{\sqrt{n^3}} \left[ \sum_{i=1}^{n-1} (n-i)^2 \right]^{1/2} N(0,1) = \frac{1}{\sqrt{n^3}} \left[ \sum_{i=1}^{n-1} i^2 \right]^{1/2} N(0,1).
\]

Since \( \sum_{i=1}^{n} i^2 \sim \frac{1}{3} n^3 \), it follows that
\[
(2.8) \quad \psi(n) \overset{D}{\rightarrow} \sqrt{\frac{1}{3}} N(0,1) \quad \text{as} \quad n \rightarrow \infty.
\]

But \( v_n \Rightarrow \xi \) by Donsker's Theorem, and \( \psi \) is continuous, so by the Continuous Mapping Theorem
\[
(2.9) \quad \psi(v_n) \overset{D}{\rightarrow} \psi(\xi) = \int_0^1 \xi(t) dt \quad \text{as} \quad n \rightarrow \infty.
\]

Combining (2.8) and (2.9), we have proved the following result.

**Theorem 6:** \( \int_0^1 \xi(t) dt \) is normally distributed with mean zero and variance \( 1/3 \).

One ramification of Theorem 2 for the particular metric space \((\mathbb{D}[0,1], d)\) is the following very useful result.

**Theorem 7:** Let \( \{U_n\} \) and \( \{V_n\} \) be random variables and define the random functions
\[
\zeta_n(t) = \frac{1}{\sqrt{n}} U_{[nt]} , \quad 0 \leq t \leq 1 , \quad n = 1,2,\ldots
\]
\[ x_n(t) = \frac{1}{\sqrt{n}} v_{nt}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots. \]

If \( \xi_n \Rightarrow \xi \) as \( n \to \infty \) and \( V_n = U_n + \Delta(n^{1/2}) \), then \( X_n \Rightarrow \xi \) as \( n \to \infty \) as well.

\textbf{Proof}: By Theorem 2 we need only show that \( d(\xi_n, X_n) \overset{P}{\rightarrow} 0 \) as \( n \to \infty \).

Note that

\[ d(\xi_n, X_n) \leq \sup_{0 \leq t \leq 1} |\xi_n(t) - x_n(t)| = \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |u_i - v_i|. \]

Since \( V_n = U_n + \Delta(n^{1/2}) \), \( \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |u_i - v_i| \overset{P}{\rightarrow} 0 \) by the definition of our \( \Delta(n^p) \) notation.

Q.E.D.

In order to develop functional limit theorems for sequences of random vectors, we need a multi-dimensional version of Donsker's Theorem. Toward that end, we introduce the product space \( D[0,1]^K \).

\textbf{Definition}: Let \( D[0,1]^K \) denote the product of \( K \) copies of \( D[0,1] \), endowed with the product Skorohod topology. Specifically, we take the metric \( d^K \) on \( D[0,1]^K \) to be \( d^K(x,y) = \max_{1 \leq k \leq K} \{d(x^k, y^k)\} \).

Let \( \{X_k\} \) be a sequence of random \( K \)-vectors (i.e., random elements of \( \mathbb{R}^K \)), and as usual define the (vector) partial sums \( S_n = X_1 + \cdots + X_n \) \( (S_0 = 0) \). Just as a sequence of random variables generates a natural sequence of random functions in \( D[0,1] \), this sequence of random \( K \)-vectors generates the following sequence of random functions in \( D[0,1]^K \). Let
\[ v_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}, \quad 0 < t < 1, \quad n = 1, 2, \ldots \]

For each \( n = 1, 2, \ldots \), \( v_n(t) \) is a piecewise constant vector-valued function with jumps only at values of \( t \) which are integer multiples of \( 1/n \). (Figure 3 illustrates a typical sample path of one coordinate function \( v^k_n(\cdot) \) for \( n = 8 \).) The partial sum \( S_i \) may be thought of as the position after \( i \) steps of a random walk in \( K \) dimensions. The following generalization of Donsker's Theorem, proved by Iglehart (1968), shows that if the step-size vectors \( \{X_i\} \) satisfy certain conditions, then the sample path of the random walk (properly scaled) converges weakly to that of \( K \)-dimensional Brownian Motion as the number of steps becomes large. Recall that \( \xi^K \) denotes standard \( K \)-dimensional Brownian Motion over the unit interval.

Theorem 8: If \( \{X_i\} \) is an i.i.d. sequence of random \( K \)-vectors with mean 0 and covariance matrix \( \mathbb{I} \), then

\[ v_n \Rightarrow \mathbb{I}^{1/2} \xi^K \quad \text{as} \quad n \to \infty, \]

where the random functions \( v_n \) are defined by (2.10).

We shall now state multi-dimensional versions of Theorems 5, 6 and 7. These results will not be proved, since each follows easily from its one-dimensional analog. In the usual way, (Riemann) integration of vector functions is defined as component-wise integration.

Theorem 9:
(a) Let \( \varphi: \mathbb{R}^K \to \mathbb{R}^L \) be continuous. Then the mapping
\[ F: D[0,1]^K \rightarrow D[0,1]^L \] defined by
\[ [F(x)](t) = (\varphi x)(t), \quad 0 \leq t \leq 1, \quad x \in D[0,1]^K \]
is continuous.

(b) The mapping \( G = (g^1, \ldots, g^K): D[0,1]^K \rightarrow D[0,1]^K \) defined by
\[ [G^K(x)](t) = \sup_{0 \leq s \leq t} x^k(s), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K, \quad x \in D[0,1]^K \]
is continuous.

(c) The projection \( \pi: D[0,1]^K \rightarrow \mathbb{R}^K \) defined by
\[ \pi(x) = x(1), \quad x \in D[0,1]^K \]
is continuous.

(d) The (Riemann) integration functional \( \psi: D[0,1]^K \rightarrow \mathbb{R}^K \) defined by
\[ \psi(x) = \int_0^1 x(t) dt, \quad x \in D[0,1]^K \]
is continuous.

Theorem 10: \( \int_0^1 \xi^K(t) dt \) is normally distributed with mean \( 0 \) and covariance matrix \( \frac{1}{2} I \).

Theorem 11: Let \( \{U_n\} = \{(U^K_n)\} \) and \( \{V_n\} = \{(V^K_n)\} \) be random K-vectors, and define
\[ \zeta_n(t) = \frac{1}{\sqrt{n}} U_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots \]
\[ x_n(t) = \frac{1}{\sqrt{n}} V_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots \]
If \( \xi_n \Rightarrow \xi \) as \( n \to \infty \) and \( \nu_n^k = U_n^k + \Delta(n^{1/2}) \) for all \( k = 1, \ldots, K \) then \( \chi_n \Rightarrow \xi_n \) as \( n \to \infty \) as well.

Thus far we have discussed only one type of random element of \( D[0,1] \) which arises in application, the functions \( \nu_n(\cdot) \) generated by a discrete-parameter stochastic process \( \{S_n\} \). The theory of weak convergence in \( D[0,1] \) is also very useful, however, in characterizing the asymptotic behavior of certain continuous-parameter stochastic processes. For purposes of illustration, suppose that \( \{T_n\} \) is a sequence of random variables satisfying \( 0 = T_0 \leq T_1 \leq T_2 \leq \cdots \) almost everywhere. Define the corresponding counting process

\[
N(t) = \max\{n: T_n \leq t\}, \quad 0 \leq t < \infty,
\]

where \( N(t) \) is taken to be \( \infty \) if \( T_n \leq t \) for all \( n \). If \( \{X_i\} \) is an i.i.d. sequence of non-negative random variables and \( T_n = X_1 + \cdots + X_n \), then of course \( N(t) \) is a renewal process. Generalizing a result of Billingsley, Iglehart and Whitt (1969) have shown the essential equivalence of functional limit theorems for the discrete-parameter process \( \{T_n; n = 0, 1, \ldots\} \) and the associated continuous-parameter counting process \( \{N(t); 0 \leq t < \infty\} \). The result is as follows. Letting \( \mu > 0 \) be an unspecified constant, define the translated r.v.'s

\[
\hat{T}_n = T_n - n\mu, \quad n = 0, 1, \ldots
\]

and the random functions in \( D[0,1] \) generated by \( \{\hat{T}_n\} \),

\[
(2.11) \quad \nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots.
\]
Corresponding to \( \{v_n\} \) is the sequence of random functions in \( D[0,1] \) generated by \( N(\cdot) \),

\[
(2.12) \quad \zeta_n(t) = \frac{1}{\sqrt{n}} [N(nt) - nt/\mu], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots
\]

A typical sample path of \( \zeta_n(\cdot) \) is shown in Figure 4 below. The graph is obtained by taking the sample path of the translated process \( N(t) - t/\mu \) over the interval \( t \in [0,n] \), adjusting the time scale by a factor of \( \frac{1}{n} \), and adjusting the vertical scale by a factor of \( \mu^2/n \).

\[
\begin{array}{c}
\zeta_n(t) \\
\downarrow \\
0 \\
\downarrow \\
1 \\
\downarrow \\
t
\end{array}
\]

Figure 4: A typical sample path of \( \zeta_n(\cdot) \)

The result of Iglehart and Whitt is as follows.

**Theorem 12:** Suppose that \( \nu \) is a random element of \( D[0,1] \) whose sample paths are almost surely continuous. Then \( \nu_n \Rightarrow \nu \) as \( n \to \infty \) if and only if \( \zeta_n \Rightarrow -\mu^{3/2} \nu \) as \( n \to \infty \).

In the renewal case mentioned above, Donsker's Theorem shows that \( \nu_n \Rightarrow \sigma^2 \), where \( \sigma^2 = \text{Var}(X) \) and \( \mu \) is taken to be \( E(X) \). Then by Theorem 12,
\[ \zeta_n \Rightarrow -\mu^{-3/2} \sigma_x \overset{D}{=} \mu^{-3/2} \sigma_x, \]

and we have obtained a functional version of the well known central limit theorem for renewal processes (cf. Feller (1957), p. 297).

We shall now state a multi-dimensional version of the "if" part of Theorem 12. Although this proposition is somewhat ungainly in appearance, it will prove to be very useful in the analysis of assembly-like queueing systems. Let \( \{n_1, \ldots, n_K\} \) be r.v.'s satisfying \( 0 = T_0 \leq T_1 \leq \cdots \) almost everywhere \( (1 \leq k \leq K) \), and let \( \mu_1, \ldots, \mu_K \) be positive constants. Define

\[
\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix}, \quad T_n = \begin{pmatrix} T_1^n \\ \vdots \\ T_K^n \end{pmatrix}
\]

\[ \hat{T}_n = T_n - n\mu, \quad n = 0,1, \ldots \]

\[ \nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1,2, \ldots \]

Now let

\[ N^k(t) = \max\{n : T_n^k \leq t\}, \quad 0 \leq t < \infty, \quad 1 \leq k \leq K, \]

\[ \alpha = \begin{pmatrix} \mu_1^{-1} \\ \vdots \\ \mu_K^{-1} \end{pmatrix}, \quad N(t) = \begin{bmatrix} N^1(t) \\ \vdots \\ N^K(t) \end{bmatrix}, \quad 0 \leq t < \infty, \]

\[ \zeta_n(t) = \frac{1}{\sqrt{n}} [N(nt) - \alpha nt], \quad 0 \leq t \leq 1, \quad n \geq 1. \]
Note that the processes \( \{T_n^k\} \) are not assumed to be independent, nor are the constants \( \mu_1, \ldots, \mu_K \) assumed to be equal. We define the \( K \times K \) diagonal matrix

\[
A = \begin{pmatrix}
1 & & & & \mu_1 & & & \\
& 1 & & & & & & \mu_2 & \\
& & 1 & & & & & & \mu_3 \\
& & & 1 & & & & & \mu_4 \\
& & & & 1 & & \vdots & & \\
& & & & & 1 & \mu_K & & \\
& & & & & & & 1 & \\
\end{pmatrix}.
\]

We assume without loss of generality that \( \mu_k > 1 \) for all \( k = 1, \ldots, K \).
(This is just a matter of scale, of course.)

**Theorem 13:** If \( \nu_n \Rightarrow \nu \) as \( n \to \infty \), where \( \nu = (\nu^1, \ldots, \nu^K) \) is a random element of \( D[0,1]^K \) whose sample paths are almost surely continuous, the \( \xi_n \Rightarrow -A\xi \) as \( n \to \infty \), where \( \xi_K = (\xi^1, \ldots, \xi^K) \) is a random element of \( D[0,1]^K \) defined by

\[
\xi^k(t) = \nu^k(\mu_k^{-1}t), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K.
\]

Since the proof of this theorem is somewhat involved it is relegated to an appendix.

We conclude this section with some remarks on notational conventions which have been observed thus far and will be maintained throughout. Random elements of \( D[0,1] \) and \( D[0,1]^K \) will always be denoted by lower case Greek letters. The term "random function" is understood hereafter to mean a random element of \( D[0,1] \) or \( D[0,1]^K \). The letter \( \pi \) is used throughout to denote the projection \( D[0,1]^K \to \mathbb{R}^K \) at the point \( t = 1 \), as in Theorem 9. Mappings from \( D[0,1]^K \) into \( D[0,1]^L \) will be denoted by upper case Roman letters. Lower case Roman letters are normally used for mappings from \( D[0,1]^K \) into \( \mathbb{R}^L \), and we shall often let \( f = \pi \circ F, \quad g = \pi \circ G \), etc.
2.3 Convergence of Expectations

A primary objective throughout Part I and Part II will be to show weak convergence for sequences of random variables, i.e., to show that $X_n \xrightarrow{D} X$ as $n \to \infty$, where $\{X_n\}$ is some (possibly normalized) sequence of r.v.'s associated with an assembly-like system and $X$ is a non-degenerate r.v. Given such a statement, one would also like to assert that $E(X_n) \to E(X)$ as $n \to \infty$, and in this regard the notion of uniform integrability is key.

**Definition:** A sequence of r.v.'s $\{X_n\}$ with d.f.'s $\{F_n\}$ is called uniformly integrable (u.i.) if

$$\sup_n \int_{|x| > A} x dF_n \to 0 \text{ as } A \to \infty.$$

The following result, which appears as Theorem 5.4 in Billingsley (1968), shows the intimate relationship between uniform integrability and convergence of first moments.

**Theorem 14:** Suppose that $X_n \xrightarrow{D} X$ as $n \to \infty$. If $\{X_n\}$ is uniformly integrable then $E(X_n) \to E(X)$ as $n \to \infty$. If $X$ and $\{X_n\}$ are non-negative and integrable, then $E(X_n) \to E(X)$ as $n \to \infty$ if and only if $\{X_n\}$ is u.i.

The following sufficient conditions for uniform integrability are elementary and will not be proved. Condition (a) is trivial, (b) follows immediately from the definition of uniform integrability, (c) and (d) appear as problems in Chung (1968), p. 93, and (e) is easily established from (a), (b) and (d).
Theorem 15:
(a) \( \{X_n\} \) is u.i. if and only if \( \{|X_n|\} \) is.
(b) If \( \{X_n\} \) is u.i. and \( |Y_n| \leq |X_n| \) for all \( n \), then \( \{Y_n\} \) is u.i. Thus if \( E|X| < \infty \) and \( |Y_n| \leq |X| \) for all \( n \), then \( \{Y_n\} \) is u.i.
(c) If \( E(X_n)^2 \) is uniformly bounded in \( n \), then \( \{X_n\} \) is u.i.
(d) If \( \{X_n\} \) and \( \{Y_n\} \) are u.i., then so are \( \{X_n+Y_n\} \) and \( \{X_n-Y_n\} \).
(e) If \( \{X_n\} \) and \( \{Y_n\} \) are u.i., then so are \( \{\max(X_n,Y_n)\} \) and \( \{\min(X_n,Y_n)\} \).

For the remainder of this section let \( \{X_n\} \) be an i.i.d. sequence and define the partial sums

\[
S_0 = 0 \\
S_n = \sum_{i=1}^{n} X_i \quad (n = 1,2,\ldots)
\]

Theorem 16: If \( E(X_n) = 0 \) and \( E(X_n)^2 = \sigma^2 < \infty \), then \( \left\{ \frac{1}{\sqrt{n}} S_n \right\} \) is u.i.

Proof: \( E\left(\frac{1}{\sqrt{n}} S_n\right)^2 = \frac{1}{n} E(S_n)^2 = \frac{1}{n} (n\sigma^2) = \sigma^2 \) for all \( n \), and hence the proposition follows from (c) of Theorem 15.

Q.E.D.

In contrast to the preceding proposition, the following result is definitely non-trivial. Its proof is given by Whitt (1968), pp. 80-83, and will not be repeated.

Theorem 17: If \( E(X_n) = 0 \) and \( E(X_n)^2 = \sigma^2 < \infty \), then
\[ \left\{ \frac{1}{\sqrt{n}} \max_{0 \leq j \leq n} (S_j) \right\} \text{ is u.i.} \text{ Hence } \left\{ \frac{1}{\sqrt{n}} \min_{0 \leq j \leq n} (S_j) \right\} \text{ is also u.i.} \]

We conclude this section with another elementary proposition which will prove useful in later chapters.

**Theorem 18:** If \( E(X_1)^2 < \infty \), then \( \left\{ \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} X_i \right\} \) is u.i.

**Proof:** Clearly \( \left( \max_{1 \leq i \leq n} X_i \right)^2 \leq \sum_{i=1}^{n} X_i^2 \), so

\[
E\left( \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} X_i \right)^2 \leq \frac{1}{n} E\left( \sum_{i=1}^{n} X_i^2 \right) = E(X_1^2) .
\]

The proposition then follows from part (c) of Theorem 15.

Q.E.D.

2.4 Some Miscellaneous Lemmas

In this section some lemmas which will be used repeatedly in subsequent chapters are stated and proved. They are collected here in order to avoid what are essentially technical distractions in the main arguments.

**Lemma 19:** Let \( \{X_n\} \) be i.i.d. random variables with d.f. \( F(\cdot) \). Suppose \( E(X_1)^2 < \infty \). Then

\[
\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} X_i \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty ,
\]

implying that \( X_n = \Delta(n^{1/2}) \).

**Proof:** Following the proof of Whitt (1968), p. 50, we note that if
\[ \epsilon > 0 \] then by Chebycheff's inequality
\[ P\left( \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |X_i| > \epsilon \right) = P\left( \max_{1 \leq i \leq n} |X_i| > \epsilon \sqrt{n} \right) \]
\[ \leq n P\left( |X_1| > \epsilon \sqrt{n} \right) \leq n \left[ -\frac{1}{n \epsilon^2} \int \{ |y| > \epsilon \sqrt{n} \} y^2 dF \right] \]
\[ = \frac{1}{\epsilon^2} \int \{ |y| > \epsilon \sqrt{n} \} y^2 dF \to 0 \text{ as } n \to \infty , \]

since \( E(X_1)^2 < \infty \).

\[ \text{Q.E.D.} \]

Lemma 20: Let \( \{X_n\} \) be i.i.d. random variables with \( E(X_1) = \mu < 0 \) and \( E(X_1)^2 < \infty \). Define the partial sums \( S_0 = 0, S_n = X_1 + \cdots + X_n \) \( (n = 1, 2, \ldots) \) as usual. Then
\[ \frac{1}{\sqrt{n}} \max_{0 \leq i \leq j \leq n} (S_j - S_i) \to 0 \text{ as } n \to \infty , \]

implying that \( \max_{0 \leq i \leq n} (S_n - S_i) = O(n^{1/2}) \).

Proof: We define the sequence \( \{n_i; i = 0, 1, \ldots\} \) of "descending ladder indices" in the following recursive fashion. Let
\[ n_0 = 0 , \text{ and} \]
\[ n_i = \min\{ j > n_{i-1}; S_j \leq S_{n_{i-1}} \} , \quad i = 1, 2, \ldots . \]

Further define
\[ \alpha_i = n_i - n_{i-1} , \quad i = 1, 2, \ldots , \]
\[ N(n) = \max\{i \geq 0: n_1 \leq n\}, \quad n = 1,2,\ldots , \]
\[ Y_i = \max_{n_{i-1} \leq j < n_i} (S_j - S_{n_{i-1}}), \quad i = 1,2,\ldots . \]

Both \( \{\alpha_i; i = 1,2,\ldots \} \) and \( \{Y_i; i = 1,2,\ldots \} \) are non-negative sequences of i.i.d. random variables, and the condition \( \mathbb{E}(X_1) < 0 \) implies that \( \alpha_1 \) is finite almost everywhere and \( \mathbb{E}(\alpha_1) < \infty \) (cf. Chung (1968), p. 260). Also, clearly

\[
0 \leq \max_{0 \leq i \leq j \leq n} (S_j - S_i) \leq \max_{1 \leq i \leq N(n) + 1} (Y_i), \quad n = 1,2,\ldots .
\]

Since \( N(n) \leq n \), we then have

\[
0 \leq \frac{1}{\sqrt{n}} \max_{0 \leq i \leq j \leq n} (S_j - S_i) \leq \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n+1} (Y_i) .
\]

Thus, if it can be shown that \( \mathbb{E}(Y_1)^2 < \infty \), the desired result follows from Lemma 19.

To prove that \( \mathbb{E}(Y_1)^2 < \infty \), we first note that

\[
0 \leq Y_1 \leq \sum_{i=1}^{\alpha_1} |X_i| .
\]

It follows directly from the results of Chow, Robbins and Teicher (1965) that

\[
\mathbb{E}\left(\sum_{i=1}^{\alpha_1} |X_i|\right)^2 = \mathbb{E}(\alpha_1)\text{Var}(|X_1|) + (\mathbb{E}(\alpha_1)\mathbb{E}|X_1|)^2 < \infty .
\]

Thus \( \mathbb{E}(Y_1)^2 < \infty \) also, and the proof is complete.

Q.E.D.
Lemma 21: Let \( \{u_n\} \) and \( \{v_n\} \) be independent i.i.d. sequences with \( \mathbb{E}(u_n) = a, \mathbb{E}(v_n) = b, \mathbb{E}(u_n)^2 < \infty, \) and \( \mathbb{E}(v_n)^2 < \infty. \) Define the partial sums

\[
U_0 = 0, \quad U_n = \sum_{i=1}^{n} u_i \quad (n = 1, 2, \ldots) ,
\]

\[
V_0 = 0, \quad V_n = \sum_{i=1}^{n} v_i \quad (n = 1, 2, \ldots) ,
\]

and let

\[
X_n = \max_{0 \leq i \leq n} [U_i + (V_n - V_i)] \quad (n = 1, 2, \ldots) .
\]

(a) \( \frac{1}{n} X_n \overset{a.s.}{\to} \max(a, b) \) as \( n \to \infty \)

(b) If \( a < b, \) then \( X_n = V_n + \Delta(n^{1/2}) \)

(c) If \( a > b, \) then \( X_n = U_n + \Delta(n^{1/2}) \)

Symmetrically, each statement continues to hold if the direction of each inequality is reversed and "max" is replaced by "min".

Proof: For the case \( a = b, \) write

\[
X_n = V_n + \max_{1 \leq i \leq n} (U_i - V_i) .
\]

Since \( \frac{1}{n} V_n \overset{a.s.}{\to} b = \max(a, b) \) by the strong law of large numbers (S.L.L.N.), we need only show that

\[
\frac{1}{n} \max_{1 \leq i \leq n} (U_i - V_i) \overset{a.s.}{\to} 0 .
\]

Using the fact that \( \frac{1}{n} (U_n - V_n) \overset{a.s.}{\to} 0, \) this is an easy exercise.

For the case \( a < b, \) let \( S_j = V_j - U_j \quad (j = 0, 1, \ldots) \) and write
\[ X_n = V_n - \min_{0 \leq i \leq n} S_i. \]

Since \( \frac{1}{n} V_n \xrightarrow{a.e.} b = \max(a, b) \), we need only show that

\[ \frac{1}{n} \min_{0 \leq i \leq n} S_i \xrightarrow{a.e.} 0, \quad \min_{0 \leq i \leq n} S_i = \Delta(n^{1/2}). \]

The latter is equivalently stated as

\[ \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} \left| \min_{0 \leq j \leq i} S_j \right| = \frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} \left| S_i \right| \rightarrow 0. \]

By the S.L.L.N., \( \frac{1}{n} S_n \xrightarrow{a.e.} (b-a) > 0 \), implying that \( S_n \xrightarrow{a.e.} +\infty \) and hence that \( \min_{0 \leq i \leq n} S_i \) converges a.e. to a finite limit. Thus

\[ \frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} S_i \xrightarrow{a.e.} 0, \]

which implies each of the desired statements above.

For the case \( a > b \), the desired strong convergence again follows from the S.L.L.N., and to show that \( X_n = U_n + \Delta(n^{1/2}) \) we defined \[ \{S_j\} \] as before and write

\[ X_n = U_n + \max_{0 \leq i \leq n} (S_n - S_i). \]

It is a direct consequence of Lemma 20 that

\[ \max_{0 \leq i \leq n} (S_n - S_i) = \Delta(n^{1/2}). \]

Q.E.D.
Lemma 22: Let \( \{u_n\}, \{v_n\}, \{U_n\}, \{V_n\}, \text{ and } \{X_n\} \) be as in the preceding lemma. Then

\[
\{ \frac{1}{\sqrt{n}} [X_n - n \max(a,b)] \}
\]

is a uniformly integrable sequence.

Proof: As in the proof of the preceding lemma, let

\[
S_n = V_n - U_n, \quad n = 0,1,2,\ldots
\]

noting that \( S_n \) is the sum of i.i.d. random variables with mean \( (b-a) \) and finite variance. Again note that \( X_n \) can be re-expressed as

\[
(2.13) \quad X_n = V_n - \min_{0 \leq i \leq n} (S_i) = U_n + \max_{0 \leq i \leq n} (S_n - S_i).
\]

If \( a = b \), then we write

\[
(2.14) \quad \frac{1}{\sqrt{n}} [X_n - n \max(a,b)] = \frac{1}{\sqrt{n}} (V_n - nb) - \frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} (S_i).
\]

Since \( \{\frac{1}{\sqrt{n}} (V_n - nb)\} \) is u.i. by Theorem 16 and \( \{\frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} S_i\} \) is u.i. by Theorem 17, the desired result follows from Theorem 15.

If \( a < b \), let \( \hat{S}_n = n(b-a) - S_n \) and note that

\[
0 \leq -\frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} S_i = \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} (-S_i) \leq \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} \hat{S}_i.
\]

Since \( \{\frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} \hat{S}_i\} \) is u.i. by Theorem 17, it follows from Theorem 15 that \( \{-\frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} S_i\} \) is also. Since (2.14) is still
valid when \( a < b \), the proposition follows as before.

Finally, if \( a > b \) we use the last part of (2.13) to write

\[
\frac{1}{\sqrt{n}} [X_n - n \max(a, b)] = \frac{1}{\sqrt{n}} (U_n - na) + \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} (S_n - S_i) .
\]

Now let \( \hat{S}_n = S_n - n(b - a) \) and note that

\[
0 \leq \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} (S_n - S_i) \leq \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} (\hat{S}_n - \hat{S}_i) .
\]

\[
eq \frac{1}{\sqrt{n}} \hat{S}_n - \frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} \hat{S}_i .
\]

Since \( \{\frac{1}{\sqrt{n}} \hat{S}_n\} \) is u.i. by Theorem 16 and \( \{\frac{1}{\sqrt{n}} \min_{0 \leq i \leq n} \hat{S}_i\} \) is u.i. by Theorem 17, it follows from Theorem 15 that \( \{\frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} (S_n - S_i)\} \) is u.i. The proposition then follows as before.

Q.E.D.

Lemma 23: Let \( \{X_n\} \) and \( \{Y_n\} \) be such that

\[
\frac{1}{n} X_n \overset{a.s.}{\to} a , \quad \frac{1}{n} Y_n \overset{a.s.}{\to} b \leq a \text{ as } n \to \infty .
\]

Then

(a) \( \frac{1}{n} \max(X_n, Y_n) \overset{a.s.}{\to} a \) as \( n \to \infty . \)

(b) \( \max(X_n, Y_n) = X_n + \Delta(n^{1/2}) \) if \( b < a \).

Symmetrically, each statement continues to hold if the direction of each inequality is reversed and "max" is replaced by "min".

Proof: Note that \( \max(X_n, Y_n) = X_n + (Y_n - X_n)^+ \). By hypothesis,
\[ \frac{1}{n} X_n \stackrel{a.e.}{\to} a, \text{ and} \]

\[ \frac{1}{n} (Y_n - X_n)^+ = \left( \frac{1}{n} Y_n - \frac{1}{n} X_n \right)^{a.e.} (b-a)^+ = 0, \]

so part (a) is proved. Moreover, if \( b < a \) then \( \frac{1}{n} (Y_n - X_n)^{a.e.} (b-a) < 0 \), implying that

\( (Y_n - X_n)^{a.e.} \to \infty. \)

Thus \( (Y_n - X_n)^+ \) converges almost everywhere to a finite limit, implying that

\[ \frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} (Y_i - X_i)^+ \to 0. \]

It follows immediately that \( (Y_n - X_n)^+ = \Delta(n^{1/2}) \), and the proof of part (b) is complete.

Q.E.D.

**Lemma 24:** Let \( \{X_n^1, \ldots, X_n^K\} \) be r.v.'s and \( a_1, \ldots, a_K \) constants such that \( \{\frac{1}{\sqrt{n}} (X_n^k - na_k)\} \) is uniformly integrable for each \( k = 1, \ldots, K \). Then

\[ \frac{1}{\sqrt{n}} \left[ \max_{1 \leq k \leq K} (X_n^k) - n \max_{1 \leq k \leq K} (a_k) \right] \]

is u.i. as well. The same statement holds when "max" is replaced by "min".

**Proof:** We shall prove only the "max" version of the proposition since the "min" version follows in identical fashion. If \( a_1 = \cdots = a_K \), then
\[
\frac{1}{\sqrt{n}} \left[ \max_{1 \leq k \leq K} \left( x_n^k \right) - \max_{1 \leq k \leq K} \left( a_k \right) \right] = \max_{1 \leq k \leq K} \left[ \frac{1}{\sqrt{n}} \left( x_n^k - na_k \right) \right]
\]

and the proposition follows immediately from part (e) of Theorem 15 and induction. Now suppose on the other hand that \( a_1 = \cdots = a_{K^*} > a_{K^*+1} \geq \cdots \geq a_K \), where \( 1 \leq K^* < K \). Then

\[
\frac{1}{\sqrt{n}} \left[ \max_{1 \leq k \leq K} \left( x_n^k \right) - n \max_{1 \leq k \leq K} \left( a_k \right) \right]
= \frac{1}{\sqrt{n}} \left[ \max_{1 \leq k \leq K^*} \left( x_n^k \right) + \left( \max_{K^* < k \leq K} \left( x_n^k \right) - \max_{1 \leq k \leq K^*} \left( x_n^k \right) \right) \right] - na_1
\]

\[
= \frac{1}{\sqrt{n}} \left[ \max_{1 \leq k \leq K^*} \left( x_n^k \right) - na_1 \right] + \delta_n
\]

where

\[
\delta_n = \left[ \max_{K^* < k \leq K} \left( \frac{1}{\sqrt{n}} x_n^k \right) - \max_{1 \leq j \leq K^*} \left( \frac{1}{\sqrt{n}} x_n^j \right) \right]^+.\]

Since \( \left\{ \max_{1 \leq k \leq K^*} \left( \frac{1}{\sqrt{n}} \left( x_n^k - na_k \right) \right) \right\} \) is u.i. by Theorem 15 (e) and induction, we need only show that \( \left\{ \delta_n \right\} \) is u.i. to establish the proposition. Clearly

\[
0 \leq \delta_n \leq \sum \left[ \frac{1}{\sqrt{n}} \left( x_n^k - \frac{1}{\sqrt{n}} x_n^j \right) \right]^+
\]

\[
\leq \sum \left[ \frac{1}{\sqrt{n}} \left( x_n^k - na_k \right) - \frac{1}{\sqrt{n}} \left( x_n^j - na_j \right) \right]^+
\]

\[
\leq \sum \left[ \frac{1}{\sqrt{n}} \left( x_n^k - na_k \right) - \frac{1}{\sqrt{n}} \left( x_n^j - na_j \right) \right],
\]

57
where the summation is taken over all $j$ and $k$ satisfying $K^* < k \leq K$ and $1 \leq j \leq K^*$. Now each summand in the last line of (2.15) is u.i. by Theorem 15, and hence $\{5_n\}$ is u.i. also by that same theorem.

The numbering of the sequences $\{X_n^k\}$ and corresponding constants $a_k$ is arbitrary, so we can assume without loss of generality that $a_1 \geq \cdots \geq a_K$. Thus the cases $a_1 = \cdots = a_K$ and $a_1 = \cdots = a_{K^*} \geq \cdots \geq a_K$ are exhaustive.

Q.E.D.

**Lemma 25:** Let $\{X_n\}$ be a sequence of random variables satisfying

$$\frac{1}{n} X_n \xrightarrow{a.e.} a, \quad \mathbb{E}\left(\frac{1}{\sqrt{n}} X_n\right) \to c \text{ as } n \to \infty.$$ 

Then

$$\frac{1}{n^2} \sum_{i=1}^{n} X_i \xrightarrow{a.e.} \frac{1}{2} a, \quad \mathbb{E}\left(\frac{1}{\sqrt{n^3}} \sum_{i=1}^{n} X_i\right) \to \frac{2}{3} c \text{ as } n \to \infty.$$ 

**Proof:** It is clearly sufficient to show that if $\{a_n\}$ and $\{c_n\}$ are real numbers satisfying

$$\frac{1}{n} a_n \to a, \quad \frac{1}{\sqrt{n}} c_n \to c,$$

then

$$\frac{1}{n^2} \sum_{i=1}^{n} a_i \to \frac{1}{2} a, \quad \frac{1}{\sqrt{n^3}} \sum_{i=1}^{n} c_i \to \frac{2}{3} c.$$ 

That these statements are true follows easily from the well known facts that
\[
\sum_{i=1}^{n} i \sim \frac{1}{2} n^2, \quad \sum_{i=1}^{n} \sqrt{i} \sim \frac{2}{3} \sqrt{n^3}.
\]
Q.E.D.

2.5 Single-Server Queues

Consider a queueing system in which customers arrive individually at a single-server facility and are processed (or served) in the order of their arrival. If an arriving customer finds the server occupied, then he joins a queue and waits for admission to service; otherwise his service begins immediately. Such a queueing system will be referred to throughout this paper as a single-server queue.

Suppose that, beginning from time zero, the server suffers a period of "initial paralysis" of duration \( v_0 \), but that there are no customers initially present in the system. (One could equivalently imagine that the system initially contains some number of customers whose combined service times total \( v_0 \), but the notion of a period of "initial paralysis" will later prove to be convenient.) When the server recovers from this "paralysis" he is ready to begin his first service and will do so immediately if a customer has arrived in the interval \((0,v_0]\). Otherwise he is idle until the first customer arrives. Let

\[ u_0 = \text{the time at which the first customer arrives}, \]
\[ u_n = \text{the time between the arrival of the } n^{th} \text{ customer and the arrival of the } (n+1)^{st} \text{ customer } (n = 1,2,\ldots), \]
\[ v_0 = \text{duration of the server’s initial "paralysis"} \]
\[ v_n = \text{service time of } n^{th} \text{ customer } (n = 1,2,\ldots). \]
Further let

\[ V_0 = 0, \quad U_0 = 0 \]
\[ V_n = \sum_{j=0}^{n-1} v_j \quad (n = 1, 2, \ldots) \]
\[ U_n = \sum_{j=0}^{n-1} u_j = \text{arrival time of } n^{th} \text{ customer} \quad (n = 1, 2, \ldots). \]

Assuming the times \( \{u_n; n = 0, 1, \ldots\} \) and \( \{v_n; n = 0, 1, \ldots\} \) to be stochastically variable, it is clear that the distributions of all random variables associated with the queueing system are completely determined by the joint distributions of the basic sequences \( \{u_n\} \) and \( \{v_n\} \). A stochastic process of fundamental interest in this system is the discrete-parameter process \( \{w_n; n = 1, 2, \ldots\} \), where

\[ w_n = \text{the waiting time of the } n^{th} \text{ arriving customer, exclusive of service time} \quad (n = 1, 2, \ldots). \]

Let us now define

\[ S_0 = 0 \]
\[ S_n = V_n - U_n = \sum_{j=0}^{n-1} (v_j - u_j), \quad n = 1, 2, \ldots. \]

The following observation, originally due to Lindley (1952), gives an explicit and compact representation for \( w_n \) in terms of the basic sequences \( \{u_n\} \) and \( \{v_n\} \).

**Lemma 26**: For all \( n = 1, 2, \ldots \)

\[ w_n = \max_{0 \leq j \leq n} (S_n - S_j) = S_n - \min_{0 \leq j \leq n} S_j. \]
Note that this relationship is a deterministic one, holding for every \( n \) and every realization of the processes \( \{u_n\} \) and \( \{v_n\} \).

An important and still very general type of single-server queue is that in which \( \{u_n; n = 0,1,2,\ldots\} \) and \( \{v_n; n = 0,1,2,\ldots\} \) are independent sequences of i.i.d. non-negative random variables. Following Kendall's (1953) notation, we call such a system a \( GI/G/1 \) queue. For a \( GI/G/1 \) queue we define

\[
a = \mathbb{E}[u_0], \quad b = \mathbb{E}[v_0], \quad \rho = b/a.
\]

The quantity \( \rho \) is often called the system's "traffic intensity", and a fundamental result of queueing theory is that the \( GI/G/1 \) system will approach "statistical equilibrium" if \( \rho < 1 \). In order to make this statement precise for the waiting time process \( \{w_n\} \), let us first define

\[
M_n = \max_{0 \leq j \leq n} S_j, \quad n = 0,1,2,\ldots
\]

\[
M = \sup_{j \geq 0} S_j \quad \text{(possible $\pm \infty$)}.
\]

Clearly \( M_n \) is non-negative for all \( n \) and increases monotonically to \( M \) a.e. The next two results are due to Lindley (1952) and will be proved in order to demonstrate in this simple setting a procedure which will be used later for much more complicated systems.

**Lemma 27:** If \( \{u_n\} \) and \( \{v_n\} \) are independent i.i.d. sequences, then

\[
w_n \overset{d}{=} M_n \quad \text{for all} \quad n = 1,2,\ldots.
\]

**Proof:** Let \( x_i = v_i - u_i \) for all \( i = 0,1,\ldots \). Then \( \{x_i\} \) is an
i.i.d. sequence and \( S_n = \sum_{i=0}^{n-1} x_i \) for all \( n = 1, 2, \ldots \). If \( f \) is an arbitrary function from \( \mathbb{R}^n \) into \( \mathbb{R} \), then clearly
\[
f(x_0, \ldots, x_{n-1}) \overset{D}{=} f(x_{n-1}, \ldots, x_0) \quad .
\]

A particular application of this general premise is that
\[
f(S_0, S_1, \ldots, S_n) \overset{D}{=} f(S_{n-S_n}, S_{n-1}, \ldots, S_0) \quad .
\]
or
\[
(2.16) \quad f(S_j; j = 0, 1, \ldots, n) \overset{D}{=} f(S_{n-S_n-j}; j = 0, 1, \ldots, n) \quad .
\]

Now combining Lemma 26 and (2.16),
\[
v_n = \max_{0 \leq j \leq n} (S_{n-S_n-j}) \overset{D}{=} \max_{0 \leq j \leq n} [S_n - (S_{n-S_n-j})]
\]
\[
= \max_{0 \leq j \leq n} (S_n-j) = \max_{0 \leq j \leq n} S_j = M_n
\]
Q.E.D.

