AN ABSTRACT OF THE THESIS OF

GEORGE REILLY FEGAN for the DOCTOR OF PHILOSOPHY
(Name) (Degree)

in STATISTICS presented on July 20, 1973
(Major) (Date)

Title: THE PURE BIRTH PROCESS WITH TIME-INDEPENDENT, UNEQUAL TRANSITION PARAMETERS: MAXIMUM LIKELIHOOD ESTIMATION; EXACT AND COMPUTATIONAL FORMS FOR THE EXPECTATION AND THE VARIANCE OF THE PROCESS.

Abstract approved: Redacted for privacy

John P. Mullooly

For the pure birth process with independent waiting times having distinct transition parameters, let \( \{X(0,t)=k\} \), \( k \geq 0 \), be the event that \( k \) occurrences have taken place in the interval \((0,t)\). For \( k \geq 1 \) the probability function \( \Pr[X(0,t)=k] \) is shown to be the product of a constant and the density function of \( T_{k+1} \) the sum of \((k+1)\) waiting times, a random variable which has the general Erlang distribution. This form is then used to show that the maximum likelihood estimate for \( \theta t \), where \( \theta \) is the intensity parameter of the process, is the unique modal point of the density function of the distribution of \( T_{k+1} \). By means of a result due to J.B.S. Haldane, an approximation to this modal point is given. For various examples of the
process, this approximation is compared to the exact maximum
likelihood estimate with favorable results.

It is then shown that for any positive numbers,

\[ p_1, p_2, p_3, \ldots, p_s \]

\[
\sum_{i=1}^{s} \frac{p_i^r}{\prod_{j \neq i} (p_i - p_j)} \quad (r \geq s)
\]

is equal to the product of \( \frac{1}{(r-s+1)!} \) and the \((r-s+1)\)-st raw
moment of an Erlang distribution with parameters \((\frac{1}{p_1}, \frac{1}{p_2}, \frac{1}{p_3}, \ldots, \frac{1}{p_s})\).

The result is then used to develop a Maclaurin series for \( E(X(0,t))^{n} \).

The series is then applied to various examples in order to compute
their expectations and variance.

Finally in an appendix analogous results are derived for the pure
birth process with independent waiting times having equal transition
parameters, namely the Poisson process.
The Pure Birth Process with Time-Independent, Unequal Transition Parameters: Maximum Likelihood Estimation; Exact and Computational Forms for the Expectation and the Variance of the Process.

by

George Reilly Fegan

A THESIS submitted to Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

June 1974
APPROVED:

Redacted for privacy

Assistant Professor of Statistics
in charge of major

Redacted for privacy

Chairman of Department of Statistics

Redacted for privacy

Dean of Graduate School

Date thesis is presented

Typed by Clover Redfern for George Reilly Fegan

July 20, 1973
ACKNOWLEDGMENT

The author wishes to express his thanks to his parents who in one application of the process discussed herein made it all possible.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. THE GENERAL PROCESS AND PARTICULAR MODELS</td>
<td>1</td>
</tr>
<tr>
<td>Definitions and General Assumptions</td>
<td>1</td>
</tr>
<tr>
<td>Example 1: The Truncated Epidemic Model S → I</td>
<td>2</td>
</tr>
<tr>
<td>Example 2: The r-th Order Kinetic Model Using Combinatorial Transition Rates A → B</td>
<td>3</td>
</tr>
<tr>
<td>Example 3: The Blowfly Mating Model V → P</td>
<td>4</td>
</tr>
<tr>
<td>The Probability Density Function of X(0, t) as a Function of the Difference Between the Cumulative Distribution Functions of the Sums of the Waiting Times</td>
<td>5</td>
</tr>
<tr>
<td>II. THE MAXIMUM LIKELIHOOD ESTIMATION OF THE INTENSITY PARAMETER</td>
<td>11</td>
</tr>
<tr>
<td>Existence of a Unique Solution to the Likelihood Equation</td>
<td>16</td>
</tr>
<tr>
<td>Approximations to the Maximum Likelihood Estimate Using Haldane's Result in Examples 1, 2, and 3</td>
<td>23</td>
</tr>
<tr>
<td>III. EXTENSION OF CHIANG'S LEMMA</td>
<td>34</td>
</tr>
<tr>
<td>IV. EXPECTATION AND VARIANCE OF THE PROCESS</td>
<td>40</td>
</tr>
<tr>
<td>Results of Evaluating the Expectation and the Variance of the Truncated Pure Birth Process by Means of a Maclaurin Series</td>
<td>52</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>64</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>66</td>
</tr>
<tr>
<td>Appendix I</td>
<td>66</td>
</tr>
<tr>
<td>Appendix II</td>
<td>71</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Epidemic model: initial number of susceptibles (N = 10, 20, 30); the initial number of infectives (y = 1).</td>
<td>32</td>
</tr>
<tr>
<td>2. Kinetic model using combinatorial transition parameters: the initial number of molecules in A, (N = 30), and (r = 2).</td>
<td>33</td>
</tr>
<tr>
<td>3. Blowfly mating model: the initial number of males (M = 20), the initial number of females (N = 40).</td>
<td>33</td>
</tr>
<tr>
<td>4. Example 3: the blowfly mating model: initial population (M = 20, N = 40), (\max{E(X(0, t))} = 20), (\max{\lambda_i} = \lambda_1 = 800).</td>
<td>53</td>
</tr>
<tr>
<td>5. Example 2: (r)-th order kinetic model using combinatorial transition rates: initial population (N = 10), (r = 2), (\max{E(X(0, t))} = 5), (\max{\lambda_i} = \lambda_1 = 45).</td>
<td>54</td>
</tr>
<tr>
<td>6. Example 2: (r)-th order kinetic model using combinatorial transition rates: initial population (N = 50), (r = 2), (\max{E(X(0, t))} = 25), (\max{\lambda_i} = \lambda_1 = 1225).</td>
<td>55</td>
</tr>
<tr>
<td>7. Example 2: (r)-th order kinetic model using combinatorial transition rates: initial population (N = 100), (r = 2), (\max{E(X(0, t))} = 50), (\max{\lambda_i} = \lambda_1 = 4950).</td>
<td>56</td>
</tr>
<tr>
<td>8. Example 3: epidemic model: initial population (N = 10), (\max{E(X(0, t))} = 5), (\max{\lambda_i} = \lambda_5 = 30).</td>
<td>57</td>
</tr>
<tr>
<td>9. Example 3: epidemic model: initial population (N = 20), (\max{E(X(0, t))} = 10), (\max{\lambda_i} = \lambda_{10} = 110).</td>
<td>58</td>
</tr>
<tr>
<td>10. Example 3: epidemic model: initial population (N = 30), (\max{E(X(0, t))} = 15), (\max{\lambda_i} = \lambda_{15} = 240).</td>
<td>59</td>
</tr>
</tbody>
</table>
THE PURE BIRTH PROCESS WITH TIME-INDEPENDENT, UNEQUAL TRANSITION PARAMETERS: MAXIMUM LIKELIHOOD ESTIMATION; EXACT AND COMPUTATIONAL FORMS FOR THE EXPECTATION AND VARIANCE OF THE PROCESS

I. THE GENERAL PROCESS AND PARTICULAR MODELS

Definitions and General Assumptions

For the pure birth process considered here, let the random variable $X(t_1, t_2)$ be defined as the number of events which occur in the interval $(t_1, t_2)$. A birth or event will be an irreversible transference of a member or set of members of a parent population to a second population. The parent population will be considered to be a countable collection of undefined elements. Then $\{X(t_1, t_2) = k\}$, $k$ a non-negative integer, is defined as $k$ events occurring in $(t_1, t_2)$ and hence an increase in the second population by an amount which is a function of $k$.

Assumptions:

(i) $X(t_1, t_2)$ and $X(t_2, t_3)$, $t_1 < t_2 < t_3$, are independent random variables

(ii) $\Pr[X(0, t+\Delta t) = j+1/X(0, t) = j] = \lambda_{j+1} \theta \Delta t + o(\Delta t)$

(iii) $\Pr[X(0, t+\Delta t) = j+r/X(0, t) = j] = o(\Delta t)$, $r \neq 0, 1$

(iv) $\Pr[X(t, t) = 0] = 1$

(v) $\lambda_i \neq \lambda_j$ for all $i, j$
The $\lambda_i$ must be non-negative and independent of time and will be
denoted as transition parameters; $\theta$ will be denoted as the intensity
parameter of the process and to avoid trivial results $\theta$ will be
taken to be strictly positive.

As is generally known (Chiang, 1968), these assumptions imply the results

$$\Pr[X(0,0) = k] = 0 \quad \text{for } k \neq 0,$$
$$\Pr[X(0,t) = 0] = e^{-\lambda_1 \theta t}$$

$$\Pr[X(0,t) = k] = \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \ldots \lambda_k}{(\lambda_j - \lambda_i)} e^{-\lambda_i \theta t} \quad \text{for } k \geq 1.$$

(1.1)

Three particular examples of the process will be considered. However, in all three cases the parent population will be finite and
decreased by the occurrence of an event; hence the set $\{\lambda_i\}$ will be
taken to be a finite set of positive, distinct quantities and the particular processes being described should be regarded as truncated pure
birth processes.

Example 1: The Truncated Epidemic Model $S \rightarrow I$

The parent population $S$ consists of $N$ susceptibles. A second class $I$ of infectives exists and for any $t$ has a specified
number of members, \( y \), and \( y \) will be constrained so that \( y \leq N \). An event is the infecting of a member of class \( S \) and thereby irreversibly transferring this member to class \( I \). The \( \lambda_i \) are defined as follows:

\[
\lambda_1 = (N)(y), \lambda_2 = (N-1)(y+1), \lambda_3 = (N-2)(y+2), \ldots, \lambda_j = (N-j+1)(y+j-1), \ldots, (j \leq \frac{N-y}{2} + 1).
\]

The \( \lambda_i \) are increasing in \( i \) in this model.

The restriction on \( j \) accounts for the word truncated in the name of the model. While all three examples are truncated pure birth processes, this particular model is also a truncated epidemic model.

The usual epidemic model has the set of transition parameters \( \{\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n\} \) where \( \lambda_N = (1)(y+N-1) \). In order to keep the \( \lambda_i \) distinct the number of possible events has been constrained and hence \( X(0,t) = \frac{N-y}{2} + 1 \), a.e. In the case where the class of infectives initially consists of one member, the maximum number of events is the greatest integer less than or equal to \( \frac{N+1}{2} \).