**Theorem 28:** If \( \rho < 1 \) in a GI/G/1 queue, then \( M \) exists and is finite almost everywhere and
\[
w_n \overset{D}{\to} M \text{ as } n \to \infty \quad .
\]

**Proof:** If \( \rho < 1 \), then \( E(x_1) = E(v_i - u_i) = b-a < 0 \). Then by the strong law of large numbers, \( \frac{1}{n} S_n \overset{a.e.}{\to} (b-a) \), implying that \( S_n \overset{a.e.}{\to} \infty \). Hence \( M \) is finite almost everywhere. By Lemma 27, \( w_n \overset{D}{=} M_n \overset{a.e.}{\to} M \), so \( w_n \overset{D}{\to} M \). Q.E.D.
Although the distribution of $M$ is guaranteed to have no mass at infinity, there still remains the question of whether $E(M) = \lim E(M_n) = \lim E(v_n)$ is finite. The following proposition, proved by Kiefer and Wolfowitz (1956), provides the answer.

**Theorem 29:** If $E(v_n)^2 < \infty$ in a GI/G/1 queueing system with $\rho < 1$, then $E(M) < \infty$.

What Kiefer and Wolfowitz actually show is that $E(M) < \infty$ if and only if $E[(v_n - u_n)^2] < \infty$, so the requirement that $E(v_n)^2$ be finite is clearly sufficient. From this result and Lemma 27, the following proposition is immediate.

**Theorem 30:** If $\rho < 1$ and $E(v_n)^2 < \infty$ in a GI/G/1 queueing system, then $E(w_n) \to E(M) < \infty$ as $n \to \infty$.

We conclude this section by stating some known results for GI/G/1 queueing systems in the "heavy traffic" case of $\rho \geq 1$. Although none of these results are actually needed in succeeding chapters, they are included in order to provide examples of functional limit theorems arising in the theory of queues.

Suppose first that $\rho = 1$, and assume that $\sigma_A^2 = \text{Var}(u_n) < \infty$ and $\sigma_S^2 = \text{Var}(v_n) < \infty$. Define the random functions generated by $\{w_n\}$,

$$v_n(t) = \frac{1}{\sqrt{(\sigma_A^2 + \sigma_S^2)n}} w_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots$$

Define $F : D[0,1] \to D[0,1]$ by letting

63
\[ [F(x)](t) = x(t) - \inf_{0 \leq s \leq t} x(s), \ 0 \leq t \leq 1, \ x \in D[0,1]. \]

Letting \( \pi:D[0,1] \to \mathbb{R} \) be as in Theorem 5, we define \( f:D[0,1] \to \mathbb{R} \) by

\[ f(x) = (\pi \circ F)(x) = x(1) - \inf_{0 \leq t \leq 1} x(t), \ x \in D[0,1]. \]

The following functional limit theorem for \( \{w_n\} \) when \( \rho = 1 \) is due to Whitt (1968). Its corollary, showing ordinary convergence in distribution for the normalized waiting times \( \frac{1}{\sqrt{n}} w_n \), was known previously, however, being a direct consequence of Lemma 27 and an early result by Erdös and Kac (1946) for the maxima of partial sums.

**Theorem 31:** If \( \rho = 1 \), then \( v_n \Rightarrow F(\xi) \) as \( n \to \infty \).

**Corollary 32:** If \( \rho = 1 \), then

\[
\lim_{n \to \infty} \mathbb{P}\left\{ \frac{w_n}{\sqrt{(\sigma_A^2 + \sigma_S^2)n}} \leq x \right\} = \mathbb{P}\{f(\xi) \leq x \} = \begin{cases} 0 & \text{if } x \leq 0 \\ \sqrt{\frac{2}{\pi}} \int_0^x e^{-y^2/2} \, dy & \text{if } x > 0 \end{cases}
\]

The diffusion process \( F(\xi) \) obtained as a weak limit in Theorem 31 has the same distribution as ordinary Brownian Motion with a reflecting barrier at the origin (cf. Cox and Miller (1965), p. 233). Corollary 32 is obtained from Theorem 31 by applying the continuous projection \( \pi \) to the random functions \( v_n(\cdot) \).

The following functional limit theorem for \( \{w_n\} \) when \( \rho > 1 \) is
also due to Whitt (1968), although its corollary was obtained earlier by Kingman (1962). Let

\[ \hat{\xi}_n(t) = \frac{1}{\sqrt{\sigma^2_A + \sigma^2_S n}} \xi([nt]), \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots \]

**Theorem 33:** If \( \rho > 1 \), then \( \xi_n \overset{d}{\rightarrow} \xi \) as \( n \to \infty \).

**Corollary 34:** If \( \rho > 1 \), then \( \frac{\xi_n - n(b-a)}{\sqrt{\sigma^2_A + \sigma^2_S n}} \overset{d}{\rightarrow} N(0,1) \) as \( n \to \infty \).

As was indicated in Section 2.2, the usefulness of functional results like Theorems 31 and 33 lies largely in the wealth of corollaries that can be obtained from them through application of the Continuous Mapping Theorem. For example, by applying the integration mapping \( \Psi \) to \( \nu_n(\cdot) \) and \( \xi_n(\cdot) \), limit theorems can be obtained for the accumulated waiting times

\[ W_n = \sum_{i=1}^{n} W_i \]

From that piecewise constant nature of \( \nu_n(\cdot) \) we have that

\[ \Psi(\nu_n) = \int_0^1 \nu_n(t) = \frac{1}{\sqrt{\sigma^2_A + \sigma^2_S n^3}} W_{n-1} \]

Since \( \nu_n \overset{d}{\rightarrow} F(\xi) \) and \( \Psi \) is continuous, the Continuous Mapping Theorem establishes that \( \Psi(\nu_n) \overset{d}{\rightarrow} (\Psi \circ F)(\xi) \) as \( n \to \infty \). Using the fact that \( \frac{1}{\sqrt{n^3}} (W_n - W_{n-1}) = \frac{1}{\sqrt{n^3}} W_n \overset{d}{\rightarrow} 0 \) by Corollary 32, we then have that
\[ \frac{W_n}{\sqrt{(\sigma_A^2 + \sigma_S^2)n^3}} \xrightarrow{D} \frac{1}{\mathbb{D}} \int_0^1 [F(\xi)](t) dt \text{ when } \rho = 1. \]

Applying \( \psi \) to \( \xi_n(\cdot) \) and recalling that \( \int_0^1 \xi(t) dt \xrightarrow{D} N(0, \frac{1}{2}) \) by Theorem 6, we also find that

\[ \frac{W_n - \frac{n}{2} (b-a)}{\sqrt{\frac{1}{2} \sigma_A^2 + \sigma_S^2 n^3}} \xrightarrow{D} N(0,1) \text{ when } \rho > 1. \]
PART I
LIMIT THEOREMS FOR THE SINGLE-STATION MODEL

CHAPTER 3
Introduction

In this chapter a mathematical model of an assembly-like operation is developed. The model is a multiple input generalization of the GI/G/1 queueing system. The basic stochastic processes associated with the model are identified here and studied in Chapters 4-6 below.

3.1 Formulation and Notation

We suppose that \( K \) different types of items (indexed by \( k = 1, 2, \ldots, K \)) arrive separately at a single-server facility. The server does not process items individually but rather in batches of \( K \), each batch consisting of precisely one item of each type. A complete set of \( K \) items, one of each type, will be called a unit, and the set consisting of the \( n^{th} \) arriving item of each type will be called the \( n^{th} \) arriving unit. We assume that the input processes begin at time zero, the system being initially empty, and that items of each type are used in the order of their arrival. Thus the \( n^{th} \) batch to undergo service is the \( n^{th} \) arriving unit (\( n = 1, 2, \ldots \)). It will be convenient to assume that beginning from time zero the server suffers a period of "initial paralysis" of (random) duration \( v_0 \). When the server recovers from this "paralysis", he is ready to begin his first service and will do so immediately if at least one item of each type has arrived in the interval \((0, v_0] \). Otherwise he is idle until the last member of the
first unit arrives. Having completed service of the $n^{th}$ unit
$(n = 1, 2, \ldots)$, the server immediately begins service of the $(n+1)^{st}$
unit if all of its members are present. Otherwise he is idle until
the last member of the $(n+1)^{st}$ unit arrives. The time that an individual
item waits before admission to service can be divided into two parts,
the time (if any) that it waits for the other members of its unit to
arrive, and the time (if any) that the completed unit subsequently waits
before admission to service. Generalizing the notation developed for
single-server queues in section 2.3, we establish the following
definitions.

$$u^k_{o} = \text{the time at which the first item of type } k \text{ arrives}
(k = 1, \ldots, K).$$

$$u^k_{n} = \text{the time between the arrival of the } n^{th} \text{ item of type } k
\text{ and the } (n+1)^{st} \text{ item of type } k \quad (k = 1, 2, \ldots, K;
\quad n = 1, 2, \ldots).$$

$$v_{o} = \text{duration of the server's "initial paralysis".}$$

$$v_{n} = \text{the service time of the } n^{th} \text{ arriving unit } (n = 1, 2, \ldots).$$

$$V_{o} = 0, \quad V^k_{o} = 0 \quad (k = 1, \ldots, K).$$

$$V_{n} = \sum_{j=0}^{n-1} v_{j} \quad (n = 1, 2, \ldots), \quad u^k_{n} = \sum_{j=0}^{n-1} u^k_{n} \quad (k = 1, \ldots, K;
\quad n = 1, 2, \ldots).$$

Further defining

$$U^*_{o} = 0, \quad U^*_{n} = \max_{1 \leq k \leq K} u^k_{n} \quad (n = 1, 2, \ldots),$$
we note that $U^*_{n}$ is the arrival time of the $n^{th}$ unit. It is then natural to define

$$u^*_{\frac{n}{n+1}} = U^*_{n+1} - U^*_{n} \quad (n = 0, 1, \ldots)$$

so that $\{u^*_{\frac{n}{n+1}}\}$ and $\{v_{\frac{n}{n+1}}\}$ represent the interarrival and service times respectively for a single-server queueing system (in the formal sense of Section 2.5) which is "imbedded" within the assembly-like operation.

We assume throughout that $\{u^*_{\frac{n}{n+1}}; n = 0, 1, \ldots\}, \{v_{\frac{n}{n+1}}; n = 0, 1, \ldots\}$, and $\{v_{\frac{n}{n+1}}; n = 0, 1, \ldots\}$ are mutually independent sequences of non-negative i.i.d. random variables with

$$0 < b = E[v_{\frac{n}{n+1}}] < \infty,$$

$$0 < \sigma^2_s = \text{Var}[v_{\frac{n}{n+1}}] < \infty,$$

$$0 < a_k = E[u^*_{\frac{n}{n+1}}] < \infty \quad (k = 1, \ldots, K),$$

$$0 < \sigma^2_k = \text{Var}[u^*_{\frac{n}{n+1}}] < \infty \quad (k = 1, \ldots, K).$$

(Note that $\{u^*_{\frac{n}{n+1}}\}$ is not in general an i.i.d. sequence.) By assuming all variances positive, of course, we rule out the possibility that either the service process or any one of the input processes is deterministic.

It will be convenient to assume that the input processes are numbered so that $a_1 \geq a_2 \geq \cdots \geq a_K$ (meaning that the highest index $K$ corresponds to the "fastest" input process, etc.), and we let

$$K^* = \max\{k: a_k = a_1\},$$

so that
\[ a_1 = \cdots = a_{K^*} > a_{K^*+1} \geq \cdots \geq a_K. \]

We define the system traffic intensity

\[ \rho = b/a_1. \]

The basic stochastic processes associated with the model are as follows. Let

\[ w_n^k = \text{the time that the } n^{th} \text{ arriving item of type } k \text{ waits before being admitted to service} \quad (k = 1, \ldots, K; n = 1, 2, \ldots), \]

\[ w_n^* = \text{the time that the members of the } n^{th} \text{ unit wait together before being admitted to service} \quad (n = 1, 2, \ldots). \]

Clearly \( \{w_n^*\} \) is the waiting time process in the "imbedded" single-server queueing system mentioned above, and for each \( n = 1, 2, \ldots \)

\[ w_n^* = \min_{1 \leq k \leq K} w_n^k. \]

Further define

\[ w_n^k = \sum_{i=1}^{n} w_i^k \quad (k = 1, \ldots, K; n = 1, 2, \ldots), \]

so that \( w_n^k \) represents the cumulative waiting time (exclusive of service time) suffered by the first \( n \) items of type \( k \). Rather than studying the individual processes \( \{w_n^k\} \) and \( \{w_n^*\} \), we shall more generally treat the vector processes \( \{w_n\} \) and \( \{W_n\} \), where
\[ w_n = \begin{pmatrix} w_{n1} \\ \vdots \\ w_{nK} \end{pmatrix} \quad (n = 1, 2, \ldots) , \quad \text{and} \]
\[ W_n = \begin{pmatrix} w_{n1}^1 \\ \vdots \\ w_{nK}^1 \end{pmatrix} = \sum_{i=1}^{n} w_i^k \quad (n = 1, 2, \ldots) . \]

We shall study one more discrete-parameter process. Let
\[ T_n = \text{the time at which the server completes his } n^{th} \text{ service} \]
\[ (n = 1, 2, \ldots) . \]

Clearly \( 0 \leq T_1 \leq T_2 \leq \cdots \) and
\[ T_n = U_n^* + w_n^* + v_n \quad (n = 1, 2, \ldots) . \]

The continuous-parameter processes to be studied are

\[
D(t) = \text{the number of service completions by time } t \\
(0 \leq t < \infty),
\]

\[
Q^k(t) = \text{the number of type } k \text{ items (including possibly one undergoing service) present in the system at time } t \\
(k = 1, \ldots, K; 0 \leq t < \infty),
\]

\[
Q(t) = \begin{bmatrix} Q^1(t) \\ \vdots \\ Q^K(t) \end{bmatrix} , \quad 0 \leq t < \infty .
\]
Note that \( \{D(t)\} \) is the counting process, corresponding to the non-decreasing sequence of r.v.'s \( \{T_n\} \). That is, if we let \( T_0 = 0 \), then
\[
D(t) = \max\{n: T_n \leq t\}, \quad 0 \leq t < \infty.
\]

In the following section, explicit representations are developed for \( w_n^*, w_n^k \) and \( T_n \) in terms of the basic sequences \( \{u_n^k\} \) and \( \{v_n\} \). In order to express those representations as compactly as possible, it will be convenient to establish the following further notation. Let
\[
S_n^k = v_n - u_n^k \quad (k = 1, \ldots, K; n = 0, 1, 2, \ldots),
\]
\[
S_n^* = v_n - u_n^* = \min_{1 \leq k \leq K} S_n^k \quad (n = 0, 1, 2, \ldots).
\]

It will also be convenient to define the following random vectors. For each \( n = 0, 1, 2, \ldots \) let
\[
U_n = \begin{pmatrix} u_n^1 \\ \vdots \\ u_n^K \end{pmatrix}, \quad U_n = \begin{pmatrix} u_n^1 \\ \vdots \\ u_n^K \end{pmatrix} = \sum_{i=0}^{n-1} u_i^1,
\]
\[
x_n = \begin{pmatrix} v_n - u_n^1 \\ \vdots \\ v_n - u_n^K \end{pmatrix}, \quad S_n = \begin{pmatrix} s_n^1 \\ \vdots \\ s_n^K \end{pmatrix} = \sum_{i=0}^{n-1} x_i^1.
\]

We also define

72
\[ a = \mathbf{E}(u_n) = \left( \begin{array}{c} a_1 \\ \vdots \\ a_K \end{array} \right), \]

\[ \mathbf{D}^2 = \text{Cov}(u_n) = \left( \begin{array}{c} \sigma^2_1 \\ \vdots \\ \sigma^2_K \end{array} \right), \]

\[ \mathbf{\Phi} = \text{Cov}(X_n) = (s_{ij}), \quad s_{ij} = \begin{cases} \frac{\sigma^2_i + \sigma^2_S}{2} & \text{if } i = j \\ \frac{\sigma^2_S}{2} & \text{if } i \neq j \end{cases}. \]

In a final generalization of the notation developed for GI/G/1 queueing systems in Section 2.5, we let

\[ M_n^k = \max_{0 \leq j \leq n} S_j^k \quad (k = 1, \ldots, K; n = 0, 1, 2, \ldots) ; \]

\[ M^k = \sup_{j \geq 0} S_j^k \quad (\text{possibly } +\infty) , \quad (k = 1, \ldots, K) . \]

3.2 Basic Representations and Preliminary Results

We have spoken in the preceding section of a single-server queueing system (in the formal sense of Section 2.5) which is "imbedded" within an assembly-like operation. This imbedded system takes complete units (K-tuples) as its basis:"customers." The duration of the server's initial paralysis in the imbedded system is \( v_0 \), the sequence of service times is \( \{v_n; n = 1, 2, \ldots\} \), the sequence of interarrival times
is \( \{u_n^*: n = 0, 1, \ldots\} \), and the sequence of "customer" waiting times (exclusive of service time) is \( \{w^*_n; n = 1, 2, \ldots\} \). Since \( \{u_n^*\} \) is not an i.i.d. sequence, the imbedded system is not GI/G/1. Lemma 2.26 still applies, however, and hence the following representation for \( w^*_n \) is immediate.

**Lemma 1:** \( w^*_n = S^*_n - \min_{0 \leq j \leq n} S^*_j \) for all \( n = 1, 2, \ldots \).

We can now obtain explicit representations for \( w^*_n, w^k_n \) and \( T_n \) in terms of the basic sequences \( \{u_n^k\} \) and \( \{v_n\} \).

**Lemma 2:** For all \( n = 1, 2, \ldots \),

(a) \( w^*_n = \min_{1 \leq k \leq K} S^k_n - \min_{0 \leq j \leq n} S^k_j \); 

(b) \( w^k_n = S^k_n - \min_{1 \leq i \leq K} S^i_j \) for all \( k = 1, \ldots, K \); 

(c) \( T_n = \max_{0 \leq j \leq n} \left[ u^k_j + (v_n - v_j) \right] + v_n \).

**Proof:** (a) From Lemma 1,

\[ w^*_n = S^*_n - \min_{0 \leq j \leq n} S^*_j = \min_{1 \leq k \leq K} S^k_n - \min_{0 \leq j \leq n} \left[ \min_{1 \leq k \leq K} S^k_j \right]. \]

(b) From our basic definitions it is clear that

\[ w^k_n = w^*_n + (U^*_n - U^k_n) \]

Then substituting from Lemma 1,
\[ w_n^k = (S^{*}_n - \min_{0 \leq j \leq n} S^*_j) + (U^* - U^k_n) \]
\[ = (V_n - U^* - \min_{0 \leq j \leq n} S^*_j) + (U^* - U^k_n) \]
\[ = S^{*}_n - \min_{0 \leq j \leq n} \left[ \min_{1 \leq i \leq K} S^i_j \right] . \]

(c) \[ T_n = U^*_n + w^*_n + v_n = U^*_n + S^*_n - \min_{0 \leq j \leq n} S^*_j + v_n \]
\[ = V_n - \min_{0 \leq j \leq n} S^*_j + v_n = \max_{0 \leq j \leq n} [V_n - (V_j - U^*_j)] + v_n \]
\[ = \max_{0 \leq j \leq n} [\max_{1 \leq k \leq K} U^k_j + (V_n - V_j)] + v_n . \]

Q.E.D.

From these representations some important properties of \( T_n \), \( w_n^k \) and \( w^*_n \) can now be proved. The purpose of Lemmas 3-5 is to show that each of the representations above can be "boiled down" to an important part plus an asymptotically negligible remainder in each of the cases \( \rho < 1 \), \( \rho = 1 \) and \( \rho > 1 \). Lemma 6 then shows the uniform integrability of certain normalized sequences, so that statements of convergence of expectations can be obtained as immediate corollaries to subsequent results establishing convergence in distribution.

\textbf{Lemma 3:} (a) \( \frac{1}{n} T_n \overset{a.s.}{\longrightarrow} \max(a_1, b) \) as \( n \to \infty \).

(b) \( T_n = \max_{1 \leq k \leq K^*} U^k_n + \Delta(n^{1/2}) \) if \( \rho < 1 \),
\[ T_n = \max_{1 \leq k \leq K^*} [U^k_j + (V_n - V_j)] + \Delta(n^{1/2}) \] if \( \rho = 1 \),
\[ 0 \leq j \leq n \]

75
\[ T_n = v_n + \Delta(n^{1/2}) \quad \text{if } \rho > 1. \]

**Proof:** For each \( k = 1, \ldots, K \) and \( n = 1, 2, \ldots \) let

\[ X_n^k = \max_{0 \leq j \leq n} [U_n^k + (V_n - V_j)]. \]

Lemma 2 established that

\[ T_n = \max_{1 \leq k \leq K} X_n^k + v_n. \]

From Lemma 2.21 (a) it is immediate that

\[ \frac{1}{n} X_n^k \xrightarrow{a.s.} \max(a_k, b) \quad \text{as } n \to \infty \]

for each \( k = 1, \ldots, K. \) Then using Lemma 2.23 and induction yields

\[ \frac{1}{n} \left[ \max_{1 \leq k \leq K} X_n^k \right] \xrightarrow{a.s.} \max_{1 \leq k \leq K} [\max(a_k, b)] = \max(a_\perp, b). \]

Since \( \frac{1}{n} v_n \xrightarrow{a.s.} 0 \) as an elementary consequence of the strong law of large numbers, the proof of (a) is complete.

To prove the first statement in (b), we first note that if \( \rho < 1 \) then \( \max(a_k, b) = a_\perp \) for \( k \leq K^* \), while \( \max(a_k, b) < a_\perp \) for \( k > K^* \). Thus, using Lemma 2.23 (a) and induction,

\[ \max_{1 \leq k \leq K} X_n^k = \max_{1 \leq k \leq K} X_n^k + \Delta(n^{1/2}). \]

Moreover, Lemma 2.21 (c) shows that \( X_n^k = U_n + \Delta(n^{1/2}) \) for all
k = 1, \ldots, K^*, \text{ so we have}

\begin{align*}
\max_{1 \leq k \leq K^*} \chi^k_n &= \max_{1 \leq k \leq K^*} \left[ U^k_n + \Delta(n^{1/2}) \right] + \Delta(n^{1/2}) \\
&= \max_{1 \leq k \leq K^*} U^k_n + \Delta(n^{1/2}).
\end{align*}

Finally, Lemma 2.19 shows that \( v_n = \Delta(n^{1/2}) \), so the first statement in (b) is proved. The remaining statements follow similarly from Lemmas 2.19, 2.21 and 2.23.

Q.E.D.

Lemma 4: For each \( k = 1, \ldots, K \)

(a) \( \frac{1}{n} w^k_n \xrightarrow{a.e^*} \max(a_{\perp}, b) - a_k \) as \( n \to \infty \).

(b) \( w^k_n = \max_{1 \leq i \leq K^*} u^i_n - U^k_n + \Delta(n^{1/2}) \) if \( \rho < 1 \),

\[ w^k_n = s^k_n - \min_{0 \leq j \leq K^*} s^j_n + \Delta(n^{1/2}) \] if \( \rho = 1 \),

\[ w^k_n = s^k_n + \Delta(n^{1/2}) \] if \( \rho > 1 \).

Proof: Each part of this proposition follows easily from the corresponding part of Lemma 4 and the observation that \( w^k_n = T_n - U^k_n - v_n = T_n - U^k_n + \Delta(n^{1/2}) \).

Q.E.D.

Lemma 5: (a) \( \frac{1}{n} w^*_n \xrightarrow{a.e^*} (b - a_{\perp})^+ \) as \( n \to \infty \).

(b) \( w^*_n = \min_{1 \leq k \leq K^*} w^k_n + \Delta(n^{1/2}) \).
Proof: From Lemma 4 (a), Lemma 2.23 (a) and the observation that
\[ w_n^* = \min_{1 \leq k \leq K} w_n^k, \] it follows that
\[ \frac{1}{n} w_n^{*, e.} \min_{1 \leq k \leq K} [\max(a_l, b)-a_k] = (b-a_1)^+. \]

So part (a) is proved. Part (b) follows easily from Lemma 2.23 (b) and Lemma 4 by considering separately the cases \( \rho < 1 \), \( \rho = 1 \) and \( \rho > 1 \).

Q.E.D.

Lemma 6: Each of the sequences
\[ \left\{ \frac{1}{\sqrt{n}} [T_{n} - n \max(a_l, b)] \right\}, \]
\[ \left\{ \frac{1}{\sqrt{n}} (v_n^k - n[\max(a_l, b) - a_k]) \right\}, \quad k = 1, \ldots, K, \]
\[ \left\{ \frac{1}{\sqrt{n}} [w_n^{*, n}(b-a_1)^+] \right\} \]

is uniformly integrable.

Proof: As before, we let
\[ x_n^k = \max_{0 \leq j \leq n} [u_j^k + (v_{nj} - v_j)], \quad 1 \leq k \leq K, \]

and note that
\[ (3.1) \quad T_n = \max_{1 \leq k \leq K} x_n^k + v_n \]

by Lemma 2. By Lemma 2.22,
\[ \left\{ \frac{1}{\sqrt{n}} \max X_{n}^{k} - \text{max}(a_{k}, b) \right\} \]

is u.i. for each \( k = 1, \ldots, K \), and hence

\[ (3.2) \quad \left\{ \frac{1}{\sqrt{n}} \max \max_{1 \leq k \leq K} X_{n}^{k} - \text{max}(a_{1}, b) \right\} \text{ is u.i.} \]

by Lemma 2.24. Combining (3.1), (3.2) and the obvious fact that \( \left\{ \frac{1}{\sqrt{n}} v_{n} \right\} \) is u.i., we have proved the uniform integrability of the first sequence above. The uniform integrability of the second then follows from the observation that \( v_{n}^{k} = T_{n}^{k} - U_{n}^{k} - v_{n} \) and the fact that \( \left\{ \frac{1}{\sqrt{n}} (U_{n}^{k} - \text{na}_{k}) \right\} \) is u.i. by Theorem 2.16. The uniform integrability of the third sequence then follows from Lemma 2.24 and the fact that \( \hat{w}_{n}^{*} = \min_{1 \leq k \leq K} w_{n}^{k} \).

Q.E.D.
CHAPTER 4
The Case \( \rho < 1 \)

It is assumed throughout this chapter that

\[
b < a_1 = \cdots = a_{K^*} > a_{K^* + 1} \geq \cdots \geq a_K.
\]

Limit theorems are developed for \( w_n, w_n, W_n \) and \( T_n \) as \( n \to \infty \). The methods used in the analysis of \( \{w_n^*\} \) differ greatly from those used in the analysis of the other processes, and for this reason its treatment is presented alone in Section 4.1. The other processes are then discussed together in Section 4.2. None of the results obtained in Sections 4.1 and 4.2 are stated in a completely explicit form. However, Section 4.3 deals with some important special cases (e.g., \( K^* = 1, K^* = 2 \), all input processes Poisson) in which some or all of the general theorems simplify.

4.1 The Process \( \{w_n^*\} \)

Our first lemma shows that the "faster" input processes \( K^* + 1, \ldots, K \) play no role in determining the asymptotic behavior of \( w_n^* \).

Lemma 1: \( w_n^* = \bar{w}_n^* + \delta_n \), where \( \delta_n \xrightarrow{p} 0 \) as \( n \to \infty \) and

\[
(4.1) \quad \bar{w}_n^* = \min_{1 \leq k \leq K^*} S_n^k - \min_{j \in \Lambda} S_n^j.
\]

Proof: We first observe that \( w_n^* = T_n - U_n^* - v_n \). Let
\[
X^k_n = \max_{0 \leq j \leq n} \left[ U^k_n + (V_n^j - V_j) \right], \quad 1 \leq k \leq K, \quad n = 0, 1, \ldots
\]

Lemma 3.2 shows that \( T_n - v_n = \max_{1 \leq k \leq K} X^k_n \). Correspondingly, \( U^*_n = \max_{1 \leq k \leq K} U^k_n \). Lemma 2.21 shows that

\[
\frac{1}{n} X^k_n \overset{a.e.}{\to} a_1 \quad \text{if} \quad 1 \leq k \leq K^*,
\]

\[
\frac{1}{n} X^k_n \overset{a.e.}{\to} \max(a_k, b) < a_1 \quad \text{if} \quad K^* < k \leq K,
\]

It follows easily then that

\[
0 \leq \delta'_n = [(T_n - v_n) - \max_{1 \leq k \leq K^*} X^k_n] \overset{a.e.}{\to} 0.
\]

Similarly, by the strong law of large numbers,

\[
\frac{1}{n} U^k_n \overset{a.e.}{\to} a_1 \quad \text{if} \quad 1 \leq k \leq K^*,
\]

\[
\frac{1}{n} U^k_n \overset{a.e.}{\to} a_k < a_1 \quad \text{if} \quad K^* < k \leq K,
\]

from which it follows that

\[
0 \leq \delta''_n = [U^*_n - \max_{1 \leq k \leq K^*} U^k_n] \overset{a.e.}{\to} 0.
\]

One may easily verify that

\[
\bar{w}^*_n = \max \left\{ X^k_n - \max_{1 \leq k \leq K^*} U^k_n \right\},
\]

so we have shown that \( |v_n^* - \bar{w}^*_n| \leq \delta'_n + \delta''_n \overset{a.e.}{\to} 0 \), which completes the proof.

Q.E.D.
The following proposition is key in analyzing the asymptotic behavior of \( w_n^* \). It is completely analogous to Lemma 2.27 for GI/G/1 queueing systems and is proved in a precisely parallel fashion.

**Lemma 2:** For each \( n = 1, 2, \ldots \)

\[
\underline{w}^*_n \equiv \underline{w}^*_n = \max_{1 \leq k \leq K^*} (U_{n}^{k} + M_{n}^{k}) - \max_{1 \leq k \leq K^*} (U_{n}^{k}),
\]

where \( \underline{w}^*_n \) is defined by (4.1).

**Proof:** If \( f \) is an arbitrary function from \( \mathbb{R}^{(n+1)(K^*+1)} \) into \( \mathbb{R} \), then clearly

\[
f(v_0, \ldots, v_n, u_0^1, \ldots, u_n^1, \ldots, u_0^{K^*}, \ldots, u_n^{K^*})
\]

\[
\equiv f(v_n, \ldots, v_0, u_n^1, \ldots, u_0^1, \ldots, u_n^{K^*}, \ldots, u_0^{K^*}),
\]

or, more compactly stated,

\[
f(v_j, u_j^k; 0 \leq j \leq n, 1 \leq k \leq K^*) \equiv f(v_n-j, u_n-j; 0 \leq j \leq n, 1 \leq k \leq K^*).
\]

A particular application of this general premise is that

\[
g(V_j, U_j^k; 0 \leq j \leq n, 1 \leq k \leq K^*)
\]

\[
\equiv g(V_n-j, U_n-j; 0 \leq j \leq n, 1 \leq k \leq K^*)
\]

for arbitrary \( g: \mathbb{R}^{n(K^*+1)} \to \mathbb{R} \) and all \( n = 1, 2, \ldots \). Now applying this to the expression (4.1) which defines \( \underline{w}^*_n \), we have
\[ \tilde{w}_n^* = \min_{1 \leq k \leq K^*} (V_n - U_n^k) - \min_{0 \leq j \leq n} \min_{1 \leq k \leq K^*} (V_j - U_j^k) \]

\[ D = \min_{1 \leq k \leq K^*} (V_n - U_n^k) - \min_{0 \leq j \leq n} \min_{1 \leq k \leq K^*} [(V_n - V_{n-j}) - (U_n^k - U_{n-j}^k)] \]

\[ = \min_{1 \leq k \leq K^*} (-U_n^k) - \min_{0 \leq j \leq n} \min_{1 \leq k \leq K^*} [-U_n^k - (V_{n-j} - U_{n-j}^k)] \]

\[ = \max_{1 \leq k \leq K^*} \min_{0 \leq j \leq n} [(U_n^k - (V_{n-j} - U_{n-j}^k))] - \max_{1 \leq k \leq K^*} (U_n^k) \]

\[ = \max_{1 \leq k \leq K^*} \min_{0 \leq j \leq n} [(U_n^k + (V_{n-j} - U_{n-j}^k))] - \max_{1 \leq k \leq K^*} (U_n^k) \]

\[ = \max_{1 \leq k \leq K^*} \max_{0 \leq j \leq n} (S_n^k) - \max_{1 \leq k \leq K^*} (U_n^k) \]

\[ = \max_{1 \leq k \leq K^*} (U_n^k + M_n^k) - \max_{1 \leq k \leq K^*} (U_n^k) \]

\[ = \max_{1 \leq k \leq K^*} (U_n^k + M_n^k) - \max_{1 \leq k \leq K^*} (U_n^k) \]

Q.E.D.

Before proceeding to the main theorem for \( w_n^* \), we first prove the following preliminary proposition.

**Lemma 3**: Let \( N \) be any fixed positive integer. Then for all \( x \in \mathbb{R} \),

\[
\lim_{n \to \infty} P\{ \max_{1 \leq k \leq K^*} (U_n^k + M_n^k) + \max_{1 \leq k \leq K^*} (U_n^k) \leq x \}
\]

\[ = \sum_{k=1}^{K^*} P[X_k = \max(X_1, \ldots, X_{K^*})] P[M_n^k \leq x], \]

where \( X_1, \ldots, X_{K^*} \) are independent normal r.v.'s with mean zero and variance \( \sigma_1^2, \ldots, \sigma_{K^*}^2 \) respectively (i.e., \( \text{Var}(X_k) = \text{Var}(u_n^k), 1 \leq k \leq K^* \)).
Proof: For each \( n > N \) and each \( k = 1, \ldots, K^* \) define

\[
\tilde{U}_n^k = U_n^k - U_N^k, \quad \hat{U}_n^k = \hat{U}_n^k - (n-N)a_k,
\]

and let \( \tilde{U}_n = (\tilde{U}_n^1, \ldots, \tilde{U}_n^{K^*}) \). By the Central Limit Theorem

\[
(4.2) \quad \frac{1}{\sqrt{n}} \tilde{U}_n \xrightarrow{D} X = (X_1, \ldots, X_{K^*}),
\]

and thus (by the Continuous Mapping Theorem)

\[
(4.3) \quad P \{ \hat{U}_n^k = \max_{1 \leq j \leq K^*} (\hat{U}_n^j) \} = P \{ \frac{1}{\sqrt{n}} \tilde{U}_n^k = \max_{1 \leq j \leq K^*} (\frac{1}{\sqrt{n}} \tilde{U}_n^j) \} \rightarrow P \{ X_k = \max_{1 \leq j \leq K^*} X_j \} \text{ as } n \to \infty.
\]

Moreover, since the joint normal distribution is absolutely continuous, (4.2) implies that for any constant \( L \)

\[
(4.4) \quad P \left\{ \min_{1 \leq j,k \leq K^*} \left| \tilde{U}_n^j - \tilde{U}_n^k \right| > L \right\} = P \left\{ \min_{1 \leq j,k \leq K^*} \left| \frac{1}{\sqrt{n}} \tilde{U}_n^j - \frac{1}{\sqrt{n}} \tilde{U}_n^k \right| > \frac{1}{\sqrt{n}} L \right\} \to 1 \text{ as } n \to \infty.
\]

Now define a random index \( k^*_n \) by

\[
\hat{U}_n^{k^*_n} = \max_{1 \leq k \leq K^*} \hat{U}_n^k, \quad n > N.
\]

(Since the probability approaches zero that the maximum is not uniquely achieved, it is immaterial how \( k^*_n \) is defined in that case.) Now note
that the event

\[ A_n \equiv \{ \max_{1 \leq k \leq K^*} (U_n^{k,k} + M_n^k) - \max_{1 \leq k \leq K^*} (U_n^k) = M_n^{k^*} \} \]

is implied by the event

\[ B_n \equiv \{ \min_{1 \leq j,k \leq K^*} |\hat{U}_n^j - \hat{U}_n^k| \geq \max_{1 \leq k \leq K^*} (U_n^{k,k} + M_n^k) \} \]

Since the r.v.'s $U_n^k$ and $M_n^k$ are bounded in probability, it follows from (4.4) that $P[B_n] \to 1$ as $n \to \infty$, so we conclude that $P[A_n] \to 1$ as $n \to \infty$. Thus for any $x \in \mathbb{R}$

\[ \lim_{n \to \infty} P\left[ \max_{1 \leq k \leq K^*} (U_n^{k,k} + M_n^k) - \max_{1 \leq k \leq K^*} (U_n^k) \leq x \right] = \lim_{n \to \infty} P[M_n^{k^*} \leq x] \]

if the limit on the right exists. But now we note that for $n > N$ the random vector $(M_N^{1}, \ldots, M_N^{K^*})$ and the random index $k^*_n$ are independent. Combining this with (4.3) and the fact that $k^*_n = k$ if and only if $\hat{U}_n^k = \max(\hat{U}_n^1, \ldots, \hat{U}_n^k)$,

\[ P[M_n^{k^*} \leq x] = \sum_{k=1}^{K^*} P[U_n^k = \max_{1 \leq j \leq K^*} \hat{U}_n^j] P[M_n^k \leq x] \to \sum_{k=1}^{K^*} P[X_k = \max_{1 \leq j \leq K^*} X_j] P[M_n^k \leq x], \]

which completes the proof.

Q.E.D.

With Lemmas 1-3 in hand, the main result for $w^*_n$ can now be proved. Recall that if $b < a_k$ then $M_n^k$ is finite almost everywhere.
by Theorem 2.28. Moreover, the limiting waiting time distribution in
a GI/G/1 queueing system with interarrival times \( \{u_{n}^{k}\} \) and service
times \( \{v_{n}^{k}\} \) is the same as the distribution of \( m^{k} \).

**Theorem 4:** Let \( X_{1}, \ldots, X_{K^{*}} \) be as in Lemma 3. Then
\[
\widetilde{w}_{n}^{*} \xrightarrow{D} w^{*} \text{ as } n \to \infty ,
\]
where \( w^{*} \) is a non-negative r.v. whose distribution is specified by
\[
P[w^{*} \leq x] = \sum_{k=1}^{K^{*}} P[X_{k} = \max(X_{1}, \ldots, X_{K^{*}})] P[M_{n}^{k} \leq x] , \quad x \geq 0 .
\]

**Proof:** By Lemmas 1 and 2 it suffices to show that \( \widetilde{w}_{n}^{*} \xrightarrow{D} w^{*} \) as \( n \to \infty \).
To do so, let \( \epsilon > 0 \) be arbitrary and choose \( N \) large enough that
\[
(4.5) \quad P[M_{n}^{k} = M^{k} \text{ for each } k = 1, \ldots, K^{*}] \geq 1 - \epsilon .
\]
(Such an \( N \) must exist since \( S_{n}^{k} \xrightarrow{s} \infty \) as \( n \to \infty \) for each
\( k = 1, \ldots, K^{*} \)). Now recall that \( \widetilde{w}_{n}^{*} \) is defined by
\[
\widetilde{w}_{n}^{*} = \max_{1 \leq k \leq K^{*}} (U_{n}^{k} + M_{n}^{k}) - \max_{1 \leq k \leq K^{*}} (U_{n}^{k}) , \quad n = 1, 2, \ldots .
\]
Thus, from (4.5),
\[
P[\widetilde{w}_{n}^{*} = \max_{1 \leq k \leq K^{*}} (U_{n}^{k} + M_{n}^{k}) - \max_{1 \leq k \leq K^{*}} (U_{n}^{k})] \geq 1 - \epsilon
\]
for all \( n > N \), and hence
\[
(4.6) \quad |P[\widetilde{w}_{n}^{*} \leq x] - P[\max_{1 \leq k \leq K^{*}} (U_{n}^{k} + M_{n}^{k}) - \max_{1 \leq k \leq K^{*}} (U_{n}^{k}) \leq x]| \leq \epsilon
\]
for all \( n > N \) and \( x \in \mathbb{R} \). Now letting \( n \to \infty \) in (4.6) and using Lemma 3, we have
\[
\lim_{n \to \infty} \left| P\{\tilde{w}_n^* \leq x\} - \sum_{k=1}^{K^*} P\{X_k = \max_{1 \leq j \leq K^*} X_j\} P\{M_n^k \leq x\} \right| \leq \epsilon
\]
for all \( x \in \mathbb{R} \). Then by (4.5)
\[
\lim_{n \to \infty} \left| P\{\tilde{w}_n^* \leq x\} - \sum_{k=1}^{K^*} P\{X_k = \max_{1 \leq j \leq K^*} X_j\} P\{M_n^k \leq x\} \right| \leq 2\epsilon
\]
for all \( x \in \mathbb{R} \), or equivalently stated,
\[
\lim_{n \to \infty} \left| P\{\tilde{w}_n^* \leq x\} - P\{w^* \leq x\} \right| \leq 2\epsilon , \; x \in \mathbb{R} .
\]
Since \( \epsilon \) was chosen arbitrarily, we conclude that \( \tilde{w}_n^* \overset{D}{\to} w^* \) and the proof is complete.

Q.E.D.

Theorem 2.29 establishes that \( E(M_n^k) < \infty \) for all \( k = 1, \ldots, K^* \) and hence
\[
E(w^*) = \sum_{k=1}^{K^*} P\{X_k = \max_{1 \leq j \leq K^*} X_j\} E(M_n^k) < \infty .
\]
It would be of interest to show that \( E(w_n^*) \to E(w^*) \) as \( n \to \infty \), but this writer has been unable to do so. Note that, from the definition of \( \tilde{w}_n^* \),
\[
\tilde{w}_n^* \leq \max_{1 \leq k \leq K^*} M_n^k \leq \sum_{k=1}^{K^*} M_n^k \leq \sum_{k=1}^{K^*} M^k .
\]
Thus \( \{\hat{w}_n^*\} \) is uniformly integrable by part (b) of Theorem 2.15. Since we have shown that \( \hat{w}_n^* \Rightarrow w^* \), it follows from Theorem 2.14 that \( E(\hat{w}_n^*) \Rightarrow E(w^*) \). Thus, by Lemmas 1 and 2, one need only show that \( E(\hat{w}_n) = E(w_n^* - \hat{w}_n) \rightarrow 0 \) in order to establish that \( E(w_n^*) \rightarrow E(w^*) \).

Our limit theorem for \( w_n^* \) has the following physical interpretation. As \( n \) becomes large, the probability approaches one that the last arriving member of the \( n^{th} \) unit is of one of the types \( 1, \ldots, K^* \). Moreover, regardless of the type of that item, the probability approaches one that it arrives much later than the other items comprising the \( n^{th} \) unit. As in the proof of Lemma 3, let us denote by \( k_n^* \) the type of this last arriving member of the \( n^{th} \) unit. Then \( U_n^* = U_n^k \) and \( w_n^* = w_n^k \), and one might say that input process \( k_n^* \) is "pacing" the operation at the time of the \( n^{th} \) service. As \( n \) becomes large, the probability that \( k_n^* = k \) approaches \( P(X_k = \max_{1 \leq j \leq K^*} X_j) \) for \( k = 1, \ldots, K^* \). Given that input process \( k \) is pacing the operation, \( w_n^* = w_n^k \) tends to behave (for \( n \) large) as the limiting waiting time distribution in a GI/G/1 queue with interarrival times \( \{u_n^k\} \) and service times \( \{v_n\} \), i.e., as the distribution of \( M^k \).

4.2 The Processes \( \{T_n\}, \{w_n\} \) and \( \{W_n\} \)

The principle results of this section are functional limit theorems for \( \{T_n\} \) and \( \{w_n\} \). From them all of the other results follow as easy corollaries.

We define (for purposes of this section only) the translated r.v.'s...
\[ \hat{T}_n = T_n - n \alpha_1, \quad n = 1, 2, \ldots \]

(for completeness we take \( \hat{T}_0 = T_0 = 0 \)) and the random functions generated by \( \{\hat{T}_n\} \),

\[ \nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}[nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots \]

Define \( G: D[0, 1]^{K^*} \rightarrow D[0, 1] \) by

\[ [F^k(x)](t) = \max_{1 \leq k \leq K^*} x^k(t), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K^*, \quad x \in D[0, 1]^{K^*}. \]

Theorem 5: \( \nu \Rightarrow G(D^*_{K^*}) \) as \( n \rightarrow \infty \), where \( D^* \) is the \( K^* \times K^* \) diagonal matrix with diagonal elements \( \sigma_1, \ldots, \sigma_K^* \).

**Proof:** Let \( \hat{U}_n^k = U_n^k - n \alpha_k = \hat{U}_n^k - n \alpha_1 \) for all \( k = 1, \ldots, K^* \). Then let \( \hat{U}_n = (\hat{U}_n^1, \ldots, \hat{U}_n^{K^*}) \). According to Lemma 3.3,

\[ T_n = \max_{1 \leq k \leq K^*} U_n^k + \Delta(n^{1/2}) \]

and thus

\[ \hat{T}_n = \max_{1 \leq k \leq K^*} \hat{U}_n^k + \Delta(n^{1/2}) \]

Then by Theorem 2.7 it suffices to show that \( \gamma \Rightarrow G[D^*_{K^*}] \), where

\[ \gamma_n(t) = \frac{1}{\sqrt{n}} \max_{1 \leq k \leq K^*} \hat{U}_n^k [nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots \]
Let us define $\xi_n = (\xi_{n1}, \ldots, \xi_{nk^*}) \in D[0,1]^{K^*}$ by

$$\xi_n(t) = \frac{1}{\sqrt{n}} \hat{U}_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots.$$ 

Then for all $n = 1, 2, \ldots$ and $t \in [0,1]$,

$$(4.7) \quad \gamma_n(t) = \max_{1 \leq k \leq K^*} [\xi_n(t)] = [G(\xi_n)](t).$$

Now $\xi_n \Rightarrow D^*_{\xi_i}^{K^*}$ as $n \to \infty$ as an immediate consequence of the multi-dimensional version of Donsker's Theorem. (Henceforth we shall omit the modifying "multi-dimensional version of"). Since $G$ is obviously continuous (using part (a) of Theorem 2.9), we have from (4.7) and the Continuous Mapping Theorem

$$\gamma_n = G(\xi_n) \Rightarrow G(D^*_{\xi_i}^{K^*}),$$

which completes the proof. Q.E.D.

**Corollary 6:** Let $X_1, \ldots, X_{K^*}$ be independent normal r.v.'s with mean zero and variance $\sigma_1^2, \ldots, \sigma_{K^*}^2$ respectively. Then

(a) $\frac{1}{\sqrt{n}} (T_n - n\lambda) \Rightarrow \max_{1 \leq k \leq K^*} X_k$ as $n \to \infty$,

(b) $E[\frac{1}{\sqrt{n}} (T_n - n\lambda)] \to E[\max_{1 \leq k \leq K^*} X_k]$ as $n \to \infty$,

(c) $\frac{1}{n} T_n \Rightarrow a$, e. a.e., $a_1$ as $n \to \infty$.

**Proof:** Letting $\pi$ denote the usual projection, we have from Theorem 5

90
and the Continuous Mapping Theorem

\[ \pi(v_n) = v_n(1) = \frac{1}{\sqrt{n}} (T_n - na_1) \overset{D}{\to} [G(D^{K*})](1) . \]

But \( D^{K*}(1) \overset{D}{=} (X_{1*}, \ldots, X_{K*}) \), so part (a) follows. Part (b) then follows from Theorem 2.14 and the uniform integrability shown in Lemma 3.6. Part (c) actually has no relation to the preceding theorem but is just a restatement of Lemma 3.3 (a).

Q.E.D.

To obtain corresponding results for the vector process \( \{v_n\} \), we define (for purposes of this section only)

\[ d = \begin{pmatrix} d_1 \\ \vdots \\ d_K \end{pmatrix}, \quad d_k = a_l - a_k, \quad 1 \leq k \leq K, \]

\[ \hat{w}_n = w_n - nd, \quad n = 1, 2, \ldots \]

(for completeness we take \( \hat{w}_0 = w_0 = 0 \)), and the random functions in \( D[0,1]^K \) generated by \( \{\hat{w}_n\} \),

\[ X_n(t) = \frac{1}{\sqrt{n}} \hat{w}_n[nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots . \]

Define \( F : D[0,1]^K \to D[0,1]^K \) by letting

\[ [F(x)](t) = \max_{1 \leq j \leq K*} [x_j(t)] - x_k(t), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K, \]

91
for $x \in D[0,1]^K$. Recall from Chapter 3 that $D^2$ is defined to be the
$K \times K$ diagonal matrix with diagonal elements $u_1^2, \ldots, u_K^2$. We let $D$
denote its square root.

Theorem 7: $X_n \Rightarrow F(D^K)$ as $n \to \infty$.

Proof: Let $\hat{U}_n = U_n - ma_k$ as before, but now take $\hat{U}_n = (\hat{U}_n^1, \ldots, \hat{U}_n^K)$.
According to Lemma 3.4

$$w_n^k = \max_{1 \leq j \leq K^*} U_n^j - U_n^k + \Delta(n^{1/2}), \quad 1 \leq k \leq K$$

and hence

$$\hat{w}_n^k = \max_{1 \leq j \leq K^*} \hat{U}_n^j - \hat{U}_n^k + \Delta(n^{1/2}), \quad 1 \leq k \leq K .$$

Then by Theorem 2.11 it suffices to prove that $\gamma_n = (\gamma_n^1, \ldots, \gamma_n^K) \Rightarrow F(D^K)$, where

$$\gamma_n^k(t) = \max_{1 \leq j \leq K^*} \hat{U}_n^j [nt] - \hat{U}_n^k [nt], \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K, \quad n = 1, 2, \ldots .$$

Define $\zeta_n \in D[0,1]^K$ by

$$\zeta_n(t) = \frac{1}{\sqrt{n}} \hat{U}_n [nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots .$$

Then for each $k = 1, \ldots, K$ and $t \in [0,1]$

$$\gamma_n^k(t) = \max_{1 \leq j \leq K^*} \zeta_n^j(t) - \zeta_n^k(t) = [F^k(\zeta_n)](t) .$$

Hence $\gamma_n = F(\zeta_n)$. Since $\zeta_n \Rightarrow D^K$ by Donsker's Theorem and $F$ is
clearly continuous, we have \( \gamma_n = F(\zeta_n) \Rightarrow F(D^K) \) by the Continuous Mapping Theorem.

Q.E.D.

**Corollary 8:** Let \( X_1, \ldots, X_K \) be independent normal r.v.'s with mean zero and variance \( \sigma_1^2, \ldots, \sigma_K^2 \) respectively. Let \( Y_k = \max_{1 \leq j \leq k*} X_j - X_k \) \((k = 1, \ldots, K)\) and \( Y = (Y_1, \ldots, Y_K) \).

(a) \( \frac{1}{\sqrt{n}} (w_n - \mu) \xrightarrow{D} Y \) as \( n \to \infty \),

(b) \( E[\frac{1}{\sqrt{n}} (w_n - \mu)] \xrightarrow{D} E(\max_{1 \leq j \leq K*} X_j) \) as \( n \to \infty \), \( 1 \leq k \leq K \),

(c) \( \frac{1}{n^k} \xrightarrow{a.s.} d_k \) as \( n \to \infty \), \( 1 \leq k \leq K \).

**Proof:** As before, we have from the Continuous Mapping Theorem

\[
\pi(X_n) = X_n(1) = \frac{1}{\sqrt{n}} (w_n - \mu) \xrightarrow{D} [F(D^K)](1).
\]

But \( D^K(1) \xrightarrow{D} (X_1, \ldots, X_K) \), so part (a) follows. Part (b) is then established by Lemma 3.6 and the observation that \( E(Y_k) = E(\max_{1 \leq j \leq K*} X_j) \) for all \( k \). Part (c) is just a restatement of Lemma 3.4 (a).

Q.E.D.

Note that for \( k = 1, \ldots, K* \) the limiting distribution of \( \frac{1}{\sqrt{n}} w_n^k \) is concentrated on \( \mathbb{R}^+ \) and has a probability mass of \( P[X_k = \max_{1 \leq j \leq K*} X_j] \) at the origin. This is of course the limiting probability that input process \( k \) is "pacing" the operation at the
time of the $n^\text{th}$ service (and hence that $w_n^k = w_n^\ast$). Using the methods
of the preceding section, it can be shown directly that for $k = 1, \ldots, K_\ast$

$$
\lim_{n \to \infty} P[w_n^k \leq x] = P[X_k = \max_{1 \leq j \leq K_\ast} X_j]P[M^k \leq x], \quad x \geq 0.
$$

If $K_\ast = 1$, then for $k = 1$ the preceding corollary gives us only
the uninteresting fact that $\frac{1}{\sqrt{n}}w_n^1 P \to 0$ as $n \to \infty$. In the following
section it will be shown that in this event $w_n^1$ and $w_n^\ast$ are
asymptotically identical and distributed as $M^1$. For $K_\ast < k \leq K$ the
limiting distribution shown above for $\frac{1}{\sqrt{n}}[w_n^k - n(a_1 - a_k)]$ is not
concentrated on $\mathbb{R}^+$ and does not have positive probability mass at any
single point.

**Corollary 9:** Define $X_1, \ldots, X_K$ as in the preceding corollary.

(a) $\frac{1}{\sqrt{n}^3} \left( w_n^1 - \frac{n^2}{2} d_k \right) D \to \int_0^1 [F(D^k)](t) dt \quad \text{as } n \to \infty ;$

(b) $E\left[ \frac{1}{\sqrt{n}^3} \left( w_n^k - \frac{n^2}{2} d_k \right) \right] \to \frac{2}{3} E\left( \max_{1 \leq j \leq K_\ast} X_j \right) \quad \text{as } n \to \infty , \quad 1 \leq k \leq K ;$

(c) $\frac{1}{n} w_n^k \xrightarrow{a.e.} \frac{1}{2} d_k \quad \text{as } n \to \infty , \quad 1 \leq k \leq K .$

**Proof:** From the piecewise-constant nature of $X_n(t)$ it is clear that

$$
\int_0^1 X_n(t) dt = \sum_{i=1}^{n-1} \left( \frac{1}{n} \right) \int_0^1 \hat{w}_i dt = \frac{1}{\sqrt{n}^3} \sum_{i=1}^{n-1} (w_i-\text{id})
$$

$$
= \frac{1}{\sqrt{n}^3} \left( w_{n-1}-d \sum_{i=1}^{n-1} 1 \right) .
$$
Since \( X_n \rightarrow F(D_{\xi}^K) \) and the integration functional is continuous by Theorem 2.9, we have from the Continuous Mapping Theorem

\[
\int_0^1 x_n(t) \, dt \xrightarrow{D} \int_0^1 [F(D_{\xi}^K)](t) \, dt .
\]

Combining (4.7) and (4.8) with the facts that \( \sum_{i=1}^n i \sim \frac{1}{2} n^2 \) and \( \frac{1}{\sqrt{n^3}} w_n \xrightarrow{P} 0 \), part (a) is proved. Parts (b) and (c) follow directly from Corollary 8 and Lemma 2.25.

Q.E.D.

In applications to actual assembly operations, a quantity likely to be of great interest is

\[
C_n = c \cdot w_n = \sum_{k=1}^K c_k w_n^k = \sum_{k=1}^K c_k \sum_{i=1}^n w_i^k,
\]

where \( c = (c_1, \ldots, c_K) \) is a vector of real constants. If \( c_k \) represents a cost per unit time of holding a type \( k \) item in inventory, then \( C_n \) gives the total cost of carrying in-process inventory during a production run of size \( n \). Applying the continuous operation of vector multiplication to the random vectors \( W_n \), we have from the preceding corollary

\[
c \cdot \left[ \frac{1}{\sqrt{n^3}} (W_n - \frac{n}{2} d) \right] \xrightarrow{D} c \cdot \int_0^1 [F(D_{\xi}^K)](t) \, dt ,
\]

or equivalently,

\[
\frac{1}{\sqrt{n^3}} [C_n - \frac{n}{2} \sum_{k=K+1}^K c_k (a_{-1} - a_k)] \xrightarrow{D} \sum_{k=1}^K c_k \int_0^1 [F^k(D_{\xi})](t) \, dt .
\]
Also

\[ (4.12) \mathbb{E}(\frac{1}{\sqrt{n^3}} [C_n - \frac{2}{n} \sum_{k=K^*+1}^{K} c_k(a_{1-k} - a_k)]) \rightarrow \frac{2}{3} \mathbb{E}(\max_{1 \leq k \leq K^*} X_k) \sum_{k=1}^{K} c_k. \]

Of course the limiting distribution in (4.11) is not known explicitly, nor is the right-hand side of (4.12). In several special cases, however, these expressions simplify nicely, as will be shown next.

4.3 Special Cases

If \( K^* = 1 \) (i.e., there exists a unique "slowest" input process), then our mapping \( G:D[0,1]^K \rightarrow D[0,1] \) reduces to the identity map from \( D[0,1] \) onto itself. Similarly, \( F:D[0,1]^K \rightarrow D[0,1]^K \) becomes

\[ [F^k(x)](t) = x^1(t) - x^k(t), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K, \]

or equivalently, \( [F(x)](\cdot) = Ax(\cdot) \), where \( A = (\alpha_{ij}) \) is a \( K \times K \) matrix with

\[
\alpha_{ij} = \begin{cases} 
1 & \text{if } 2 \leq i \leq K \text{ and } j = 1 \\
-1 & \text{if } 2 \leq i = j \leq K \\
0 & \text{otherwise}
\end{cases}
\]

It is easily verified that \( AD = Q^{1/2} \), where \( Q = (q_{ij}) \) is a \( K \times K \) matrix with

\[
q_{ij} = \begin{cases} 
0 & \text{if } i = 1 \text{ or } j = 1 \\
\sigma_i^2 + \sigma_j^2 & \text{if } i = j > 1 \\
\sigma_i^2 & \text{otherwise}
\end{cases}
\]
Thus Theorem 7 becomes $\chi_n \Rightarrow q_{1/2}^{1/2}$, and Corollary 9 reduces to

$$
\frac{1}{\sqrt{n^3}} (W_n - \frac{n^2}{2} d) \overset{D}{\rightarrow} \int_0^1 q_{1/2}^{1/2} (t) dt \overset{D}{\rightarrow} \sqrt{\frac{1}{2}} q_{1/2}^{1/2} N(0, 1),
$$

since $\int_0^1 q_{1/2}^{1/2} (t) dt \overset{D}{\rightarrow} \sqrt{\frac{1}{2}} N(0, 1)$ by Theorem 2.10. From these facts, all parts of the following proposition except the last follow directly.

**Corollary 10:** If $K^* = 1$, then

(a) $\nu_n \Rightarrow \sigma_{1}^{1/2}$ as $n \rightarrow \infty$.

(b) $\frac{1}{\sigma_{1}^{1/2}} (T_n - na) \overset{D}{\rightarrow} N(0, 1)$ as $n \rightarrow \infty$.

E[$\frac{1}{\sqrt{n}} (T_n - na)$] $\rightarrow$ 0 as $n \rightarrow \infty$.

(c) $\chi_n \Rightarrow q_{1/2}^{1/2} K$ as $n \rightarrow \infty$.

(d) $\frac{1}{\sqrt{n}} (w_n - nd) \overset{D}{\rightarrow} q_{1/2}^{1/2} N(0, 1)$ as $n \rightarrow \infty$.

E[$\frac{1}{\sqrt{n}} (w_n^k - nd_k)$] $\rightarrow$ 0 as $n \rightarrow \infty$, $1 \leq k \leq K$.

(e) $\sqrt{\frac{3}{n}} (\bar{w}_n - \frac{n^2}{2} d) \overset{D}{\rightarrow} q_{1/2}^{1/2} N(0, 1)$ as $n \rightarrow \infty$.

E[$\frac{1}{\sqrt{n}} (\bar{w}_n^k - \frac{n^2}{2} d_k)$] $\rightarrow$ 0 as $n \rightarrow \infty$, $1 \leq k \leq K$.

Thus, if $C_n$ is defined by (4.9),

$$
C_n - \frac{n^2}{2} \sum_{k=2}^{K} c_k(a_{1-k} - a_k)
$$

$\overset{D}{\rightarrow} N(0, 1)$ as $n \rightarrow \infty$, 

$$
\sqrt{\frac{1}{3} n^3 \left[ \sum_{k=2}^{K} c_k^2 (a_{1-k} - a_k)^2 + \sigma_{1}^{2} \left( \sum_{k=2}^{K} c_k \right)^2 \right]} 
$$

97
\[ E\left( \frac{1}{\sqrt{n}} \left[ C_n - \frac{n}{2} \sum_{k=2}^{K} c_k(a_{\ell k} - a_k) \right] \right) \to 0 \quad \text{as} \quad n \to \infty. \]

\( w_n \xrightarrow{D} M \) as \( n \to \infty \) and \( w_n \xrightarrow{D} M \) as \( n \to \infty \).

**Proof of (f):** That \( w_n \xrightarrow{D} M \) follows by specializing Theorem 4. To see that the same is true for \( w_n \), note that from our basic definitions

\[ w_n = w_n^* + (U_n^k - U_n^k_\perp) \, . \]

Thus

\[ 0 \leq w_n^* - w_n = \max_{1 \leq k \leq K} U_n^k - U_n^k_\perp \leq \sum_{k=2}^{K} (U_n^k - U_n^k_\perp)^\perp \, . \]

But \( \frac{1}{n} (U_n^k - U_n^k_\perp) \xrightarrow{a.e.} (a_{\ell k} - a_k) < 0 \) for each \( k = 2, \ldots, K \). This implies that \( (U_n^k - U_n^k_\perp)^\perp \xrightarrow{a.e.} 0 \) and hence \( (U_n^k - U_n^k_\perp)^\perp \xrightarrow{a.e.} 0 \). So

\[ (w_n^* - w_n) \xrightarrow{a.e.} 0 \, , \]

which completes the proof.

Q.E.D.

If \( K^* = 2 \), then some but not all of the general theorems can be made explicit. We shall mention only those results which simplify significantly from their original form. The important points to note are that if \( X = (X_1, \ldots, X_K) \) is distributed \( N(0, D^2) \), then

\[ \max(X_1, X_2) - X_1 = (X_2 - X_1)^\perp \xrightarrow{D} [N(0, \sigma_1^2 + \sigma_2^2)]^\perp \, , \]

\[ \max(X_1, X_2) - X_2 = (X_1 - X_2)^\perp \xrightarrow{D} [N(0, \sigma_1^2 + \sigma_2^2)]^\perp \, , \]

98
\[ P[X_1 = \max(X_1, X_2)] = P[X_2 = \max(X_1, X_2)] = P[[N(0, \sigma_1^2 + \sigma_2^2)]^+ > 0] = \frac{1}{2}, \]

\[ E[\max(X_1, X_2)] = E[N(0, \sigma_1^2 + \sigma_2^2)]^+ = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2\pi}}. \]

Using these facts to specialize Theorem 4 and Corollaries 6, 8 and 9, the following proposition is immediate.

**Corollary 11:** If \( K^* = 2 \), then

(a) \( w_n^* \overset{D}{\to} w^* \), where

\[ P[w^* \leq x] = \frac{1}{2} P[\bar{M}_1^2 \leq x] + \frac{1}{2} P[M_2^2 \leq x], \quad x \geq 0. \]

(b) \( \lim_{n \to \infty} P\left( \frac{1}{\sqrt{n}} \left( T_n - na_\perp \right) \leq x \right) = \Phi\left( \frac{x}{\sigma_\perp} \right) \Phi\left( \frac{x}{\sigma_\parallel} \right) \),

where \( \Phi(\cdot) \) is the standard normal d.f. Also

\[ E\left( \frac{1}{\sqrt{n}} \left( T_n - na_\perp \right) \right) \to \sqrt{\frac{\sigma_\perp^2 + \sigma_\parallel^2}{2\pi}} \quad \text{as} \quad n \to \infty. \]

(c) For \( k = 1, 2 \)

\[ \lim_{n \to \infty} P\left( \frac{w_n^k}{\sqrt{(\sigma_\perp^2 + \sigma_\parallel^2)n}} \leq x \right) = \begin{cases} \frac{1}{2} & \text{if } x < 0 \\ \Phi(x) & \text{if } x \geq 0 \end{cases}. \]

For \( k = 1, \ldots, K \)

\[ E\left( \frac{1}{\sqrt{n}} \left[ w_n^k - n(a_\perp - a_\parallel) \right] \right) \to \sqrt{\frac{\sigma_\perp^2 + \sigma_\parallel^2}{2\pi}} \quad \text{as} \quad n \to \infty. \]

(d) \( E\left( \frac{1}{\sqrt{n^2}} \left[ w_n^k - \frac{n^2}{2} (a_\perp - a_\parallel) \right] \right) \to \sqrt{\frac{\sigma_\perp^2 + \sigma_\parallel^2}{2\pi}} \quad \text{as} \quad n \to \infty, \quad 1 \leq k \leq K. \)
Thus, if $C_n$ is defined by (4.9),

$$E\left(\frac{1}{\sqrt{n}} \left[ C_n - \frac{n^2}{2} \sum_{k=1}^K c_k (a_1 - a_k) \right] \right) \to \frac{2}{\pi} \frac{\sigma_1^2 + \sigma_2^2}{\int} \sum_{k=1}^K c_k.$$ 

The formula given for the asymptotic distribution of $T_n$ above is of course not peculiar to the case $K^* = 2$. From Corollary 6 we have for general $K^*$ that

$$\lim_{n \to \infty} P\left\{ \frac{1}{\sqrt{n}} (T_n - n a_1) \leq \right\} = \prod_{k=1}^{K^*} \Phi\left( \frac{X_k}{\sigma_k} \right), \quad x \in \mathbb{R},$$

which can be obtained from standard tables. Considering the interpretation of $T_n$ as the completion time for a production run, this formula is of particular interest.

There is another special case in which Theorem 4 for the $\{V_n\}$ process simplifies partially. If $\sigma_1 = \ldots = \sigma_{K^*}$ ($K^*$ being arbitrary), then the r.v.'s $X_1, \ldots, X_{K^*}$ are i.i.d. and by symmetry

$$P\{X_k = \max(X_1, \ldots, X_{K^*})\} = \frac{1}{K^*}, \quad 1 \leq k \leq K^*.$$ 

The following proposition is then immediate.

**Corollary 12:** If $\rho < 1$ and $\sigma_1 = \ldots = \sigma_{K^*}$, then $w_n \overset{D}{\to} w^*$ as $n \to \infty$, where

$$P\{w^* \leq x\} = \frac{1}{K^*} \sum_{k=1}^K P\{M_k \leq x\}.$$ 

None of the three special cases treated thus far involve any assumptions as to the specific form of the interarrival time or service
time distributions. There are a variety of such assumptions under which the exact distributions of the maxima $M^k$ are known (i.e., under which the limiting waiting time distributions in the corresponding GI/G/1 queues are known). Given the exact distributions of $M^1, \ldots, M^{K^*}$, Theorem 4 can of course be made explicit, but we shall not attempt to catalog all such results here. Rather, one particularly simple example will be treated by way of illustration. Readers familiar with queueing theory will recall that if $\{u^k_n\}$ and $\{v_n\}$, are both exponentially distributed and $b < a_k$ then (cf. Prabhu (1965), p. 18)

$$P\{M^k \leq x\} = \left(1 - \frac{b}{a_k}\right) + \frac{b}{a_k} \left[1 - e^{-\frac{a_k-b}{a_k}x}\right], \quad x \geq 0.$$  

The following proposition is then immediate.

**Corollary 13:** If $u^1_n, \ldots, u^{K^*}_n$ are all exponentially distributed with mean $a > b$ (so that $\sigma_i^2 = \cdots = \sigma_{K^*}^2 = a^2$), and if $v_n$ is also exponentially distributed, then

$$\lim_{n \to \infty} P\{w^*_n \leq x\} = \left(1 - \frac{b}{a}\right) + \frac{b}{a} \left[1 - e^{-\frac{a-b}{ab}x}\right], \quad x \geq 0.$$
CHAPTER 5

The Case $c > 1$

It is assumed throughout this chapter that

\[ b > a_1 = \cdots = a_{K*} > a_{K*+1} \geq \cdots \geq a_K. \]

Our principle results are again functional limit theorems for $\{T_n\}$ and $\{W_n\}$. Corollaries dealing with $\{V_n^*\}$ and $\{W_n\}$ are obtained from the latter. In contrast both to the results presented in the last chapter for $p < 1$ and to those of the next chapter for $p = 1$, all of the limit theorems obtained here can be stated in a completely explicit form.

To obtain a functional limit theorem for $\{T_n\}$ we define (for purposes of this chapter)

\[ \hat{T}_n = T_n - nb, \quad n = 1, 2, \ldots \]

and the random functions in $D[0, 1]$ generated by $\{T_n\}$,

\[ v_n(t) = \frac{1}{\sqrt{n}} \hat{T}_{\lfloor nt \rfloor}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots. \]

**Theorem 1:** $v_n \Rightarrow \sigma_S$ as $n \to \infty$.

**Proof:** Let $\hat{V}_n = V_n - nb, \quad n = 0, 1, \ldots$. According to Lemma 3.3,

\[ T_n = V_n + \Delta(n^{1/2}), \]

and hence
\( \hat{T}_n = \hat{V}_n + \Delta(n^{1/2}) \).

Then by Theorem 2.7 it suffices to show that \( \xi_n \Rightarrow \sigma_S \), where

\[ \xi_n = \frac{1}{\sqrt{n}} \hat{V}_{[nt]} , \quad 0 \leq t \leq 1 , \quad n = 1,2,\ldots \]

But this is a direct consequence of Donsker's Theorem, so the proof is complete.

Q.E.D.

**Corollary 2:** (a) \( \frac{1}{\sqrt{n}} (T_n - nb) \overset{D}{\rightarrow} N(0, \sigma_S^2) \) as \( n \rightarrow \infty \).

(b) \( E[\frac{1}{\sqrt{n}} (T_n - nb)] \rightarrow 0 \) as \( n \rightarrow \infty \).

(c) \( \frac{1}{n} T_n \overset{a.s.}{\rightarrow} b \) as \( n \rightarrow \infty \).

**Proof:** Part (a) follows as usual from the Continuous Mapping Theorem by applying the projection \( \pi \) to \( V_n \). Part (b) is then immediate from Lemma 3.6, and part (c) is just a restatement of Lemma 3.3 (a).

Q.E.D.

To obtain corresponding results for the vector process \( \{\hat{\omega}_n\} \), we define (for purposes of this chapter)

\[ d = \begin{pmatrix} b-a_1 \\ \vdots \\ b-a_K \end{pmatrix} , \]

\[ \hat{\omega}_n = \omega_n - nd , \quad n = 1,2,\ldots \]

\[ \chi_n(t) = \frac{1}{\sqrt{n}} \hat{W}_{[nt]} , \quad 0 \leq t \leq 1 , \quad n = 1,2,\ldots \]
Recall that in Chapter 3 we defined \( \Phi = \text{Cov}(X_n) \).

**Theorem 3:** \( X_n \Rightarrow \Phi^{1/2} \mathbf{Z} \) as \( n \to \infty \).

**Proof:** Let \( \hat{S}_n = S_n - \text{nd} \). According to Lemma 3.4,

\[
\hat{w}_n^k = \hat{S}_n^k + \Delta(n^{1/2}) , \quad 1 \leq k \leq K ,
\]

and hence

\[
\hat{w}_n^k = \hat{S}_n^k + \Delta(n^{1/2}) , \quad 1 \leq k \leq K .
\]

So by Theorem 2.11 it suffices to show that \( \zeta_n \Rightarrow \Phi^{1/2} \mathbf{Z} \), where

\[
\zeta_n(t) = \frac{1}{\sqrt{n}} \hat{S}_{[nt]} , \quad 0 \leq t \leq 1 , \quad n = 1,2,...
\]

But this is a direct consequence of Donsker's Theorem, so the proof is complete.

Q.E.D.

**Corollary 4:**

(a) \( \frac{1}{\sqrt{n}} (w_n - \text{nd}) \overset{D}{\to} N(0, \Phi) \) as \( n \to \infty \).

(b) \( E[\frac{1}{\sqrt{n}} (w_n^k - \text{nd}_k)] \to 0 \) as \( n \to \infty \), \( 1 \leq k \leq K \).

(c) \( \frac{1}{n} w_n \overset{a.s.}{\to} d_k = (b - a_k) \) as \( n \to \infty \), \( 1 \leq k \leq K \).

**Proof:** Immediate from the Continuous Mapping Theorem and Lemmas 3.4 and 3.6.

Q.E.D.

**Corollary 5:**

(a) \( \frac{\sqrt{3}}{n^3} (w_n - \frac{n^2}{2} d) \overset{D}{\to} N(0, \Phi) \) as \( n \to \infty \).
(b) \( \sqrt{n} \left( W_n^k - \frac{n}{2} \delta_k \right) \) \( \overset{a.e.}{\rightarrow} \{\frac{1}{2} \delta_k \} \) as \( n \rightarrow \infty \), \( 1 \leq k \leq K \).

(c) \( \frac{1}{n^2} \bar{W}_n^k \overset{a.e.}{\rightarrow} \frac{1}{2} \delta_k \) as \( n \rightarrow \infty \), \( 1 \leq k \leq K \).

Proof: Applying the continuous integration mapping to \( \chi_n(\cdot) \) as before, we have

\[
\frac{1}{\sqrt{n}} \left[ W_{n-1} - \sum_{i=1}^{n-1} i \right] = \int_0^1 \chi_n(t) \, dt \overset{D}{\rightarrow} \int_0^1 \frac{1}{\sqrt{2}} \xi K(t) \, dt .
\]

Observing that \( \frac{1}{\sqrt{n^3}} W_n \overset{P}{\rightarrow} 0 \) by Corollary 4, that \( \sum_{i=1}^n i \sim \frac{1}{2} n^2 \), and that

\[
\int_0^1 \frac{1}{\sqrt{2}} \xi K(t) \, dt = \frac{1}{\sqrt{2}} \int_0^1 \xi K(t) \, dt = \sqrt{\frac{1}{2}} \sqrt{n} \overset{D}{\rightarrow} N(0, I)
\]

by Theorem 2.10, part (a) is proved. Parts (b) and (c) follow from Corollary 4 and Lemma 2.25.

Q.E.D.

As was true in the case of \( \rho < 1 \) and \( K^* = 1 \), we have found that \( W_n \) converges in distribution to a multivariate normal after proper translation and scaling. Thus explicit results are again obtainable for the quantity \( C_n \) defined by (4.9). Specifically, in the present case we have

\[
c \left[ \sqrt{\frac{n}{3}} (W_n - \frac{n}{2} \delta) \right] \overset{D}{\rightarrow} N(0, c \xi \xi^T) \text{ as } n \rightarrow \infty ,
\]

or equivalently,
\[
\frac{c_n - \frac{n^2}{2} \sum_{k=1}^{K} c_k (b-a_k)}{\sqrt{\frac{1}{3} n^3 \left[ \sum_{k=1}^{K} c_k^2 \sigma_k^2 + \sum_{k=1}^{K} c_k^2 \right]}} \overset{D}{\to} N(0,1) \text{ as } n \to \infty.
\]

Moreover, \( \frac{1}{n^2} c_n \overset{a.s.}{\to} \frac{1}{2} \sum_{k=1}^{K} c_k (b-a_k) \) as \( n \to \infty \), and
\[
E\left( \frac{1}{n^3} \left[ c_n - \frac{n^2}{2} \sum_{k=1}^{K} c_k (b-a_k) \right] \right) \to 0 \text{ as } n \to \infty.
\]

Treatment of the case \( \rho > 1 \) is now complete except for the units' waiting time process \( \{w_n^x\} \). Since \( w_n^x \) is of relatively little significance in most applications, we shall not bother to state a functional limit theorem for this process, although such is easily obtainable by generalizing the proof of the following proposition. Let the covariance matrix \( \hat{\Phi} \) be partitioned
\[
\hat{\Phi} = \begin{bmatrix}
\hat{\Phi} & \hat{\Phi}' \\
\hat{\Phi}' & \hat{\Phi}''
\end{bmatrix},
\]

where \( \hat{\Phi} \) is \( K^x \times K^x \). For ease of expression, let us agree that if \( X = (X_1, \ldots, X_{K^x}) \), then \( \min\{X\} \) will be taken to mean \( \min_{1 \leq k \leq K^x} X_k \).

**Theorem 6:**

(a) \( \frac{1}{\sqrt{n}} (w_n^x - nd_1) \overset{D}{\to} \min\{N(0, \hat{\Phi})\} \) as \( n \to \infty \).

(b) \( E\left[ \frac{1}{\sqrt{n}} (w_n^x - nd_1) \right] \to E[\min\{N(0, \hat{\Phi})\}] \) as \( n \to \infty \).

**Proof:** Let
\[
Y_n = \begin{pmatrix}
w_n^x \\
\vdots \\
K^x
\end{pmatrix}, \quad Y_n = \begin{pmatrix}
w_n^x - nd_1 \\
\vdots \\
K^x - nd_1
\end{pmatrix}.
\]
According to Lemma 3.5 (b),

\[ w^*_n = \min \{ y_n \} + \Delta(n^{1/2}) \]

and hence

\[ \frac{1}{\sqrt{n}} (w^*_n - n \Delta) = \min \left\{ \frac{1}{\sqrt{n}} \hat{y}_n \right\} + \frac{1}{\sqrt{n}} \Delta(n^{1/2}) . \]

Now \( \frac{1}{\sqrt{n}} \hat{y}_n \xrightarrow{D} N(0, \theta) \) by Corollary 4, and \( \min: \mathbb{R}^X \rightarrow \mathbb{R} \) is obviously continuous, so

\[ \min \left\{ \frac{1}{\sqrt{n}} \hat{y}_n \right\} \xrightarrow{D} \min \{ N(0, \theta) \} . \]

Since \( \frac{1}{\sqrt{n}} \Delta(n^{1/2}) \xrightarrow{P} 0 \), the proof of (a) is complete. Part (b) then follows from Lemma 3.6.

Q.E.D.
CHAPTER 6

The Case \( \rho = 1 \)

It is assumed throughout this chapter that

\[
b = a_1 = \cdots = a_{k^*} > a_{k^*+1} \geq \cdots \geq a_K.
\]

Functional limit theorems are developed for \( \{T_n\} \) and \( \{w_n\} \), and from the latter result corollaries are obtained dealing with \( \{W_n\} \) and \( \{w^*_n\} \). None of the results obtained can be stated in an explicit form for arbitrary \( K^* \), but when \( k^* = 1 \) they simplify considerably. This special case is treated in Section 6.2.