If \( Y(t) \) is defined as the size of \( S \) at time \( t \), then \( Y(t) = N - X(0,t) \).

**Example 2: The r-th Order Kinetic Model Using Combinatorial Transition Rates A \( \rightarrow \) B**

The parent population \( A \) consists of \( N \) molecules. An event is a chemical reaction in which \( r \) molecules of \( A \) react and
thereby are irreversibly transferred into a single unit of $B$. Let $j$ be the greatest integer such that $j \leq \frac{N}{r}$. Then there can only be $j$ events and the $\lambda_i$ are defined as

$$\lambda_1 = \binom{N}{r}, \lambda_2 = \binom{N-1}{r}, \lambda_3 = \binom{N-2}{r}, \ldots, \lambda_j = \binom{N-(j-1)r}{r}.$$ 

The $\lambda_i$ are decreasing in $i$ in this model.

If $Y(t)$ is defined as the number of molecules in the unreacted class $A$ at time $t$, then $Y(t) = N - rX(0, t)$.

Example 3: The Blowfly Mating Model $V \rightarrow P$

Bartlett, Brennan, and Pollack (1971) have described an experiment which utilized a parent population $V$ made up of $N$ female and $M$ male blowflies. The parent population was placed in an observation cage; an event was a mating, and the mated pair was removed and placed in a holding cage. The population in the holding cage will be designated $P$. The $\lambda_i$ are decreasing in $i$ and are defined as follows:

$$\lambda_1 = (M)(N), \lambda_2 = (M-1)(N-1), \lambda_3 = (M-2)(N-2), \ldots, \lambda_j = (M-j)(N-j);$$

$$j \leq \min\{M, N\}; \quad X(0, t) \leq \min\{M, N\}, \quad \text{a.e.}$$

If $Y(t)$ is the size of $V$ at time $t$, $M(t)$ is the size of the male population in $V$ at time $t$, and $N(t)$ is the size of the
female population in $V$ at time $t$, then $Y(t) = M + N - 2X(0, t)$, 
$M(t) = M - X(0, t)$, and $N(t) = N - X(0, t)$.

The Probability Density Function of $X(0, t)$ as a Function of the Difference Between the Cumulative Distribution Functions of the Sums of the Waiting Times

Let $W_1, W_2, W_3, \ldots, W_i, \ldots$ be a sequence of independent, non-negative random variables, where $W_i$ is the waiting time between the $(i-1)$st event and the $i$-th event. Let $T_k = \sum_{i=1}^{k} W_i$.

The following well-known theorem gives the relationship between the probability of exactly $k$ events and the cumulative distribution functions of $T_k$ and $T_{k+1}$:

Theorem (1.1).

$$\Pr[\text{Exactly } k \text{ events in } (0, t)] = \Pr[T_k \leq t] - \Pr[T_{k+1} \leq t]$$

Proof: The event \{$T_k \leq t$\} can be written as the union of two mutually exclusive events, i.e.,

$$\{T_k \leq t\} = \{T_k \leq t \text{ and } T_{k+1} \leq t\} \cup \{T_k \leq t \text{ and } T_{k+1} > t\}.$$

But \{$T_k \leq t \text{ and } T_{k+1} \leq t$\} = \{$T_k \leq t$\} and \{$T_k \leq t \text{ and } T_{k+1} > t$\} = \{Exactly $k$ events in $(0, t)$\}. 

Hence

\[ \Pr[T_k \leq t] = \Pr[T_{k+1} \leq t] + \Pr[T_k \leq t \text{ and } T_{k+1} > t] \]

or

\[ \Pr[\text{Exactly } k \text{ events in } (0, t)] = \Pr[T_k \leq t] - \Pr[T_{k+1} \leq t]. \]

Now if the \( W_i \) are independent, non-identically distributed exponentials with parameters \( \lambda_i \theta \), \( T_k = \sum_{i=1}^{k} W_i \) is a general Erlang random variable. Writing the c.d.f. of \( T_k \) as

\[ F_k(t) = \Pr[T_k \leq t] = 1 - \sum_{i=1}^{k} \left( \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} \right) e^{-\lambda_i \theta t} \]

and expressing \( \{\text{Exactly } k \text{ events in } (0, t)\} \) as \( \{X(0, t) = k\} \), we get for the pure birth processes under consideration

\[ \Pr[X(0, t) = k] = F_k(t) - F_{k+1}(t) = \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_k}{k+1} e^{-\lambda_i \theta t} \cdot \prod_{j \neq i} (\lambda_j - \lambda_i) \]

Since the probability density function of \( T_{k+1} \) is

\[ \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_{k+1} \theta}{(k+1)} e^{-\lambda_i \theta t} \cdot \prod_{j \neq i} (\lambda_j - \lambda_i) \]

the following lemma is immediate:
Lemma (1.1). Denoting the probability density function of $T_{k+1}$ as $F_{k+1}^{(1)}(t)$, we may write for any positive $k$

$$
\Pr[X(0, t) = k] = \frac{1}{\lambda_{k+1} \theta} F_{k+1}^{(1)}(t).
$$

Proof: Multiplying and dividing $\Pr[X(0, t) = k]$ by $\lambda_{k+1} \theta$, we get

$$
\Pr[X(0, t) = k] = \frac{\lambda_{k+1} \theta}{\lambda_{k+1} \theta} \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \ldots \lambda_k}{(\lambda_j - \lambda_i)} e^{-\lambda_i \theta t}
$$

$$
= \frac{1}{\lambda_{k+1} \theta} \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \ldots \lambda_k \theta}{(\lambda_j - \lambda_i)} e^{-\lambda_i \theta t}
$$

$$
= \frac{1}{\lambda_{k+1} \theta} F_{k+1}^{(1)}(t).
$$

Using this lemma, we may now write the following important equation:

\begin{equation}
Pr[X(0, t) = k] = \frac{1}{\lambda_{k+1} \theta} F_{k+1}^{(1)}(t) = F_k(t) - F_{k+1}(t)
\end{equation}

In Appendix II we will show that for the Poisson process in which $\lambda_i = \lambda_j = \lambda$ for all $i$ and $j$, we may also write
Pr\{X(0, t)=k\} = \frac{1}{\lambda_k+\theta} F_k^{(1)}(t) = F_k(t) - F_{k+1}(t)

where \( F_k(t) \) is the c.d.f. of \( T_k = \sum_{i=1}^{k} W_i \) and the \( W_i \) are exponentially distributed with parameter \( \lambda_i\theta = \lambda\theta \).

The thesis will not explore the remaining case where \( \lambda_i = \lambda_j \) for only some \( i \) and \( j \). However, the following lemma shows that if \( F_k^{(1)}(t) = \lambda_k(F_k(t) - F_{k+1}(t)) \) then the individual waiting times are exponentially distributed.

**Lemma (1.2).** Let \( T_k = \sum_{i=1}^{k} W_i \) where the \( W_i \) are independent, non-negative random variables. Let \( F_0(t) = 1 \) and \( F_k(t), \ k \geq 1, \) be the c.d.f. of \( T_k \). If the probability density function \( F_k^{(1)}(t) \) can be written as \( F_k^{(1)}(t) = \lambda_k\theta(F_k(t) - F_{k+1}(t)) \), for all \( k \geq 1 \), then \( W_i \) is distributed exponentially with parameter \( \lambda_i\theta, \ i \leq k \).

**Induction Proof:** For \( k = 1 \) if \( F_1^{(1)}(t) = \lambda_1\theta(1-F_1(t)) \), then multiplying by the integrating factor \( e^{\lambda_1\theta t} \), we get

\[
\frac{d}{dt}(F_1(t)e^{\lambda_1\theta t}) = \frac{d}{dt} \int_0^t \lambda_1 e^{\lambda_1\theta x} \, dx.
\]

This implies \( F_1(t)e^{\lambda_1\theta t} = e^{-\lambda_1\theta t} \) or \( F_1(t) = 1 - e^{-\lambda_1\theta t} \) and hence
$W_1$ is exponentially distributed with parameter $\lambda_1 \theta$. Assume that $W_i$ is exponentially distributed with parameter $\lambda_i \theta$, $i \leq k$. For $k+1$ if $F^{(1)}_{k+1}(t) = \lambda_{k+1} \theta (F_k(t) - F_{k+1}(t))$, then multiplying by the integrating factor $e^{\lambda_{k+1} \theta t}$ and noting that $F_k(t)$ is Riemann integrable, we get

$$\frac{d}{dt} (F_{k+1}(t) e^{\lambda_{k+1} \theta t}) = \frac{d}{dt} \int_0^t \lambda_{k+1} \theta e^{\lambda_{k+1} \theta x} F_k(x) dx$$

which implies

$$F_{k+1}(t) = e^{-\lambda_{k+1} \theta t} \int_0^t \lambda_{k+1} \theta e^{\lambda_{k+1} \theta x} F_k(x) dx$$

$$= \int_0^t e^{-\lambda_{k+1} \theta (t-x)} F_k(x) dx.$$

Since $T_k$ and $W_{k+1}$ are non-negative random variables, this last equation implies that $F^{(1)}_{k+1}(t)$ is the convolution of the densities $-\lambda_k \theta t$ and $F^{(1)}_k(t)$ and hence $W_{k+1}$ is exponentially distributed with parameter $\lambda_{k+1} \theta$.

In order to gain general results for the estimation of the intensity parameter based on the occurrence of $k$ events in $(0, t)$, it is standard practice to reparameterize the domain variable and then estimate $\theta t$. Hence if we transform the sum of the waiting times by
means of \( \overline{T}_{k+1} = 0 \overline{T}_{k+1} \) and denote the c.d.f. of \( \overline{T}_{k+1} \) by

\[ G_{k+1}(t) \]

then \( G_{k+1}(t) = F_{k+1}(t) \) and the derivative of \( G_{k+1}(t) \)
is

\[ G_{k+1}^{(1)}(t) = \frac{d(F_{k}(t))}{dt} = \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_{k+1}}{k+1} \prod_{j \neq i} (\lambda_j - \lambda_i) e^{-\lambda_i t}. \]

Writing \( \theta t \) for \( t \), we get the very useful representation

\[ \Pr[X(0, t) = k] = \frac{1}{\lambda_{k+1}} G_{k+1}^{(1)}(\theta t) = G_k(\theta t) - G_{k+1}(\theta t), \quad (k \geq 1). \]

In order to complete our forms we note that for \( k = 0 \)

\[ \Pr[X(0, t) = 0] = 1 - F_1(t) = 1 - G_1(\theta t). \]
II. THE MAXIMUM LIKELIHOOD ESTIMATION OF THE INTENSITY PARAMETER

Differentiating Equation (1.1)

\[ \Pr[X(0, t) = k] = \sum_{i=1}^{k+1} \frac{\lambda_1 \lambda_2 \lambda_3 \ldots \lambda_k}{k+1} \frac{-\lambda_i \theta t}{\prod_{j \neq i} (\lambda_j - \lambda_i)} e^{\lambda_i t} \]

with respect to \( \theta t \), we find that the likelihood equation for the estimation of \( \theta t \) is

\[ \sum_{i=1}^{k+1} \frac{\lambda_i e^{-\lambda_i \theta t}}{(k+1) \prod_{j \neq i} (\lambda_j - \lambda_i)} = 0 \] for \( k > 0 \).