6.1 General Results

Following a now familiar procedure, we define (for purposes of this section)

\[
\hat{T}_n = T_n - nb, \quad n = 1, 2, \ldots
\]

\[
\nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_n [nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots
\]

We further define mappings \( G: \mathcal{D}[0,1]^{K^*+1} \to \mathcal{D}[0,1] \) and \( g: \mathcal{D}[0,1]^{K^*+1} \to \mathbb{R} \) as follows. For \( x \in \mathcal{D}[0,1]^{K^*+1} \) and \( t \in [0,1] \), let

\[
[G(x)](t) = \max_{1 \leq k \leq K^*} \left\{ \sup_{0 \leq \tau \leq t} \left[ x^k(\tau) + x^{K^*+1}(t) - x^{K^*+1}(\tau) \right] \right\}.
\]

\[
g(x) = (\varphi \circ G)(x) = \max_{1 \leq k \leq K^*} \left\{ \sup_{0 \leq t \leq 1} \left[ x^k(t) + x^{K^*+1}(1) - x^{K^*+1}(t) \right] \right\}.
\]

108
Theorem 1: \( \gamma_n \to G(D^{K+1}_n) \) as \( n \to \infty \), where \( D \) is a \((K+1) \times (K+1)\) diagonal matrix with diagonal elements \( \sigma_1, \ldots, \sigma_K, \sigma_S \).

Proof: Let \( \hat{V}_n = V_n - nb \) and \( \hat{U}_n^k = U_n^k - nb \), \( 1 \leq k \leq K^* \). Then define

\[
\hat{Y}_n = \begin{pmatrix}
\hat{U}_n^1 \\
\vdots \\
\hat{U}_n^{K^*} \\
\hat{V}_n
\end{pmatrix}, \quad n = 0, 1, 2, \ldots
\]

According to Lemma 3.3,

\[
T_n = \max_{1 \leq k \leq K^*} \left\{ \max_{0 \leq j \leq n} \left[ (U_n^k + \hat{V}_n - V_j) \right] + \Delta(n^{1/2}) \right\},
\]

and hence

\[
\hat{T}_n = \max_{1 \leq k \leq K^*} \left\{ \max_{0 \leq j \leq n} \left[ (\hat{U}_n^k + \hat{V}_n - \hat{V}_j) \right] + \Delta(n^{1/2}) \right\}.
\]

So by Theorem 2.7 it suffices to show that \( \gamma_n \to G(D^{K+1}_n) \), where

\[
(6.1) \quad \gamma_n(t) = \frac{1}{\sqrt{n}} \max_{1 \leq k \leq K^*} \left\{ \max_{0 \leq j \leq \lfloor nt \rfloor} \left( \hat{U}_n^k + \hat{V}_n - V_j \right) \right\}
\]

\[
= \max_{1 \leq k \leq K^*} \left\{ \sup_{0 \leq \tau \leq t} \left( \frac{1}{\sqrt{n}} V_{\lfloor nt \rfloor} + \hat{V}_n - V_{\lfloor nt \rfloor} \right) \right\}.
\]

Now define

\[
\zeta_n(t) = \frac{1}{\sqrt{n}} Y_{\lfloor nt \rfloor}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots
\]
Then (6.1) can be re-expressed as $\gamma_n = G(\xi_n)$. By Donsker's Theorem we have that $\xi_n \Rightarrow D^*_K \Rightarrow 1$, and it follows directly from Theorem 2.9 that $G$ is continuous. Thus, by the Continuous Mapping Theorem,

$$\gamma_n = G(\xi_n) \Rightarrow G(D^*_K \Rightarrow 1),$$

and the proof is complete.

Q.E.D.

Corollary 2: (a) $\frac{1}{\sqrt{n}} (T_n - nb) \Rightarrow D(\hat{D}^*_K \Rightarrow 1)$ as $n \to \infty$.

(b) $E[\frac{1}{\sqrt{n}} (T_n - nb)] \Rightarrow E[G(\hat{D}^*_K \Rightarrow 1)]$ as $n \to \infty$.

(c) $\frac{1}{\sqrt{n}} T_n \Rightarrow b$ as $n \to \infty$.

Proof: Part (a) follows from the Continuous Mapping Theorem by applying the projection $\pi$ to $\nu_n$. Parts (b) and (c) then follow from Lemmas 6 and 3.3 (a).

Q.E.D.

In order to treat the vector process \{\bar{w}_n\}, we define

$$d = \begin{pmatrix} d_1 \\ \vdots \\ d_K \end{pmatrix}, \quad d_k = b - s_k, \quad 1 \leq k \leq K,$$

$$\hat{w}_n = w_n - nd, \quad n = 1, 2, \ldots$$

$$\chi_n(t) = \frac{1}{\sqrt{n}} \hat{w}_{nt}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots$$

Further define mappings $F: D[0,1]^K \to D[0,1]^K$ and $f: D[0,1]^K \to \mathbb{R}^K$ as

110
follows. For \( x \in D[0,1]^K \) and \( t \in [0,1] \), let

\[
[F^k(x)](t) = x^k(t) - \min_{1 \leq j \leq K^*} \{ \inf_{0 \leq \tau \leq t} [x^j(\tau)] \}, \quad 1 \leq k \leq K,
\]
\[
f^k(x) = (\pi^o F)(x) = x^k(1) - \min_{1 \leq j \leq K^*} \{ \inf_{0 \leq \tau \leq 1} [x^j(\tau)] \}, \quad 1 \leq k \leq K.
\]

**Theorem 3:** \( \chi_n \Rightarrow F(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \) as \( n \to \infty \).

**Proof:** Let \( \hat{s}^k_n = s^k_n - n(b-a_k) \), \( 1 \leq k \leq K \), and

\[
\hat{s}_n = \left( \begin{array}{c}
\hat{s}_n^1 \\
\vdots \\
\hat{s}_n^K 
\end{array} \right), \quad n = 1, 2, \ldots .
\]

According to Lemma 3.4,

\[
\omega_n^k = s^k_n - \min_{1 \leq j \leq K^*} \{ s^j_i + G(n^{1/2}) \}, \quad 1 \leq k \leq K,
\]
\[
0 \leq i \leq n
\]
and hence

\[
\omega_n^k = \hat{s}_n^k - \min_{1 \leq j \leq K^*} \{ \hat{s}_i^j + G(n^{1/2}) \}, \quad 1 \leq k \leq K.
\]
\[
0 \leq i \leq n
\]

Then by Theorem 2.11 it suffices to show that \( \gamma_n \Rightarrow F(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \), where

\[
(6.2) \quad \gamma^k_n(t) = \frac{1}{\sqrt{n}} \hat{s}^k_{[nt]} - \min_{1 \leq j \leq K^*} \{ \inf_{0 \leq \tau \leq t} \hat{s}^j_{[nt]} \}, \quad 1 \leq k \leq K.
\]
If we define
\[
\zeta_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots
\]
then (6.2) can be simply re-expressed as \( \gamma_n = F(\zeta_n) \). Since
\( \zeta_n \Rightarrow \frac{1}{\frac{1}{2}} \xi^K \) by Donsker's Theorem and \( F \) is clearly continuous by
Theorem 2.9, it follows from the Continuous Mapping Theorem that
\[
\gamma_n = F(\zeta_n) \Rightarrow F(\frac{1}{\frac{1}{2}} \xi^K),
\]
which completes the proof.

Q.E.D.

Corollary 4: (a) \( \frac{1}{\sqrt{n}} (w_n - \text{nd}) D, f(\frac{1}{2} \xi^K) \) as \( n \to \infty \).

(b) \( \mathbb{E}[\frac{1}{\sqrt{n}} (w_n^k - \text{nd}_k)] \to \mathbb{E}[f(\frac{1}{2} \xi^K)] \) as \( n \to \infty, \ 1 \leq k \leq K \).

(c) \( \frac{1}{n} w_n^k \to d_k \) a.s. as \( n \to \infty, \ 1 \leq k \leq K \).

Proof: Part (a) follows as usual from the Continuous Mapping Theorem, and parts (b) and (c) then follow from Lemmas 3.6 and 3.4.

Q.E.D.

Corollary 5: (a) \( \sqrt{\frac{3}{n^3}} (W_n - \frac{n^2}{2} \text{d}) D, \int_0^1 [F(\frac{1}{2} \xi^K)](t)dt \) as \( n \to \infty \).

(b) \( \mathbb{E}[\frac{1}{\sqrt{n}} (W_n^k - \frac{n^2}{2} \text{d}_k)] \to \frac{2}{3} \mathbb{E}[f(\frac{1}{2} \xi^K)] \) as \( n \to \infty, \ 1 \leq k \leq K \).

(c) \( \frac{1}{n^2} W_n^k \to \frac{1}{2} d_k \) a.s. as \( n \to \infty, \ 1 \leq k \leq K \).

Proof: As in the cases \( \rho < 1 \) and \( \rho > 1 \), part (a) follows from the
Continuous Mapping Theorem by applying the integration functional to
\( X_n(\cdot) \). Parts (b) and (c) follow directly from Lemma 2.25.

Q.E.D.

Corollary 6: (a) \( \frac{1}{\sqrt{n}} w^*_n \Rightarrow \min_{1 \leq k \leq K^*} [f^k(\frac{1}{\sqrt{2}} K)] \) as \( n \to \infty \).

(b) \( E(\frac{1}{\sqrt{n}} w^*_n) \to E\{ \min_{1 \leq k \leq K^*} [f^k(\frac{1}{\sqrt{2}} K)] \} \) as \( n \to \infty \).

Proof: According to Lemma 3.5

\[
 w^*_n = \min_{1 \leq k \leq K^*} \hat{w}^k_n + \Delta(n^{1/2}) = \min_{1 \leq k \leq K^*} \hat{w}^k_n + \Delta(n^{1/2}) .
\]

Now

\[
 \min_{1 \leq k \leq K^*} \left[ \frac{1}{\sqrt{n}} \hat{w}^k_n \right] \Rightarrow \min_{1 \leq k \leq K^*} [f^k(\frac{1}{\sqrt{2}} K)]
\]

by Corollary 4 and the Continuous Mapping Theorem. The proof of (a) is then completed by the observation that \( \frac{1}{\sqrt{n}} \Delta(n^{1/2}) \Rightarrow 0 \). Part (b) then follows from Lemma 3.6.

Q.E.D.

Again we observe that a functional limit theorem for \( w^*_n \) can be obtained easily by generalizing the argument above.

6.2 The Special Case \( K^* = 1 \)

If \( K^* = 1 \), then our mapping \( f: D[0,1]^K \to \mathbb{R}^K \) becomes

\[
f^k(x) = x^k(1) - \inf_{0 \leq t \leq 1} [x^1(t)] , \quad 1 \leq k \leq K .
\]

Now the first component process of \( \frac{1}{\sqrt{2}} \xi^* K(\cdot) \) has the same distribution as \( \sqrt{\frac{\sigma^2 + \sigma_x^2}{2}} \xi(\cdot) \), so in particular
\[ \mathbb{I}^{1/2, K} \sqrt{\frac{\sigma^2_s + \sigma^2_l}{\sigma_s + \sigma_l}} [\xi(1) - \inf_{0 \leq t \leq 1} \xi(t)] \]

\[ = \sqrt{\frac{\sigma^2_s + \sigma^2_l}{\sigma_s + \sigma_l}} \sup_{0 \leq t \leq 1} [\xi(1) - \xi(t)] \mathbb{I}^{1/2, K} \sqrt{\frac{\sigma^2_s + \sigma^2_l}{\sigma_s + \sigma_l}} \sup_{0 \leq t \leq 1} \xi(t). \]

Also, for each \( k = 1, \ldots, K \)

\[ \mathbb{E}[\mathbb{I}^{k/2, K}] = \mathbb{E}\left[ (\inf_{0 \leq t \leq 1} \left[ \sqrt{\frac{\sigma^2_s + \sigma^2_l}{\sigma_s + \sigma_l}} \xi(t) \right]) \right] \]

\[ = \sqrt{\frac{\sigma^2_s + \sigma^2_l}{\sigma_s + \sigma_l}} \mathbb{E}\left[ \sup_{0 \leq t \leq 1} \xi(t) \right]. \]

Similarly, the mapping \( g: \mathbb{D}[0, 1]^2 \rightarrow \mathbb{R} \) simplifies so that

\[ \mathbb{E}[g(D\xi^2)] = \sqrt{\frac{\sigma^2_s + \sigma^2_l}{\sigma_s + \sigma_l}} \mathbb{E}\left[ \sup_{0 \leq t \leq 1} \xi(t) \right]. \]

Using these observations and the well-known fact that the supremum of Brownian Motion has a standard half-normal distribution, the following proposition is established.

**Corollary 7:** If \( K^* = 1 \), then

(a) \( \lim_{n \to \infty} \mathbb{P}\left[ \frac{w}{\sqrt{(\sigma^2_s + \sigma^2_l)n}} \leq x \right] = \lim_{n \to \infty} \mathbb{P}\left[ \frac{w^*}{\sqrt{(\sigma^2_s + \sigma^2_l)n}} \leq x \right] \)

\[ = \begin{cases} 
0 & \text{if } x < 0 \\
\sqrt{\frac{2}{\pi}} \int_0^x e^{-y^2/2} dy & \text{if } x \geq 0 
\end{cases} \]
(b) \( E \left[ \frac{1}{\sqrt{n}} (w_n - nd_k) \right] \to \sqrt{\frac{2(\sigma^2 + \sigma^2_0)}{\pi}} \) as \( n \to \infty \), \( 1 \leq k \leq K \).

(c) \( E \left[ \frac{1}{\sqrt{n^3}} (w_n^k - \frac{n^2}{2} d_k) \right] \to \frac{2}{3} \sqrt{\frac{2(\sigma^2 + \sigma^2_0)}{\pi}} \) as \( n \to \infty \), \( 1 \leq k \leq K \).

Thus, if \( C_n \) is defined by (4.9),

\[
E \left[ \frac{1}{\sqrt{n^3}} (C_n - \frac{n^2}{2} \sum_{k=1}^{K} c_k d_k) \right] \to \frac{2}{3} \sqrt{\frac{2(\sigma^2 + \sigma^2_0)}{\pi}} \sum_{k=1}^{K} c_k \text{ as } n \to \infty
\]

(d) \( E \left[ \frac{1}{\sqrt{n}} (T_n - nb) \right] \to \sqrt{\frac{2(\sigma^2 + \sigma^2_0)}{\pi}} \) as \( n \to \infty \).
CHAPTER 7

Additional Remarks, Examples and Applications

7.1 The Processes $D(t)$ and $Q(t)$

For each of the cases $\rho < 1$, $\rho = 1$ and $\rho > 1$ we have found a random element $\nu$ of $D[0,1]$ such that

$$\nu_n \Rightarrow \nu \text{ as } n \to \infty,$$

where

$$\nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_n[nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots,$$

$$\hat{T}_n = T_n - n\mu, \quad n = 1, 2, \ldots,$$

$$\mu = \max(a, b) > 0.$$

Moreover, it is easily verifiable from the sample path continuity of Brownian Motion that in each case the random function $\nu$ has sample paths which are almost surely continuous. Combining this observation with the fact that $\{D(t); 0 \leq t < \infty\}$ is the counting process corresponding to $\{T_n; n = 0, 1, \ldots\}$, the following proposition is immediate from Theorem 2.15. Let $\mu$ be as above and define

$$\xi_n(t) = \frac{1}{\sqrt{n}} [D(nt) - nt/\mu], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots.$$

**Theorem 1:** If $\nu_n \Rightarrow \nu$ as $n \to \infty$, then $\xi_n \Rightarrow -\mu^{-3/2} \nu$ as $n \to \infty$.

In each of the cases $\rho < 1$, $\rho > 1$ and $\rho = 1$ we also have found a random element $X$ of $D[0,1]^K$ such that
\[ X_n \Rightarrow X \text{ as } n \to \infty, \]

where
\[ X_n(t) = \frac{1}{\sqrt{n}} \hat{w}_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1,2,\ldots, \]
\[ \hat{v}_n = w_n - nd, \quad n = 1,2,\ldots \]
\[ d = \begin{pmatrix}
\mu - a_1 \\
\vdots \\
\mu - a_K
\end{pmatrix}. \]

Now define the random functions \( \gamma_n \in \mathbb{D}[0,1]^K \) generated by \( \{Q(t)\} \) by letting
\[ \gamma_n(t) = \frac{1}{\sqrt{n}} [Q(nt) - ntc], \quad 0 \leq t \leq 1, \quad n = 1,2,\ldots \]

where
\[ c = \begin{pmatrix}
\frac{\mu-a_1}{\mu a_1} \\
\vdots \\
\frac{\mu-a_K}{\mu a_K}
\end{pmatrix}. \]

Using Theorem 2.13, the multi-dimensional generalization of Theorem 2.12, one can show that a correspondence exists between weak convergence theorems for \( X_n \) and weak convergence theorems for \( \gamma_n \). In particular, a random element \( \gamma \) of \( \mathbb{D}[0,1]^K \) can be found for each of the cases \( \rho < 1, \rho = 1 \) and \( \rho > 1 \) such that \( \gamma_n \Rightarrow \gamma \), the distribution of \( \gamma \) bearing a certain relationship to the distribution of \( X \). Unfortunately,
Chapters 4, 5 and 6 are organized in a way which makes it difficult both to state and to prove this general relationship. We shall now state the weak convergence theorem for $\gamma_n$ for each of the three cases. The relationship between $\gamma$ and $\chi$ will then be pointed out, but the theorem will not be proved. Rather, the reader is referred to a more general proposition for assembly-like networks appearing in Chapter 10.

Corresponding to the matrices $D^2$ and $\tilde{\mathcal{F}}$ which have appeared earlier in our functional limit theorems for $\{w_n\}$, we define the $K \times K$ matrices

$$D^2 = \begin{pmatrix}
\sigma_1^2 & 0 \\
\frac{\sigma_2^2}{a_2} & 0 \\
\vdots & \ddots \\
0 & \frac{\sigma_{K-1}^2}{a_{K-1}} & \frac{\sigma_K^2}{a_K}
\end{pmatrix},$$

$$\tilde{\mathcal{F}} = (\tilde{s}_{ij}), \quad \tilde{s}_{ij} = \begin{cases}
\frac{\sigma_1^2}{a_1} + \frac{\sigma_2^2}{b_2} & \text{if } i = j \\
\frac{\sigma_{K-1}^2}{a_{K-1}} & \text{if } i = j
\end{cases}.$$

**Theorem 2:** (a) If $\rho < 1$, then

$$\gamma_n \Rightarrow F(D^2_K) \quad \text{as } n \to \infty,$$

where $F:D[0,1]^K \to D[0,1]^K$ is defined as in Chapter 4.

(b) If $\rho > 1$, then
\[ \gamma_n \xrightarrow{\text{as } n \to \infty} \chi_n^{2/3} \]

(c) If \( \rho = 1 \), then

\[ \gamma_n \xrightarrow{\text{as } n \to \infty} F(\chi_n^{2/3}) \]

where \( F: [0,1]^K \to [0,1]^K \) is defined as in Chapter 6.

It is left to the reader, following the format of Chapters 4-6, to derive from Theorems 1 and 2 the various corollaries and special cases of interest.

By comparing Theorem 2 with our various earlier results for \( X_n \), the following relationship is seen to exist between \( X \) and \( \gamma \). In each of the cases \( \rho < 1 \), \( \rho > 1 \) and \( \rho = 1 \) the distribution of \( X \) depends upon (or is parameterized by) some or all of the constants \( \sigma_1^2, \ldots, \sigma_K^2, \sigma_S^2 \).

In each case the distribution of \( \gamma \) depends in exactly the same way on the constants \( \sigma_1^2/a_1^3, \ldots, \sigma_K^2/a_K^3, \sigma_S^2/b_3^3 \). This relationship is proved in Theorem 10.7 for an assembly-like network. Since our single-station model is just a special type of network, Theorem 2 above follows as a corollary, and we shall not provide a separate proof here.

We conclude this section with some remarks concerning an alternate mode of presentation which could have been used in each of Chapters 4, 5 and 6 and actually will be adopted in the analysis of assembly-like networks later. One purpose of these remarks is to demonstrate at least roughly how functional limit theorems for \( \{Q(t)\} \) can be obtained using Theorem 2.13.

The perceptive reader may have noticed that the development in
each of Chapters 4, 5 and 6 could be further unified in the following way. Rather than beginning each chapter with a functional limit theorem for \( \{T_n\} \), one can with very little additional effort develop a functional limit theorem for the vector process \( \{Y_n\} \), where

\[
Y_n = \begin{pmatrix}
U_1 \\
 \vdots \\
U_n \\
 \vdots \\
K \\
T_n
\end{pmatrix}, \quad n = 1, 2, \ldots
\]

Now noting that

\[
(7.1) \quad w^k_n = T_n - U^k_n - V_n = T_n - U^k_n + \Delta(n^{1/2}), \quad 1 \leq k \leq K,
\]

it is apparent that the random functions in \( D[0,1]^K \) generated by \( \{w_n\} \) can effectively be obtained directly from those generated by \( \{Y_n\} \) in \( D[0,1]^{K+1} \) through appropriate addition and subtraction of component functions. Thus the functional limit theorem for \( \{w_n\} \) follows as a corollary to that for \( \{Y_n\} \), meaning that each chapter would consist of one theorem and a group of corollaries. An additional advantage of this approach is that, given a functional limit theorem for \( \{Y_n\} \), we immediately obtain from Theorem 2.1\( 3 \) a functional limit theorem for the continuous parameter vector process

\[
W(t) = \begin{pmatrix}
A^1(t) \\
 \vdots \\
A^K(t) \\
D(t)
\end{pmatrix}, \quad 0 \leq t < \infty
\]
where $A^k(t)$ is the number of type $k$ items arriving in $(0,t)$.

Now noting that

\[(7.2) \quad Q^k(t) = A^k(t) - D(t), \quad 1 \leq k \leq K,\]

it is apparent that the random functions generated by $\{Q(t)\}$ in $D[0,1]^K$ can be obtained directly from those generated by $\{N(t)\}$ in $D[0,1]^{K+1}$ through appropriate addition and subtraction of component functions. So the functional limit theorem for $\{Q(t)\}$ is obtained as a corollary to that for $\{N(t)\}$. Moreover, given the nature of Theorem 2.13 and the similarity of (7.1) and (7.2), it is clear that the functional limit theorem for $\{Q(t)\}$ must bear a close relationship to that for $\{w_n\}$.

In Part II we shall identify (for an assembly-like network) vector processes $\{T_n\}$ and $\{D(t)\}$ which are actually generalized versions of $\{Y_n\}$ and $\{N(t)\}$. The approach taken in analyzing $\{w_n\}$ and $\{Q(t)\}$ is then precisely that outlined above.

7.2 Initial Conditions

The formulation presented in Section 3.1 contains some very unrealistic assumptions as to the initial state of the system. Specifically, we have assumed that the system is initially devoid of items, that beginning from time zero the server suffers a period of "initial paralysis" of (random) duration $\nu_0$, that the first item of type $k$ arrives after a (random) length of time $u^k_0$ ($k = 1, \ldots, K$), and that $\{v_0, u^1_0, \ldots, u^K_0\}$ are independent and distributed exactly as
\{v_1, u^1_1, \ldots, u^K_1\}. With these assumptions, the sequences \{v_n; n = 0, 1, \ldots\},
\{u^1_n; n = 0, 1, \ldots\}, \ldots, \{u^K_n; n = 0, 1, \ldots\} are independent i.i.d. streams,
a fact which of course has been used often. (In particular, Lemma 4.2,
establishing the equivalence in distribution of \( \tilde{w}_n^k \) and the more easily
studied \( \hat{w}_n^k \), is not valid unless exactly this condition holds.) In
this section it will be shown that if we allow \{v_o, u^1_o, \ldots, u^K_o\} to be
dependent and arbitrarily distributed, then almost any physically
imaginable initial conditions can be accommodated by the model, and it
will be indicated why such a change in the assumptions has no bearing
on any of the limit theorems presented in Chapters 4-6.

Suppose that the system initially (i.e., at time zero) contains \( m^k \)
items of type \( k \) (\( k = 1, \ldots, K \)), and for ease of discussion assume
that \( m^1 \geq m^2 \geq \ldots \geq m^K \). It is clear that from the viewpoint of the
first item of type 1 to arrive after time zero, the server is unavailable
(or occupied, or effectively "paralyzed") until he has completed the
service of all the \( m^K \) complete units that were initially present plus
the service of another \( m^1 - m^K \) units, at least one of whose members
actually arrived after time zero. Thus the following redefinitions
suggest themselves. Let \( u^1_o \) denote the time at which the first item
of type 1 arrives from outside the system. For \( k = 2, \ldots, K \), let
\( u^K_o \) denote the time at which the \((m^1 - m^K + 1)\)th item of type \( k \) arrives
from outside the system. Finally let \( v_o \) denote the time at which the
server completes his \((m^1)\)th service. Henceforth let us agree that the
term "\( n \)th arriving item of type \( k \)" will in fact mean the \((m^1 - m^K + n)\)th
item of type \( k \) to arrive from outside the system (\( k = 1, \ldots, K \)). Thus
for each \( k = 2, \ldots, K \) we do not "count" any type \( k \) item whose arrival
is required to complete a unit part of whose members are initially present. The arrival time of the first item of type $k$ who "counts" is $u_{o}^{k}$, as it has been redefined above. With the redefinitions suggested here, $\{v_{o}, u_{o}^{1}, \ldots, u_{o}^{K}\}$ are in general dependent, and their exact distributions may be extremely complex (especially that of $v_{o}$, since it is not merely the sum of $m$ consecutive service times but may also include some server idleness).

The question now arises as to whether deviant behavior of the random variables $\{v_{o}, u_{o}^{1}, \ldots, u_{o}^{K}\}$ can affect the limit theorems that have been developed in Chapters 4-6. Assuming that each of these r.v.'s has its first two moments finite, the answer is no. Excepting for the moment the treatment of $w_{n}^{x}$ given in Section 4.1 for the case $\rho < 1$, a careful review of the proofs in preceding chapters will show that the only results pertaining to the sequences $\{v_{n}\}$ and $\{u_{n}^{k}\}$ which need hold are the strong law of large numbers, Donsker's Theorem, and a few miscellaneous lemmas involving uniform integrability. None of these results is affected by a change in the distributions of the initial r.v.'s $v_{o}, u_{o}^{1}, \ldots, u_{o}^{K}$ (assuming that their second moments remain finite).

In Section 4.1, Lemma 4.2 must be discarded if the initial conditions are made arbitrary, but a parallel results is obtainable using only the interchangeability of $\{v_{1}, \ldots, v_{n}\}$, $\{u_{1}^{1}, \ldots, u_{n}^{1}\}$, $\ldots$, and $\{u_{1}^{K}, \ldots, u_{n}^{K}\}$, and from it Theorem 4.4 can be established in the same way.

7.3 Deterministic Streams

It has been assumed throughout that both the input processes and the service process are non-deterministic (i.e., that all of the variances
\( \sigma_1^2, \ldots, \sigma_K^2, \sigma_S^2 \) are positive). This assumption is not critical to any of our results, however. If deterministic streams are allowed, then some of the propositions in Section 4.1 must be altered slightly, but in each case the appropriate revision is obvious. With a few trivial exceptions, all other results actually continue to hold in the stated form. */ By reviewing the proofs in Chapters 4-6, the interested reader may verify this fact, noting that Theorem 2.8, the multi-dimensional version of Donsker's Theorem, continues to hold when the covariance matrix \( \Psi \) is positive semi-definite but not positive definite.

What can change when deterministic streams are allowed are the qualitative conclusions derived from our limit theorems. Assuming that \( K \geq 2 \), the following statements concerning the "stability" of the system are easily proved.

(i) If \( \rho < 1 \) and \( \sigma_1^2 = \cdots = \sigma_{K^*}^2 = 0 \), then

\[
\begin{align*}
\omega^*_n &= \omega^*_n = \cdots = \omega^{K^*}_n \quad \text{for all } n \\
\omega^*_n &\rightarrow M^1 \quad \text{as } n \rightarrow \infty. \quad \text{For each } k = K^*+1, \ldots, K, \quad \omega^k_n \text{ blows up as } n \rightarrow \infty.
\end{align*}
\]

(ii) If \( b = a_1 = \cdots = a_{K^*} \) and \( \sigma_S^2 = \sigma_1^2 = \cdots = \sigma_{K^*}^2 = 0 \), then

\[
\begin{align*}
\omega^*_n &= \omega^*_n = \cdots = \omega^{K^*}_n = 0 \quad \text{for all } n,
\end{align*}
\]

*/ The only exceptions are results like Corollary 4.10 (b), stated in a form involving division by a constant which may vanish.
but \( w_n^k \) blows up as \( n \to \infty \) for all \( k = K^*+1, \ldots, K \).

(iii) The entire waiting time vector \( w_n \) converges in distribution as \( n \to \infty \) if and only if the condition in (i) is satisfied with \( K^* = K \) or the condition in (ii) is satisfied with \( K^* = K \).

In each of statements (i)-(iii) we assume the usual initial conditions, i.e., that \( (v_o^1, u_o^1, \ldots, u_o^K) \) is distributed as \( (v_n^1, u_n^1, \ldots, u_n^K) \), \( n \geq 1 \).

The necessary modifications under non-standard initial conditions are obvious.

7.4 Examples and Applications

In order to demonstrate in a concrete way the practical usefulness of the limit theorems in Chapters 4-6, consider the following problem.

A production manager has been ordered to manufacture some relative large (but known and fixed) number of finished products of a given type. He must choose between two different assembly-like system designs, each of which is capable of producing the desired finished product. Schematic representation and relevant data for these competing designs are given in Figures 5 and 6 below.

![Diagram](image)

- \( a_1 = 1.9 \)  \( a_2 = 1.9 \)  \( b = 2.1 \)
- \( \sigma_1 = 0.6 \)  \( \sigma_2 = 0.7 \)  \( \sigma_S = 0.3 \)
- \( v_1 = 74 \)  \( v_2 = 54 \)  \( v_S = 80 \)
- \( c_1 = 13.0 \)  \( c_2 = 12.0 \)  \( f = 5000 \)

Figure 5: Schematic representation and relevant data for configuration I.
Figure 6: Schematic representation and relevant data for configuration II.

Each of the possible system configurations happens to consist of two "input stations", at which basic components or subassemblies are manufactured, plus a single station at which the input items are assembled into finished products. (In general, of course, systems capable of producing the same finished product need not have the same number of stations.) As is common in inventory control models, we assume that there is associated with each of the two possible system configurations a cost structure of the following linear type. A fixed cost of $f$, including such one-time expenses as equipment purchases and operator training, is incurred before the production run even begins. A variable operating cost of $v_k$ is incurred for each unit of time that input station $k$ remains in operation ($k = 1,2$), and a corresponding variable cost of $v_S$ is incurred for each unit of time that the assembly station remains in operation. These variable cost rates include hourly wages of operators, power costs, etc. Finally, an inventory carrying cost of $c_k$ is incurred for each unit of time that a type $k$ item is held in inventory (i.e., for each unit of time that a type $k$ item "waits" plus each unit of time that a type $k$ item spends undergoing "service"). Assuming that production times at each input station are
i.i.d. and assembly times are i.i.d., other relevant information for each configuration are the means $a_1$, $a_2$ and $b$ and the variances $\sigma_1^2$, $\sigma_2^2$ and $\sigma_s^2$.

As in earlier chapters, we let

$$C_n = \sum_{k=1}^{2} c_k W_n^k, \quad n = 1, 2, \ldots$$

Then, for a given configuration, the total cost incurred during a production run in which $n$ finished products are manufactured is

$$TC(n) = f + \sum_{k=1}^{2} v_k U_n^k + v_s T_n^s + C_n + \left( \sum_{k=1}^{2} c_k \right) V_n,$$

where $U_n^k$, $V_n^k$, and $T_n^s$ are defined as in Chapter 3. Consequently, the expected total cost of the production run is

$$(7.3) \quad E[TC(n)] = f + n \sum_{k=1}^{2} v_k a_k + v_s E(T_n^s) + E(C_n) + nb \sum_{k=1}^{2} c_k.$$

Thus, in order to estimate the expected total cost of the run with each configuration, we need only estimate $E(T_n^s)$ and $E(C_n)$ for each configuration. Using Corollary 4.11, Corollary 5.2, and the remarks following Corollary 5.5, we have:

(i) For configuration I

$$E(T_n^s) = nb + o(\sqrt{n}) \preceq nb,$$

$$E(C_n) = \frac{n^2}{2} \sum_{k=1}^{2} c_k (b-a_k) + o(\sqrt{n^3}) \preceq \frac{n^2}{2} \sum_{k=1}^{2} c_k (b-a_k),$$

so from (7.3)
\[ E[TC(n)] = f + n[v_Sa + \sum_{k=1}^{2} (a_k v_k + c_k b)] + \frac{n^2}{2} \sum_{k=1}^{2} c_k (b-a_k) \]

\[ = 5000 + 463.4 n + 2.5 n^2. \]

(ii) For configuration II

\[ E(T_n) = na_1 + \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)n}{2\pi}} + o(\sqrt{n}) \approx na_1 + \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)n}{2\pi}} \]

\[ E(C_n) = \frac{2}{3} n \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)n^3}{2\pi}} \sum_{k=1}^{2} c_k + o(\sqrt{n^3}) \approx \frac{2}{3} n \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)n^3}{2\pi}} \sum_{k=1}^{2} c_k \]

so from (7.3)

\[ E[TC(n)] \approx f + \sqrt{n} \left[ \frac{1}{2\pi} \right]^{1/2} \frac{(\sigma_1^2 + \sigma_2^2)}{2\pi} \frac{1}{2} \left[ \frac{1}{2\pi} \right] n^3 \sum_{k=1}^{2} c_k \]

\[ = 15000 + 46.8 \sqrt{n} + 508.9 n + 0.9 \frac{n^3}{2} \]

Now let us suppose that it is desired to choose that configuration which will minimize expected total cost for a production run of size \( n = 100 \), subject to the constraint that the probability of completing the run within 220 time units is at least 0.90. Using Corollary 4.11 and Corollary 5.2 we have that:

(i) For configuration I,

\[ \lim_{n \to \infty} \frac{1}{\sigma_S \sqrt{n}} (T_n - nb) \leq x = \Phi(x), \]

so that

128
\[ P\{T_{100} \leq 220\} = P\{ \frac{1}{\sigma_S \sqrt{100}} (T_{100} - 100b) \leq 3.33\} \]
\[ \approx \Phi(3.33) = .99+ \]

(ii) For configuration II,

\[ \lim_{n \to \infty} P\{ \frac{1}{\sqrt{n}} (T_n - na_1) \leq x\} = \Phi\left( \frac{x}{\sigma_1}\right) \Phi\left( \frac{x}{\sigma_2}\right), \]

so that

\[ P\{T_{100} \leq 220\} = P\{ \frac{1}{\sqrt{100}} (T_{100} - 100a_1) \leq 2.0\} \]
\[ \approx \Phi\left( \frac{2.0}{0.5}\right) \Phi\left( \frac{2.0}{1.2}\right) = .94. \]

Thus the completion time constraint is satisfied by both configurations, and, substituting \( n = 100 \) into the expressions developed above, we find that \( E[TC(100)] \approx 76,340 \) with configuration I, while \( E'[TC(100)] \approx 72,300 \) with configuration II. So configuration II is seen to be slightly more favorable, having an expected total cost which is about 4% lower than that for configuration I.

Let us now consider the problem of choosing between these same two system designs under somewhat different circumstances. Suppose that finished products are needed at a constant rate for an extended period of time. In particular, one finished product is needed every \( r = 2.2 \) units of time. It has been decided that these products will be manufactured in a sequence of production runs of equal size. The fixed cost \( f \) is now interpreted as a set-up cost incurred at the beginning of each production run, and the variable operating costs and in-process
inventory carrying costs incurred during each run are as before. At the beginning of the first run, there is no inventory of finished products. Assuming that the chosen configuration can produce finished assemblies at a mean rate which exceeds the demand rate, an inventory of finished products develops as time passes. When the production run is completed, the assembly-like system is shut down until the ending inventory of finished products has been dissipated. At that point in time another production run is begun and the cycle repeats itself. Let \( N \) denote the size of the production run and define

\[
R_n = r_n = 2.2n, \quad n = 1, 2, \ldots, N.
\]

We assume that a cost of \( c_S = 5 \) is incurred for each unit of time that a finished product is held in inventory. Thus, during a given cycle, the total cost of carrying finished product inventories is

\[
I(N) = c_S \sum_{n=1}^{N} (R_n - T_n)^+.
\]

The decision-maker's problem is to decide which system configuration should be used and how large the production runs should be.

Using the fact that

\[
\frac{1}{n} T_n \xrightarrow{a.s.} \max(a_1, b) < r
\]

for each configuration, it can easily be shown that

\[
E(R_n - T_n)^+ = r_n - E(T_n) + o(\sqrt{n})
\]

implying that
\[ E[I(N)] = c_s \left( \frac{N^2}{2} r - \frac{N}{2} \right) E[T_n] + o(\sqrt{N^3}) \, . \]

Thus we find that for configuration I

\[ E[I(N)] = c_s \left( \frac{N^2}{2} r - \frac{N}{2} \right) b + o(\sqrt{N^3}) \]
\[ = 0.25 N^2 + o(\sqrt{N^3}) \, , \]

and for configuration II

\[ E[I(N)] = c_s \left( \frac{N^2}{2} r - \frac{N}{2} a_1 - \frac{2}{3} \sqrt{\left( \frac{\sigma_1^2 + \sigma_2^2}{3} \right) N^3} \right) + o(\sqrt{N^3}) \]
\[ = 0.5 N^2 - 1.75 \sqrt{N^3} + o(\sqrt{N^3}) \, . \]

In each case we shall approximate \( E[I(N)] \) by dropping the \( o(\sqrt{N^3}) \) term. For each configuration the optimal production run size \( N^* \) is clearly that value of \( N \) which minimizes the expected average cost per finished product,

\[ E[AC(N)] = \frac{E[TC(N)] + E[I(N)]}{N} \]

Combining the approximations developed above for \( E[TC(N)] \) and \( E[I(N)] \), we find that for configuration I

\[(7.4) \quad E[AC(N)] \approx 5000 N^{-1} + 463.4 + 2.75 N \, , \]

and for configuration II

\[(7.5) \quad E[AC(N)] \approx 15000 N^{-1} + 46.8 N^{-1/2} + 508.9 + 4.15 N^{1/2} + 0.5 N \, . \]
For configuration I, the expression for $E[AC(N)]$ is sufficiently simple that we can just differentiate with respect to $N$ and set the derivative equal to zero to obtain

$$N^* \approx \sqrt{\frac{5000}{2.75}} \approx 43,$$

$$E[AC(N^*)] \approx 697$$

Since the expression for $E[AC(N)]$ under configuration II is so much more complicated, it is easier to just draw a graph of $E[AC(N)]$ and find the approximate minimizing value of $N$ visually. Upon so doing, the approximate optimal run size and corresponding average cost are found to be

$$N^* \approx 150$$

$$E[AC(N^*)] \approx 738$$

So in this case configuration I is seen to be more favorable, having a minimal expected average cost which is almost 6% lower than that for configuration II.