For \( k = 0 \), \( \Pr[X(0, t) = 0] = 1 - G_1(\theta t) \) and the likelihood function is maximized by \( \hat{\theta} t = 0 \).

For \( k = 1 \), Equation (2.1) has the unique solution

\[ \hat{\theta} t = \frac{1}{\lambda_1 - \lambda_2} \ln \frac{\lambda_1}{\lambda_2} \]

Since, as we will later see, the likelihood function is unimodal, this is the maximum likelihood estimate.

For \( k > 1 \), there is no known closed form solution for Equation (2.1) for an arbitrary set \( \{\lambda_i \mid \lambda_i > 0 \text{ for all } i\} \).
The main procedures are to approximate the maximum likelihood estimate by discovering an asymptotic solution to Equation (2.1) or to solve the likelihood equation based on an approximate distribution of $T_k$. In either case the resulting approximate solution to Equation (2.1) is then usually tested in a few cases against the actual maximum likelihood estimate calculated by means of numerical methods.

Exact solutions for the first order and asymptotic solutions for the r-th order kinetic model using combinatorial transition rates have been found by Bartholomay (1959) and Mullooly (1972) respectively. Mullooly (1973) has also shown that for the r-th order kinetic model with combinatorial transition rates $T_k$, the sum of $k$ waiting times, possesses asymptotic normality provided: (1) $N$, the initial number of molecules in $A$, is large; and (2) the reaction is not near completion. The author using methods used by Mullooly (1972) has found an asymptotic solution to Equation (2.1) for the truncated epidemic model with an initial number of infectives $y = 1$. The results of using this estimator are found in Table 1; the development of the estimator is to be found in Appendix I.

The problems with these procedures are the following: (1) an asymptotic solution for general conditions is difficult to develop; (2) in many interesting cases the distribution of the sum of the waiting times is difficult to approximate. Asymptotic normality does not seem
to be the general situation; McGill and Gibbon (1965) have suggested that the gamma distribution may serve as an approximating distribution for the sum of the waiting times.

An illustration of the truth of point (2) is found by examining the distribution of the sum of the waiting times in the truncated epidemic model with $y = 1$. Since $T_k = \sum_{i=1}^{k} W_i$ has a general Erlang distribution, then

$$E(T_k) = \sum_{i=1}^{k} \frac{1}{\lambda_i \theta}$$

and

$$\sigma^2(T_k) = \sum_{i=1}^{k} \frac{1}{(\lambda_i \theta)^2}$$

and so we can show that

$$\lim_{k \to \infty} \max_{1 \leq j \leq k} \frac{\sigma(W_j)}{\sigma(T_k)} \to 0$$

and hence the distribution of $T_k$ is not asymptotically normal in the usual sense.

**Lemma (2.1).** Let $W_i$ be the waiting time between the $(i-1)$-st and $i$-th event for the truncated epidemic model with
$y = 1$. Then the distribution of $T_k = \sum_{i=1}^{k} W_i$ is not asymptotically normal in the usual sense.

**Proof:** Since $(N)(1) \theta \leq \lambda_1 \theta$ for all $i \leq \max j \leq k$ 

Also since

$$\sigma^2(T_k) \leq \sigma^2(T_{N/2}) \quad \text{or} \quad \sigma^2(T_k) \leq \sigma^2(T_{(N+1)/2})$$

depending on whether $N$ is even or odd, $\sigma^2(T_k)$ is bounded by

$$\sigma^2(T_{N/2}) = \frac{1}{\theta^2} \left( \frac{1}{(1)(N)} \right)^2 + \frac{1}{((2)(N-1))^2} + \ldots + \frac{1}{((N)(\frac{1}{2})(\frac{N+1}{2}))^2}$$

$$\leq \frac{1}{\theta^2} \left( \frac{1}{N+2} \right)^2 \left( \frac{1}{1^2} + \frac{1}{2^2} + \ldots + \frac{1}{(N)^2} \right)$$

or

$$\sigma^2(T_{(N+1)/2}) = \frac{1}{\theta^2} \left( \frac{1}{(1)(N)} \right)^2 + \frac{1}{((2)(N-1))^2} + \ldots + \frac{1}{((N+1)(\frac{1}{2})(\frac{N+1}{2}))^2}$$

$$\leq \frac{1}{\theta^2} \left( \frac{1}{N+2} \right)^2 \left( \frac{1}{1^2} + \frac{1}{2^2} + \ldots + \frac{1}{(N+1)^2} \right) .$$

The sum $\sum_{i=1}^{j} \frac{1}{i^2}$ can be bounded by means of the integral:

$$\sum_{i=1}^{j} \frac{1}{i^2} < \int_{1}^{j} \frac{1}{i^2} \, di \leq \sum_{i=1}^{j-1} \frac{1}{i^2} \quad (j, \text{any positive integer})$$
or
\[
\sum_{i=1}^{j} \frac{1}{i^2} \leq 2 - \frac{1}{j}.
\]

Hence if \( N \) is odd or even
\[
\sigma^2(T_k) \leq \frac{1}{\theta^2 \left( \frac{N+1}{2} \right)^2} \left( 2 - \frac{1}{N+1} \right).
\]

So
\[
\max_{1 \leq j \leq k} \frac{\sigma(W_j)}{\sigma(T_k)} \geq \frac{1}{N\theta} \sqrt{2 - \frac{2}{N+1}} = \frac{1 + \frac{1}{N}}{2 \sqrt{2 - \frac{2}{N+1}}},
\]

and
\[
\lim_{k \to \infty} \max_{1 \leq j \leq k} \frac{\sigma(W_j)}{\sigma(T_k)} \geq \frac{1}{2\sqrt{2}} \quad \text{since} \quad k \leq \frac{N+1}{2}.
\]

Hence the distribution of \( T_k \) is not asymptotically normal in the usual sense.

The main result of the chapter then is to give a good approximation to the maximum likelihood estimate for any \( k > 0 \) which allows a great deal of generality without making too many demands on the form of the distribution of \( T_k \). We will first give a few preliminary results concerning the solution to the likelihood equation.
Existence of a Unique Solution to the Likelihood Equation

We state a lemma, a proof for which may be found in Chiang (1968).

**Lemma (2.2) (Chiang).** For any distinct numbers, \( p_1', p_2', p_3', \ldots, p_s' \), we have

\[
\sum_{i=1}^{s} \frac{p_i^r}{\prod_{j=1}^{s} (p_i - p_j)} = 0 \quad \text{for} \quad 0 < r < s - 1
\]

\[
\sum_{i=1}^{s} \frac{p_i^r}{\prod_{j=1}^{s} (p_i - p_j)} = 1 \quad \text{for} \quad r = s - 1
\]

An immediate consequence of this lemma is the following:

**Corollary.** Let \( G_k^{(1)}(\theta t) \) be the reparameterized probability density function for \( T_k = \sum_{i=1}^{k} W_i \), where the \( W_i \) are independent, non-identically distributed exponentials with parameters \( \lambda_i, \theta \). Then for \( k > 1 \), \( G_k^{(1)}(0) = 0 \).

**Proof:**

\[
G_k^{(1)}(\theta t) = \sum_{i=1}^{k} \frac{\prod_{j \neq i} (\lambda_i - \lambda_j)}{\prod_{j=1}^{k} (\lambda_i - \lambda_j)} \lambda_i^{\lambda_i - 1} \theta^t e^{-\lambda_i \theta t}
\]

Hence

\[
G_k^{(1)}(0) = \lambda_1 \lambda_2 \lambda_3 \ldots \lambda_k \sum_{i=1}^{k} \frac{1}{\prod_{j \neq i} (\lambda_i - \lambda_j)} (-1)^{k-1}
\]
and the result follows from the lemma.

This corollary has some interesting consequences for asymptotic solutions to Equation (2.1). By Equation (1.3)

\[
\frac{d\Pr[X(0,t) = k]}{d(\theta t)} = \frac{d}{d(\theta t)} \frac{1}{\lambda_{k+1}} G^{(1)}_{k+1}(\theta t) = \frac{d}{d(\theta t)} (G^{(1)}_k(\theta t) - G^{(1)}_{k+1}(\theta t))
\]

Then let \( \hat{\theta} t \) be any function of \( k \) and \( N \) such that

\( \hat{\theta} t = h(k, N) > 0 \) where \( k \) is the number of events observed and \( N \) is the initial size of the parent population; and let \( h(k, N) \) be such that for fixed \( k > 1 \) \( \lim_{N \to \infty} h(k, N) = 0 \). Then by the corollary \( \hat{\theta} t N \to \infty \) is an asymptotic solution for Equation (2.1) since

\[
\lim_{N \to \infty} (G^{(1)}_{k+1}(\theta t) - G^{(1)}_{k+1}(\theta t)) = 0.
\]

As we will see by means of a result by Haldane, it is reasonable to believe that the maximum likelihood estimate \( \hat{\theta} t \) based on \( k \) observations in \((0, t)\) is less than \( E(T_{k+1}) \). Assume that \( \lambda_i \) increases with \( N \) for \( 1 \leq i \leq k+1 \); it then follows that

\[
\lim_{N \to \infty} \sum_{i=1}^{k+1} \frac{1}{\lambda_i} = 0,
\]

and consequently for \( \theta > 0 \) with \( t \) fixed we have for any \( k \)

\[
\lim_{N \to \infty} E(T_{k+1}) = \frac{1}{\theta} \lim_{N \to \infty} \sum_{i=1}^{k+1} \frac{1}{\lambda_i} = 0.
\]

It is then reasonable to expect \( \hat{\theta} t \) to be very close to zero for large \( N \).
Hence under these conditions the difference between $\hat{\theta}$ and $\hat{\theta}$ would with high probability be very small. This is an unsatisfactory result since the class of asymptotic solutions to Equation (2.1) is extremely large and the choice of a particular solution seems to imply a certain arbitrariness. The procedure used by Mullooly (1972) to derive an asymptotic maximum likelihood estimator is to derive a particular asymptotic solution to Equation (2.1) by means of a carefully chosen auxiliary likelihood equation and then to compare the results from this solution with exact maximum likelihood estimates found by numerical methods for a few cases of small $N$.

We will now show that the distribution function of $T_k$ is unimodal. To do so we will use a theorem by A. Ibragimov (1956) which gives conditions for a distribution function to be strongly unimodal. Ibragimov calls a distribution function strongly unimodal if the distribution function is unimodal and if its convolution with any unimodal distribution function is unimodal.

Our first step will be to show that the distribution function of $W_j$, the exponentially distributed random variable, is unimodal. Although this is rather obvious and can be easily done by differentiation, we will employ an interesting proof using characteristic functions. A well-known theorem due to A. Ya. Khinchine, a proof for which can be found in Lukacs (1970) gives necessary and sufficient conditions for a distribution function to be unimodal.
Theorem (2.1) (Khinchine). A distribution function $F(t)$ is unimodal with vertex at $t = 0$ if, and only if, its characteristic function $f(s)$ can be represented as

$$f(s) = \frac{1}{s} \int_{0}^{s} g(u)du \quad (-\infty < s < \infty)$$

where $g(u)$ is a characteristic function.

Corollary. Let $H_j(t)$ be the distribution function of $W_j$, where $W_j$ is distributed exponentially with parameter $\lambda_j \theta$. Then $H_j(t)$ is a unimodal function with vertex at $t = 0$.

Proof: The characteristic function for the distribution function $H_j(t)$ is

$$f_{W_j}(s) = \frac{1}{(1-\lambda_j \theta is)}.$$

Differentiating $(sf_{W_j}(s))$ with respect to $s$, we get

$$\frac{d}{ds} (sf_{W_j}(s)) = \frac{1}{(1-\lambda_j \theta is)} + \frac{\lambda_j \theta is}{(1-\lambda_j \theta is)^2} = \frac{1}{(1-\lambda_j \theta is)^2}.$$

But $\frac{1}{(1-\lambda_j \theta is)^2}$ is the characteristic function of the convolution of $H_j(t)$ with $H_j(t)$; hence $H_j(t)$ is unimodal.
The next theorem due to Ibragimov gives necessary and sufficient conditions for a distribution function to be strongly unimodal. In the proof the author uses the following condition for concavity: a function \( v(x) \) is concave on an interval \( I \) if \( v''(x) \leq 0 \) at each point of \( I \).

**Theorem (2.2) (Ibragimov).** A (non-degenerate) unimodal distribution function \( F(t) \) is strongly unimodal if, and only if, \( F(t) \) is continuous and if \( \log F^{(1)}(t) \) is concave on the set of points at which neither the right-hand nor the left-hand derivative of \( F(t) \) vanishes.

**Corollary.** Let \( H_j(t) \) be the distribution function of \( W_j \), where \( W_j \) is distributed exponentially with parameter \( \lambda_j \). Then \( H_j(t) \) is strongly unimodal.

**Proof:** \( H_j(t) \) is continuous everywhere; and

\[
H_j^{(1)}(t) = \begin{cases} 
0 & t < 0 \\
-\lambda_j \theta t & 0 \leq t < \infty 
\end{cases}
\]

\[
= \lambda_j \theta e^{-\lambda_j \theta t} 
\]

\[
\log H_j^{(1)}(t) = \log \lambda_j \theta - \lambda_j \theta t 
\]

and

\[
\frac{d^2}{dt^2} \log H_j^{(1)}(t) = 0 
\]
and hence \( \log H_j^{(1)}(t) \) is concave. Therefore \( H_j(t) \) is strongly unimodal.

**Lemma (2.3).** For \( k \geq 1 \) the distribution function of \( T_k \) is unimodal.

**Induction Proof:** For \( k = 1 \) we employ the corollary to Theorem (2.1). For \( j \leq k-1 \) we assume the distribution function for \( T_j \) is unimodal. For \( j = k \) we note that the distribution function of \( T_k \) is the convolution of the distribution function of \( T_{k-1} \) and the unimodal distribution function of \( W_k \), a strongly unimodal function. By the corollary to Theorem (2.2) and the definition of a strongly unimodal function, the result follows.

**Corollary.** For \( k \geq 1 \), the distribution function for the reparameterized random variable \( \overline{T}_k = \theta T_k \) is unimodal.

**Proof:** The proof is immediate since \( F_k(t) = G_k(\theta t) \).

**Lemma (2.4).** For all \( k \geq 1 \) let \( (\theta t)_k \) be the point at which \( G_k^{(1)}(\theta t) \) achieves its maximum value. Then \( (\theta t)_k \leq (\theta t)_{k+1} \).

**Proof:** We first show that for any \( k \geq 1 \) \( (\theta t)_k \) is the unique point at which \( G_k(\theta t) \) achieves its maximum value. Otherwise

1. \( G_k(\theta t) \) is not unimodal; or
2. \( G_k^{(1)}(\theta t) = G_k^{(1)}(\theta t)_k \) for all \( \theta t \)
in some neighborhood of \((\theta t)_k\). But (1) is false since \(G_k(\theta t)\) is unimodal; (2) is false since \(G_k^{(1)}(\theta t)\) is analytic on \((0, \infty)\) and \(G_k^{(1)}(\theta t) = G_k^{(1)}(\theta t)_k\) for all \(\theta t\) in some neighborhood of \((\theta t)_k\) implies that \(G_k^{(1)}(\theta t)\) is a constant on \((0, \infty)\).

We now prove the lemma by contradiction. Assume
\[(\theta t)_k > (\theta t)_{k+1}\] for some \(k \geq 1\). Since \(G_k^{(1)}(\theta t)\) achieves its maximum value at the unique point \((\theta t)_{k+1}\), then \(G_k^{(2)}(\theta t)_{k+1} = 0\) and \(G_{k+1}^{(2)}(\theta t) \leq 0\) for all \((\theta t) > (\theta t)_{k+1}\). Then by Equation (1.3) and the assumption \(G_k^{(1)}(\theta t)_{k+1} = G_{k+1}^{(1)}(\theta t)_{k+1}\) and \(G_k^{(1)}(\theta t)_k \leq G_{k+1}^{(1)}(\theta t)_k\).

Since \((\theta t)_{k+1}\) is the unique point at which \(G_k^{(1)}(\theta t)\) achieves its maximum \(G_k^{(1)}(\theta t)_k \leq G_{k+1}^{(1)}(\theta t)_k < G_{k+1}^{(1)}(\theta t)_{k+1}\) and since \(G_k^{(1)}(\theta t)_{k+1} = G_{k+1}^{(1)}(\theta t)_{k+1}\) we have a contradiction that \((\theta t)_k\) is the maximum point for \(G_k^{(1)}(\theta t)\).

**Corollary.** Let \((\theta t)_k\) be the unique point at which \(G_k^{(1)}(\theta t)\) achieves its maximum value. Then \(G_k^{(1)}(\theta t)_k \geq G_{k+1}^{(1)}(\theta t)\) for all \((\theta t) \geq 0\) and all \(k \geq 1\).

**Proof:** For any \((\theta t) \neq (\theta t)_{k+1}\) since \((\theta t)_{k+1}\) is the unique modal point of \(G_k^{(1)}(\theta t)\), we have \(G_k^{(1)}(\theta t)_{k+1} > G_{k+1}^{(1)}(\theta t)\). But \(G_{k+1}^{(2)}(\theta t)_{k+1} = 0\) implies \(G_k^{(1)}(\theta t)_{k+1} = G_{k+1}^{(1)}(\theta t)_{k+1}\) and since \((\theta t)_k\) is the modal point of \(G_k^{(1)}(\theta t)\) we get the following inequalities
\[G_k^{(1)}(\theta t)_k \geq G_k^{(1)}(\theta t)_{k+1} \geq G_{k+1}^{(1)}(\theta t)_{k+1} > G_{k+1}^{(1)}(\theta t)\).
The result upon which our approximation to the maximum likelihood estimate is to be based is found in the next theorem.

**Theorem (2.3).** Let \( T_{k+1} = \sum_{i=1}^{k+1} W_i \), the \( W_i \) being independent, non-identically distributed exponentials with distinct parameters \( \lambda_i \theta \). Let \( G_{k+1}^{(1)}(\theta t) \) be the reparameterized density function of \( \overline{T_{k+1}} \). The unique solution to the likelihood Equation (2.1)

\[
\sum_{i=1}^{k+1} \frac{\lambda_i e^{-\lambda_i \theta t}}{\prod_{j \neq i} (\lambda_j - \lambda_i)} = 0 \quad \text{for} \quad k \geq 1
\]

is the modal point of \( G_{k+1}^{(1)}(\theta t) \).

**Proof:** By Equation (1.3) we can write

\[
\Pr[X(0,t)=k] = \frac{1}{\lambda_{k+1}^{(1)}} G_{k+1}^{(1)}(\theta t).
\]

Hence the \( \theta t \) which maximizes \( G_{k+1}^{(1)}(\theta t) \) maximizes \( \Pr[X(0,t)-k] \).

The uniqueness follows from the fact that \( G_{k+1}^{(1)}(\theta t) \) is unimodal.

**Approximations to the Maximum Likelihood Estimate Using Haldane's Result in Examples 1, 2, and 3**

The problem then of finding the maximum likelihood estimate for \( \theta t \) in the pure birth process with unequal parameters is now placed
in a new context: instead of trying to find a solution to the likelihood Equation (2.1) we can now concentrate on finding at least a good approximation to the modal point of the probability density function of \( T_{k+1} \). Such an approximation under rather general conditions has been given by J. B. S. Haldane, whose discoveries in human genetics are fundamental to the science. Haldane (1942) suggested the following approximation to the modal point of the probability density function of a distribution which is "nearly normal":

\[
\text{the modal point} = \mu'_1 - \frac{\mu_3}{2\mu_2} + \frac{\mu_5}{8\mu_2} - \frac{5\mu_3\mu_4}{12\mu_2^3} + \frac{\mu_3^3}{4\mu_2^4} + O(n^{-2})
\]

where \( \mu_i \) is the \( i \)-th central moment, \( \mu'_1 \) is the mean, \( O(n^{-2}) \) is read as "terms of at most order \( n \)," and \( n \) is some parameter of the distribution. To derive his result Haldane expanded the density function by means of Hermitian polynomials, differentiated the function, and set the result equal to zero. Applying the restriction that the \( r \)-th cumulant exists and be of order \( n \) for some parameter \( n \) and all \( r \), he used an auxiliary equation in which he ignored all terms in his original equation of power greater than 2. He then solved the resulting quadratic equation and presented his estimate for the modal point.