The reader may verify that for each configuration the U-shaped graph of $E[AC(N)]$, as approximated by (7.4) and (7.5), is very flat at the bottom. Thus, in each case, run sizes which are quite different from the approximate "optimal" value $N^*$ may have an associated expected average cost which is only slightly larger than $E[AC(N^*)]$. Moreover, the lower order terms which we have dropped in approximating $E[AC(N)]$ might significantly change the value of $N^*$ if they were known. Thus the optimal run sizes which we have calculated may be very crude.
approximations, but the corresponding approximations for \( E[AC(W^*)] \) should be reasonably accurate.

7.5 **Additional Remarks**

As the preceding section indicates, a primary application of limit theorems like those developed in Chapters 4-6 is as approximation theorems for large but finite values of \( n \). A natural question to ask then is how large \( n \) must be for these approximations to be reliable. Since most of our results are derived from Donsker's Theorem by way of the Continuous Mapping Theorem, the question reduces essentially to one of speed of convergence for Donsker's Theorem. Some theoretical results in this direction have been obtained, most notably by Rosenkrantz (1967), but they are not directly useful in application. However as computational experience with diffusion approximations in congestion theory increases, cf. Gaver (1968), there is reason to hope that ready guidelines can be developed as to what constitutes a "sufficiently large" value of \( n \) for approximation theorems like ours.

We have mentioned earlier that many useful results can be obtained by employing the Continuous Mapping Theorem in conjunction with functional limit theorems for processes like \( \{w_n\} \) and \( \{Q(t)\} \). In particular, we have derived limit theorems for \( W_n \) by applying the integration mapping to the random functions generated by \( \{w_n\} \). For a discussion of other mappings which arise naturally in queuing theory, the reader is referred to Whitt (1968), Chapter 9.

In choosing examples for discussion in the preceding section, we have conspicuously restricted attention to special cases for which
completely explicit limit theorems were developed earlier. For most other cases, however, limiting distributions have been specified only indirectly, usually in terms of the (unknown) distributions of certain functions of a standard normal vector or functionals of Brownian Motion. For example, Corollary 4.6 states that if $K = 3$ and $b < a_1 = a_2 = a_3$, then

$$(7.6) \quad \frac{1}{\sqrt{n}} w_n^\perp Y = [ \max_{1 \leq k \leq 3} \{ \sigma_k Z_k \} - \sigma_1 Z_1 ] .$$

where $Z_1, Z_2$ and $Z_3$ are i.i.d. and distributed $N(0,1)$. Although such a result has no direct computational significance (since the distribution of $Y$ is not known explicitly), it is of interest for a variety of reasons. First is that it contains a great deal of qualitative information about $w_n^\perp$, showing how much of the associated probability mass escapes to infinity as $n \to \infty$ and the speed at which it does so. Moreover, it is clear that the distribution of $Y$ can be simulated (for given values of $\sigma_1, \sigma_2$ and $\sigma_3$), making (7.6) useful for application. Finally, such a result is of considerable theoretical interest, and perhaps further research into the distribution of $Y$ will be stimulated.

If $\rho$ is close to the critical value of 1 (with either $\rho > 1$ or $\rho < 1$), then in general $n$ must be very large for the limit theorems in Chapter 4 or Chapter 6 to serve as good approximations. To approximate system behavior in such cases for moderate values of $n$, one needs limit theorems of the type developed by Prohorov (1963), Viskov (1964) and Whitt (1968) for $GI/G/1$ queues. These theorems consider a
sequence of queueing systems (indexed by \(i\), say) with traffic intensities \(\rho_i\) satisfying

\[
\rho_i \to 1 \quad \text{as} \quad n \to \infty,
\]

and then investigate the behavior of various processes as \(i \to \infty\) and \(n \to \infty\) simultaneously in such a way that

\[
(1-\rho_i) \sqrt{n} \to c, \quad \text{a constant}.
\]
PART II
LIMIT THEOREMS FOR THE NETWORK MODEL

CHAPTER 8
Single-Server Queues in Series

We wish to consider a finite collection of single-server queueing facilities arranged in series. Rather than imagining that customers demanding service arrive from outside the system, it will be convenient (and clearly equivalent), to assume that the system is comprised of stations (or servers, or facilities) numbered 1,2,...,K, the first of which serves only to generate customers as input to subsequent stations. Such a system can be schematically represented as in Figure 8.1 below.

![Diagram of single-server queues in series]

Figure 7. Single-server queues in series.

The server at station 1 is never idle. He begins his first service at time zero, his second immediately upon completion of the first, etc. When server 1 completes his \( n \)th service \((n = 1,2,...)\), a customer numbered \( n \) is generated and proceeds immediately to station 2. If the customer finds the server there to be occupied, then he joins a queue and waits for admission to service; otherwise his service at station 2 begins immediately. After completing service at station 2 the customer proceeds immediately to station 3, etc. After completing service at
station $K$, the customer departs the system. It is assumed that customers are served in the order of their arrival at each station $k = 2, \ldots, K$ and that all waiting rooms are of infinite capacity. We let

$$v^k_0 = \text{the duration of server } k's \ "\text{initial paralysis}"$$

$$(k = 2, \ldots, K), \text{ and}$$

$$v^k_n = \text{the service time of customer } n \text{ at station } k$$

$$(k = 2, \ldots, K; \ n = 1, 2, \ldots).$$

Generalizing rather loosely the terminology used in Section 2.5, we shall say that a collection of single-server queues in series is a GI/G/1 series queue (or queueing system) if \( \{v^1_n; n = 0, 1, \ldots\}, \ldots, \{v^K_n; n = 0, 1, \ldots\} \) are mutually independent i.i.d. sequences of non-negative random variables with

$$0 < b_k = E(v^k_n) < \infty, \ 1 \leq k \leq K$$

$$0 < \sigma^2_k = \text{Var}(v^k_n) < \infty, \ 1 \leq k \leq K.$$ 

All of the results obtained in this chapter except Lemma 1 are for GI/G/1 series queues only. The initial conditions assumed here are convenient but essentially irrelevant. If \( v^1_0, \ldots, v^K_0 \) are allowed to have arbitrary distributions, then virtually any initial conditions can be accomodated by the model and the reader will note that none of
our limit theorems are affected. */

The basic stochastic processes with which we shall deal are

\[ \text{w}_{\text{n}}^{\text{k}} = \text{the waiting time (exclusive of service time) of customer n at station k (k = 2, \ldots, K; n = 1, 2, \ldots)} , \]

\[ \text{t}_{\text{n}}^{\text{k}} = \text{the time at which customer n departs from station k (k = 1, \ldots, K; n = 1, 2, \ldots)} . \]

For completeness, we take

\[ \frac{\text{t}_{\text{n}}^{\text{k}}}{\tau_{\text{o}}} = 0 , \quad k = 1, \ldots, K . \]

Of course \( \frac{\text{t}_{\text{n}}^{\text{k}}}{\tau_{\text{o}}} \) could be alternately defined as the time at which server k completes his \( n \)th service, and thus

\[ \frac{\text{t}_{\text{n}}^{1}}{\tau_{\text{o}}} = \frac{\text{v}_{\text{n}}^{1}}{\tau_{\text{o}}} , \quad n = 1, 2, \ldots \]

\[ \frac{\text{t}_{\text{n}}^{\text{k}}}{\tau_{\text{o}}} = \frac{\text{t}_{\text{n}}^{\text{k-1}}}{\tau_{\text{o}}} + \frac{\text{v}_{\text{n}}^{\text{k}}}{\tau_{\text{o}}} + \frac{\text{v}_{\text{n}}^{\text{k}}}{\tau_{\text{o}}} , \quad 2 \leq k \leq K , \quad n = 1, 2, \ldots . \]

It is clear that the departure times \( \{ \text{v}_{\text{n}}^{\text{k}} \} \) blow up as \( n \to \infty \), regardless of what the system parameters may be, but the asymptotic behavior of the waiting time processes \( \{ \text{w}_{\text{n}}^{\text{k}} \} \) is critically dependent upon the mean

*\ Also, note that by requiring \( c_{\text{k}}^{2} > 0 \) for all \( k \) we rule out the possibility of deterministic service times at any station. The only reason for this is to avoid tedious differentiations between degenerate and non-degenerate cases in discussing the results that will be obtained. All of the results actually apply equally to the case of some or all \( \{ \text{v}_{\text{n}}^{\text{k}} \} \) deterministic.
service times \( b_1, \ldots, b_K \). We define

\[
a_k = \max_{1 \leq j \leq k-1} (b_j), \quad 2 \leq k \leq K,
\]

\[
\rho_k = \frac{b_k}{a_k}, \quad 2 \leq k \leq K.
\]

The quantity \( \rho_k \) is called the traffic intensity at station \( k \), and it has been shown by Sacks (1960) that the entire waiting time vector

\[
\mathbf{w}_n = \left( \begin{array}{c}
\mathbf{w}_1^n \\
\vdots \\
\mathbf{w}_K^n
\end{array} \right), \quad n = 1, 2, \ldots
\]

converges in distribution as \( n \to \infty \) if and only if \( \rho_k < 1 \) for all \( k = 2, \ldots, K \). A similar result was proved for more general systems by Loynes (1962). In this chapter we shall derive functional limit theorems for the vector processes \( \{w_n\} \) and \( \{T_n\} \), where

\[
T_n = \left( \begin{array}{c}
T_1^n \\
\vdots \\
T_K^n
\end{array} \right), \quad n = 1, 2, \ldots
\]

Although each of these limit theorems is of some interest in itself and new in at least some regard, the basic purpose of this chapter is to lay some necessary groundwork for the study of assembly-like networks. At the end of this chapter it will be indicated how our theorems relate to previous results for series queueing systems.

Our first order of business is to develop an explicit representation
for the departure times $t_n^k$ in terms of the elemental sequences $\{v_n^k\}$.

In this regard Lemma 2.26 for single-server queues is of critical importance. Note that each station $k = 2, ..., K$ of our series arrangement is, when viewed in isolation, a single-server queueing system in the formal sense of Section 2.5. The times at which customers arrive in this system are $\{t_n^{k-1}\}$, the duration of the server's "initial paralysis" is $v_o^k$, and the sequence of service times is $\{v_n^k; n = 1, 2, ...\}$. Thus the following proposition is an immediate application of Lemma 2.26.

**Lemma 1:** For each $k = 2, ..., K$ and $n \geq 1$,

$$w_n^k = (v_n^k - t_n^{k-1}) - \min_{0 \leq j \leq n} (v_{j}^{k} - t_{j}^{k-1}) .$$

**Lemma 2:** For each $k = 2, ..., K$ and $n \geq 1$,

$$t_n^k = \max_{0 \leq j \leq n} [t_j^{k-1} + (v_j^k - v_{j}^{k-1}) + v_n^k .$$

**Proof:** Using the fact that

$$t_n^k = t_n^{k-1} + w_n^k + v_n^k ,$$

the proposition is immediate from Lemma 1.

Q.E.D.

Using Lemma 2 and the fact that $t_n^1 = v_n^1$, an exact representation for $t_n^k$ can be developed inductively. With the initial conditions assumed
here, however, the exact representation is rather messy. For our purposes, the following approximation will prove to be sufficiently accurate. By dropping the final term \( v_n^k \) in Lemma 2, one can write

\[
T_n^k \cong \max_{0 \leq j \leq n} [T_n^{k-1} + (v_n^k - v_j^k)] .
\]

(8.1)

Proceeding inductively, we can use (8.1) and the fact that \( T_n^1 = v_n^1 \) to write

\[
T_n^1 = v_n^1 ,
\]

\[
T_n^2 \cong \max_{0 \leq j \leq n} [v_n^1 + (v_n^2 - v_j^2)] ,
\]

\[
T_n^3 \cong \max_{0 \leq i \leq n} [v_i^1 + (v_j^2 - v_j^3) + (v_n^3 - v_j^3)] ,
\]

\[
\vdots
\]

(8.2)

\[
T_n^k \cong \max_{0 = j_0 \leq \cdots \leq j_k = n} [\sum_{i=1}^{k} (v_j^i - v_j^{i-1})] .
\]

Before stating formally the sense in which (8.2) gives a "sufficiently accurate" approximation for \( T_n^k \), we shall introduce some additional notation. By making this notation slightly more general than is actually required for present purposes, we shall be able to maintain it through Chapters 9 and 10.

*/ In Chapter 13 we shall indicate why the precise representation is not of a compact form, and it will be shown how the difficulty can be resolved by altering our initial conditions.

141
Definition: For each \( n = 0,1,2, \ldots \) and each non-empty ordered set \( c = (i_1, \ldots, i_m) \) of distinct indices \( i_j \) satisfying \( 1 \leq i_j \leq K \) for all \( j = 1, \ldots, m \) we define

\[
X_n(c) = \max_{0 = r_0 \leq \cdots \leq r_m = n} \left[ \sum_{j=1}^{m} \frac{i_j}{r_j - r_{j-1}} \right],
\]

the maximization being over integer values of \( r_1, \ldots, r_{m-1} \) which satisfy the indicated constraint. Note that \( X_0(c) = 0 \), and if \( m = 1 \) then \( X_n(c) = \frac{i_1}{n} \) for all \( n \geq 0 \).

Definition: For each \( k = 1, \ldots, K \) let \( c_k \) denote the non-empty ordered set of distinct indices \( (1, \ldots, k) \).

With this notation, (8.1) can be rewritten as

\[
T_n^k \simeq X_n(c_k).
\]

We shall now demonstrate the sense in which this approximation is "sufficiently accurate".

Lemma 3: For each \( k = 1, \ldots, K \)

(a) \( \frac{1}{n} [T_n^k - X_n(c_k)] \xrightarrow{\mathbb{P}} 0 \) as \( n \to \infty \),

(b) \( [T_n^k - X_n(c_k)] = \Delta(n^{1/2}) \),

(c) \( \frac{1}{\sqrt{n}} [T_n^k - X_n(c_k)] \) is uniformly integrable.

Proof: For each \( k = 1, \ldots, K \) and \( n \geq 0 \), let

\[
\delta_n^k = T_n^k - X_n(c_k).
\]
We first prove by induction that for all $k = 1, \ldots, K$

\[(8.2) \quad 0 \leq \delta^k_n \leq \sum_{j=1}^{k} \max_{0 \leq r \leq n} v^j_r, \quad n \geq 0.\]

By definition, $T^l_n = V^l_n = X_n(c^l_1)$ for all $n \geq 0$, so (8.2) is trivially satisfied for $k = 1$. Now suppose that $2 \leq k \leq K$ and that (8.2) is satisfied when $k$ is replaced by $k-1$. Since $T^k_0 \equiv 0$ and $X_0(c^k_1) \equiv 0$, $\delta^k_0 \equiv 0$. Moreover, for $n \geq 1$ we have from Lemma 2 that

\[(8.3) \quad \delta^k_n = \max_{0 \leq r \leq n} \left[ \tau^r_{k-1} + (v^r_k - v^r_n) \right] + v^k_n - X_n(c^k_k)\]

\[= \max_{0 \leq r \leq n} \left[ X_r(c^k_{k-1}) + \delta^k_r + (v^r_k - v^r_n) \right] + X_n(c^k_k) + v^k_n.\]

It is easily verified that

\[X_n(c^k_k) = \max_{0 \leq r \leq n} \left[ X_r(c^k_{k-1}) + (v^r_k - v^r_n) \right], \quad n \geq 0,\]

so by the induction hypothesis

\[(8.4) \quad 0 \leq \max_{0 \leq r \leq n} \left[ X_r(c^k_{k-1}) + \delta^k_r + (v^r_k - v^r_n) \right] - X_n(c^k_k)\]

\[\leq \max_{0 \leq r \leq n} \delta^k_r \leq \max_{0 \leq r \leq n} \left[ \sum_{j=1}^{k-1} \max_{0 \leq i \leq r} v^j_i \right]\]

\[= \sum_{j=1}^{k-1} \max_{0 \leq r \leq n} v^j_r.\]

Combining (8.3) and (8.4) we have
\[ 0 \leq \sum_{n=1}^{k-l} \max_{j=1}^{k} v_{r}^{j} + v_{n}^{j} \leq \sum_{j=1}^{k} \max_{0 \leq r \leq n} v_{r}^{j} . \]

So the inductive proof that (8.2) holds for all \( k = 1, \ldots, K \) is complete.

It is an immediate consequence of the strong law of large numbers that

\[ (8.5) \quad \frac{1}{n} \max_{0 \leq r \leq n} v_{r}^{j} \sim_{a.e.} 0 , \quad 1 \leq j \leq K . \]

Combining (8.2) with (8.5), Theorem 2.18 and Lemma 2.19, each of the desired results follows easily.

Q.E.D.

In order to develop a limit theorem for the departure times \( T_{n}^{k} \), we first prove some important properties of the random variables \( X_{n}(c_{k}) \).

**Definition:** For each \( k = 1, \ldots, K \) let \( \hat{c}_{k} \) denote the non-empty ordered set of distinct indices obtained by deleting from \( c_{k} \) all indices \( j \) such that

\[ b_{j} < \max(b_{1}, \ldots, b_{k}) , \]

maintaining the original ordering of the remaining indices.

**Lemma 4:** For each \( k = 1, \ldots, K \)

(a) \( \frac{1}{n} X_{n}(c_{k}) \sim_{a.e.} \max(b_{1}, \ldots, b_{k}) \) as \( n \to \infty \).
(b) \( X_n(c_k) = X_n(\tilde{c}_k) + \Delta(n^{1/2}) \).

**Proof:** If \( k = 1 \), then \( c_k = \tilde{c}_k = (1) \) and statement (b) is trivially true. Moreover, \[
\frac{1}{n} X_n(c_1) = \frac{1}{n} \sum_{i=1}^{n} a_i e_i b_1 \quad \text{as} \quad n \to \infty
\]

by the strong law of large numbers, so (a) holds also. Proceeding inductively, suppose that \( 1 \leq k \leq K \) and that (a) and (b) both hold when \( k \) is replaced by \( k-1 \), i.e., that

\[
\frac{1}{n} X_n(c_{k-1}) \overset{a.e.}{\to} \max(b_1, \ldots, b_{k-1}) = a_{k-1},
\]

and

\[
X_n(c_{k-1}) = X_n(\tilde{c}_{k-1}) + \Delta(n^{1/2})
\]

Now \( X_n(c_k) \) can be re-expressed as

\[
X_n(c_k) = \max_{0 \leq r \leq n} \left[ X_r(c_{k-1}) + (V^k_n - V^r_n) \right]
\]

\[
= V^k_n + \max_{0 \leq r \leq n} \left[ X_r(c_{k-1}) - V^r_n \right]
\]

Since \( \frac{1}{n} V^k_n \overset{a.e.}{\to} b_k \) by the strong law, we have from (8.6) that

\[
\frac{1}{n} [X_n(c_{k-1}) - V^k_n] \overset{a.e.}{\to} (a_k - b_k)
\]

from which it follows easily that

145
(8.9) \[ \frac{1}{n} \max_{0 \leq r \leq n} [X_r(c_{k-1}) - V_r] \xrightarrow{a.e.} (a_k - b_k)^+ . \]

Combining (8.8) and (8.9),

\[ \frac{1}{n} X_n(c_k) \xrightarrow{a.e.} b_k + (a_k - b_k)^+ = \max(a_k, b_k) , \]

so statement (a) holds for \( k \).

To prove that (b) also holds for \( k \), we consider separately the cases \( a_k = b_k \), \( a_k < b_k \) and \( a_k > b_k \). If \( a_k = b_k \), then \( \tilde{c}_k = (c_{k-1}, k) \) and hence, using (8.7),

\[ X_n(c_k) = \max_{0 \leq r \leq n} [X_r(c_{k-1}) + (V_n^k - V_r^k)] \]

\[ = \max_{0 \leq r \leq n} [X_r(\tilde{c}_{k-1}) + \Delta(n^{1/2}) + (V_n^k - V_r^k)] \]

\[ = \max_{0 \leq r \leq n} [X_r(\tilde{c}_{k-1}) + (V_n^k - V_r^k)] + \Delta(n^{1/2}) \]

\[ = X_n(\tilde{c}_{k-1}, k) + \Delta(n^{1/2}) = X_n(\tilde{c}_k) + \Delta(n^{1/2}) . \]

If \( a_k < b_k \), then \( \tilde{c}_k = (k) \), and again we write

(8.10) \[ X_n(c_k) = V_n^k + \max_{0 \leq r \leq n} [X_r(\tilde{c}_{k-1}) - V_r^k] + \Delta(n^{1/2}) . \]

Since

\[ \frac{1}{n} [X_n(\tilde{c}_{k-1}) - V_n^k] \xrightarrow{a.e.} (a_k - b_k) < 0 \ , \]

it follows that \( \max_{0 \leq r \leq n} [X_r(\tilde{c}_{k-1}) - V_r^k] \) converges almost everywhere to a finite limit as \( n \to \infty \), implying that
\begin{equation}
\max_{0 \leq r \leq n} [X_r(c_{k-1}) - v_r^k] = \Delta(n^{1/2}) .
\end{equation}

Then, combining (8.10) and (8.11),

\[ X_n(c_k) = v_n^k + \Delta(n^{1/2}) + \Delta(n^{1/2}) = X_n(c_{k-1}^\circ) + \Delta(n^{1/2}) . \]

Finally, for the case \( a_k > b_k \), we can again use (8.7) to write

\begin{equation}
X_n(c_k) = \max_{0 \leq r \leq n} [X_r(c_{k-1}) + (v_n^k - v_r^k)]
\end{equation}

\[ = \max_{0 \leq r \leq n} [X_r(c_{k-1}) + (v_n^k - v_r^k) + \Delta(n^{1/2})]
\end{equation}

\[ = X_n(c_{k-1}^\circ) + \delta_n + \Delta(n^{1/2}) , \]

where

\[ \delta_n = \max_{0 \leq r \leq n} \{ [v_n^k - v_r^k] - [X_n(c_{k-1}^\circ) - X_r(c_{k-1}^\circ)] \} . \]

Now let us suppose that \( c_{k-1}^\circ = (i_1, \ldots, i_m) \). A careful examination of the definition of \( X_n(c) \) shows that for any \( r \) satisfying \( 0 \leq r \leq n \)

\[ X_n(c_{k-1}^\circ) - X_r(c_{k-1}^\circ) \geq v_n^i - v_r^i \geq 0 . \]

Consequently, letting \( S_j = v_j^k - v_j^i \) \( (j = 0, 1, \ldots) \), we have for each \( n \geq 0 \)

\begin{equation}
0 \leq \delta_n \leq \max_{0 \leq r \leq n} [(v_n^k - v_r^k) - (v_n^m - v_r^m)]
\end{equation}

\[ = \max_{0 \leq r \leq n} (S_n - S_r) . \]
From the definition of $\tilde{c}_{k-1}$ we know that $b_j = a_k$, so $S_n$ is the sum of i.i.d. random variables with mean $b_k - a_k < 0$ and finite variance. Then by Lemma 2.20,

\[ \max_{0 \leq r \leq n} (S_n - S_r) = \Delta(n^{1/2}) . \]

Combining (8.12), (8.13), and (8.14), we have

\[ X_n(c_k) = X_n(\tilde{c}_{k-1}) + \Delta(n^{1/2}) = X_n(\tilde{c}_k) + \Delta(n^{1/2}) , \]

which completes the proof.

Q.E.D.

Lemma 5: For each $k = 1, \ldots, K$

\[ \{ \frac{1}{\sqrt{n}} [X_n(c_k) - n \max(b_1, \ldots, b_k)] \}

is a uniformly integrable sequence.

Proof: Let $k$ be fixed. It will be convenient to single out one index $j^* \in c_k$ such that

\[ b_{j^*} = \max(b_1, \ldots, b_k) , \]

i.e., such that $j^* \in \tilde{c}_k$. For expository simplicity, we shall assume that $j^*$ can be taken equal to $k$. (The nature of the argument is the same whatever the value of $j^*$, but it is bothersome to maintain a general notation.) If we define

\[ S_r^j = v_r^j - v_r^{j^*} = v_r^j - v_r^k \quad (1 \leq j \leq k-1, \ r \geq 0) , \]

148
then

\[ X_n(c_k) = \max_{0 = r_0 \leq \cdots \leq r_k = n} \left[ \frac{1}{r_{j-l}} \left( \sum_{j=1}^{k} Y_j - Y_j^{(j-1)} \right) \right] \]

\[ = v_n^{k} + \max_{0 = r_0 \leq \cdots \leq r_k = n} \left[ \frac{1}{r_{j-l}} \left( \sum_{j=1}^{k-1} (S_j - S_j^{(j-1)}) \right) \right] \]

Consequently, with our assumption that \( b_k = \max(b_1, \ldots, b_k) \),

\[ \frac{1}{\sqrt{n}} \left[ X_n(c_k) - n \max(b_1, \ldots, b_k) \right] = \frac{1}{\sqrt{n}} \left( v_n^{k} - nb_k \right) + \delta_n , \]

where

\[ \delta_n = \max_{0 = r_0 \leq \cdots \leq r_{k-1} \leq n} \left[ \sum_{j=1}^{k-1} \frac{1}{\sqrt{n}} \left( S_j - S_j^{(j-1)} \right) \right] . \]

Since \( \left\{ \frac{1}{\sqrt{n}} (v_n^{k} - nb_k) \right\} \) is u.i. by Theorem 2.16, it suffices to show that \( \{ \delta_n \} \) is u.i. Note that

\[ \delta_n \leq 0 \leq \delta_n \leq \sum_{j=1}^{k-1} \frac{1}{\sqrt{n}} \max_{0 \leq i \leq r \leq n} \left( S_j^{(j-1)} - S_i \right) . \]  

(8.15)

Now let us define

\[ \hat{S}_r^j = S_r^j - r(b_j - b_k), \quad 1 \leq j \leq k-1, \quad n \geq 0 . \]

Since \( b_k \geq b_j \) for all \( j = 1, \ldots, k-1 \), it is immediate that
\[(S^j_r - S^j_i) \leq (\hat{S}^j_r - \hat{S}^j_i) \text{ if } 1 \leq j \leq k - 1, \ 0 \leq i \leq r,\]

so from (8.15)

\[(8.16) \quad 0 \leq \hat{\delta}_n \leq k - 1 \sum_{j=1}^{k-1} \frac{1}{\sqrt{n}} \max_{0 \leq i \leq r \leq n} (\hat{S}^j_r - \hat{S}^j_i)\]

\[\leq \sum_{j=1}^{k-1} \frac{1}{\sqrt{n}} \left[ \max_{0 \leq r \leq n} (\hat{S}^j_r) - \min_{0 \leq r \leq n} (\hat{S}^j_r) \right].\]

Now note that \(\hat{S}^j_r\) is the sum of i.i.d. random variables with mean zero and finite variance, so by Theorem 2.17

\[\left\{ \frac{1}{\sqrt{n}} \max_{0 \leq r \leq n} (\hat{S}^j_r) \right\} \text{ and } \left\{ \frac{1}{\sqrt{n}} \min_{0 \leq r \leq n} (\hat{S}^j_r) \right\}\]

are both uniformly integrable sequences for each \(j = 1, \ldots, k - 1\). It follows then from (8.16) and Theorem 2.15 that \(\{\hat{\delta}_n\}\) is u.i., which completes the proof.

Q.E.D.

Most of the work required to obtain a functional limit theorem for the vector process \(\{T_n\}\) has now been done. Let

\[\mu_k = \max(b_1, \ldots, b_k), \quad 1 \leq k \leq K,\]

\[\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix} .\]

Now define

150
\[ \hat{T}_n = T_n - n\mu, \quad n \geq 0, \]

and the random functions in \( D[0,1]_K \) generated by \( \{\hat{T}_n\} \),

\[ \nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_{[nt]}, \quad 0 \leq t \leq 1, \quad n \geq 1. \]

Before stating the main result for \( \nu_n(\cdot) \), some additional notation is required. Again we shall make this notation slightly more general than is actually required for present purposes.

**Definition:** For each non-empty ordered set \( c = (i_1, \ldots, i_m) \) of distinct indices satisfying \( 1 \leq i_j \leq K \) for all \( j = 1, \ldots, m \) we define mappings

\[ F^c : D[0,1]_K \to D[0,1]_K, \quad f^c : D[0,1]_K \to \mathbb{R} \]

as follows. For \( x = (x^1, \ldots, x^K) \in D[0,1]^K \), let

\[ [F^c(x)](t) = \sup_{0 = \tau_0 \leq \cdots \leq \tau_m = t} \left\{ \sum_{j=1}^{m} \left[ x^{i_j}(\tau_j) - x^{i_j}(\tau_{j-1}) \right] \right\}, \]

\[ f^c(x) = (\pi \circ F^c)(x) = \sup_{0 = \tau_0 \leq \cdots \leq \tau_m = 1} \left\{ \sum_{j=1}^{m} \left[ x^{i_j}(\tau_j) - x^{i_j}(\tau_{j-1}) \right] \right\}. \]

To state a weak convergence theorem for \( \nu_n(\cdot) \), we define mappings \( G : D[0,1]_K \to D[0,1]_K \) and \( g : D[0,1]_K \to \mathbb{R}^K \) by letting

\[ G^k(x) = F^k(x), \quad g^k(x) = (\pi \circ G^k)(x) = f^k(x), \quad 1 \leq k \leq K. \]

**Theorem 6:** \( \nu_n \Rightarrow G(D^k) \) as \( n \to \infty \), where \( D \) is a \( K \times K \) diagonal
matrix with diagonal elements $\sigma_1, \ldots, \sigma_k$.

Proof: Let

$$\hat{X}_n^{(c_k)} = X_n^{(c_k)} - n\mu_k, \quad 1 \leq k \leq K, \quad n \geq 0.$$ 

Combining Lemma 3(b) and Lemma 4(b), we have

$$T_n^k = X_n^{(c_k)} + \Delta(n^{1/2}), \quad 1 \leq k \leq K,$$

implying that

$$\hat{\eta}_n^k = \hat{X}_n^{(c_k)} + \Delta(n^{1/2}), \quad 1 \leq k \leq K,$$

Thus, by Theorem 2.11, it suffices to show that $\gamma_n \Rightarrow G(D^{k})$, where

$$\gamma_n^k(t) = \frac{1}{\sqrt{n}} \hat{X}_n^{[nt]}(\tilde{c}_k), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K, \quad n \geq 1.$$

Now let

$$V_n^k = V_n^k - n\mu_k, \quad 1 \leq k \leq K, \quad n \geq 0,$$

$$\hat{V}_n = \begin{pmatrix} \hat{V}_n^1 \\ \vdots \\ \hat{V}_n^K \end{pmatrix}, \quad n \geq 0,$$

$$\xi_n(t) = \frac{1}{\sqrt{n}} \hat{V}_n^{[nt]}, \quad 0 \leq t \leq 1, \quad n \geq 1.$$

Noting the important fact that

$$b_j = \mu_k \quad \text{for all} \quad j \in \tilde{c}_k, \quad 1 \leq k \leq K,$$

152
it follows that if \( \bar{c}_k = (i_1, \ldots, i_m) \), then

\[
\hat{X}_n(\bar{c}_k) = \max_{0 = r_0 \leq \cdots \leq r_m = n} \left[ \sum_{j=1}^{m} \left( \frac{i_j}{r_j} - \frac{i_j}{r_{j-1}} \right) \right].
\]

Consequently,

\[
(8.17) \quad \gamma_n^k(t) = \max_{0 = r_0 \leq \cdots \leq r_m = [nt]} \left\{ \sum_{j=1}^{m} \left( \frac{i_j}{r_j} - \frac{i_j}{r_{j-1}} \right) \right\}
\]

\[
= \sup_{0 = \tau_0 \leq \cdots \leq \tau_m = t} \left\{ \sum_{j=1}^{m} \left[ \frac{i_j}{\sqrt{n \tau_j}} - \frac{i_j}{\sqrt{n \tau_{j-1}}} \right] \right\}
\]

\[
= \sup_{0 = \tau_0 \leq \cdots \leq \tau_m = t} \left\{ \sum_{j=1}^{m} \left[ \xi_j^i(\tau_j) - \xi_j^i(\tau_{j-1}) \right] \right\}
\]

\[
= [\bar{c}^k_0(\xi_n)](t) = [g^k(\xi_n)](t).
\]

Thus \( \gamma_n = G(\xi_n) \). Now \( \xi_n \to D^K_k \) by Donsker's Theorem, and it follows easily from Theorem 2.9 that \( G \) is continuous, so

\[
\gamma_n \to G(D^K_k) \text{ as } n \to \infty
\]

by the Continuous Mapping Theorem, which completes the proof.

Q.E.D.

Corollary 7: (a) \( \frac{1}{\sqrt{n}} (T_n - nu) \to D^k_K \) as \( n \to \infty \).

(b) \( E \left[ \frac{1}{\sqrt{n}} (T^K_n - nu^K_k) \right] \to E[g^k(D^K_k)] \) as \( n \to \infty \), \( 1 \leq k \leq K \).

(c) \( \frac{1}{n} T_n \to \mu_k \) as \( n \to \infty \), \( 1 \leq k \leq K \).
Proof: Part (a) follows from the Continuous Mapping Theorem by applying
the projection \( \pi: \mathbb{D}[0,1]^K \to \mathbb{R}^K \) to \( \nu_n \). Lemma 3(c) and Lemma 5 taken
together show that \( \left\{ \frac{1}{\sqrt{n}} \left( T_n^k - n \mu_k \right) \right\} \) is u.i. for each \( k = 1, \ldots, K \), so
(b) follows from (a) and Theorem 2.14. Part (c) is not directly related
to the preceding theorem but follows directly from Lemma 3(a) and
Lemma 4(a).

Q.E.D.

There are two special cases of particular interest in which the
preceding theorem and its corollary simplify greatly. If
\( b_1 < b_2 < \cdots < b_K \), then \( G \) is just the identity mapping and we have

\[ \nu_n \Longrightarrow D_\text{sk}^K, \quad \frac{1}{\sqrt{n}} \left( T_n - n \mu \right) \overset{D}{\to} N(0, D^2) . \]

If \( b_1 < b_k \) for all \( k = 2, \ldots, K \) then

\[ G^k(x) = x^k \quad \text{for all} \quad k = 2, \ldots, K \]

and we have

\[ \nu_n \Longrightarrow \sigma_{\text{sk}}^k \cdot 1, \quad \frac{1}{\sqrt{n}} \left( T_n - n \mu \right) \overset{D}{\to} N(0, \sigma_k^2 \cdot 1) . \]

From Theorem 6 we can easily obtain a functional limit theorem
for the vector process \( \{ w_n \} \). This result really has content, however,
only in the "heavy traffic" case where \( \rho_k \geq 1 \) for at least one \( k \).

Let

\[ d_k = \mu_k - \mu_{k-1} = (b_k - a_k)^+, \quad 2 \leq k \leq K , \]
\[ d = \begin{pmatrix} d_2 \\ \vdots \\ d_K \end{pmatrix}. \]

Now define
\[ \hat{w}_n = w_n - nd, \quad n \geq 1 \]
(for completeness we take \( \hat{w}_0 = w_0 = 0 \), and the random functions in \( D[0,1]^{K-1} \) generated by \( \{\hat{w}_n\} \),
\[ X_n(t) = \frac{1}{\sqrt{n}} \hat{w}_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots. \]

Further define \( H: D[0,1]^K \to D[0,1]^{K-1} \) and \( h: D[0,1]^K \to \mathbb{R}^{K-1} \) by letting
\[ H^k(x) = G^k(x) - G^{k-1}(x) \quad , \quad 2 \leq k \leq K , \]
\[ h^k(x) = (\pi^k H^k)(x) = g^k(x) - g^{k-1}(x) , \quad 2 \leq k \leq K . \]

An important point to note is that
\[ [H^k(\cdot)](\cdot) \equiv 0 \quad \text{if and only if} \quad \rho_k < 1 , \quad 2 \leq k \leq K . \]

Thus if \( \rho_k < 1 \) for all \( k = 2, \ldots, K \) the following proposition merely states that \( X_n(\cdot) \) converges weakly to the constant function in \( D[0,1]^{K-1} \) all of whose components are everywhere zero. If such is not the case, however, it gives us a joint functional limit theorem for those of the \( \{w_n^k\} \) processes which do blow up.

**Theorem 8:** \[ X_n \Rightarrow H(D^K) \quad \text{as} \quad n \to \infty , \quad \text{where} \ D \ \text{is as in Theorem 6}. \]
Proof: It is immediate from our definitions that

\[ w_n^k = T_n^k - T_n^{k-1} - v_n^k = T_n^k - T_n^{k-1} + \Delta(n^{1/2}) \]

and hence

\[ \hat{w}_n^k = \hat{T}_n^k - \hat{T}_n^{k-1} + \Delta(n^{1/2}) \]

So by Theorem 2.11 it suffices to show that \( \gamma_n \Rightarrow H(D_n^K) \) as \( n \to \infty \),

where

\[ (8.13) \quad \gamma_n^k(t) = \frac{1}{\sqrt{n}} \hat{w}_n^{k[n]} - \frac{1}{\sqrt{n}} \hat{T}_n^{k-1} = v_n^k(t) - v_n^{k-1}(t), \quad 2 \leq k \leq K. \]

Since \( v_n \Rightarrow G(D_n^K) \) by Theorem 6, it follows easily from (8.13),
the definition of \( H \), and the Continuous Mapping Theorem that \( \gamma_n \Rightarrow H(D_n^K) \).

Q.E.D.

Corollary 9: (a) \( \frac{1}{\sqrt{n}} (w_n \cdot d) \Rightarrow h(D_n^K) \) as \( n \to \infty \).

(b) \( E[\frac{1}{\sqrt{n}} (w_n \cdot d)^k] \Rightarrow E[h(D_n^K)] \) as \( n \to \infty \), \( 2 \leq k \leq K \).

(c) \( \frac{1}{n} w_n^k \overset{a.s.}{\rightarrow} d_k \) as \( n \to \infty \), \( 2 \leq k \leq K \).

Proof: Part (a) follows as usual from the preceding theorem and the
Continuous Mapping Theorem. Parts (b) and (c) follow easily from their
counterparts in Corollary 7 and the fact that \( w_n^k = T_n^k - T_n^{k-1} - v_n^k \).

Q.E.D.

Again we remark that there are a number of special cases in which
Theorem 8 and its corollary simplify significantly, but it is left to the reader to investigate these in detail. In general, of course, it is very difficult to determine explicitly the limiting distributions indicated in Corollary 7 and Corollary 9.