In Haldane's development the author assumes that Haldane
implied that $K_r$ be of order $n$ uniformly and yet in the examples which follow the development he applies his result to the $\chi^2$ distribution with $n$ degrees of freedom and $K_r = 2^{r-1}(r-1)!n$. What is more, it appears that Haldane could have used the restriction that $K_r/r!$ must be of order $n$ since the original expansion used the standardized cumulants divided by $r!$. However, Haldane gives empirical evidence that his approximation is quite good for Type III distributions and hence we deduce that it might make a good approximation for the modal point of the general Erlang distribution.

Before applying the approximation we will investigate the form and order of the cumulants and the form of the central moments.

**Lemma (2.5).** The $r$-th cumulant for $T_k = \theta T_k = \sum_{j=1}^{k} \theta W_j$, where the $W_j$ are independent, non-identically distributed exponentials with distinct parameters $\lambda, \theta$, is

$$K_r = (r-1)! \sum_{j=1}^{k} \frac{1}{(\lambda_j)^r}.$$ 

**Proof:** Since the characteristic function of $\bar{T}_k$ is

$$\phi_{\bar{T}_k}(s) = \prod_{j=1}^{k} \frac{1}{1 - i s \lambda_j},$$

the cumulant generating function is
\[
\psi_T^n(s) = - \sum_{j=1}^{k} \log(1 - \frac{is}{\lambda_j}).
\]

Expanding each log function into a Maclaurin power series and collecting terms, we get

\[
\psi_T^n(s) = \sum_{j=1}^{k} \frac{1}{1 \lambda_j} s + \sum_{j=1}^{k} \frac{1}{2 \lambda_j^2} s^2 + \ldots + \sum_{j=1}^{k} \frac{1}{r \lambda_j^r} s^r + \ldots
\]

Hence

\[
K_r = (r-1)! \sum_{j=1}^{k} \frac{1}{(\lambda_j)^r}
\]

for the distribution of the reparameterized sum of the waiting times.

To examine the order of the cumulants for Examples 1, 2, and 3 we note that in all three cases \( \lambda_j \geq 1 \) for all \( j \). Hence

\[
\sum_{j=1}^{k} \left(\frac{1}{\lambda_j}\right)^r \leq \sum_{j=1}^{k} \frac{1}{\lambda_j} \quad (r \geq 1).
\]

For Example 1 since there are at most \( \frac{N+1}{2} \) events, \( \frac{K_r}{r!} \leq \frac{1}{r} \frac{N+1}{2} \) where \( N \) is the initial number of susceptibles.

For Example 2, there are at most \( N/r \) reactions where \( r \) represents here the order of the reaction and since \( r \geq 1 \) there can be no more than \( N \) reactions. Hence \( \frac{K_r}{r!} \leq \frac{1}{r} N \) where \( N \) is the
initial number of molecules in the unreacted class A.

For Example 3, there are at most \( \min\{M, N\} \) possible events. Hence \( \frac{K_r}{r!} \leq \frac{1}{r} \min\{M, N\} \) where \( M \) is the initial number of males and \( N \) is the initial number of females.

In each example therefore \( K_r \) is not uniformly of order \( n \) for the chosen parameter but \( \frac{K_r}{r!} \) is. Therefore from the original expansion used by Haldane, we have good reason to believe that his approximation to the modal point will be quite good for these examples.

Using standard textbook results for expressing central and raw moments in terms of cumulants we find that for the distribution of the reparameterized \( \overline{T_k} \):

\[
\mu_1' = \sum_{i=1}^{k} \frac{1}{\lambda_i}, \quad \mu_4 = 9 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 6 \sum_{i=1}^{k} \sum_{j>i} \frac{1}{\lambda_i \lambda_j}, \quad \mu_5 = 44 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 20 \sum_{i=1}^{k} \sum_{j \neq i} \frac{1}{\lambda_i \lambda_j}
\]

\[
\mu_2 = \sum_{i=1}^{k} \frac{1}{\lambda_i}, \quad \mu_3 = 2 \sum_{i=1}^{k} \frac{1}{\lambda_i}
\]

**Lemma (2.6).** The 4-th and 5-th central moments of the reparameterized \( \overline{T_k} \) can be written
(i) \[ \mu_4 = 6 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 3\mu_2^2 \]

(ii) \[ \mu_5 = 24 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 10\mu_3\mu_2 \]

**Proof:** (i) \[ \mu_2 = \sum_{i=1}^{k} \frac{1}{\lambda_i} \]

\[ \mu_2^2 = \sum_{i=1}^{k} \frac{1}{\lambda_i} + 2 \sum_{i=1}^{k} \sum_{j>i}^{k} \frac{1}{\lambda_i} \frac{1}{\lambda_j} \]

Therefore

\[ \mu_4 = 6 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 3\mu_2^2 \]

(ii) \[ \mu_3 = 2 \sum_{i=1}^{k} \frac{1}{\lambda_i} \]

\[ \mu_3\mu_2 = 2 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 2 \sum_{i=1}^{k} \sum_{j>i}^{k} \frac{1}{\lambda_i} \frac{1}{\lambda_j} \]

\[ \mu_5 = 24 \sum_{i=1}^{k} \frac{1}{\lambda_i} + 10\mu_3\mu_2 \]

From the development of Haldane's approximation and from the empirical evidence summarized in Tables 1, 2, 3, the author suggests
that the estimation of $\Theta_t$ in the pure birth process with $\lambda_j \leq 1$

for all $j$ be done by means of a very simple process. One calculates

$$\sum_{i=1}^{k} \left( \frac{1}{\lambda_i} \right)^r$$

for $r = 1, 2, 3, 4, 5$ and then uses the lower central moments to calculate $\mu_4$ and $\mu_5$. The estimate itself then is

$$\hat{\Theta}_t = \mu_1 - \frac{\mu_3}{2\mu_2} + \frac{\mu_5}{8\mu_2^2} - \frac{5\mu_3\mu_4}{12\mu_2^3} + \frac{\mu_3^3}{4\mu_2^4}.$$ 

For a reasonably small $k$, the work can be done without a calculating machine; what is more important for the examples in which the author has applied this estimator, the estimation results are excellent, using the criterion of absolute distance from the true maximum likelihood estimate.

Tables 1, 2, and 3 summarize the results of using Haldane's approximation to the mode of the density for the distribution of the sum of the waiting times in order to estimate $\Theta_t$ for $2 \leq k \leq j-1$ where $j$ is the maximum number of reactions possible given the initial conditions. The estimate for $k = 0$ is $\hat{\Theta}_t = 0$; for $k = 1$, the exact solution to Equation (2.1) is $\hat{\Theta}_t = \frac{1}{\lambda_1 - \lambda_2} \ln \frac{\lambda_1}{\lambda_2}$; and for $k = j$ in Examples 2 and 3, $\hat{\Theta}_t = \infty$. However, in Example 1, if one considers the model as an epidemic model truncated so that the $\lambda_i$ might be distinct for all $i$, then $\Theta_t$ cannot be estimated for $k = j$ by means of the procedures discussed, for it is unrealistic to
regard the infective process as being completed.

In each table comparisons are made with the exact mode $\hat{\theta}$ which was found by numerical solution of Equation (2.1)

$$\sum_{i=1}^{k+1} \frac{-\lambda_i \theta t}{\lambda_i e} \prod_{j \neq i} (\lambda_j - \lambda_i) = 0 \quad \text{for} \ k \geq 1.$$ 

In Table 1, the number of initial infectives $y$ was chosen to be 1 so that comparisons could be made with the asymptotic estimator

$$\hat{\theta} = \frac{\ln(k+1)}{N} + \frac{(k-2)+\frac{2}{k}}{N^2},$$

the development of which is given in Appendix I.

In Table 2, $r$ was chosen to be 2 and $N$ as 30; the exact solutions are taken from tables provided by Mullooly (1972).

In Table 3, the initial number of males was chosen to be 20, and the initial number of females to be 40. This was for the sake of correspondence with the experiment of Bartlett, Brennan, and Pollack (1971).

It is interesting to note that the results of using the approximation are excellent as long as the smallest $\lambda_i$ used in the calculation is somewhat larger than 1 so that $1/\lambda_i$ remains rather small.
In the kinetic model, the $\lambda_i$ decrease reaching a minimum of $\lambda_{15} = 1$. The estimates for this model are all quite good except in the case $k = 14$, the only case in which $\lambda_{15}$ is used. If one compares the moments for $\overline{T}_{15} = \sum_{i=1}^{15} \theta W_i$ with those of $\overline{T}_{14} = \sum_{i=1}^{14} \theta W_i$, the effect of $\lambda_{15}$ can be seen

<table>
<thead>
<tr>
<th></th>
<th>$\overline{T}_{14}$</th>
<th>$\overline{T}_{15}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>.3535</td>
<td>1.3535</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>.0345</td>
<td>1.0345</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>.0100</td>
<td>2.0100</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>.0083</td>
<td>9.2155</td>
</tr>
<tr>
<td>$\mu_5$</td>
<td>.0066</td>
<td>44.7970</td>
</tr>
</tbody>
</table>

The effect of the division of $\mu_3$ by $\mu_2$ and the division of $\mu_5$ by $\mu_2^2$ in the case of $\overline{T}_{15}$ produces numbers larger than 1 and affects the convergence rate of the Haldane approximation.
Table 1. Epidemic model: initial number of susceptibles $N = 10, 20, 30$; the initial number of infectives $y = 1$. A comparison is made among the exact solution, Haldane's approximation to the mode ($\hat{N}$), and the asymptotic estimator $[\hat{N}]$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$N$</th>
<th></th>
<th></th>
<th>$k$</th>
<th>$N$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>.123</td>
<td>.058</td>
<td>.038</td>
<td>8</td>
<td>.132</td>
<td>.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.124)</td>
<td>(.058)</td>
<td>(.038)</td>
<td></td>
<td>(.132)</td>
<td>(.082)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.162</td>
<td>.075</td>
<td>.049</td>
<td>9</td>
<td>.142</td>
<td>.087</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.164)</td>
<td>(.075)</td>
<td>(.049)</td>
<td></td>
<td>(.142)</td>
<td>(.087)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.199</td>
<td>.089</td>
<td>.057</td>
<td>10</td>
<td></td>
<td>.092</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.202)</td>
<td>(.089)</td>
<td>(.057)</td>
<td></td>
<td></td>
<td>(.092)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[.189]</td>
<td>[.087]</td>
<td>[.057]</td>
<td></td>
<td></td>
<td>[.089]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.101</td>
<td>.064</td>
<td></td>
<td>11</td>
<td></td>
<td>.096</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.101)</td>
<td>(.065)</td>
<td></td>
<td></td>
<td></td>
<td>(.097)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[.099]</td>
<td>[.064]</td>
<td></td>
<td></td>
<td></td>
<td>[.093]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.122</td>
<td>.071</td>
<td></td>
<td>12</td>
<td></td>
<td>.101</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.112)</td>
<td>(.071)</td>
<td></td>
<td></td>
<td></td>
<td>(.101)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[.109]</td>
<td>[.070]</td>
<td></td>
<td></td>
<td></td>
<td>[.097]</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.122</td>
<td>.077</td>
<td></td>
<td>13</td>
<td></td>
<td>.105</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.122)</td>
<td>(.077)</td>
<td></td>
<td></td>
<td></td>
<td>(.106)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[.118]</td>
<td>[.076]</td>
<td></td>
<td></td>
<td></td>
<td>[.101]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td></td>
<td>.110</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.110)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[.104]</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Kinetic model using combinatorial transition parameters: the initial number of molecules in A, \(N = 30\), and \(r = 2\). A comparison is made between the exact solution \(\hat{\theta}t\) and Haldane's approximation \(\bar{\theta}t\) ( ).

<table>
<thead>
<tr>
<th>(k)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.005</td>
<td>.009</td>
<td>.013</td>
<td>.017</td>
<td>.023</td>
<td>.030</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.005)</td>
<td>(.009)</td>
<td>(.013)</td>
<td>(.017)</td>
<td>(.023)</td>
<td>(.030)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>8</th>
<th>8</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.040</td>
<td>.052</td>
<td>.070</td>
<td>.097</td>
<td>.143</td>
<td>.239</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.040)</td>
<td>(.052)</td>
<td>(.070)</td>
<td>(.097)</td>
<td>(.143)</td>
<td>(.239)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.572</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.416)</td>
</tr>
</tbody>
</table>

Table 3. Blowfly mating model: the initial number of males \(M = 20\), the initial number of females \(N = 40\). A comparison is made between the exact solution \(\hat{\theta}t\) and Haldane's approximation \(\bar{\theta}t\) ( ).

<table>
<thead>
<tr>
<th>(k)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.0027</td>
<td>.0043</td>
<td>.0059</td>
<td>.0077</td>
<td>.0097</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.0027)</td>
<td>(.0042)</td>
<td>(.0059)</td>
<td>(.0077)</td>
<td>(.0097)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.0119</td>
<td>.0143</td>
<td>.0171</td>
<td>.0202</td>
<td>.0238</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.0119)</td>
<td>(.0143)</td>
<td>(.0171)</td>
<td>(.0202)</td>
<td>(.0238)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.0279</td>
<td>.0327</td>
<td>.0384</td>
<td>.0455</td>
<td>.0545</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.0279)</td>
<td>(.0327)</td>
<td>(.0385)</td>
<td>(.0455)</td>
<td>(.0546)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}t)</td>
<td>.0667</td>
<td>.0842</td>
<td>.1158</td>
</tr>
<tr>
<td>(\bar{\theta}t)</td>
<td>(.0667)</td>
<td>(.0846)</td>
<td>(.1154)</td>
</tr>
</tbody>
</table>
III. EXTENSION OF CHIANG'S LEMMA

In Chapter II the following lemma due to Chiang (1968) was cited:

**Lemma (2.2) (Chiang).** For any distinct numbers, \( p_1, p_2, p_3, \ldots, p_s \), we have

\[
\sum_{i=1}^{s} \frac{p_i^{r}}{\prod_{j=1, j \neq i}^{s}(p_i - p_j)} = 0 \quad \text{for} \quad 0 \leq r < s-1
\]

\[
\prod_{j=1}^{s}(p_i - p_j) = 1 \quad \text{for} \quad r = s-1
\]

The main result of this chapter is to extend this result for \( p_i > 0 \), for all \( i \), to the case where \( r \geq s \). To do so we must first reframe the lemma in terms of distribution functions.

**Lemma (3.1).** For \( k \geq 1 \), let \( H_k(t) \) be a function such that \( H_k(0) = 0 \) and let \( H_k^{(n)}(t) \) exist for all \( n \geq 0 \) with \( H_k^{(0)}(t) \) being defined as \( H_k(t) \). If \( H_k^{(1)}(t) = c(H_k(t) - H_{k+1}(t)) \) for some constant \( c \), then \( H_k^{(j)}(0) = 0 \) for \( 0 \leq j \leq k-1 \).

**Proof:** For all \( k \geq 1 \) \( H_k^{(0)}(0) = H_k(0) = 0 \). Assume that for \( k = n \) and all \( j \leq n-1 \)

\( H_n^{(j)}(0) = 0 \) (main induction hypothesis).
Assume that for \( k = n+1 \)
\[
H_{n+1}^{(m)}(0) = 0 \quad \text{for all} \quad m \leq n-1 \quad \text{(secondary hypothesis)}.
\]

Then for \( k = n+1 \) and \( m = n \)
\[
H_{n+1}^{(n)}(0) = c(H_{n}^{(n-1)}(0) - H_{n+1}^{(n-1)}(0)).
\]

From the secondary hypothesis \( H_{n}^{(n-1)}(0) = 0 \) and by the main induction hypothesis \( H_{n}^{(n-1)}(0) = 0 \).

Hence \( H_{n+1}^{(m)}(0) = 0 \) for all \( m \leq n \) and the lemma is proved.

**Lemma (3.2).** For \( k \geq 1 \) let \( H_{k}(t) \) be the cumulative distribution function of \( T_k = \sum_{i=1}^{k} W_i \), where the \( W_i \) are distributed exponentially with parameters \( p_i \) (not necessarily distinct). Let \( H_{k}^{(1)}(t) = p_k (H_{k-1}(t) - H_k(t)) \) and define \( H_{0}(t) = 1 \) for all \( t \geq 0 \).

Then
\[
H_{k}^{(j)}(0) = 0 \quad 0 \leq j < k
\]
\[
= \prod_{i=1}^{k} p_i \quad j = k
\]

**Proof:** Since \( H_{k}^{(j)}(0) \) exists for all \( j \geq 0 \) and all \( k \geq 1 \)
\( H_{k}^{(j)}(0) = 0, \quad 0 \leq j < k \) by Lemma (3.1). For \( k = 1 \) we have
\( H_1(t) = 1 - e^{-p_1 t} \). Hence \( H_1^{(1)}(0) = p_1 \). Assume for \( k = n \) that
\[
H_{n}^{(n)}(0) = \prod_{i=1}^{n} p_i.
\]
Then for \( k = n+1 \) \( H_{n+1}^{(n+1)} = p_{n+1}^{(n)}(H_n^{(n)}(0) - H_{n+1}^{(n)}(0)) \). But \( H_{n+1}^{(n)}(0) = 0 \) and by hypothesis

\[
H_n^{(n)}(0) = \prod_{i=1}^{n} p_i
\]

and the result follows.

**Lemma (3.3).** For \( k > 1 \) let \( H_k^*(t) \) be the distribution function of \( T_k^* = \sum_{i=1}^{k} W_i^* \) where the \( W_i^* \) are independent and distributed exponentially with distinct parameters \( 1/p_i \).

Then

\[
E(T_k^* j) = \sum_{i=1}^{k} \frac{p_i^{k+j-1}}{\prod_{i \neq j} (p_i - p_j)}
\]

**Proof:** Substituting \( \lambda = 1/p \) in the density function of the general Erlang distribution of \( T_k \), we get

\[
\frac{1}{p_1 p_2 p_3} \cdots \frac{1}{p_k} \sum_{i=1}^{k} \frac{e^{-t/p_i}}{\prod_{j \neq i} (\frac{1}{p_j} - \frac{1}{p_i})} = \sum_{i=1}^{k} \frac{p_i^{k-2} e^{-t/p_i}}{\prod_{j \neq i} (p_i - p_j)}
\]

So
Theorem (3.1) (Extension of Chiang's lemma for $p_i > 0$). For any distinct positive numbers, $p_1, p_2, p_3, \ldots, p_s$, we have

$$\sum_{i=1}^{s} \frac{p_i^r}{\prod_{j=1}^{s} (p_i - p_j)} = \frac{H_s^{(r+1)}(0)}{(-1)^{r-s+1} \prod_{i=1}^{s} p_i}$$

where $H_s(t)$ is the distribution function of $T_s = \sum_{i=1}^{s} W_i$ where the $W_i$ are independent and distributed exponentially with distinct parameters $p_i$. Also

$$\sum_{i=1}^{s} \frac{p_i^r}{\prod_{j=1}^{s} (p_i - p_j)} = \begin{cases} 0 & 0 \leq r < s-1 \\ 1 & r = s-1 \\ \frac{1}{(r-s+1)!} \mu^{r-s+1} & r \geq s \end{cases}$$
where $\mu'_n$ is the $n$-th raw moment of the general Erlang distribution of $T^*_s = \sum_{i=1}^{s} W^*_i$ where the $W^*_i$ are independent and distributed exponentially with distinct parameters $1/p_i$.