Note that by using Theorem 2.13 in conjunction with Theorem 6 a functional limit theorem can be obtained for the vector process

\[
D(t) = \begin{bmatrix}
D^1(t) \\
\vdots \\
D^K(t)
\end{bmatrix}
\]

where \( D^k(t) \) is the number of services completed by server \( k \) in the interval \( (0,t] \). The result which one obtains is a special case of Theorem 9.6 below. It is then an easy matter to obtain a corresponding result for the queue length process

\[
Q(t) = \begin{bmatrix}
Q^2(t) \\
\vdots \\
Q^K(t)
\end{bmatrix}, \quad t \geq 0,
\]

where \( Q^k(t) \) is defined in the obvious way. These functional results for \( \{D(t)\} \) and \( \{Q(t)\} \) have already been obtained, however, by Iglehart and Whitt (1969b). More accurately, those authors have demonstrated a general methodology which can be used to obtain such results directly. Their methodology applies equally well to other more general types of systems, including multi-server queues in series. The approach which we have taken cannot be readily generalized to deal with systems involving multi-server facilities, but it is considerably more direct and yields results in a somewhat more explicit form.
CHAPTER 9

Assembly-Like Networks

We wish to consider now a queueing system composed of many assembly-like operations arranged in a network, so that the outputs from some operations act as inputs to others. Rather than speaking explicitly of input processes, it is convenient (and clearly equivalent) to suppose that the system is composed entirely of \( J \) stations, some of which (called external stations) serve only to generate inputs, and the remainder of which (called non-external stations) are assembly-like operations. We shall assume throughout that among the non-external stations there is a unique terminal station (or "sink"), whose nature will be specified shortly. It is convenient to imagine that the stations of the network are labelled \( 1, \ldots, J \), the external stations being labelled \( 1, \ldots, J_E \) and the terminal station being labelled \( J \) \((1 \leq J_E < J)\). With these conventions and assumptions, an assembly-like network can be schematically represented as in Figure 8 below.

![Figure 8. An example of an assembly-like network.](image)

We assume that each station \( j = 1, \ldots, J-1 \) has a unique successor.
and we let

\[ S(j) = \text{the (unique) successor of station } j \quad (1 \leq j < J). \]

The output from station \( j \) acts as direct input (only) to station \( S(j) \). Station \( J \) (the terminal station) has no successor; its output simply departs the system. For each \( j = J_E + 1, \ldots, J \) we define the (necessarily non-empty) set of station \( j \)'s predecessors,

\[ P(j) = \{ i : j = S(i) \}, \quad J_E < j \leq J. \]

We assume that each station of the network is occupied by (or simply consists of) a single server, and the terms "station \( j \)" and "server \( j \)" will be used interchangeably. The terms "item" and "unit", introduced in Part I, will be used in the following generalized sense. The \( n^{th} \) item of type \( j \)" refers to the entity created by server \( j \) at the completion of his \( n^{th} \) service \( (1 \leq j \leq J; n = 1, 2, \ldots) \). If \( 1 \leq j < J \), then type \( j \) items proceed immediately to station \( S(j) \) upon their creation. Type \( J \) items depart the system immediately upon their creation. Servers at external stations are never idle, but the service discipline at each non-external station is that of an assembly-like operation. In particular, if \( J_E < j \leq J \), then server \( j \) needs exactly one item of each type \( i \in P(j) \) for each of his services, and he uses items of each type in the order of their arrival. If \( P(j) \) contains exactly \( m \) elements, then each service at station \( j \) can be viewed as converting an \( m \)-tuple of items of the specified types into a new and distinct item of type \( j \). Such an \( m \)-tuple, consisting of one
item of each type \( i \in P(j) \), will be referred to as a type \( j \) unit.

Assuming the system to be initially empty, the batch of items consumed by the \( n \)th service at station \( j \) consists precisely of the \( n \)th item of each type \( i \in P(j) \). This particular \( m \)-tuple will be called the \( n \)th unit of type \( j \) \( (J_n < j \leq J; \ n = 1, 2, \ldots) \).

We shall assume throughout that the network is acyclic in the sense that there exists no sequence of labels (or indices) \( j_1, \ldots, j_m \) such that \( j_2 = s(j_1), j_3 = s(j_2), \ldots, j_m = s(j_{m-1}) \) and \( j_1 = j_n \).

Roughly speaking, this means that the output from a given station cannot eventually return to act as input to that same station. Combined with our assumptions that there is a unique terminal station and that every other station has a unique successor, this assumption implies that the network "fans out" from the terminal station. In the language of graph theory, the network is an "arborescence". The assumption that the network contains no cycles is indispensible. It is used only once, in the proof of Lemma 3 below, but all of our subsequent results depend upon that lemma.

The assumption that each station has a unique successor is altogether natural for applications to actual assembly systems. For precedence network applications, however, it is quite restrictive. In Chapter 14 we shall indicate how the arguments and results to follow can be generalized so as to obviate this assumption.

9.1 Formulation and Notation

We shall say that section \( i \) is an antecedent of station \( j \) if the output from station \( i \) acts either directly or indirectly as
input to station \( j \). Given our assumption that the network is acyclic, the set of station \( j \)'s antecedents can be defined recursively by

\[
A(j) = \begin{cases} 
\varnothing, \text{ the empty set} & \text{if } 1 \leq j \leq J_E \\
\bigcup_{i \in P(j)} \{ [i] \cup A(i) \} & \text{if } J_E < j \leq J
\end{cases}
\]

For each \( j = 1, \ldots, J \) we define a **chain to station** \( j \) to be an ordered set of indices

\[
c = (i_1, \ldots, i_m)
\]

such that

\[
1 \leq i_1 \leq J_E,
\]

\[
i_m = j,
\]

and

\[
i_{k-1} \in P(i_k) \text{ for all } k = 2, \ldots, m.
\]

Let

\[
C(j) = \text{the (non-empty) set of all chains to station } j
\]

\( (1 \leq j \leq J) \).

If \( 1 \leq j \leq J_E \), then \( C(j) \) consists only of the singleton \( \{i\} \). For the network pictured in Figure 8,

\[
C(9) = \{(1,7,9),(2,7,9),(3,9),(4,8,9),(5,8,9),(6,8,9)\}
\]

Because our network is acyclic, the elements of a chain to station \( j \)
(1 ≤ j ≤ J) are necessarily distinct.

More generally, a subchain to station j (1 ≤ j ≤ J) is defined to be a non-empty ordered set of indices

\[ c = (i_1, \ldots, i_m) \]

such that

\[ i_m \in A(j) \]

and

\[ i_{k-1} \in A(i_k) \text{ for all } k = 2, \ldots, m \]

Any chain to station j is also a subchain to j, and if 1 ≤ j ≤ J, then the only subchain to j is the singleton \{(j)\}. Note that the elements of a subchain to station j (1 ≤ j ≤ J) are necessarily distinct. Note that each subchain to station j (1 ≤ j ≤ J) can be obtained by deleting indices (or elements, or components) from some chain to station j, maintaining the original ordering of the remaining indices.

Although chains and subchains are defined to be ordered sets (i.e., vectors), we shall occasionally speak of a subchain c which is "contained in" another subchain c'. Such a statement should always be interpreted as meaning that c is contained in c' in the ordinary sense of set inclusion. The reader will note, however, that if c and c' are both subchains to station j with c contained in c', then the elements of c appear in the same order in c' as they do in c.

Having discussed only the structure of the assembly-like network thus far, we now specify our assumptions as to the service processes
at the various stations of the system. It is assumed that the system is empty at time zero and that a service is begun at each external station at that time. We further assume that the server at each non-external station $j$ suffers a period of "initial paralysis" which begins at time zero and lasts a (random) length of time $v^j_0$. When server $j$ recovers from this paralysis, he is ready to begin his first service if the requisite set of inputs from preceding stations is available. For each external station we define

$$v^j_n = \text{the } (n+1)^{\text{st}} \text{ service time at station } j$$

$$(1 \leq j \leq J^E; \ n = 0,1,...) ,$$

and for each non-external station we define

$$v^j_n = \text{the } n^{\text{th}} \text{ service time at station } j$$

$$(J^E < j \leq J; \ n = 1,2,...) .$$

Our assumption throughout is that $\{v^1_n; \ n = 0,1,...\},...,\{v^j_n; \ n = 0,1,...\}$ are mutually independent i.i.d. sequences of non-negative random variables with

$$0 < E(v^j_n) = b_j < \infty \quad (j = 1,\ldots,J),$$

$$0 < \text{Var}(v^j_n) = \sigma^2_j < \infty \quad (j = 1,\ldots,J) ,$$

and as usual we define the partial sums

$$v^j_0 \equiv 0 \quad (j = 1,\ldots,J),$$

$$v^j_n = \sum_{k=0}^{n-1} v^j_k \quad (j = 1,\ldots,J; \ n = 1,2,...) .$$
The first family of stochastic processes with which we shall deal is \( \{ T^j_n \} \), where

\[
T^j_0 = 0 \quad (1 \leq j \leq J),
\]

\[
T^j_n = \text{the time at which server } j \text{ completes his } n^{\text{th}} \text{ service}
\]

\[
(1 \leq j \leq J; \ n = 1,2,\ldots).
\]

Clearly

\[
T^j_n = V^j_n \quad \text{for all} \quad j = 1,\ldots,J_E \quad \text{and} \quad n = 0,1,\ldots.
\]

If \( 1 \leq j < J \) and \( n \geq 1 \), then \( T^j_n \) can alternately be viewed as the time at which the \( n^{\text{th}} \) item of type \( j \) arrives at station \( S(j) \).

The second family of stochastic processes with which we shall deal is \( \{ w^j_n \} \), where

\[
w^j_n = \text{the waiting time (exclusive of service time) at station}
\]

\[
S(j) \text{ for the } n^{\text{th}} \text{ item of type } j \ (1 \leq j < J; \ n = 1,2,\ldots).
\]

Actually the phrase "at station \( S(j) \)" is superfluous, since type \( j \) items are no longer considered to exist once they have undergone service at station \( S(j) \). It is not meaningful, of course, to speak of waiting times for type \( J \) items.

As in the analysis of single-server queues in series, it is important to recognize that each non-external station \( j \) of an assembly-like network is, when viewed in isolation, a single-server queueing system in the formal sense of Section 2.5. To facilitate discussion of these "imbedded" single-server systems we define
\[ U_n^j = \max_{i \in P(j)} T_n^i, \quad J_E < j \leq J, \quad n = 0, 1, 2, \ldots, \]
\[ w_n^j = \min_{i \in P(j)} w_n^i, \quad J_E < j \leq J, \quad n = 1, 2, \ldots. \]

For each \( n \geq 1 \), \( U_n^j \) represents the time at which the last member of the \( n \)th unit of type \( j \) arrives. Thus it is effectively the arrival time for the \( n \)th unit of type \( j \), and \( \{U_n^j\} \) is the sequence of arrival times to the "imbedded" queueing system mentioned above. (Note that \( U_n^j = 0 \) for all \( j = J_E + 1, \ldots, J \).) Correspondingly, \( \{w_n^j; n = 1, 2, \ldots\} \) in the sequence of waiting times (exclusive of service time) in the "imbedded" system.

The continuous-parameter processes to be dealt with are

\[ D_n^j(t) = \text{the number of services completed by server \( j \) up to time} \quad t \quad (1 \leq j \leq J, \ 0 \leq t < \infty), \]

\[ Q_n^j(t) = \text{the number of type} \ j \ \text{items present in the system at time} \ t \quad (1 \leq j < J, \ 0 \leq t < \infty). \]

We also define the accumulated waiting times

\[ W_n^j = \sum_{i=1}^{n} w_i^j, \quad 1 \leq j < J, \quad n = 1, 2, \ldots \]

and identify the vector processes

\[ T_n = \left( \begin{array}{c} T_n^1 \\ \vdots \\ T_n^J \\ q_n^J \\ u_n^J \\ v_n^J \\ w_n^J \\ w_n^{J-1} \\ \vdots \\ w_n^1 \\ w_n^J \end{array} \right), \quad w_n = \left( \begin{array}{c} w_n^1 \\ \vdots \\ w_n^J \\ w_n^{J-1} \\ \vdots \\ w_n^1 \\ w_n^J \end{array} \right), \quad n \geq 1, \]
\[
D(t) = \begin{bmatrix}
D^1(t) \\
\vdots \\
D^j(t) \\
\vdots \\
D^{J-1}(t)
\end{bmatrix}, \quad Q(t) = \begin{bmatrix}
Q^1(t) \\
\vdots \\
Q^j(t) \\
\vdots \\
Q^{J-1}(t)
\end{bmatrix}, \quad t \geq 0 .
\]

Some obvious relationships among the various stochastic processes associated with an assembly-like network are as follows.

(i) \( t_n^j = v_n^j + \tilde{v}_n^j + v_n^j , \quad J_E < j \leq J , \quad n \geq 1 \).

(ii) \( \tilde{w}_n^j = T_n^{S(j)} - T_n^j - v_n^j , \quad 1 \leq j < J , \quad n \geq 1 . \)

(iii) \( D_j(t) = \sup\{n \geq 0 : T_n^j \leq t\} , \quad 1 \leq j < J , \quad t \geq 0 . \)

(iv) \( Q_j(t) = D_j(t) - D_j^{S(j)}(t) , \quad 1 \leq j < J , \quad t \geq 0 . \)

It is intuitively clear that the manner in which congestion develops at a non-external station \( j \) is critically dependent upon the relative magnitude of the mean service time at station \( j \) and the largest of the mean service times at stations \( i \in P(j) \). Accordingly, we define

\[
\alpha_j = \begin{cases}
0 & \text{if } 1 \leq j \leq J_E \\
\max_{i \in A(j)} b_i & \text{if } J_E < j \leq J
\end{cases}
\]

\[
\rho_j = b_j / \alpha_j , \quad J_E < j \leq J .
\]

The quantity \( \rho_j \) will be called the traffic intensity at station \( j \).

Finally we define the important set \( \mathcal{C}(j) \) of critical subchains to
station \( j \) \((1 \leq j \leq J)\) in the following constructive manner. From each chain

\[ c = (i_1, \ldots, i_m) \in C(j) \]

delete all indices \( i_k \) such that

\[ b_{i_k} < \max(a_j, b_j) , \]

maintaining the original ordering of the remaining indices. If the resulting ordered set of indices is non-empty (i.e., if it is a sub-chain to \( j \)), then it is a candidate for \( \tilde{C}(j) \). Then let \( \tilde{C}(j) \) be the set of all candidates which are not contained in any other candidate.

To repeat, we let

\[ \tilde{C}(j) = \text{the (necessarily non-empty) set of critical subchains to station } j \ (1 \leq j \leq J) . \]

For purposes of illustration, the network of Figure 8 has been redrawn in Figure 9 below with hypothetical \( b_j \) values added. It has already been pointed out that for this network

\[ C(9) = \{(1,7,9),(2,7,9),(3,4),(4,8,9),(5,8,9),(6,8,9)\} \]

![Diagram](image-url)

Figure 9. The network of Figure 8 with hypothetical \( b_j \) values added.
Thus, with the indicated $b_j$ values, the candidates for $\mathcal{C}(9)$ are
$(7,9), (2,7,9), (3,9)$ and $(9)$. Deleting the first and last of these because they are contained in the second and third, we have

$$\mathcal{C}(9) = \{(2,7,9),(3,9)\}.$$ 

Similarly, one finds that

$$\mathcal{C}(7) = \{(2,7)\}, \quad \mathcal{C}(8) = \{(4)\}.$$ 

Of course for external stations $j$ we always have $\mathcal{C}(j) = C(j) = \{(j)\}$.

9.2 Basic Representations and Preliminary Results

The initial goal of this section is to develop an explicit representation for $T_n^j$ in terms of the elemental sequences $\{v_n^j\}$. The key to success in this regard is the all-important Lemma 2.26 for single-server queueing systems. We have noted above that each non-external station $j$ in an assembly-like network is a single-server queueing system when viewed in isolation. The duration of the server's "initial paralysis" in this imbedded system is $v_0^j$ and the sequence of service times is $\{v_n^j; n = 1,2,\ldots\}$. The "customers" in the imbedded system are type $j$ units, so the sequence of arrival times is $\{u_n^j; n = 1,2,\ldots\}$ with $v_0^j = 0$. The sequence of "customer" waiting times is $\{w_n^j; n = 1,2,\ldots\}$. Thus the following proposition is immediate from Lemma 2.26.

**Lemma 1:** If $j_E < j < J$ and $n \geq 1$, then

168
\[\tilde{w}_n^j = (v_n^j - u_n^j) - \min_{0 \leq k \leq n} (v_k^j - u_k^j).\]

**Lemma 2:** If \( J_E < j \leq J \) and \( n \geq 1 \), then

\[T_n^j = \max_{i \in P(j)} \left\{ \max_{0 \leq k \leq n} [T_k^i + (v_n^j - v_k^j)] + v_n^j \right\}.\]

**Proof:** First we recall the basic relationship

\[t_n^j = u_n^j + \tilde{w}_n^j + v_n^j, \quad J_E < j \leq J, \quad n \geq 1.\]

Then using the representation for \( \tilde{w}_n^j \) given in Lemma 1 and the defining relationship

\[u_n^j = \max_{i \in P(j)} t_n^i, \quad J_E < j \leq J, \quad n \geq 0,\]

we have

\[t_n^j = v_n^j - \min_{0 \leq k \leq n} (v_k^j - u_k^j) + v_n^j\]

\[= \max_{0 \leq k \leq n} [u_k^j + (v_k^j - v_n^j)] + v_n^j\]

\[= \max_{0 \leq k \leq n} [\max_{i \in P(j)} [T_k^i + (v_n^j - v_k^j)] + v_n^j]\]

\[= \max_{i \in P(j)} \left\{ \max_{0 \leq k \leq n} [T_k^i + (v_n^j - v_k^j)] + v_n^j \right\}.

Q.E.D.

Using Lemma 2 and the fact that \( t_n^j = v_n^j \) for all external stations \( j \), an exact representation for \( T_n^j \) can be developed inductively. The
exact representation is somewhat messy, but there exists a reasonably compact approximation which will suffice for our purposes. This approximation is expressed in terms of the random variables $X_n(c)$, defined in Chapter 8. For ease of reference we repeat that definition.

**Definition:** For each $n = 0, 1, 2, \ldots$ and each non-empty ordered set $c = (i_1, \ldots, i_m)$ of distinct indices satisfying $1 \leq i_k \leq J$ for all $k = 1, \ldots, m$ we define

$$X_n(c) = \max_{0 = r_0 \leq \cdots \leq r_m = n} \left[ \sum_{j=1}^{m} (V_{i_j}^j - V_{r_j}^{j-1}) \right],$$

the maximization being over integer values of $r_1, \ldots, r_{m-1}$ which satisfy the indicated constraint. Note that $X_0(c) = 0$, and if $c = (i_1)$, a singleton, then $X_n(c) = V_{i_1}^1$ for all $n \geq 0$.

In terms of this convenient notation, our approximate representation for $T_n^j$ is

$$T_n^j \approx \max_{c \in C(j)} X_n(c), 1 \leq j \leq J, n \geq 0.$$  

The following proposition shows the sense in which this approximation is "sufficiently accurate".

**Lemma 3:** For all $j = 1, \ldots, J$

(a) $\frac{1}{n} [T_n^j - \max_{c \in C(j)} X_n(c)] \xrightarrow{a.s.} 0$ as $n \to \infty$,

(b) $[T_n^j - \max_{c \in C(j)} X_n(c)] = \Delta(n^{1/2})$,

(c) $\left\{ \frac{1}{\sqrt{n}} [T_n^j - \max_{c \in C(j)} X_n(c)] \right\}$ is uniformly integrable.
Proof: For each \( j = 1, \ldots, J \) and \( n \geq 0 \), let

\[
\delta_n^j = T_n^j - \max_{c \in C(j)} X_n(c) .
\]

We first prove by induction that for all \( j = 1, \ldots, J \)

\[
0 \leq \delta_n^j \leq \sum_{m \in A(j)} \max_{0 \leq r \leq n} v_r^m + v_n^j , \quad n \geq 0 .
\]

(9.1)

If \( 1 \leq j \leq J_E \), then \( C(j) \) contains only the singleton \( (j) \) and it is immediate that

\[
\max_{c \in C(j)} X_n(c) = v_n^j = T_n^j , \quad n \geq 0 .
\]

Thus (9.1) is satisfied for all external stations \( j \). Now proceeding inductively, suppose that \( J_E < j \leq J \) and that (9.1) is satisfied when \( j \) is replaced by any \( i \in P(j) \). It is immediate then that

\[
0 \leq \delta_n^j \leq \sum_{m \in A(j)} \max_{0 \leq r \leq n} v_r^m , \quad i \in P(j) , \quad n \geq 0 .
\]

(9.2)

Since \( X_o(c) \equiv 0 \) for all \( c \in C(j) \) and \( T_o^j = 0 \) by definition, we have \( \delta_o^j = 0 \), and (9.1) is trivially satisfied for \( n = 0 \). Moreover, for \( n \geq 1 \) we have from Lemma 2 that

\[
T_n^j = \max_{i \in P(j)} \left\{ \max_{0 \leq r \leq n} \left[ T_r^i + (v_n^j - v_r^j) \right] + v_n^j \right\}
\]

\[
= \max_{i \in P(j)} \left\{ \max_{0 \leq r \leq n} \left[ \max_{c \in C(i)} X_r(c) + \delta_r^i + (v_n^j - v_r^j) \right] + v_n^j \right\}
\]

171
Since $S_r^j \geq 0$ for all $i \in P(j)$ and $r \geq 0$ by (9.2), it follows that

\begin{equation}
0 \leq m_n^j - \max_{i \in P(j)} \left\{ \max_{0 \leq r \leq n} \left[ \max_{c \in C(i)} X_r(c) + (v_n^j - v_r^j) \right] \right\}
\end{equation}

\begin{equation}
\leq \max_{i \in P(j)} \left\{ \max_{0 \leq r \leq n} S_r^j + v_n^j \right\}.
\end{equation}

If $c$ is any chain not containing $j$, then it is immediate from the definition of $X_n(c)$ that

\begin{equation}
\max_{0 \leq r \leq n} [X_r(c) + (v_n^j - v_r^j)] = X_n(c, j), \quad n \geq 0.
\end{equation}

Moreover, it is clear that

\begin{equation}
c(j) = \bigcup_{i \in P(j)} (c, j), \quad \text{since } J_E < j \leq J.
\end{equation}

Combining (9.4) and (9.5), we have

\begin{equation}
\max_{i \in P(j)} \left\{ \max_{0 \leq r \leq n} \left[ \max_{c \in C(i)} X_r(c) + (v_n^j - v_r^j) \right] \right\}
= \max_{i \in P(j)} \left\{ \max_{0 \leq r \leq n} \left[ X_r(c) + (v_n^j - v_r^j) \right] \right\}
= \max_{i \in P(j)} \left\{ X_n(c, j) \right\} = \max_{c \in C(j)} X_n(c).
\end{equation}

Next we use (9.2) to write
(9.7) \[
\max_{i \in \mathcal{P}(j)} \left\{ \max_{0 \leq r \leq n} \mathbf{s}_r^i + v_n^j \right\} \\
\leq \max_{i \in \mathcal{P}(j)} \left\{ \max_{0 \leq r \leq n} \left[ \sum_{m \in \mathcal{A}(j)} \max_{0 \leq k \leq r} v_r^m \right] + v_n^j \right\} \\
= \sum_{m \in \mathcal{A}(j)} \max_{0 \leq r \leq n} \left( v_r^m + v_n^j \right).
\]

Combining (9.3), (9.6) and (9.7) we have that (9.1) is satisfied for the particular \( j \) in question. By induction then, (9.1) is satisfied for all \( j = 1, \ldots, J. \)

Using (9.1) in conjunction with Lemma 2.19, Theorem 2.18 and the fact that

\[
\frac{1}{n} \max_{0 \leq r \leq n} v_r^j \xrightarrow{a.e.} 0 \quad \text{as} \quad n \to \infty, \quad 1 \leq j \leq J,
\]

each of the desired results follows easily.

Q.E.D.

We conclude this chapter with two propositions which, when used in conjunction with the preceding lemma, are of crucial importance in developing limit theorems for the vector process \( \{T_n\} \). In proving these propositions there is a heavy reliance on Lemmas 8.4 and 8.5, dealing with the random variables \( X_n(c) \) in the context of a GI/G/1 series queue. We note again that if

\[
c = (i_1, \ldots, i_m) \in \mathcal{C}(j), \quad 1 \leq j \leq J,
\]

*/ The legitimacy of the general inductive step follows from our assumption that the network is acyclic.
then the indices $i_1, \ldots, i_m$ are distinct. Thus $\{v_{n_1^1}, \ldots, v_{n^m}\}$ are mutually independent i.i.d. sequences, and Lemmas 8.3 and 8.4 continue to hold (except for some minor and obvious discrepancies in notation) when $c_{k'}$ is replaced by $c$.

**Lemma 4:** For each $j = 1, \ldots, J$

(a) \( \frac{1}{n} \left[ \max_{c \in C(j)} X_n(c) \right] \overset{a.s.}{\rightarrow} \max(a_j, b_j) \) as $n \to \infty$,

(b) \( \max_{c \in C(j)} X_n(c) = \max_{\tilde{c} \in \tilde{C}(j)} X_n(\tilde{c}) + \Delta(n^{1/2}) $.

**Proof:** Let $j$ be arbitrary but fixed. If

\[ c = (i_1, \ldots, i_m) \in C(j), \]

then Lemma 8.4 establishes that

**9.8**

\[ \frac{1}{n} X_n(c) \overset{a.s.}{\rightarrow} \max_{1 \leq k \leq m} (b_{i_k}) \text{ as } n \to \infty \]

Then by Lemma 2.23 (a) and induction

\[ \frac{1}{n} \left[ \max_{c \in C(j)} X_n(c) \right] = \max_{c \in C(j)} \left[ \frac{1}{n} X_n(c) \right] \overset{a.s.}{\rightarrow} d_j , \]

where

\[ d_j = \max_{c = (i_1, \ldots, i_m) \in C(j)} \left\{ \max_{1 \leq k \leq m} (b_{i_k}) \right\} = \max(a_j, b_j). \]

So (a) is proved. To prove (b), let $C'(j)$ denote the (non-empty) set of all chains
\[ c = (i_1, \ldots, i_m) \in C(j) \]

such that

\[ \max_{1 \leq k \leq m} (b_{i_k}) = \max(a_j, b_j) . \]

From (9.8), Lemma 2.23 (b) and induction it follows easily that

\[ \max_{c \in C(j)} X_n(c) = \max_{c \in C'(j)} X_n(c') + \Delta(n^{1/2}) . \]  

(9.9)

Now form a set \( C^*(j) \) of subchains to station \( j \) by deleting from each chain \( c' \in C'(j) \) all indices \( i_k \) such that

\[ b_{i_k} < \max(a_j, b_j) . \]

Lemma 8.4 (b) shows that if \( c' \in C'(j) \) and \( c^* \) is the corresponding element of \( C^*(j) \), then

\[ X_n(c') = X_n(c^*) + \Delta(n^{1/2}) . \]  

(9.10)

Combining (9.9) and (9.10) we have

\[ \max_{c \in C(j)} X_n(c) = \max_{c^* \in C^*(j)} X_n(c^*) + \Delta(n^{1/2}) . \]  

(9.11)

Now a review of the definition of \( \bar{C}(j) \) will show that \( C^*(j) \) is what was called the set of candidates for \( \bar{C}(j) \). That is, \( C^*(j) \) consists of all critical subchains \( \bar{C} \in \bar{C}(j) \) plus possibly some additional subchains \( c^* \) such that

175
c* \subset \tilde{c} \text{ for some } \tilde{c} \in \mathcal{C}(j) .

But if \( c^* \) is contained in \( \tilde{c} \), it follows easily from the definition of \( X_n(\cdot) \) that

\[
X_n(\tilde{c}) \geq X_n(c^*) \text{ for all } n = 0, 1, 2, \ldots .
\]

Thus the maximization in (9.11) can be restricted to critical subchains \( \tilde{c} \in \mathcal{C}(j) \), which proves part (b). Q.E.D.

**Lemma 5:** For each \( j = 1, \ldots, J \)

\[
\left\{ \frac{1}{\sqrt{n}} \left[ \max_{c \in \mathcal{C}(j)} X_n(c) - n \max(a_j, b_j) \right] \right\}
\]

is a uniformly integrable sequence.

**Proof:** Let \( j \) be arbitrary but fixed and define

\[
d(c) = \max_{1 \leq k \leq m} (b_{j_k}) \text{ for each } c = (i_1, \ldots, i_m) \in \mathcal{C}(j) .
\]

Lemma 8.5 shows that

\[
\left\{ \frac{1}{\sqrt{n}} [X_n(c) - nd(c)] \right\} \text{ is u.i., } c \in \mathcal{C}(j) .
\]

So from Lemma 2.24

\[
\left\{ \frac{1}{\sqrt{n}} \left[ \max_{c \in \mathcal{C}(j)} X_n(c) - n \max_{c \in \mathcal{C}(j)} d(c) \right] \right\} \text{ is u.i.}
\]

But \( \max_{c \in \mathcal{C}(j)} d(c) = \max(a_j, b_j) \), so the proposition is proved. Q.E.D.
CHAPTER 10

Limit Theorems for a General Assembly-Like Network

We begin this chapter by developing a functional limit theorem for the vector process \( \{ T_n \} \). This result and all those to follow are expressed in terms of the mappings \( F^c \), defined in Chapter 8. For ease of reference we repeat that definition.

Definition: For each non-empty ordered set \( c = (i_1, \ldots, i_n) \) of distinct indices satisfying \( 1 \leq i_k \leq J \) for all \( k = 1, \ldots, m \) we define mappings

\[
F^c : D[0,1]^J \to D[0,1], \quad f^c : D[0,1]^J \to \mathbb{R}
\]

as follows. For \( x = (x^1, \ldots, x^J) \in D[0,1]^J \), let

\[
[F^c(x)](t) = \sup_{0 = \tau_0 \leq \cdots \leq \tau_m = t} \left\{ \sum_{k=1}^{m} [x^{i_k}(\tau_k) - x^{i_k}(\tau_{k-1})] \right\},
\]

\[
f^c(x) = (f \circ F^c)(x) = \sup_{0 = \tau_0 \leq \cdots \leq \tau_m = 1} \left\{ \sum_{k=1}^{m} [x^{i_k}(\tau_k) - x^{i_k}(\tau_{k-1})] \right\}.
\]

Letting

\[
\mu_j = \max(a_j, b_j), \quad 1 \leq j \leq J,
\]

\[
\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_J \end{pmatrix},
\]

177
we define the translated random vectors

\[ \hat{T}_n = T_n - n\mu, \quad n = 0, 1, 2, \ldots \]

and the random functions in \( D[0,1]^J \) generated by \( \{\hat{T}_n\} \),

\[ \nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_{[nt]}, \quad 0 \leq t \leq 1, \quad n = 0, 1, \ldots. \]

To state a weak convergence theorem for \( \nu_n \), we define

\[ G : D[0,1]^J \rightarrow D[0,1]^J, \quad g : D[0,1]^J \rightarrow \mathbb{R}^J \]

by letting

\[ [G^j(x)](t) = \max_{\tilde{c} \in \mathcal{C}(j)} [\tilde{f}(x)](t), \quad 0 \leq t \leq 1, \quad 1 \leq j \leq J, \]

\[ g^j(x) = (\pi^j G^j)(x) = \max_{\tilde{c} \in \mathcal{C}(j)} [\tilde{f}(x)], \quad 1 \leq j \leq J. \]

Also define the diagonal matrix

\[ D = \begin{pmatrix} 0 & & & & 0 \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ 0 & & & & 0 \end{pmatrix}. \]

**Theorem 1:** \( \nu_n \Rightarrow G(D^J) \) as \( n \rightarrow \infty. \)

**Proof:** Let
\[ \hat{X}_n(\tilde{c}) = X_n(\tilde{c}) - n\mu_j, \quad 1 \leq j \leq J, \quad \tilde{c} \in \tilde{O}(j), \quad n \geq 0. \]

Combining Lemmas 9.3 (b) and 9.4 (b), we have

\[ T_n^j = \max_{\tilde{c} \in \tilde{O}(j)} X_n(\tilde{c}) + \Delta(n^{1/2}), \quad 1 \leq j \leq J, \]
and consequently

\[ T_n^j = \max_{\tilde{c} \in \tilde{O}(j)} \hat{X}_n(\tilde{c}) + \Delta(n^{1/2}), \quad 1 \leq j \leq J. \]

Thus by Theorem 2.11 it suffices to show that \( \gamma_n \Rightarrow G(D_\xi^J) \), where

\[ \gamma_n^j(t) = \frac{1}{\sqrt{n}} \max_{\tilde{c} \in \tilde{O}(j)} X_{[nt]}(\tilde{c}), \quad 0 \leq t \leq 1, \quad 1 \leq j \leq J, \quad n \geq 1. \]

Now let

\[ \hat{V}_n^j = v_n^j - nb_j, \quad 1 \leq j \leq J, \quad n \geq 0, \]

\[ \hat{V}_n = \left( \begin{array}{c} \hat{V}_n^1 \\ \vdots \\ \hat{V}_n^J \end{array} \right), \quad n \geq 0, \]

\[ \xi_n(t) = \frac{1}{\sqrt{n}} \hat{V}_{[nt]}, \quad 0 \leq t \leq 1, \quad n \geq 1. \]

Note that if \( \tilde{c} = (i_1, \ldots, i_m) \in \tilde{O}(j) \) then

\[ b_{\tilde{c}} = \mu_j, \quad 1 \leq k \leq m, \]
and hence

\[(10.1) \quad X_n(\mathcal{C}) = \max_{0 = r_0 \leq \cdots \leq r_m = n} \left\{ \sum_{k=1}^{m} (\hat{V}_k - \hat{V}_{k-1}) \right\} \]

Using (10.1) and proceeding exactly as in the proof of Theorem 8.6, we find that for each \( j = 1, \ldots, J \) and \( \tilde{\omega} \in \mathcal{C}(j) \),

\[
\frac{1}{\sqrt{n}} X_{[nt]}(\tilde{\omega}) = [F(\hat{\xi}_n)](t), \quad 0 \leq t \leq 1, \quad n \geq 1.
\]

Thus, for each \( j = 1, \ldots, J \) and \( n \geq 1 \),

\[
\gamma_n^j(t) = \max_{\tilde{\omega} \in \mathcal{C}(j)} \left\{ \frac{1}{\sqrt{n}} X_{[nt]}(\tilde{\omega}) \right\} = \max_{\tilde{\omega} \in \mathcal{C}(j)} \left\{ [F(\hat{\xi}_n)](t) \right\} = [G^j(\hat{\xi}_n)](t), \quad 0 \leq t \leq 1,
\]

or more simply

\[
\gamma_n = G(\xi_n), \quad n = 1, 2, \ldots
\]

Now \( \xi_n \rightarrow D^J \) by Donsker's Theorem, and it follows easily from Theorem 2.9 that \( G \) is continuous, so

\[
\gamma_n \rightarrow G(D^J) \quad \text{as} \quad n \rightarrow \infty.
\]

by the Continuous Mapping Theorem.

Q.E.D.
Corollary 2: (a) \( \frac{1}{\sqrt{n}} (T_n - n\mu) \xrightarrow{D} g(D_\epsilon^J) \) as \( n \to \infty \).

(b) \( E[\frac{1}{\sqrt{n}} (T_n - n\mu_j)] \to E[g^j(D_\epsilon^J)] \) as \( n \to \infty \), \( 1 \leq j \leq J \).

(c) \( \frac{1}{n} T_n^j \xrightarrow{a.s.} \mu_j \) as \( n \to \infty \), \( 1 \leq j \leq J \).

Proof: Part (a) follows from the Continuous Mapping Theorem by applying the projection \( \pi: D[0,1]^J \to \mathbb{R}^J \) to \( \nu_n \). Lemmas 9.3 (c) and 9.5 taken together show that \( \frac{1}{\sqrt{n}} (T_n^j - n\mu_j) \) is uniformly integrable for each \( j = 1, \ldots, J \), so part (b) follows from (a) and Theorem 2.14. Part (c) follows directly from Lemmas 9.3 (a) and 9.4 (a).

Q.E.D.

We can now use Theorem 1 to obtain a functional limit theorem for the vector waiting time process \( \{w_n\} \). Let

\[
d_j = \mu_\delta(j) - \mu_j , \quad 1 \leq j < J ,
\]

\[
d = \begin{pmatrix}
d_1 \\
\vdots \\
\vdots \\
d_{J-1}
\end{pmatrix}.
\]

We define the translated random vectors

\[
\hat{v}_n = w_n - nd , \quad n = 1,2,\ldots
\]

and (letting \( \hat{w}_0 = w_0 \equiv 0 \) for completeness) the random functions in \( D[0,1]^J \) generated by \( \{\hat{w}_n\} \),
\[ X_n(t) = \frac{1}{\sqrt{n}} \hat{W}[nt], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots \]

We also define mappings

\[ H: D[0,1]^J \rightarrow D[0,1]^{J-1}, \quad h: D[0,1]^J \rightarrow D^{J-1} \]

by letting

\[ [H^j(x)](t) = [G^S(j)(x)](t) - [G^j(x)](t), \quad 1 \leq j < J, \]

\[ h^j(x) = (\pi_j H^j)(x) = g^S(j)(x) - g^j(x), \quad 1 \leq j < J. \]

**Theorem 3:** \( X_n \Rightarrow H(D_{\xi}^J) \) as \( n \rightarrow \infty \).

**Proof:** As was noted in Chapter 9,

\[ w_n^j = T_n^S(j) - \frac{T_n^j}{n} - v_n^j = T_n^S(j) - T_n^j + \Delta(n^{1/2}), \quad 1 \leq j < J, \]

and hence

\[ \hat{w}_n^j = T_n^S(j) - \frac{T_n^j}{n} + \Delta(n^{1/2}), \quad 1 \leq j < J. \]

Thus by Theorem 2.11 it suffices to show that \( \gamma_n \Rightarrow H(D_{\xi}^J) \), where

\[ (10.2) \quad \gamma_n^j(t) = \frac{1}{\sqrt{n}} \hat{T}_n^S(j) - \frac{1}{\sqrt{n}} \hat{T}_n^j \]

\[ = v_n^S(j)(t) - v_n^j(t), \quad 0 \leq t \leq 1, \quad 1 \leq j < J, \quad n \geq 1. \]
Since \( \gamma_n \Rightarrow G(D_{\xi^J}) \) by Theorem 1, it follows easily from (10.2), the definition of \( H \) and the Continuous Mapping Theorem that \( \gamma_n \Rightarrow H(D_{\xi^j}). \)

Q.E.D.

**Corollary 4:**
(a) \( \frac{1}{\sqrt{n}} (w_n - nd) \overset{D}{\to} h(D_{\xi^j}) \) as \( n \to \infty. \)

(b) \( E[\frac{1}{\sqrt{n}} (w_n - nd_j)] \to E[h(D_{\xi^j})] \) as \( n \to \infty, \ 1 \leq j < J. \)

(c) \( \frac{1}{n} w_n^j \overset{a.e.}{\to} d_j \) as \( n \to \infty, \ 1 \leq j < J. \)

**Proof:** Part (a) follows as usual from the Continuous Mapping Theorem, and parts (b) and (c) follow directly from the corresponding parts of Corollary 2 and the fact that \( w_n^j = T_n^j - \tau_n - \gamma_n^j. \)

Q.E.D.