**Proof:**

\[
H^{(r+1)}_s(t) = p_1 p_2 p_3 \cdots p_s \sum_{i=1}^{s} \frac{(-1)^r p_i e^{-p_i t}}{\prod_{j \neq i} (p_j - p_i)}
\]

\[
= (-1)^{r-s+1} p_1 p_2 p_3 \cdots p_s \sum_{i=1}^{s} \frac{p_i e^{-p_i t}}{\prod_{j \neq i} (p_i - p_j)}
\]

Then

\[
H^{(r+1)}_s(0) = (-1)^{r-s+1} \prod_{i=1}^{s} \frac{p_i}{\prod_{j \neq i} (p_i - p_j)}
\]

By Lemma (3.2)

\[
H^{(r+1)}_s(0) = 0 \text{ for } r+1 < s
\]

\[
= \prod_{i=1}^{s} p_i \text{ for } r+1 = s
\]

and by Lemma (3.3) letting $k = s$ and $j = r-s+1$ we obtain the result.
The usefulness of this extension is due to the fact that the raw
moments of $T^*_s$ are easy to obtain. Since the distribution function
of $T^*_s$ is the convolution of the distribution functions of the $W^*_i$,
the raw moments of the distribution of $T^*_s$ are the convolutions of
the raw moments of the $W^*_i$. The first three raw moments of $T^*_s$
by way of example are

$$
\mu^1_1 = \sum_{i=1}^{s} p_i
$$

$$
\mu^1_2 = 2!( \sum_{i=1}^{s} p_i^2 + \sum_{i=1}^{s} \sum_{j>i}^{s} p_i p_j )
$$

$$
\mu^1_3 = 3!( \sum_{i=1}^{s} p_i^3 + \sum_{i=1}^{s} \sum_{j>i}^{s} p_i^2 p_j + \sum_{i=1}^{s} \sum_{j>i}^{s} \sum_{l>j}^{s} p_i p_j p_l )
$$
IV. EXPECTATION AND VARIANCE OF THE PROCESS

In order to guarantee the finiteness of the expectation and variance of the process, we will restrict the set of parameters $\{\lambda_i\}$ to a finite number $N$ and consider what we have called a truncated pure birth process. In our Examples 1, 2, and 3 we have already restricted ourselves to a finite set $\{\lambda_i\}$ by restricting our population upon which reactions are to be made to a finite number and requiring $\lambda_i$ to be strictly positive for all $i$.

Because of the intractability of the forms for the probability distribution of $X(0,t)$, the number of events in $(0,t)$, no general form for the expectation and variance of the pure birth process with unequal transition parameters is known. Forms for the $r$-th order kinetic model, using combinatorial transition parameters, are known for the $r = 1, 2$ cases. In the case $r = 1$, the process is the Yule process and forms for all the moments are easily found. For $r = 2$ MacQuarrie (1967) has shown that forms for the mean and variance could be arrived at by means of Gegenbauer polynomials; Haskey (1954) derived a form for the mean of the untruncated simple stochastic model and this form can be used for the truncated case. Unfortunately these forms are restricted to the particular models by means of the transition parameters used in the models and hence will not generalize to fit a wider class of transition parameters. In the second
order kinetic model using combinatorial transition parameters, it was a remarkable feat by MacQuarrie to recognize that the solution to the differential-difference equations entails the use of Gegenbauer polynomials but it is easy to see that for \( r > 2 \), this is not the case. MacQuarrie has also given approximate methods for finding higher moments of the process but the results are not very satisfactory.

The main result of this chapter is a form by which all moments of the truncated pure birth process for distinct transition parameters can be calculated. In Appendix II we will apply this form to the case where all the transition parameters are equal.

Let \( F_k(t) \) be the distribution function of \( T_k = \sum_{i=1}^{k} W_i \) where the \( W_i \) are independent and distributed exponentially with distinct parameters \( \lambda_i, \theta \). Let \( G_k(\theta t) \) be the distribution function of the reparameterized \( \tilde{T}_k = \theta T_k \). Let the set of \( \{\lambda_i\} \) be restricted such that the largest subscript is \( N \), that is \( \{\lambda_1, \lambda_2, \ldots, \lambda_N\} \). We define \( \Pr[X(0,t)=N] = F_N(t) = G_N(\theta t) \). Then using Equation (1.2) and Equation (1.3) we may write
\[ \Pr[X(0,t)=0] = 1 \quad -F_1(t) = 1 \quad -G_1(\theta t) \]
\[ \Pr[X(0,t)=1] = F_1(t) \quad -F_2(t) = G_1(\theta t) \quad -G_2(\theta t) \]
\[ \Pr[X(0,t)=k] = F_k(t) \quad -F_{k+1}(t) = G_k(\theta t) \quad -G_{k+1}(\theta t) \]
\[ \Pr[X(0,t)=N-1] = F_{N-1}(t) \quad -F_N(t) = G_{N-1}(\theta t) \quad -G_N(\theta t) \]
\[ \Pr[X(0,t)=N] = F_N(t) \quad -G_N(\theta t) \]

Multiplying each equation above by \( k^n \) and adding all the equations we get

\[
(4.1) \quad E(X(0,t)) = \sum_{k=1}^{N} (k^n-(k-1)^n)F_k(t) = \sum_{k=1}^{N} (k^n-(k-1)^n)G_k(\theta t) 
\]

For \( k = 1 \) this becomes

\[
E(X(0,t)) = \sum_{k=1}^{N} F_k(t) = \sum_{k=1}^{N} G_k(\theta t) 
\]

and the variance of \( X(0,t) \) can be written

\[
V(X(0,t)) = \sum_{k=1}^{N} (2k-1)F_k(t) \quad - \left( \sum_{k=1}^{N} F_k(t) \right)^2 = \sum_{k=1}^{N} (2k-1)G_k(\theta t) \quad - \left( \sum_{k=1}^{N} G_k(\theta t) \right)^2 
\]

We note that \( F_k^{(m)}(0) \) exists for all \( k \geq 1 \) and all \( m \geq 0 \).
and that $F_k(0) = 0$. In order to arrive at a computational form for the moments, and in particular for the expectation and the variance, we will formally expand $E(X(0,t))$ in a Maclaurin power series, examine the individual terms to see if we can achieve some tractable form for computational purposes and then see if the formal series converges.

Formally then,

$$E(X(0,t))^n = \sum_{k=1}^{N} \left(k^n-(k-1)^n\right) F_k^{(1)}(0) \frac{t^1}{1!} + \sum_{k=1}^{N} \left(k^n-(k-1)^n\right) F_k^{(2)}(0) \frac{t^2}{2!}$$

$$+ \ldots + \sum_{k=1}^{N} \left(k^n-(k-1)^n\right) F_k^{(m)}(0) \frac{t^m}{m!} + \ldots ,$$

where $F_k^{(m)}(0)$ is the $m$-th derivative of $F_k(t)$ evaluated at $t = 0$ or

$$E(X(0,t))^n = \sum_{k=1}^{N} \left(k^n-(k-1)^n\right) G_k^{(1)}(0) \frac{(\theta t)^1}{1!} + \sum_{k=1}^{N} \left(k^n-(k-1)^n\right) G_k^{(2)}(0) \frac{(\theta t)^2}{2!}$$

$$+ \ldots + \sum_{k=1}^{N} \left(k^n-(k-1)^n\right) F_k^{(m)}(0) \frac{t^m}{m!} + \ldots ,$$

where $G_k^{(m)}(0)$ is the $m$-th derivative of $G_k(\theta t)$ evaluated at $\theta t = 0$. Since $F_k^{(m)}(0) = \theta^m G_k^{(m)}(0)$, the individual terms in the two
expansions are identical and we will therefore concentrate on the expansion in terms of \( G_k^{(m)}(0) \).

If for \( m-k \leq 1 \) we let \( \mu_{(k, m-k)}' \) denote the \((m-k)\)-th moment of the distribution of \( T_k^* = \sum_{i=1}^{k} W_i^* \) where the \( W_i^* \) are independent, exponentially distributed with distinct parameters \( 1/\lambda_i \) and define

\[
\mu_{(k, m-k)}' = \begin{cases} 
0 & \text{if } m \leq k-1 \\
1 & \text{if } m = k
\end{cases}
\]

then using Theorem (3.1) we may write this formal expansion as

\[
E(X(0, t))^n = \sum_{m=1}^{\infty} \frac{\left(\theta t\right)^m}{m!} \sum_{k=1}^{N} \left(k^n - (k-1)^n\right) \frac{(-1)^{m-k}}{(m-k)!} \left(\prod_{j=1}^{k} \lambda_j\right) \mu_{(k, m-k)}' .
\]

Since the \((m-k)\)-th raw moment of the distribution of \( T_k^* \) is the \( k \)-fold convolution of the moments of the \( W_i^* \), we expect to discover that this form will lead to a tractable computational form. Also since we do have a \( k \)-fold convolution, for \((m-k) \geq 1\)

\[
\mu_{(k, m-k)}' \leq (m-k)! (\lambda_1 + \lambda_2 + \ldots + \lambda_k)^{m-k}
\]

and hence

\[
\mu_{(k, m-k)}' \leq (m-k)! (\max_{1 \leq i \leq N} \lambda_i)^{m-k}.
\]

Therefore we have
Theorem (4.1). Let $X(0,t)$ be a random variable meeting Assumptions i-v of Chapter I; let $G_k(\theta t)$ be the distribution function for the reparameterized variable $\tilde{T}_k = \Theta T_k$, where

$$T_k = \sum_{i=1}^{k} W_i$$

and the $W_i$ are independent, exponentially distributed with distinct parameters $\lambda_i\theta$. Then

$$E(X(0,t))^n = \sum_{m=1}^{\infty} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{N} (k^n -(k-1)^n)G_k^{(m)}(0) \right)$$

for $\theta t \geq 0$.

Proof: Since for all $k$,

$$|(k^n -(k-1)^n)G_k^{(m)}(0)| \leq k^n (\max_{1 \leq i \leq N} \lambda_i)^m$$

then

$$\sum_{k=1}^{N} (k^n -(k-1)^n)G_k^{(m)}(0) \leq N^n N (\max_{1 \leq i \leq N} \lambda_i)^m$$

$$= N^{n+1} (\max_{1 \leq i \leq N} \lambda_i)^m.$$

Then the absolute value of each term in the expansion is less than or equal to the corresponding term in the expansion of $N^{n+1} e^{\theta t}$ where
A = \theta(\max_{1 \leq i \leq N} \{\lambda_i\}). Hence the sum of the terms in the expansion converges for all \( \theta t > 0 \) and the result follows.

**Corollary.** Let \( X(0, t) \) be a random variable meeting Assumptions i-iv of Chapter I; let \( F_k(t) \) be the distribution function of \( T_k = \sum_{i=1}^{N} W_i \) where the \( W_i \) are independent, exponentially distributed with distinct parameters \( \lambda_i \theta \). Then

\[
E(X(0, t))^n = \sum_{m=1}^{\infty} \frac{t^m}{m!} (\sum_{k=1}^{N} (k^n-(k-1)^n)F_k^{(m)}(0))
\]

for \( t > 0 \).

**Proof:** The proof follows the same line as the theorem with the comparison now made with the expansion of \( N^{n+1}e^{At} \) and

\[
A = 0(\max_{1 \leq i \leq N} \{\lambda_i\}).
\]

Before examining the computational forms and giving the results of applying the expansion to the moments of the distributions of our examples, we will list some immediate properties of the expectation and variance of the distribution of \( X(0, t) \).