**Corollary 5:**
(a) \( \frac{1}{\sqrt{n^3}} (W_n - \frac{n^2}{2} d) \overset{D}{\to} \int_0^1 [H(D_{\xi^j})](t)dt \) as \( n \to \infty. \)

(b) \( E[\frac{1}{\sqrt{n^3}} (W_n - \frac{n^2}{2} d_j)] \overset{D}{\to} \frac{n^2}{2} E[g(D_{\xi^j})] \) as \( n \to \infty, \ 1 \leq j < J. \)

(c) \( \frac{1}{n^2} W_n^j \overset{a.e.}{\to} \frac{1}{2} d_j \) as \( n \to \infty, \ 1 \leq j < J. \)

**Proof:** We first note that

\[
(10.3) \quad \int_0^1 x_n(t)dt = \sum_{i=1}^{n-1} \left( \frac{1}{n} \right) \left[ \frac{1}{\sqrt{n}} \left( W_i - id \right) \right] = \frac{1}{\sqrt{n^3}} [W_{n-1} - d] \sum_{i=1}^{n-1} 1.
\]

Since \( x_n \Rightarrow H(D_{\xi^j}) \) by Theorem 3 and the integration functional is continuous by Theorem 2.9, the Continuous Mapping Theorem gives

\[
(10.4) \quad \int_0^1 x_n(t)dt \overset{D}{\to} \int_0^1 [H(D_{\xi^j})](t)dt.
\]
Then noting that \[ \sum_{i=1}^{n} i \sim \frac{n^2}{2} \] and

\[ \frac{1}{\sqrt{n}} w_n \xrightarrow{P} 0 \]

by Corollary 4, part (a) follows from (10.3) and (10.4). Parts (b) and (c) follow directly from the corresponding parts of Corollary 4 and from Lemma 2.25.

Q.E.D.

In the following chapter it will be shown how the general results presented here simplify when certain relationships of particular interest exist among the mean service times \( b_j \). Even without simplification, however, these results contain a great deal of qualitative information about the asymptotic behavior of the system. First consider that the completion time processes \( \{r_n^j\} \). Note that the mapping \( G^j(\cdot) \) involves only the critical subchains to station \( j \). Consequently, we conclude that the output process from the sub-network consisting of station \( j \) and its antecedents is essentially unaffected by those stations which do not have maximal mean service times. Also, by noting that \( G^j(\cdot) \) depends only on critical subchains, one can often demonstrate the "asymptotic independence" of two departure processes which appear from the structure of the network to be closely interrelated.

The distribution of the \( J \)-dimensional diffusion process \( G(D^J_\varepsilon) \) is in general very complex. Except for a few very special cases, we cannot even state explicitly the distribution of \( g^j(D^j_\varepsilon) \) for each \( j = 1, \ldots, J \). It is easily shown, however, that for each \( j = 1, \ldots, J \)
the distribution of $g^j(D^j_0)$ is spread over the entire real line and does not have positive probability concentrated at any single point.

Turning next to the waiting time processes $\{w^j_n\}$, let us concentrate attention on the items produced by those stations $j \in \mathcal{P}(J)$. (The remarks to follow apply equally well to any other $j$, since one can simply consider the sub-network consisting only of station $S(j)$ and its antecedents.) Several different cases arise.

(i) If $\rho_j > 1$ then, for each $j \in \mathcal{P}(J)$, $d_j > 0$ and the distribution of $h^j(D^j_0)$ is spread over the whole real line with no positive probability at any single point.

(ii) If $\rho_j = 1$, let us define the critical predecessors of station $J$,

$$\overline{\mathcal{P}}(J) = \{j \in \mathcal{P}(J) : \mu_j = b_j\} .$$

If $j \in \mathcal{P}(J)$, then $d_j = 0$ and the distribution of $h^j(D^j_0)$ is concentrated on $\mathbb{R}^+$ but does not have positive probability at any single point. If $j \in [\mathcal{P}(J) - \overline{\mathcal{P}}(J)]$, then $d_j > 0$ and the distribution of $h^j(D^j_0)$ is spread over all of $\mathbb{R}$ with no positive probability at any single point.

(iii) If $\rho_j < 1$, let $\overline{\mathcal{P}}(J)$ be defined as above. If $j$ is the only element of $\overline{\mathcal{P}}(J)$ then $h^j(D^j_0) = 0$, implying that

$$\frac{1}{\sqrt{n}} w^j_n \to^n P_0 \quad \text{as} \quad n \to \infty .$$

If, however, $\overline{\mathcal{P}}(J)$ has more than one element, then for each
\( j \in \mathcal{P}(J), \quad d_j = 0 \) and the distribution of \( h^j(D_{\xi}^J) \) is concentrated on \( \mathbb{R}^+ \) with a mass of

\[
P\left( \max_{\tilde{c} \in \mathcal{C}(j)} \tilde{r}^j(D_{\xi}^J) = \max_{i \in \mathcal{P}(J)} \; \max_{\tilde{c} \in \mathcal{C}(i)} \tilde{r}^i(D_{\xi}^J) \right)
\]

concentrated at the origin. If \( j \in [\mathcal{P}(J) - \mathcal{P}(J)] \), then \( d_j > 0 \) and the distribution of \( h^j(D_{\xi}^J) \) is spread over all of \( \mathbb{R} \) with no positive probability at any single point.

An immediate consequence of these remarks is that the distribution of the random vector \( h(D_{\xi}^J) \) is degenerate at 0 if and only if the assembly-like network reduces to a GI/G/1 series queue with \( \rho_j < 1 \) for all \( j = 2, \ldots, J \). Thus the waiting time vector \( w_n \) can converge in distribution as \( n \to \infty \) only in that case.

We conclude this chapter by showing that the asymptotic behavior of the continuous-parameter processes \( \{D(t)\} \) and \( \{Q(t)\} \) is essentially equivalent to that of their discrete-parameter counterparts, \( \{T_n\} \) and \( \{w_n\} \). We shall state and prove only the functional limit theorems for the continuous-parameter processes, leaving it to the reader to extract the corollaries of interest. Let

\[
\alpha_j = \mu_j^{-1} \quad , \quad 1 \leq j \leq J ,
\]

\[
\beta_j = \alpha_j - \alpha_{S(j)} , \quad 1 \leq j < J ,
\]

\[
\alpha = \begin{pmatrix}
\alpha_1 \\
\vdots \\
\alpha_J
\end{pmatrix} , \quad \beta = \begin{pmatrix}
\beta_1 \\
\vdots \\
\beta_{J-1}
\end{pmatrix} ,
\]

186
and define the random functions in $D[0,1]^J$ and $D[0,1]^{J-1}$ generated by $\{D(t)\}$ and $\{Q(t)\}$ respectively,

$$\xi_n(t) = \frac{1}{\sqrt{n}} [D(nt) - nt], \quad 0 \leq t \leq 1, \quad n \geq 1,$$

$$\gamma_n(t) = \frac{1}{\sqrt{n}} [Q(nt) - nt], \quad 0 \leq t \leq 1, \quad n \geq 0.$$

We also define the diagonal matrix

$$D_n = \begin{pmatrix}
\sigma_1^2 \\
\frac{\lambda}{2} \\
\mu_1 \\
& \ddots \\
& & \sigma_J^2 \\
& & \frac{\lambda}{3} \\
& & \mu_J
\end{pmatrix},$$

and let $\tilde{D}$ denote its square root.

**Theorem 6:** $\xi_n \Rightarrow -G(\tilde{D}^J_n)$ as $n \to \infty$.

**Theorem 7:** $\gamma_n \Rightarrow H(\tilde{D}^J_n)$ as $n \to \infty$.

**Proof:** We have noted earlier that

$$D^J(t) = \max\{n; T_n^J \leq t\}, \quad 1 \leq J \leq J, \quad 0 \leq t < \infty,$$

i.e., $\{D^J(t)\}$ is the counting process corresponding to $\{T_n^J\}$ for each $J = 1, \ldots, J$. Also, it is easily verified from the sample path continuity of $\xi_n^J$ that the sample paths of $G(D^J_n)$ are almost surely continuous. Combining this fact with (10.5), Theorem 1 and Theorem 2.13, it is immediate that
(10.6) \[ \zeta_n \implies -A\zeta, \]

where

\[
A = \begin{pmatrix}
\mu_1^{-1} & 0 \\
\vdots & \ddots \\
0 & \mu_J^{-1}
\end{pmatrix}
\]

and \( \zeta = (\zeta^1, \ldots, \zeta^J) \) is a random element of \( D[0,1]^J \) defined by

\[ \zeta^j(t) = [g^j(D \zeta^j)](\mu_j^{-1}t), \quad 0 \leq t \leq 1, \quad 1 \leq j \leq J. \]

(Since \([g^j(\cdot)](t)\) is not formally defined for \( t > 1 \), we must assume \( \mu_j \geq 1 \) for all \( j = 1, \ldots, J \) is this definition of \( \zeta \) is to be meaningful. There is no loss of generality in doing so, however, since this is just a matter of scale.) Let us define a mapping

\[ \varphi = (\varphi^1, \ldots, \varphi^J): D[0,1]^J \to D[0,1]^J \]

by letting

\[ [\varphi^j(x)](t) = x^j(\mu_j^{-1}t), \quad 0 \leq t \leq 1, \quad 1 \leq j \leq J, \quad x \in D[0,1]^J. \]

In order to prove Theorem 6 we need to show that the J-dimensional diffusion process \( A\zeta \) has the same distribution as \( G(D \zeta^J) \). Since \( \zeta = \varphi(G(D \zeta^J)) \), this requirement is equivalently stated as

(10.7) \[ A\zeta = \varphi(G(D \zeta^J)) \overset{D}{=} G(D \zeta^J). \]
The proof of (10.7) is divided into two parts. We first show that, because of the particular nature of the mapping $G$,

\[(10.8) \quad \varphi(G(x)) = G(\varphi(x)), \quad x \in D[0,1]^J,\]

or equivalently,

\[(10.9) \quad [G^j(x)](\mu_j^{-1}t) = [G^j(\varphi(x))](t), \quad 0 \leq t \leq 1, \quad 1 \leq j \leq J.\]

Then we show that, also because of the particular nature of $G$,

\[(10.10) \quad A \cdot G(x) = G(Ax), \quad x \in D[0,1]^J,\]

or equivalently,

\[(10.11) \quad \mu_j^{-1}[G^j(x)](t) = [G^j(Ax)](t), \quad 0 \leq t \leq 1, \quad 1 \leq j \leq J.\]

To see that (10.8) and (10.10) taken together imply (10.7), one need only note the following facts. Since $D$ is a diagonal matrix, it is immediate that

\[(10.12) \quad \varphi(Dx) = D\varphi(x), \quad x \in D[0,1]^J.\]

It is a well-known property of Brownian motion that if $a > 0$, then

\[\{\xi(at); 0 \leq t \leq 1\} \overset{d}{=} \{a^{1/2}\xi(t); 0 \leq t \leq 1\}.\]

From this it is immediate that

189
Combining (10.8), (10.10), (10.12) and (10.13) with the fact that
\( \tilde{D} = ADA^{1/2} \), we have
\[
A \varphi(D \xi^J) = A \Gamma(\varphi(D \xi^J)) = G(A \varphi(D \xi^J))
\]
\[
= G(AD \varphi(\xi^J)) = G(ADA^{1/2}(\xi^J)) = G(\tilde{D}(\xi^J)).
\]

So to complete the proof of Theorem 6 it remains only to prove (10.9) and (10.11).

To prove (10.9), let \( x \in D[0,1]^J \), \( t \in [0,1] \), and \( j \in \{1, \ldots, J\} \) be arbitrary but fixed. Substituting the definition of \( G^j(\cdot) \) into (10.9), we need to show that
\[
\max_{\tilde{c} \in \tilde{C}(j)} [\tilde{F}^c(x)](\mu_j^{-1} t) = \max_{\tilde{c} \in \tilde{C}(j)} [\tilde{F}^c(\varphi(x))](t),
\]
which is implied by
\[
(10.14) \quad [\tilde{F}^c(x)](\mu_j^{-1} t) = [\tilde{F}^c(\varphi(x))](t) \text{ for all } \tilde{c} \in \tilde{C}(j).
\]

The key observation is that if station \( i \) is an element of a critical subchain \( \tilde{c} \in \tilde{C}(j) \), then necessarily \( \mu_i = b_i = \mu_j \). From this fact and the definition of \( F^c(\cdot) \) it follows easily that (10.14) holds, thus establishing (10.9).

To prove (10.11), again let \( x, t, \) and \( j \) be arbitrary but fixed. Substituting the definition of \( G^j(\cdot) \) into (10.11), we need to
show that

\[(10.15) \quad \mu_j^{-1} \max_{\tilde{c} \in \tilde{\mathcal{C}}(j)} [\tilde{F}^c(x)](t) = \max_{\tilde{c} \in \tilde{\mathcal{C}}(j)} [\tilde{F}^c(Ax)](t) .\]

It is immediate from the definition of \( F^c(\cdot) \) that

\[(10.16) \quad \mu_j^{-1} \max_{\tilde{c} \in \tilde{\mathcal{C}}(j)} [\tilde{F}^c(x)](t) = \max_{\tilde{c} \in \tilde{\mathcal{C}}(j)} \mu_j^{-1} [\tilde{F}^c(x)](t) = \max_{\tilde{c} \in \tilde{\mathcal{C}}(j)} [\tilde{F}^c(\mu_j^{-1}x)](t) .\]

Again the key observation is that if \( i \) is an element of \( \tilde{c} \in \tilde{\mathcal{C}}(j) \) then \( \mu_i = \mu_j \). It then follows from the definition of \( F^c(\cdot) \) that

\[(10.17) \quad [\tilde{F}^c(Ax)](t) = [\tilde{F}^c(\mu_j^{-1}x)](t) \quad \text{for all } \tilde{c} \in \tilde{\mathcal{C}}(j) .\]

Combining (10.16) and (10.17), we have proved (10.15) and hence (10.11).

Thus the proof of Theorem 6 is complete.

Now to prove Theorem 7 we again note the fundamental relationship

\[ Q^d_j(t) = D^d_j(t) - D^S_j(t), \quad 0 \leq t < \infty, \quad 1 \leq j \leq J .\]

Thus for any \( t \in [0,1] \), \( j \in \{1,\ldots,J\} \) and \( n \geq 1 \),
(10.18) \[ \gamma_n^J(t) = \frac{1}{\sqrt{n}} \left[ Q^J(nt) - (\alpha_j - \alpha_S(j))nt \right] \]

\[ = \frac{1}{\sqrt{n}} \left[ D^J(nt) - \alpha_jnt \right] - \frac{1}{\sqrt{n}} \left[ D^S(j)(nt) - \alpha_S(j)nt \right] \]

\[ = \zeta_n^J(t) - \zeta_n^S(j)(t). \]

Since \( \zeta_n \Rightarrow -q(D_n^J) \) by Theorem 6, it follows easily from (10.10), the definition of \( H \) and the Continuous Mapping Theorem that \( \gamma_n \Rightarrow H(D_n^J) \).

Q.E.D.
CHAPTER 11
Special Cases and Additional Remarks

The initial purpose of this chapter is to show how the general results presented in Chapter 10 simplify when certain relationships of particular interest exist among the mean service times $b_j$. For each special case treated, we shall state only those results which simplify significantly. Also, we shall restrict attention to the discrete-parameter processes $\{T_n\}$ and $\{W_n\}$, since the corresponding results for $D(t)$ and $Q(t)$ are virtually identical.

One special type of assembly-like network which comes immediately to mind is a "totally balanced" system, having

$$b_1 = b_2 = \cdots = b_J.$$ 

Our general results do not simplify in this special case, however, except in the sense that the definitions of $G$, $g$, $H$ and $h$ are easier to state. (This "simplification" of course results from the fact that $C(j) = C(j)$ for all $j = 1, \ldots, J$.) Quite to the contrary, this is the worst of all possible cases from the standpoint of determining explicitly the limiting distributions which appear in Chapter 10. The best of all cases from this standpoint is a network in which the mean service times are all different, and this is the first special structure which we shall investigate.
11.1 Totally Unbalanced Systems

We shall say that an assembly-like network is totally unbalanced if no two of the mean service times $b_1, \ldots, b_J$ are equal.

Definition: For each station $j = 1, \ldots, J$ in a totally unbalanced network, we define $M(j)$, the marker of station $j$, to be the unique index $i \in \{j \setminus A(j)\}$ such that $b_i = \max(a_{ij}, b_j)$.

It is immediate that

$$\mathcal{C}(j) = \{M(j)\} \quad \text{for all } j = 1, \ldots, J.$$

Consequently the important mappings $G$ and $H$ appearing in Chapter 10 simplify as follows for totally unbalanced systems.

(i) $[G(x)](t) = Ax(t), \quad 0 \leq t \leq 1, \quad x \in \mathbb{D}[0,1]^J,$

where $A = (\alpha_{ij})$ is a $J \times J$ matrix with

$$\alpha_{ij} = \begin{cases} 
1 & \text{if } j = M(i) \\
0 & \text{otherwise} 
\end{cases}.$$

Consequently,

$$[G(D_x^J)](t) = AD_x^J(t) = Q^{1/2} \xi_J(t), \quad 0 \leq t \leq 1,$$

where $Q = (q_{ij}^2)$ is a $J \times J$ matrix with

$$q_{ij}^2 = \begin{cases} 
\sigma_{M(j)}^2 & \text{if } M(i) = M(j) \\
0 & \text{otherwise} 
\end{cases}.$$
(ii) \( [H(x)](t) = Bx(t), \ 0 \leq t \leq 1, \ x \in \mathbb{D}[0,1]^J \),

where \( B = (\beta_{ij}) \) is a \((J-1) \times J\) matrix with

\[
\beta_{ij} = \begin{cases} 
1 & \text{if } j = M(S(i)) \prec M(1) \\
-1 & \text{if } j = M(1) \preceq M(S(i)) \\
0 & \text{otherwise}
\end{cases}
\]

Consequently,

\[
[H(D^J_x)](t) = BD^J_x(t), \ 0 \leq t \leq 1.
\]

We define the \( J \times J \) matrix \( \Phi = (s_{ij}^2) \) by

\[
\Phi = (BD)(BD)^T = BD^2B^T,
\]

noting that for \( j = 1, \ldots, J \)

\[
s_{jj}^2 = \begin{cases} 
\sigma^2_{M(j)} + \sigma^2_{M(S(j))} & \text{if } M(j) \preceq M(S(j)) \\
0 & \text{if } M(j) = M(S(j))
\end{cases}
\]

From these facts, the results of Chapter 10, and Theorem 2.10 for integrated Brownian Motion, we obtain the theorem given below. We take the random functions \( \{v_n\} \) and \( \{X_n\} \) and the constant vectors \( \mu = (\mu_j) \) and \( d = (d_j) \) to be as defined in Chapter 10. Note that for totally unbalanced systems
\[ \mu_j = b_M(j), \quad 1 \leq j \leq J, \]
\[ d_j = b_M(S(j)) - \mu_j, \quad 1 \leq j < J. \]

**Theorem 1:** For a totally unbalanced network,

(a) \[ v_n \Rightarrow Q^{1/2} J \quad \text{as} \quad n \to \infty. \]

(b) \[ \frac{1}{\sqrt{n}} (T_n - n\mu) \Rightarrow \mathcal{N}(0, \Sigma) \quad \text{as} \quad n \to \infty. \]

(c) \[ \mathbb{E}[\frac{1}{\sqrt{n}} (T_n^j - n\mu_j)] \to 0 \quad \text{as} \quad n \to \infty, \quad 1 \leq j \leq J. \]

(d) \[ \frac{1}{n} T_n^j \overset{a.s.}{\to} b_M(j) \quad \text{as} \quad n \to \infty, \quad 1 \leq j \leq J. \]

(e) \[ x_n \Rightarrow BD_s^J \quad \text{as} \quad n \to \infty. \]

(f) \[ \frac{1}{\sqrt{n}} (w_n - nd) \Rightarrow \mathcal{N}(0, \Sigma) \quad \text{as} \quad n \to \infty. \]

(g) \[ \mathbb{E}[\frac{1}{\sqrt{n}} (w_n^j - nd_j)] \to 0 \quad \text{as} \quad n \to \infty, \quad 1 \leq j < J. \]

(h) \[ \frac{1}{n} w_n^j \overset{a.s.}{\to} d_j \quad \text{as} \quad n \to \infty, \quad 1 \leq j < J. \]

(i) \[ \frac{1}{\sqrt{n^3}} (W_n - \frac{n^2}{2} d) \Rightarrow \mathcal{N}(0, \Sigma) \quad \text{as} \quad n \to \infty. \]

(j) \[ \mathbb{E}[\frac{1}{\sqrt{n^3}} (W_n^j - \frac{n^2}{2} d_j)] \to 0 \quad \text{as} \quad n \to \infty, \quad 1 \leq j < J. \]

(k) \[ \frac{1}{n} W_n^j \overset{a.s.}{\to} \frac{1}{2} d_j \quad \text{as} \quad n \to \infty, \quad 1 \leq j < J. \]

**11.2 Externally Balanced Systems**

We shall say that an assembly-like network is *externally balanced* if
\[ b_1 = b_2 = \cdots = b_{j_E} = b , \]

and

\[ b_j < b, \quad J_E < j \leq J . \]

To facilitate discussion of this special class of networks, we establish the following definition.

**Definition:** For an externally balanced network let

\[
M(j) = \begin{cases} 
\{j\} & \text{if } 1 \leq j \leq J_E \\
A(j) \cap \{1, \ldots, J_E\} & \text{if } J_E < j \leq J
\end{cases}
\]

It is clear that if a network is externally balanced then

\[
C(j) = \{(i) : i \in M(j)\}, \quad 1 \leq j \leq J ,
\]

Consequently, for this special case the mappings \( G \) and \( H \) appearing in Chapter 10 simplify as follows. For each \( t \in [0,1] \) and \( x \in \mathbb{D}[0,1]^J \)

\[
[G^j(x)](t) = \max_{i \in M(j)} x^i(t), \quad 1 \leq j \leq J ,
\]

\[
[H^j(x)](t) = \max_{i \in M(S(j))} x^i(t) - \max_{i \in M(j)} x^i(t), \quad 1 \leq j < J .
\]

Using these facts to specialize Corollaries 10.2 and 10.4, the following result is obtained.
Theorem 2: Let $X_1, \ldots, X_J$ be independent normal r.v.'s with mean zero and variance $\sigma_1^2, \ldots, \sigma_J^2$ respectively. For an externally balanced assembly-like network,

(a) $\frac{1}{\sqrt{n}} (R_n^j - nb) \xrightarrow{D} \max_{i \in M(j)} X_i$ as $n \to \infty$, $1 \leq j \leq J,$

(b) $E[\frac{1}{\sqrt{n}} (R_n^j - nb)] \to E[\max_{i \in M(j)} X_i]$ as $n \to \infty$, $1 \leq j \leq J,$

(c) $\frac{1}{n} w_n^j \xrightarrow{a.e.} 0$ as $n \to \infty$, $1 \leq j \leq J,$

(d) $\frac{1}{\sqrt{n}} w_n^j \xrightarrow{D} Y_j$ as $n \to \infty$, $1 \leq j < J,$

where

$$Y_j = \left[ \max_{i \in M(S(j))} X_i - \max_{i \in M(j)} X_i \right]$$

(e) $E(\frac{1}{\sqrt{n}} w_n^j) \to E(Y_j)$ as $n \to \infty$, $1 \leq j < J,$

(f) $\frac{1}{n} w_n^j \xrightarrow{a.e.} 0$ as $n \to \infty$, $1 \leq j < J.$

Note that the limiting distribution obtained for $\frac{1}{\sqrt{n}} w_n^j$ has positive probability concentrated at the origin, since

$$M(j) \subset M(S(j)), \quad 1 \leq j < J.$$ The only case in which it is actually degenerate at the origin is that where $M(j) = M(S(j))$, meaning that station $j$ is the only predecessor of station $S(j)$. It is easily verified that externally balanced networks are the only assembly-like networks satisfying
(i) $E(w^j_n) = o(n)$, \( l \leq j < J \), and

(ii) $\frac{1}{\sqrt{n}} w^j_n$ converges weakly to a limiting distribution which has positive probability concentrated at the origin.

Thus, such system configurations are "optimal" in the same sense as are assembly-like operations with $\rho < 1$ and $K^* = K$.

11.3 Additional Remarks

Most of the remarks made in Chapter 7 concerning initial conditions, deterministic streams, and speed of convergence for our single-station results apply equally well to the more general results for assembly-like networks. In particular, we emphasize that all of the results in Chapter 10 continue to hold in the stated form when some or all of the variances $\sigma^2_1, \ldots, \sigma^2_J$ are zero. In this case, however, one must alter in the obvious way the various qualitative remarks which have been made concerning the asymptotic behavior of the waiting time processes $\{w^j_n\}$.

In Chapter 1 it was mentioned that an assembly-like network can be interpreted as (or may represent) a "precedence network" of the type studied by PERT. In a typical PERT analysis, it is assumed that

$$v^j_o \equiv 0, \quad J_E < j \leq J,$$

and the quantity of primary interest is $T^j_1$, the time at which the first service is completed at the terminal station. In most applications, $T^j_1$ represents the completion time for some complex project. An important concept in PERT is that of the "critical path" through the
network. In terms of our notation, the critical path can be defined as follows. For each chain

\[ c = (i_1, \ldots, i_m) \in \mathcal{C}(J) \]

let

\[ b(c) = \sum_{k=1}^{m} b_{i_k} \]

The critical path is then that chain \( c^* \in \mathcal{C}(J) \) such that

\[ b(c^*) = \max_{c \in \mathcal{C}(J)} b(c) \]

(For simplicity, let us assume that the critical path is uniquely determined.) Assuming that

\[ c^* = (i_1, \ldots, i_m) \]

the PERT procedure is to approximate the distribution of \( T_{J_l} \) by that of

\[ \frac{i_1}{v_o} + \frac{i_2}{v_1} + \cdots + \frac{i_m}{v_1} \]

Equivalently stated, the PERT procedure approximates the distribution of \( T_{J_l} \) by simply ignoring the effect of those stations which do not lie on the critical path.

For the case of repeated passage through a precedence network, our results show that the critical path has no relevance. Instead, Corollary 10.2 shows that if \( n \) is relatively large, then a good
approximation for the distribution of \( T_n^J \) can be obtained by ignoring the effect of all stations which do not have maximal mean service time (i.e., by ignoring the effect of stations which do not lie on some critical subchain to station \( J \)). Thus, in the case of repeated passage through the network, the notion of a critical path is supplanted by the notion of a critical sub-network, all stations of which have the same mean service time.
PART III

RELATED TOPICS AND DIRECTIONS FOR FUTURE RESEARCH

CHAPTER 12

Design and Control of Assembly-Like Systems

In the preceding chapters we have introduced models of what might be called "standard" assembly-like systems, and limit theorems have been developed which show that such systems are inherently unstable. More particularly, it has been shown that the vector queue length process \( Q(t) \) associated with a standard assembly-like system tends to blow up as \( t \to \infty \). In this chapter we shall discuss some slightly different but closely related queueing systems in which the basic waiting time and queue length processes do not necessarily blow up (i.e., in which asymptotic system stability is achievable under appropriate conditions). These will be collectively referred to as "modified" assembly-like systems. The major questions which one naturally asks are the following.

(i) What classes of modifications can under appropriate conditions give rise to an asymptotically stable system?

(ii) What are the conditions under which a given modification does give rise to an asymptotically stable system?

(iii) Under a given cost structure and criterion of effectiveness, how can an "optimal" modification be selected from a given set of possible modifications?
Although we have given no precise meaning to the term "modification", it is apparent that the last of these questions effectively subsumes the first two, at least if one accepts the general proposition that under any "reasonable" cost structure a stable system capable of performing a given function is necessarily superior to an unstable system capable of performing the same function. The purpose of this chapter is to indicate the wide variety of "optimization" problems which arise naturally in conjunction with assembly-like systems and to demonstrate some methods which can be used in their analysis. Although some fragmentary results are presented, the chapter is essentially expository.

In the rapidly growing field of optimization of queueing systems, two basic classes of problems have been identified. First is the static decision problem of how a given system can be optimally designed, and second is the dynamic decision problem of how a given system can be optimally controlled. To investigate problems of optimal design selection, a researcher typically isolates a few system parameters which serve as decision variables (e.g., number of servers, mean service rate, size of finite waiting room, etc.) and hypothesizes a reasonable structure of costs and rewards associated with operating the system (including for example, a cost of customer waiting, a reward for completed services, a cost of operating the service facility, etc.). The problem then is to find that combination of values for the decision variables which maximizes some total measure of system performance, such as expected average net reward per unit time over an infinite planning horizon.

To investigate problems of dynamic control of queueing systems, a
much more general (and usually more realistic) decision framework is employed. In this case the researcher hypothesizes that a decision maker has the capability to dynamically alter the values of certain decision variables in response to changes in the state of the system. Given a structure of costs and rewards, the problem then is to select an optimal control policy (i.e., a rule which sets the values of the decision variables in each possible state of the system). As one might expect, problems of optimal control have generally proved to be quite difficult, and much of the early research in this area avoided the basic question of optimal policy selection, focusing instead on how one can select the best of all policies having a certain very simple (stationary) form. As we shall demonstrate by example below, simple problems of the latter type are usually very similar to problems of optimal design selection.

We shall not attempt to give a comprehensive discussion of all the practically significant problems which arise concerning the optimal design and control of assembly-like systems. Our intent rather is to exhibit some simple but representative problems which might prove fruitful topics for future research and lead naturally to more general and realistic decision models. For ease of discussion, attention is restricted to the simplest type of assembly-like configuration, a single-station system with two input processes, portrayed schematically in Figure 10 below. The reader will note, however, that many of the
Figure 10. Schematic representation of a single-station system with two input processes.

remarks to follow extend readily to more general types of systems.

The inherent instability which we have demonstrated previously for the system portrayed in Figure 10 is essentially traceable to two basic assumptions of our standard model.

(a) The input processes are completely independent of one another, and each involves some stochastic variability.

(b) No restriction is placed upon the number of input items of either type which may accumulate over time. Equivalently stated, both "waiting rooms" are of infinite capacity.

The standard model contains a number of other specific assumptions, of course, but they play no role in determining the stability of the system. In particular, it is easily verified that the basic queue length and waiting time processes blow up regardless of the service
process and queue discipline. */ Thus in formulating optimization
problems for assembly-like systems, it is most natural to examine
modifications of the standard model under which conditions (a) and (b)
do not both hold, i.e., modifications in the input processes. Some
obvious possibilities are the following.

(1) Let both of the waiting rooms be of finite capacity.
(2) Let just one of the waiting rooms be of finite capacity.
(3) Let input process 1 be turned off whenever \( n_1 - n_2 \geq c_1 \) and
    let input process 2 be turned off whenever \( n_2 - n_1 \geq c_2 \), where
    \( c_1 \) and \( c_2 \) are non-negative integers not both equal to zero,
    and \( n_1 \) and \( n_2 \) denote the number of type 1 items and number
    of type 2 items present in the system respectively.
(4) Let one (but not both) of the barriers \( c_k \) in (3) be infinite.

In the case of both waiting rooms finite, it is certainly clear
that the system is asymptotically stable, regardless of the inter-
arrival time and service time distributions. To demonstrate how
one might determine optimal waiting room sizes, we hypothesize a cost
structure of the following type. Suppose that a reward of \( r > 0 \) is
received for each completed service and a cost \( h_k > 0 \) is incurred
*/ By "service process" we mean the joint distribution of the service
times \( \{v_n\} \), and by "queue discipline" we mean the order in which
input items of the various types are used. We continue to assume
throughout that exactly one input item of each type is required for
each service, i.e., that the assembly proportions are fixed.
for every unit of time that a type \( k \) item spends waiting or in service. Let \( s_1 \) and \( s_2 \) be generic notations for the size of the type 1 waiting room and the size of the type 2 waiting room respectively, meaning that the number of type \( k \) items waiting and in service may not exceed \( s_k > 0 \) \((k = 1, 2)\). Let

\[
R_k(s_1, s_2) = \lim_{t \to \infty} E[n_k(t)] ,
\]

\[
R_k(s_1, s_2) = \lim_{t \to \infty} E[n_k(t)] ,
\]

\[
p(s_1, s_2) = \lim_{t \to \infty} P\{\min[n_1(t), n_2(t)] > 0\} .
\]

Letting \( b > 0 \) denote the mean service time, the long run expected net reward per unit time with waiting rooms of size \( s_1 \) and \( s_2 \) is

\[
C(s_1, s_2) = r[b^{-1}p(s_1, s_2)] + \sum_{k=1}^{2} h_k R_k(s_1, s_2) .
\]

We assume the interarrival time and service time distributions to be given and fixed, and we seek the optimal waiting room sizes \( \hat{s}_1 \) and \( \hat{s}_2 \) such that

\[
C(\hat{s}_1, \hat{s}_2) = \max_{0 \leq s_1, s_2 < \infty} [C(s_1, s_2)] .
\]

In order to determine these optimal values we must determine the functional dependence of \( R_1, R_2 \) and \( p \) on \( s_1 \) and \( s_2 \), for which it is sufficient of course to know the limiting (joint) queue length distribution as a function of \( s_1 \) and \( s_2 \). Thus the optimization problem reduces to a descriptive problem of the type
treated by classical queueing theoretic methods. This situation is fairly typical of optimal design problems encountered in queueing theory. If the descriptive theory for the system under study is sufficiently well developed (with the given interarrival time and service time distribution) then the optimization problem is trivial, at least conceptually. We emphasize, however, that unless very strong distributional assumptions are made, the required descriptive results are likely to be very difficult to obtain.

It is reasonably clear that a modified system of type (2), (3) or (4) will be asymptotically stable if the proper relationship exists among the interarrival and service time distributions and the other system parameters ($s_1$ or $s_2$, $c_1$ and/or $c_2$). The problem of determining this "proper relationship" is also of a descriptive nature, and one can expect it to be quite difficult unless strong distributional assumptions are made. Assuming that the stability condition can be found, the problem of selecting optimal design parameters for a modified system of type (2), (3) or (4) is very similar to that described earlier for a modified system of type (1), descriptive results providing the key for analysis. It should be pointed out that the problem of determining optimal parameter values $\hat{c}_1$ and/or $\hat{c}_2$ for a system of type (3) or (4) is not really a "design" problem in the usual physical sense. Rather it is a problem of selecting an optimal stationary control policy from a class of such policies having a stated form. From an analytical viewpoint, however, this distinction is unimportant.

We have emphasized the importance of distributional assumptions in obtaining descriptive results for modified systems. In Section 1
below some fragmentary descriptive results are presented for various
types of systems under the assumption that both input processes are
Poisson and service times have a negative exponential distribution.
This represents the easiest case for analysis, since the process
$\{[n_1(t), n_2(t)], t \geq 0\}$ is then a continuous time parameter Markov
chain. Even with these very strong assumptions, we have not been able
to find the equilibrium queue length distributions for arbitrary values
of the parameters $s_1$, $s_2$, $c_1$ and $c_2$. It seems likely that such
general results can be obtained with sufficient effort, but the mere
fact that the totally Markovian cases are non-trivial indicates that
descriptive results may be very difficult to obtain when less convenient
distributional assumptions are made.

The reader may justifiably feel that in seeking an optimal
modification of some given type (or even of any one of several types)
we are solving a problem which is too highly structured. For example,
rather than seeking an optimal modification of some type discussed
above, it is probably more realistic to consider the following (still
very simple) control problem. Let the cost structure be as before,
and suppose that a decision-maker has the capability to turn each input
process on or off at each point in time, there being no penalty for
changing the mode of control. This problem will be discussed in Section
2 below, also under the assumption of Poisson inputs and exponentially
distributed service times. Again these assumptions represent the
easiest possible case, since the control process is then a continuous
time Markov decision process. The treatment given in Section 2 is by
no means complete, and again the fact that the totally Markovian case
is non-trivial gives us reason to believe that more complex and realistic control problems may be very difficult to solve.

12.1 Some Equilibrium Distributions for Modified Markovian Systems

In this section we shall investigate the asymptotic behavior of the assembly-like system portrayed in Figure 10 when various simple type of modifications are made in the standard assumptions. For each case considered, we define the state of the system at time $t \geq 0$ to be

$$s(t) = [n_1(t), n_2(t)],$$

where as before

$$n_k(t) = \text{the number of type } k \text{ items present in the system at time } t \quad (k = 1, 2).$$

It is assumed hereafter that input process $k$ is Poisson with mean interarrival time $\lambda_k^{-1} > 0$ ($k = 1, 2$) and that service times have a negative exponential distribution with mean $\mu^{-1} > 0$. Under this assumption the process $\{s(t); t \geq 0\}$ is a conservative continuous time Markov chain* with state space $N^2$, where

$$N = \{0, 1, 2, \ldots\}, \quad N^2 = N \times N.$$

* Rather than attempting to justify this statement, we refer the reader to Syski (1965) for an excellent discussion of the theory of continuous time Markov chains and its applications in queueing.
For each case that we shall consider, however, one can easily identify an irreducible class of states \( S \) such that all states \( s \in \mathbb{N}^2 - S \) are inaccessible from \( S \) and obviously transient. To simplify discussion, we shall assume that the initial state of the system is contained in \( S \), so the process \( \{s(t); t \geq 0\} \) can be viewed as an irreducible Markov chain with minimal state space \( S \).

For each case considered we shall denote the transition intensity matrix for the chain \( \{s(t); t \geq 0\} \) by \( Q = (q_{s,s'}) \). That is, letting

\[
P(s',t|s) = \text{Prob}\{s(t) = s'|s(0) = s\},
\]

we define

\[
q_{s,s} = \lim_{t \downarrow 0} \frac{1 - P(s,t|s)}{t} \quad \text{for all } s \in S,
\]

\[
q_{s,s'} = \lim_{t \downarrow 0} \frac{P(s',t|s)}{t} \quad \text{for all } s,s' \in S \text{ with } s \neq s'.
\]

It is well known, cf. Miller (1963), that for chains of the type discussed here the system of linear difference equations

\[
(12.1) \quad \sum_{s' \in S} y_{s'} q_{s,s'} = 0, \quad s \in S,
\]

has a non-negative solution \( \{y_s; s \in S\} \) which is unique up to a scale factor. If

\[
\sum_{s \in S} y_s = \infty,
\]
then the chain is transient or null recurrent. Otherwise, the chain is positive recurrent and the limiting state probability distribution

\[ P(s') = \lim_{t \to \infty} P(s', t | s), \quad s' \in S, \]

is given by

\[ P(s') = \left[ \sum_{s \in S} y_s \right]^{-1} y_{s'}, \quad s' \in S. \]

If both waiting rooms are finite, then of course the state space is finite and the chain is positive recurrent regardless of the parameter values \( \lambda_1, \lambda_2 \) and \( \mu \). To find the equilibrium distribution for given parameter values and waiting room sizes, one need only solve the finite system of linear difference equations (12.1), but we have been unable to determine the form of this distribution for general waiting room sizes.

Consider now the case in which the waiting room for type 2 items is of finite capacity \( c > 0 \) and the waiting room for type 1 items is of infinite capacity. In this case the minimal state space is

\[ S = \{(i, j) \in \mathbb{N}^2 : 0 \leq i \leq c\}, \]

and the transition intensities are easily seen to be as follows. If \( s = (n_1, n_2) \) and \( s' = (n_1', n_2') \) are different states in \( S \), then
\[ q_{s,s'} = \begin{cases} 
\lambda_1 & \text{if } n_2' = n_2 \text{ and } n_1' = n_1 + 1 \\
\lambda_2 & \text{if } n_1' = n_1 \text{ and } n_2' = n_2 + 1 \\
\mu & \text{if } n_1' = n_1 - 1 \text{ and } n_2' = n_2 - 1 \\
0 & \text{otherwise} \end{cases} \]

Correspondingly, if \( s = (n_1, n_2) \in S \) then
\[ q_{s,s} = \begin{cases} 
-(\lambda_1 + \lambda_2 + \mu) & \text{if } 0 < n_1 \text{ and } 0 < n_2 < c \\
-(\lambda_1 + \mu) & \text{if } 0 < n_1 \text{ and } 0 < n_2 = c \\
-(\lambda_1 + \lambda_2) & \text{if } 0 = n_1 \text{ and } 0 \leq n_2 < c \\
& \text{or } 0 \leq n_1 \text{ and } 0 = n_2 \\
-\lambda_1 & \text{if } 0 = n_2 \text{ and } n_1 = c 
\end{cases} \]

The state space and transition structure for this chain are portrayed graphically in Figure 11 below for the case \( c = 2 \).

![State space and transition structure](Figure 11.png)

The state space and transition structure for the case \( c = 2 \).

Figure 11. State space and transition structure for the case \( c = 2 \).
In order to determine the conditions under which this chain is positive recurrent, it is useful to observe that it is equivalent to another continuous time Markov chain arising in queueing theory. Suppose that two single-server queueing facilities are arranged in series with a finite waiting room of capacity \( c > 0 \) at the second station, meaning that the number of customers waiting and in service at the second station may not exceed \( c \). Suppose that input to the first station is Poisson with mean interarrival time \( \lambda_1^{-1} \), that service times at the first station have a negative exponential distribution with mean \( \mu_1^{-1} \), that service times at the second station have a negative exponential distribution with mean \( \lambda_2^{-1} \), and that the server at the first station cannot begin a service unless (strictly) fewer than \( c \) customers are present at the second station. If we define the state of this system at time \( t \) to be

\[
s'(t) = [n_1(t), n_2(t)]
\]

where

\[
n_1(t) = \text{the number of customers present at the first station at time } t,
\]

\[
n_2(t) = c - \text{the number of customers present at the second station at time } t,
\]

then \( \{s'(t); t \geq 0\} \) is a continuous time Markov chain with exactly the same state space and transition structure as our process \( \{s(t); t \geq 0\} \). Hunt (1956) has studied the process \( \{s'(t); t \geq 0\} \)
in a slightly different but equivalent form, and due to the equivalence just mentioned, we may invoke his results directly to make the following statements concerning our modified assembly-like system.

(i) Assuming $\mu = \lambda_2$, the chain \{s(t); t \geq 0\} is positive recurrent if and only if

\[
\frac{(c+1)\lambda_1}{c\lambda_2} < 1.
\]

(ii) Assuming $\mu \neq \lambda_2$, the chain \{s(t); t \geq 1\} is positive recurrent if and only if

\[
\frac{\rho_1}{\rho_2} \left[ \frac{1 - \rho_2^{c+1}}{1 - \rho_2^c} \right] < 1
\]

where $\rho_1 = \lambda_1/\mu$ and $\rho_2 = \lambda_2/\mu$.

(iii) If $c = 1$ and the chain is positive recurrent (i.e.,

$\lambda_1^{-1} < \lambda_2^{-1} + \mu^{-1}$) then the equilibrium distribution is given by

\[
P(n,0) = K(w_1^{n+1} - w_2^{n+1}), \quad n \geq 0,
\]

\[
P(n,1) = K[(\rho_1 + \rho_2 - \rho_1/w_1)w_1^n - (\rho_1 + \rho_2 - \rho_2/w_2)w_2^n], \quad n \geq 0,
\]

where

\[
\rho_1 = \lambda_1/\mu, \quad \rho_2 = \lambda_2/\mu
\]

\[
w_1 = \frac{1}{2} (\rho_1/\rho_2) \left[ (1+p_1+p_2) + [(1+p_1+p_2)^2 - 4\rho_2]^{1/2} \right],
\]

\[
w_2 = \frac{1}{2} (\rho_1/\rho_2) \left[ (1+p_1+p_2) - [(1+p_1+p_2)^2 - 4\rho_2]^{1/2} \right],
\]

\[K = (\rho_1/\rho_2)[1-(\rho_1/\rho_2)(1+p_2)]/(w_1-w_2).
\]
The equilibrium distribution is not known for larger values of \( c \).

Let us now consider a modification of the third type discussed above. Letting the input and service processes be as before, we assume that input process 1 is turned off whenever \( n_1 - n_2 \geq c_1 \) and input process 2 is turned off whenever \( n_2 - n_1 \geq c_2 \), where \( c_1 \) and \( c_2 \) are non-negative integers not both equal to zero. The minimal state space in this case is

\[
S = \{(n_1, n_2) \in \mathbb{N}^2 : c_1 \geq n_1 - n_2 \geq -c_2 \}
\]

and the infinitesimal generator \( Q \) is as follows. If \( s = (n_1, n_2) \) and \( s' = (n'_1, n'_2) \) are different states in \( S \), then

\[
q_{s, s'} = \begin{cases} 
\lambda_1 & \text{if } n'_2 = n_2 \text{ and } n'_1 = n_1 + 1 \\
\lambda_2 & \text{if } n'_1 = n_1 \text{ and } n'_2 = n_2 + 1 \\
\mu & \text{if } n'_1 = n_1 - 1 \text{ and } n'_2 = n_2 - 1 \\
0 & \text{otherwise} 
\end{cases}
\]

If \( s = (n_1, n_2) \in S \) then

\[
q_{s, s} = \begin{cases} 
-(\lambda_1 + \lambda_2 + \mu) & \text{if } c_1 < n_1 - n_2 < -c_2, \min(n_1, n_2) > 0 \\
-(\lambda_1 + \lambda_2) & \text{if } c_1 < n_1 - n_2 < -c_2, \min(n_1, n_2) = 0 \\
-(\lambda_2 + \mu) & \text{if } c_1 = n_1 - n_2 < -c_2, \min(n_1, n_2) > 0 \\
-\lambda_2 & \text{if } c_1 = n_1 - n_2 < -c_2, \min(n_1, n_2) = 0 \\
-(\lambda_1 + \mu) & \text{if } c_1 < n_1 - n_2 = -c_2, \min(n_1, n_2) > 0 \\
-\lambda_1 & \text{if } c_1 < n_1 - n_2 = -c_2, \min(n_1, n_2) = 0 
\end{cases}
\]
The state space and transition structure for this chain are shown graphically in Figure 12 below for the case $c_1 = 2, c_2 = 1$.

![Diagram of state space and transition structure]

**Figure 12.** State space and transition structure for the case $c_1 = 2, c_2 = 1$.

From the basic equations of state (12.1) it is easily deduced that the limiting marginal distribution of $n_1(t) - n_2(t)$ for chains of this class is given by
\[
\lim_{t \to \infty} \mathbb{P}\{n_1(t) - n_2(t) = c_1 - j\}
\]
\[
= \left\{ \begin{array}{ll}
\left( \frac{1-r}{c_1 + c_2 + 1} \right)^r, & 0 \leq j \leq c_1 + c_2, \text{ if } r \neq 1,
\frac{1}{c_1 + c_2 + 1}, & 0 \leq j \leq c_1 + c_2, \text{ if } r = 1,
\end{array} \right.
\]
where \( r = \frac{\lambda_2}{\lambda_1} \). From this it follows easily that the long-run average arrival rate for input items of both types is
\[
\lambda = \left\{ \begin{array}{ll}
\lambda_2 \left[ \frac{1-r}{c_1 + c_2 + 1} \right], & \text{items/unit time if } r \neq 1,
\frac{1}{c_1 + c_2 + 1}, & \text{items/unit time if } r = 1.
\end{array} \right.
\]
It is then intuitively clear (although we shall not attempt to justify the statement rigorously) that the chain is positive recurrent if and only if \( \mu > \hat{\lambda} \).

If one of the barriers \( c_k \) is set at infinity, then the limiting marginal distribution of \( n_1(t) - n_2(t) \) is even easier to calculate. Suppose, for example, that \( c_2 = \infty \). We find then that the equilibrium distribution of \( n_1 - n_2 \) is non-existent if \( r \geq 1 \) (implying immediately that the entire chain is either null recurrent or transient), and that
\[
\lim_{t \to \infty} P\{n_1(t) - n_2(t) = c_1 - j\} = (1-r)^{-1}r^j, \quad j \geq 0, \text{ if } r < 1.
\]

The long-run average arrival rate for items of each type is then \(\lambda_2\), and the apparent condition required for positive recurrence of the entire chain is \(\mu > \lambda_2 < \lambda_1\). We conjecture, incidentally, that this is the general stability condition under any distributional assumptions.

The only modified system of type (3) or (4) for which we have found the limiting state probability distribution is that with \(c_1 = c_2 = 1\). For this case, we find that if the chain is positive recurrent (i.e., \(\mu > \hat{\lambda}_2\)) then

\[
P(0,0) = \frac{K}{(\rho_1 + \rho_2)},
\]

\[
P(0,1) = \rho_2 K/\rho_1 (\rho_1 + \rho_2) + K b_1/\rho_1,
\]

\[
P(1,0) = \rho_2 K/\rho_2 (\rho_1 + \rho_2) + K b_2/\rho_2,
\]

\[
P(n,n) = K w^{n-1}, \quad n \geq 1,
\]

\[
P(n,n+1) = K b_1 w^{n-1}, \quad n \geq 1,
\]

\[
P(n+1,n) = K b_2 w^{n-1}, \quad n \geq 1,
\]

where \(\rho_1 = \lambda_1/\mu\), \(\rho_2 = \lambda_2/\mu\) and

\[
b_1 = \rho_2 / (1 + \rho_1 - \mu), \quad b_2 = \rho_1 / (1 + \rho_2 - \mu),
\]

219
\[ K = (1-w)\gamma/[(\rho_1 + \rho_2) + (1-w)] , \]
\[ \gamma = (\rho_1 + \rho_2)/[1+\rho_1/\rho_2 + \rho_1/\rho_2] , \]
and \( w \) is the unique root in \((0,1)\) of
\[ w^3 - 2(1+\rho_1 + \rho_2)w^2 + [1+2(\rho_1 + \rho_2) + \rho_1 + \rho_2 + (\rho_1 + \rho_2)^2]w = \rho_1^2(2+\rho_1 + \rho_2). \]

12.2 Dynamic Control of Markovian Systems

In this section we again consider the assembly-like system portrayed in Figure 10, assuming input process \( k \) to be Poisson with mean interarrival time \( \lambda_k^{-1} \) \((k = 1,2)\) and service times to have a negative exponential distribution with mean \( \mu^{-1} \). We suppose now that there exists a decision maker who has the capability to control the probabilistic evolution of the system by turning the input processes on and off. Specifically, the decision maker must select one of four possible modes of control at each point in time. Mode \((1,1)\) corresponds to both input processes being turned on, mode \((1,0)\) corresponds to input process 1 turned on and input process 2 turned off, mode \((0,1)\) corresponds to input process 1 turned off and input process 2 turned on, and mode \((0,0)\) corresponds to both input processes turned off. The decision maker receives a reward of \( r > 0 \) for each completed service and incurs a cost of \( c_k > 0 \) for each unit of time that a type \( k \) item spends waiting or in service. We assume that all costs and rewards are discounted continuously using some interest rate \( \gamma > 0 \), so that a reward of \( x \) received at time \( t > 0 \) is equivalent in value to a
reward of $xe^{-\gamma t}$ received at time zero. The decision maker's problem
is to select a policy for dynamic control which maximizes the expected
total discounted reward earned by the system over an infinite planning
horizon.

Because of the "memoryless" property of the Poisson process,
this problem may be equivalently viewed as follows. We may suppose that
both input processes generate input items continuously through time,
but the decision maker has the capability to accept or reject any
arriving item. If an arriving item is rejected, then it departs the
system immediately and no cost or reward results, but if an item is
accepted, then it can depart the system only when its service is com-
pleted. Viewing the control problem in this way, we can immediately
derive a rather weak but still quite useful characterization of the
decision maker's optimal policy. Let us define the state of the system
at time $t$ to be

$$s(t) = [n_1(t), n_2(t)]$$
as before.

**Proposition**: There exist positive integers $n_1^*$ and $n_2^*$ such that the
decision maker, if he is following an optimal policy, will never accept
an arriving type $k$ item if $n_k(t) \leq n_k^*$ ($k = 1, 2$).

The following argument demonstrates the truth of this statement and
shows how upper bounds can be calculated for $n_1^*$ and $n_2^*$. Suppose
that a type 1 item arrives at time $t$ and that $n_1(t^-) = n$. For ease
of discussion we assume hereafter that \( t = 0 \). We shall assume that the sequence of future decisions as to whether or not arriving items will be accepted is fixed. Our objective then is to derive an upper bound for the net marginal benefit associated with acceptance of the type 1 item which arrives at time zero. Let \( v_1, v_2, \ldots \) denote the durations of future service times. Because of the memoryless property of the exponential distribution, we need not worry about whether a service is initially underway and can simply assert that \( v_1, v_2, \ldots \) are i.i.d. random variables. Let \( T(n) \) denote the time at which the arriving type 1 item, if accepted, will complete service, given the future policy to be followed and the fact that \( n_1(0-) = n \). Although the distribution of \( T(n) \) is very much dependent upon the future policy and the specific value of \( n_2(0) \), we know that

\[
(12.2) \quad T(n) \geq \sum_{i=1}^{n+1} v_i.
\]

If the type 1 item is accepted, then a (discounted) holding cost of

\[
(12.3) \quad h(n) = \int_{0}^{T(n)} e^{-\gamma t} c_n dt
\]

will be incurred, which can be avoided by rejecting the item. On the other hand, two possible benefits accrue from accepting the item. First is that a reward of \( r \) may be received at time \( T(n) \), which might otherwise have been foregone or at least delayed. Second is that the value of \( n_2(t) \) may be decreased by as much as one for all
t \geq T(n), thus saving some inventory holding costs. Thus the (discounted) total marginal benefit from accepting the item cannot exceed

\begin{equation}
(12.4) \quad g(n) = e^{-\gamma T(n)} r + \int_{T(n)}^{\infty} c_2 e^{-\gamma t} dt = (r + c_2/\gamma)e^{-\gamma T(n)} .
\end{equation}

Combining (12.2)-(12.4), the expected net (discounted) marginal benefit from accepting the item cannot exceed

\[ E[g(n) - h(n)] \leq \left( \frac{\mu}{\mu + \gamma} \right)^{n+1} (r + c_2/\gamma) - (c_1/\gamma)[1 - \left( \frac{\mu}{\mu + \gamma} \right)^{n+1}] , \]

regardless of what future actions may be taken. Thus letting

\[ \hat{n}_1 = \min\{ n : \left( \frac{\mu}{\mu + \gamma} \right)^{n+1} (r + c_2/\gamma) - (c_1/\gamma)[1 - \left( \frac{\mu}{\mu + \gamma} \right)^{n+1}] \leq 0 \} = \min\{ n \geq \log\left( \frac{c_1/\gamma}{c_1/\gamma + c_2/\gamma + r} \right) / \log(\frac{\mu}{\mu + \gamma}) - 1 \} , \]

it is clear that

\[ n^*_1 \leq \hat{n}_1 . \]

An upper bound \( \hat{n}_2 \) for \( n^*_2 \) can be calculated in a symmetric fashion.

Under the assumptions enumerated above, the decision maker's problem can be formulated as a continuous time Markov decision process, cf. Howard (1960), Chapter 8. Assuming for convenience that the
system is initially in some state \((n_1, n_2)\) with \(n_1 \leq \hat{n}_1\) and \(n_2 \leq \hat{n}_2\), the discussion above shows that we may take the state space for the decision process to be

\[
S = \{ (n_1, n_2) \in \mathbb{N}^2 : 0 \leq n_1 \leq \hat{n}_1, 0 \leq n_2 \leq \hat{n}_2 \}.
\]

The set of "actions" available to the decision maker when the system is in state \(s = (n_1, n_2) \in S\) is

\[
A_s = \begin{cases} 
(1, 1), (0, 1), (1, 0), (0, 0) & \text{if } n_1 < \hat{n}_1, n_2 < \hat{n}_2, \\
(0, 1), (0, 0) & \text{if } n_1 = \hat{n}_1, n_2 < \hat{n}_2, \\
(1, 0), (0, 0) & \text{if } n_1 < \hat{n}_1, n_2 = \hat{n}_2, \\
(0, 0) & \text{if } n_1 = \hat{n}_1, n_2 = \hat{n}_2.
\end{cases}
\]

In Howard's general formulation the system is allowed to earn rewards at a rate which is dependent both upon the current state of the system and the action (or control mode) currently employed, but in our case the earning rate is independent of the action. In particular, the effective rate \(r(s)\) at which the system earns rewards when in state \(s = (n_1, n_2)\) is given by

\[
r(s) = r(n_1, n_2) = \begin{cases} 
\mu r - n_1 c_1 - n_2 c_2 & \text{if } \min(n_1, n_2) > 0, \\
-n_1 c_1 - n_2 c_2 & \text{if } \min(n_1, n_2) = 0.
\end{cases}
\]

Here we have converted the reward \(r\) associated with transition from
state \((n_1, n_2)\) to state \((n_1-1, n_2-1)\) into an equivalent "earning rate" \(\mu\) associated with occupancy of state \((n_1, n_2)\), cf. Howard (1960), p. 100. The "transition rate" \(q(s' / s, a)\) to state \(s'\), given that the system is currently in state \(s\) and action \(a \in A_s\) is being employed, is as follows. If \(s = (n_1, n_2)\), and \(s' = (n'_1, n'_2)\) are different elements of \(S\) and \(a = (I_1, I_2) \in A_s\), then

\[
q(s' / s, a) = \begin{cases} 
\lambda_1 & \text{if } (n'_1, n'_2) = (n_1+1, n_2), \ I_1 = 1 . \\
\lambda_2 & \text{if } (n'_1, n'_2) = (n_1, n_2+1), \ I_2 = 1 . \\
\mu & \text{if } (n'_1, n'_2) = (n_1-1, n_2-1) \\
0 & \text{otherwise} 
\end{cases}
\]

Correspondingly,

\[
q(s / s, a) = - \sum_{s' \in S} q(s' / s, a), \quad s \in S, \quad a \in A_s
\]

Since we have shown the decision process to have an effectively finite state space, it is immediate, cf. Miller (1967), p. 51, that there exists a stationary optimal policy. Moreover, the policy improvement algorithm of Howard (1960) can be used to compute such an optimal policy for given values of the problem parameters. It would be very useful, however, to show that there exists an optimal policy of some particularly simple form, depending perhaps on just a few critical numbers. It might be, for example, that there must exist a stationary optimal policy which is equivalent to the imposition of two finite waiting rooms. We have attempted without great success
to develop such characterizations. In the remainder of this section the fragmentary results which have been obtained will be stated but not proved. In each case the proof follows easily from the familiar optimality criterion around which the policy improvement algorithm is built. We shall be brief in these remarks, assuming that the reader is familiar both with the various properties of the exponential distribution which are used and with Howard's algorithm.

Let us define

\[ R_1 = \{(n_1, n_2) \in S : n_1 \geq n_2 \} \]

\[ R_2 = \{(n_1, n_2) \in S : n_2 \geq n_1 \} \]

**Definition:** We shall say that a control policy is of type \( A \) if it rejects arriving type 1 items whenever the system is in a state \( s \in R_1 \) and rejects arriving type 2 items whenever the system is in a state \( s \in R_2 \).

The transition structure for the system under a typical stationary policy of type \( A \) is shown graphically in Figure 13 below for the case \( \hat{n}_1 = 4, \hat{n}_2 = 3 \). If the system begins in a "diagonal" state \((n,n)\) and a stationary type \( A \) policy is followed, then neither input process is ever turned on, and absorption at \((0,0)\) eventually occurs. If the initial state is in \( R_1 \) and a stationary policy of type \( A \) is followed, then the process is either absorbed eventually in some "border state" \((n,0)\) with \( n > 0 \) or else evolves to a diagonal state \((n,n)\) and is eventually absorbed at \((0,0)\). A
Figure 13. Transition structure under a typical stationary policy of type A for the case \( \hat{n}_1 = 4, \hat{n}_2 = 3 \).

The corresponding statement holds if the initial state is in \( R_2 \). A type A policy represents a scheme for simply dissipating whatever inventories may be on hand initially, and it breaks into two independent parts, one being a plan to use if the initial state is in \( R_1 \) and the other being a plan to use if the initial state is in \( R_2 \). The best stationary policy of type A can be calculated by an ordinary dynamic programming recursion, since each state can be visited at most once. The following proposition is then easily proved.
Proposition 1: There exists an optimal policy which is of type A if and only if

\[(\mu r - c_1 - c_2)/(\mu + \gamma) \leq \min\{c_1/\lambda_1, c_2/\lambda_1\}\]

If this condition is met, then there is an optimal stationary type A policy, and the expected discounted reward starting from any initial state and following any policy is non-positive.

An interpretation of the condition above is readily available. Suppose that the system is initially in state \((0,0)\), that process 1 is turned on until a single item is produced, that process 2 is then turned on until a single item is produced, and then that both input processes are turned off until a service is completed. Alternately, one may suppose that the input processes are turned on in the reverse order. The condition in Proposition 1 is met if and only if the expected discounted net reward from each of these "circuits" is non-positive. One might characterize this situation by saying that the decision process is simply unprofitable.

For purposes of computing the best type A policy, the following observation is useful. Suppose that the initial state is in \(R_1\). Then the optimal policy for those states which lie in \(R_1\) is unchanged if the original reward function is replaced by a new one, \(r'(s) = r'(n_1, n_2)\), given by

228
\[ r'(n_1, n_2) = \begin{cases} 
0 & \text{if } n_2 = 0 \\
\frac{\mu(r+c_1)}{\gamma} - n_2c_2 & \text{if } n_2 > 0 
\end{cases} \]

A symmetric statement holds for computation of the optimal (type A) policy to follow in \( R_2 \).

We are unable to give any characterization of the optimal policy when it is not of type A. On the basis of computational experience, however, we conjecture that there does not in general exist an optimal policy which depends on only a few critical numbers.
CHAPTER 13
Other Directions for Future Research

Certainly the outstanding unsolved problems raised in Parts I and II are the distributions of the multi-dimensional diffusion processes obtained as weak limits in Chapters 4, 6, 8 and 10. Some of these, such as the general processes $G(\xi^J)$ and $H(\xi^J)$ appearing in Chapter 10, are of so complex a form that one cannot reasonably expect to find complete solutions. It does seem likely, however, that at least some of the important marginal distributions can be found explicitly.

As was mentioned in Section 7.5, much more work is needed on rates of convergence and numerical comparisons in order to determine conditions under which limit theorems of the type developed here provide accurate approximations to the behavior of real queueing systems.

In Chapter 8 an "approximate representation" was developed for the service completion time $T_n^k$ in a series queueing system. If the initial conditions assumed in Chapter 8 are altered slightly, then a reasonably compact exact representation for $T_n^k$ can be developed, from which an exact representation for the customer waiting time $w_n^k$ is immediately obtainable. Specifically, suppose that $\{v_n^1; n = 0,1,\ldots\}, \ldots, \{v_n^K; n = 0,1,\ldots\}$ are independent i.i.d. sequences as before, but let us change the interpretation of $v_0^1, \ldots, v_0^K$. We imagine now that there is a customer (numbered zero) initially present at station 2 and that his service time at station $k$ is $v_o^k (2 \leq k \leq K)$. Thus, from the
viewpoint of subsequently arriving customers, server \( k \) is initially "paralyzed" (i.e., occupied) for a period of duration \( \frac{v_2^k}{v_0^2} + \cdots + \frac{v_k^k}{v_0^k} \) \( (2 \leq k \leq K) \). The reader may verify inductively that under these initial conditions

\[
T_n^k = \max_{j_o \leq \cdots \leq j_k = n} \left[ \sum_{i=1}^{k} (v_{j_i}^i - v_{j_i}^{i-1}) + \sum_{i=1}^{k} v_{j_i}^i \right]
\]

for all \( n \geq 1 \) and \( k = 2, \ldots, K \). The representation for \( w_n^k \) is then obtained from the basic relationship

\[
w_n^k = T_n^k - T_n^{k-1} - v_n^k, \quad 2 \leq k \leq K, \quad n \geq 1.
\]

By replacing \( V_j^k \) with \( v_n^k - V_{n-j}^k \) in these representations, one can show that \( w_n = (w_n^1, \ldots, w_n^K) \) has the same distribution as another more easily studied random vector \( \tilde{w}_n \). This proposition, which constitutes a generalization of Lemma 2.27 for GI/G/1 queues, is very useful in the analysis of series queueing systems, as the author intends to show in a subsequent paper.

One of the very appealing features of limit theorems which have been obtained for various types of queueing systems in heavy traffic is their robustness under changes in either the distributional assumptions or the physical characteristics of the model being studied. Iglehart and Whitt (1969 a, c), for example, have shown that the limit theorems developed by Whitt (1968) for GI/G/1 queues in heavy traffic continue to hold in essentially the same form when there is a compound input process, when there are several servers,
or when there is batch servicing. Similarly Whitt (1968), p. 67, has shown that certain types of dependence among the interarrival and service times can be allowed without changing the basic results. It seems very likely that the formulations of assembly-like systems presented in Chapters 3 and 9 can be generalized in similar ways without changing the essential character of the limit theorems developed in Parts I and II. In particular, consider an assembly-like network where servers require different (but fixed) numbers of input items of the various types for each assembly. A system characterized by this very special type of batch servicing lies outside the current formulation, but it seems likely that the results in Chapter 10 can be extended to include this more general case.

Consider the PERT network pictured in Figure 13 below. The system consists of four "workers". Worker number 1 is free to begin his task immediately, workers number 2 and 3 cannot begin their respective tasks until worker number 1 has completed his, and worker number 4 cannot begin his task until workers number 2 and 3 have both completed theirs. This network does not satisfy the structural assumptions made in Chapter 3, because station 1 has two separate successors. This situation, in

![Figure 13. Schematic representation of an assembly-like network where one station has two successors.](image-url)
which several component tasks cannot be begun until a single preceding
task is completed, is the rule rather than the exception in PERT
networks arising in practice. The system portrayed in Figure 13 can be
alternately viewed as an assembly network by imagining that each "item"
produced at station 1 actually consists of two parts, one of which
acts as an input item to station 2 and the other of which acts as an
input item to station 3.

Although it may not be immediately apparent, the treatment of
assembly-like networks given in Part II can be quite easily generalized
to allow for the possibility of multiple successors (in the sense
discussed above). The notation must be generalized, of course, since
one must then identify more types of "items" than there are stations.
The definition of "predecessors" must be changed slightly to

\[ P(j) = \{i: j \in S(i)\} \]

and then the definition of a "chain to station \( j \)" can be left
exactly as before. With these changes, an "approximate representation"
for \( T^k_n \) can be developed which is almost identical to that given in
Lemma 9.3. Generalized versions of all the limit theorems in Chapter
10 then follow as before. Although these limit theorems are quite
close in form to those in Chapter 10, the qualitative conclusions
derived from them may change drastically. In particular, it is
intuitively clear that if the system has just one external station,
like that in Figure 13, then it may be stable (in the usual sense),
APPENDIX

Proof of Theorem 2.13

We begin by stating two additional results from the theory of weak convergence which will be required. Let \((M, \rho)\) and \((M', \rho')\) be metric spaces and \(h: M \rightarrow M'\) a measurable mapping. Let

\[D_n = \{x \in M: h \text{ is discontinuous at } x\} \, .\]

The following result is a generalization of the Continuous Mapping Theorem and is proved by Billingsley (1968), p. 31.

**Lemma 1:** If \(\{X_n\}\) and \(X\) are random elements of \(M\) such that \(X_n \Rightarrow X\) and \(P\{X \in D_n\} = 0\), then \(h(X_n) \Rightarrow h(X)\).

For the next proposition we assume \(M\) and \(M'\) to be separable. For a proof of this result see Billingsley (1968), p. 27.

**Lemma 2:** Let \(\{X_n\}\) and \(X\) be random elements of \(M\), \(\{X'_n\}\) random elements of \(M'\), and \(X'\) a constant element of \(M'\). If \(X_n \Rightarrow X\) in \(M\) and \(X'_n \Rightarrow X'\) in \(M'\), then \((X_n, X'_n) \Rightarrow (X, X')\) in \(M \times M'\).

For ease of reference we now repeat the set-up for Theorem 2.13. We assume that \(\{T_n^1, \ldots, T_n^K\}\) are random variables satisfying

\[0 \leq T_n^1 \leq T_n^2 \leq \cdots\] almost everywhere \((1 \leq k \leq K)\) and that \(\mu_1, \ldots, \mu_K\) are positive constants. It is assumed for convenience that

\[\mu_k > 1, \quad 1 \leq k \leq K\, .\]
customers of class 1 demanding a single product of the first type and customers of class 2 demanding one product of each type. Assuming that customers of each type are turned away unless their entire order can be met from current inventories, the behavior of this system is clearly similar to that of the system mentioned in the previous paragraph. Of primary interest are the conditions under which inventory levels stabilize (rather than blow up) and the expected inventory levels and fraction of lost sales under stable conditions.

We have demonstrated in Chapter 10 the essential equivalence of functional limit theorems for the (vector) waiting time process and those for the (vector) queue length process in an assembly-like network. It seems likely that this equivalence can be shown to hold for much more general types of queueing systems as well (under appropriate "heavy traffic" conditions, if the system is potentially stable). Such a result would be of considerable usefulness, since the methods used in the analysis of continuous-parameter processes are typically quite different from those used to study discrete-parameter processes. The use made of Theorem 2.13 in the proof of Theorems 10.6 and 10.7 indicates its potential value in proving general equivalences of this type. The time scale transformation appearing in the current version of Theorem 2.13 makes it somewhat difficult to work with, however, and perhaps a form more convenient for queueing applications can be found.
APPENDIX

Proof of Theorem 2.13

We begin by stating two additional results from the theory of weak convergence which will be required. Let \((M, \rho)\) and \((M', \rho')\) be metric spaces and \(h: M \to M'\) a measurable mapping. Let
\[
D_\rho = \{x \in M : h \text{ is discontinuous at } x\}.
\]
The following result is a generalization of the Continuous Mapping Theorem and is proved by Billingsley (1968), p. 31.

**Lemma 1:** If \(\{X_n\}\) and \(X\) are random elements of \(M\) such that \(X_n \Rightarrow X\) and \(P\{X \in D_\rho\} = 0\), then \(h(X_n) \Rightarrow h(x)\).

For the next proposition we assume \(M\) and \(M'\) to be separable. For a proof of this result see Billingsley (1968), p. 27.

**Lemma 2:** Let \(\{X_n\}\) and \(X\) be random elements of \(M\), \(\{X'_n\}\) random elements of \(M'\), and \(X'\) a constant element of \(M'\). If \(X_n \Rightarrow X\) in \(M\) and \(X'_n \Rightarrow X'\) in \(M'\), then \((X_n, X'_n) \Rightarrow (X, X')\) in \(M \times M'\).

For ease of reference we now repeat the set-up for Theorem 2.13. We assume that \(\{T_{n,1}^1, \ldots, T_{n,K}^K\}\) are random variables satisfying \(0 \leq T_{n,1}^1 \leq T_{n,2}^2 \leq \cdots\) almost everywhere \((1 \leq k \leq K)\) and that \(\mu_1, \ldots, \mu_K\) are positive constants. It is assumed for convenience that
\[
\mu_k > 1, \quad 1 \leq k \leq K.
\]
We define the translated random vectors \( T_n = (T_n^1, \ldots, T_n^K) \) and the corresponding random functions \( \nu_n = (\nu_n^1, \ldots, \nu_n^K) \) in \( D[0,1]^K \) by letting

\[
\mu = \left( \begin{array}{c} 
\mu_1 \\
\vdots \\
\mu_K 
\end{array} \right), \quad T_n = \left( \begin{array}{c} 
T_n^1 \\
\vdots \\
T_n^K 
\end{array} \right),
\]

\[
\hat{T}_n = T_n - n\mu, \quad n = 0, 1, \ldots,
\]

\[
\nu_n(t) = \frac{1}{\sqrt{n}} \hat{T}_{nt}, \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots.
\]

Then we define a \( K \)-dimensional counting process \( N(t) \) and the corresponding random functions \( \zeta_n = (\zeta_n^1, \ldots, \zeta_n^K) \) in \( D[0,1]^K \) by letting

\[
N^k(t) = \max\{n:T_n^k \leq t\}, \quad 0 \leq t < \infty, \quad 1 \leq k \leq K,
\]

\[
\alpha = \left( \begin{array}{c} 
\alpha_1 \\
\vdots \\
\alpha_K 
\end{array} \right), \quad N(t) = \left[ \begin{array}{c} 
N^1(t) \\
\vdots \\
N^K(t) 
\end{array} \right],
\]

\[
\zeta_n(t) = \frac{1}{\sqrt{n}} [N(nt)-nt\alpha], \quad 0 \leq t \leq 1, \quad n = 1, 2, \ldots.
\]

The result to be proved is then as follows.

**Theorem:** If \( \nu_n \Rightarrow \nu \), where \( \nu \) is a random element of \( D[0,1] \)
whose sample paths are almost surely continuous, then $\xi_n \Rightarrow -A\xi$,

where

$$
A = 
\begin{pmatrix}
-1 \\
\mu_1^{-1} \\
\vdots \\
\mu_K^{-1}
\end{pmatrix}
$$

and $\xi = (\xi_1, \ldots, \xi_K)$ is a random element of $D[0,1]^K$ defined by

$$
\xi^k(t) = \nu^k(\mu_k^{-1}t), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K.
$$

This result has been proved for the special case where $K = 1$ and $\nu$ is Brownian Motion by Billingsley (1968), pp. 148-150. Our proof is a straightforward and essentially trivial extension of his. The key idea is that of a random change in the time scale for the random functions $\nu_n$. Let $D_o[0,1]$ consist of those elements $x(\cdot)$ of $D[0,1]$ that are nondecreasing and satisfy $0 \leq x(t) \leq 1$ for all $t \in [0,1]$. We endow $D_o[0,1]$ with the relativized Skorohod topology and let $D_o[0,1]^K$ denote the corresponding $K$-fold product space, endowed with the product topology. We define a sequence $\phi_n = (\phi_n^1, \ldots, \phi_n^K)$ of random elements of $D_o[0,1]^K$ by letting

*/ Using the fact that $\{\xi(at); 0 \leq t \leq 1\}$ has the same distribution as $\{a^{1/2}\xi(t); 0 \leq t \leq 1\}$, the reader may verify that in this special case, Theorem 2.13 reduces to the "if" part of Theorem 2.12.
\[
\varphi_n^k(t) = \begin{cases} 
\frac{1}{n} N_k^k(nt) & \text{if } \frac{1}{n} N_k^k(nt) \leq 1 \\
\mu_k^{-1} t & \text{otherwise}
\end{cases}
\]

Also define a constant element \( \varphi = (\varphi^1, \ldots, \varphi^K) \) of \( D_0[0,1]^K \) by letting
\[
\varphi^k(t) = \mu_k^{-1} t, \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K.
\]

Finally we define a mapping
\[
\psi = (\psi^1, \ldots, \psi^K): D_0[0,1]^K \times D[0,1]^K \to D[0,1]^K
\]
by letting
\[
[\psi^k(y,x)](t) = (x^k \circ y^k)(t) = x^k(y^k(t)), \quad 0 \leq t \leq 1, \quad 1 \leq k \leq K,
\]
for \( y = (y^1, \ldots, y^K) \in D_0[0,1]^K \) and \( x = (x^1, \ldots, x^K) \in D[0,1]^K \).

Let \( k \) be fixed but arbitrary. Given that \( \mu_k > 1 \) and \( \nu_n^k \Rightarrow \nu_k^k \) (which is implied by our assumption that \( \nu_n \Rightarrow \nu \)), the argument of Billingsley (1968), p. 149, shows that
\[
\delta(\varphi_n^k, \varphi^k) \overset{P}{\to} 0 \quad \text{as } n \to \infty.
\]

It is immediate then that
\[
\delta^K(\varphi_n, \varphi) \overset{P}{\to} 0 \quad \text{as } n \to \infty,
\]
and hence by Theorem 2.2, \( \varphi_n \Rightarrow \varphi \). Thus we have from Lemma 2 that
\[(\varphi_n, \nu_n) \Rightarrow (\varphi, \nu) \text{ in } D_o[0,1]^K \times D[0,1]^K.\]

Again let \(k\) be fixed but arbitrary. The argument of Billingsley (1968), p. 145, shows that if \(y \in D_o[0,1]\) is continuous and \(x \in D[0,1]\) is continuous, then the mapping \(\psi^k\) defined above is continuous at the point \((y, x)\). It follows that if \(y \in D_o[0,1]^K\) is continuous then and \(x \in D[0,1]^K\) is continuous then \(\psi\) is continuous at the point \((y, x)\). Since \(\varphi\) is a continuous element of \(D_o[0,1]^K\) and we have assumed \(\nu\) to be almost surely continuous, this means that

\[P[(\varphi, \nu) \in D_\psi] = 0,\]

so by Lemma 1

\[(A.1) \quad \psi(\varphi_n, \nu_n) \Rightarrow \psi(\varphi, \nu).\]

Finally, the argument of Billingsley (1968), p. 149 shows directly that for each \(k = 1, \ldots, K\)

\[d[\psi^k(\varphi_n, \nu_n), -\mu_n^k] = d(\nu_n^k \varphi_n^k, -\mu_n^k) \xrightarrow{P} 0,\]

implying immediately that

\[(A.2) \quad d^K[\psi(\varphi_n, \nu_n), -A^{-1} \zeta_n] \xrightarrow{P} 0.\]

Combining (A.1) and (A.2) with Theorem 2.2 we have that \(-A^{-1} \zeta_n \Rightarrow \psi(\varphi, \nu)\), or equivalently stated,
\[ \zeta_n \implies -A (\varphi, \nu) \].

The proof is then completed by the observation that the random function \( \zeta \) appearing in the theorem statement is just \( \Psi(\varphi, \nu) \).
REFERENCES


Queueing models for assembly-like systems are formulated and limit theorems are developed for the stochastic processes associated with these models. The basic single-station model, intended to be representative of an assembly operation, is studied in Part I. It consists of several input processes, each of which delivers a different type of "component part" or "input item", and a server who assembles these items into finished products, requiring exactly one input item of each type per assembly. Assuming the input processes to be renewal processes and service times (or assembly times) to be i.i.d. random variables, it is shown that such a system is inherently unstable, and "functional" limit theorems are developed for properly normalized versions of the basic queue length and waiting time processes. In Part II a similar model of a more complex assembly network is introduced, and generalizations are obtained for all of the results developed in Part I. In Part III some related models and possible directions for future research, including some natural problems of optimal design and control, are discussed briefly.
1. Assembly-Like Operations
2. Assembly-Like Network
3. Weak Convergence
4. Functional Limit Theorem
5. Queueing Theory
6. In-Process Inventory
7. Waiting Time
8. Precedence Network
9. Stochastic Process
10. System Stability

**INSTRUCTIONS**

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether “Restricted Data” is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 8570.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHORS(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal publication procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b. 8c. & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system number, task number, etc.

9a. ORIGINATOR’S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   (1) “Qualified requesters may obtain copies of this report from DDC.”

   (2) “Foreign announcement and dissemination of this report by DDC is not authorized.”

   (3) “U.S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _______.

   (4) “U.S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _______.

   (5) “All distribution of this report is controlled. Qualified DDC users shall request through _______.

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, relics, and weights is optional.