**Lemma (4.1).** Let \( X(0, t) \) be a random variable meeting Assumptions i-iv of Chapter I; let the transition parameter set be the finite set \( \{\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_N\} \). Then
(i) \( 0 \leq E(X(0, t))^n \leq \sum_{k=1}^{N} (k^n - (k-1)^n) \); in particular
\[ 0 \leq E(X(0, t)) \leq N \text{ and } 0 \leq E(X(0, t))^2 \leq N^2 \]

(ii) \( E(X(0, 0))^n = 0 \) and \( \lim_{t \to \infty} E(X(0, t))^n = \sum_{k=1}^{N} (k^n - (k-1)^n) \);
\[ \text{in particular } \lim_{t \to \infty} E(X(0, t))^2 = N^2 \]

(iii) \( V(X(0, 0)) = 0 \) and \( \lim_{t \to \infty} V(X(0, t)) = 0 \)

(iv) For any \( t \) such that \( F_i(t) \neq 1 \) for some \( i \) \( V(X(0, t)) > 0 \)

**Proof:**

(i) \( E(X(0, t))^n = \sum_{k=1}^{N} (k^n - (k-1)^n) F_k(t) \) but \( 0 \leq F_k(t) \leq 1 \) for all \( k \). For \( n = 2 \) we note that
\[ \sum_{k=1}^{N} (k^n - (k-1)^n) = \sum_{k=1}^{N} (2k-1) = N^2 \]

(ii) From Assumption iv \( F_k(0) = 0 \); so for all \( n \)
\[ E(X(0, 0))^n = 0. \] Since \( F_k(t) \) is a c.d.f., \( \lim_{t \to \infty} F_k(t) = 1. \)
Hence

\[ \lim_{t \to \infty} E(X(0, t))^n = \lim_{t \to \infty} \sum_{k=1}^{N} (k^n - (k-1)^n) F_k(t) = \sum_{k=1}^{N} (k^n - (k-1)^n) \]
(iii) Since \( V(X(0,t)) = \sum_{k=1}^{N} (2k-1)F_k(t) - (\sum_{k=1}^{N} F_k(t))^2 \), the result is implied by (i) and (ii) above.

(iv) We note that we can write

\[
\left( \sum_{k=1}^{N} F_k(t) \right)^2 = \sum_{k=1}^{N} \sum_{j=1}^{N} F_k(t)F_j(t)
\]

and then regrouping the terms of the right-hand side we get

\[
\left( \sum_{k=1}^{N} F_k(t) \right)^2 = F_1(t)F_1(t) + \sum_{k=2}^{N} F_k(t)(F_k(t) + \sum_{j=1}^{k-1} 2F_j(t))
\]

Since for all \( k \), \( F_k(t) \leq 1 \), then \( F_1(t) \geq F_1(t)F_1(t) \) and

\[
2(k-1) \geq 2 \sum_{j=1}^{k-1} F_j(t).
\]

So

\[
1 + 2(k-1) = (2k-1) \geq F_k(t) + 2 \sum_{j=1}^{k-1} F_j(t).
\]

Therefore multiplying by \( F_k(t) \) we get

\[
(2k-1)F_k(t) \geq F_k(t)(F_k(t) + 2 \sum_{j=1}^{k-1} F_j(t))
\]

where this inequality is strict for any \( k \) such that
\( F_k(t) \neq 1 \). Now summing over all \( k \), we get

\[
\sum_{k=1}^{N} (2k-1)F_k(t) > \left( \sum_{k=1}^{N} F(t) \right)^2
\]

and hence the variance must be strictly positive.

To show that the Maclaurin expansion facilitates the computation of the raw moments of the distributions, we will write out \( \sum_{k=1}^{N} G_k^{(m)}(0) \) for \( m = 1, 2, 3, 4 \) and show that the forms are quite simple.

\[
\sum_{k=1}^{N} G_k^{(1)}(0) = \lambda_1
\]

\[
\sum_{k=1}^{N} G_k^{(2)}(0) = -\lambda_1^2 + \lambda_1 \lambda_2
\]

\[
\sum_{k=1}^{N} G_k^{(3)}(0) = \lambda_1^3 + \lambda_1 \lambda_2^2 + \lambda_1^2 \lambda_2 + \lambda_1 \lambda_3
\]

\[
\sum_{k=1}^{N} G_k^{(4)}(0) = -\lambda_1^4 + \lambda_1^2 \lambda_2^2 + \lambda_1 \lambda_2^3 + \lambda_1^2 \lambda_3 + \lambda_1 \lambda_2 \lambda_3 + \lambda_1 \lambda_2 \lambda_3
\]

In writing a computer routine for the expansion, the author found that the relationship found in the following lemma made the
program very easy to write.

Lemma (4.2). Define $\mu_{(k, 0)}'$ to be equal to 1 for all $k \geq 1$.

For all $k$ and $i$ greater than or equal to one, let $\mu_{(k, i)}'$ be the $i$-th raw moment of the distribution of $T_k = \theta T_k = \sum_{i=1}^{k} W_i'$, where the $W_i$ are independent, exponentially distributed with distinct parameters $\lambda_i \theta$. Then

$$\mu_{(k, i)}' = \mu_{(k-1, i)}' + \lambda_k \mu_{(k, i-1)}'. $$

Proof: For $k \geq 1 \ G_k^{(1)}(\theta t) = \lambda_k (G_{k-1}(\theta t)-G_k(\theta t))$. Then

$$G_k^{(m)}(0) = \lambda_k (G_{k-1}^{(m-1)}(0) - G_k^{(m-1)}(0)).$$

By Theorem 3.1 and for $m \geq k$ this equation is equivalent to

$$(-1)^{m-k} \prod_{i=1}^{k} \lambda_i \mu_{(k, m-k)}'$$

$$= \lambda_k ((-1)^{m-k} \prod_{i=1}^{k-1} \lambda_i) \mu_{(k-1, m-k)}' = (-1)^{m-k} \prod_{i=1}^{k-1} \lambda_i \mu_{(k, m-k-1)}'.$$

Dividing by $(-1)^{m-k} \prod_{i=1}^{k} \lambda_i$ and letting $m-k = i$, we get

$$\mu_{(k, i)}' = \mu_{(k-1, i)}' + \lambda_k \mu_{(k, i-1)}'. $$
We must note that the same problems inherent in the expansion of \( e^{-x} \) in a Maclaurin series occur in the expansion of \( \mathbb{E}(X(0,t)) \) since we are in fact expanding a sum of terms of the form \( K_i e^{-\lambda_i t} \).

For a set \( \{\lambda_i\} \) the individual values of which are relatively large, the number of terms in the expansion needed for convergence should be very large. It is worth noting that in our examples three situations do arise: (i) the \( \lambda_i \) decrease in \( i \); the initial transition parameter, \( \lambda_1 = 800 \), is fairly large; yet good results for the expectation are achieved using only the first five terms of the Maclaurin expansion; (ii) the \( \lambda_i \) decrease in \( i \); the initial transition parameter, \( \lambda_1 = 4950 \), is quite large; good results for the expectation are achieved using the first ten terms of the Maclaurin expansion; (iii) the \( \lambda_i \) increase in \( i \); \( \max\{\lambda_i\} = 240 \); good results for the expectation cannot be achieved using the first 40 terms in the Maclaurin expansion. The criterion of goodness was that at least 25 percent of \( \max\{\mathbb{E}(X(0,t))\} \) could be calculated using \( n \) terms in the series.

The reason that so relatively few terms are needed in cases (i) and (ii) is that the sum \( \sum_{k=1}^{N} G_{k}^{(m)}(0) \) which constitutes the coefficient for \( (\theta t)^m / m! \) in the series is composed of positive and negative terms thereby reducing the size of the coefficient and increasing the rate of convergence. The situation in case (iii) is due to the increasing \( \lambda_i \).

The effect of this on the size of the coefficients will be discussed.
further in the epidemic example.

**Results of Evaluating the Expectation and the Variance of the Truncated Pure Birth Process by Means of a Maclaurin Series**

In the following tables comparisons are made for particular models for Examples 1, 2, and 3. The exact expectation and variance were calculated using numerical methods on

\[
E(X(0, t)) = \sum_{k=1}^{N} G_k(\theta t)
\]

and

\[
V(X(0, t)) = \sum_{k=1}^{N} (2k-1)G_k(\theta t) - \left( \sum_{k=1}^{N} G_k(\theta t) \right)^2
\]

The approximations were found by using a routine which gave partial sums of the Maclaurin series up to 40 terms. The intent was to discover the maximum number of terms necessary for convergence since after a certain number of terms it is arguable whether anything is gained by not using the form \( \sum_{k=1}^{N} G_k(\theta t) \). When a reasonable convergence was not achieved by 40 terms, asterisks appear in the tables.

We will give the results for \( E(X(0, t)) \) and \( V(X(0, t)) \) in Examples 1, 2, and 3 in reverse order since Example 1 presents problems peculiar to the situation in which the set \( \{\lambda_i\} \) increases in \( i \).
Table 4. Example 3: the blowfly mating model: initial population
\[ M = 20, \quad N = 40, \quad \max \{E(X(0, t))\} = 20, \quad \max \{\lambda_i\} = \lambda_1 = 800. \]

<table>
<thead>
<tr>
<th>( \theta t )</th>
<th>Approximate ( E(X(0, t)) )</th>
<th>Exact ( E(X(0, t)) )</th>
<th>Approximate ( V(X(0, t)) )</th>
<th>Exact ( V(X(0, t)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>.777</td>
<td>.777</td>
<td>.733</td>
<td>.733</td>
</tr>
<tr>
<td>.005</td>
<td>3.483</td>
<td>3.483</td>
<td>2.652</td>
<td>2.652</td>
</tr>
<tr>
<td>.009</td>
<td>5.674</td>
<td>5.674</td>
<td>3.570</td>
<td>3.570</td>
</tr>
<tr>
<td>.013</td>
<td>7.480</td>
<td>7.480</td>
<td>3.962</td>
<td>3.962</td>
</tr>
<tr>
<td>.017</td>
<td>8.988</td>
<td>8.988</td>
<td>4.071</td>
<td>4.071</td>
</tr>
<tr>
<td>.021</td>
<td>10.263</td>
<td>10.263</td>
<td>4.024</td>
<td>4.024</td>
</tr>
<tr>
<td>.025</td>
<td>11.350</td>
<td>11.350</td>
<td>3.893</td>
<td>3.893</td>
</tr>
<tr>
<td>.029</td>
<td>12.286</td>
<td>12.286</td>
<td>3.689*</td>
<td>3.718</td>
</tr>
<tr>
<td>.033</td>
<td>13.097*</td>
<td>13.098*</td>
<td>*</td>
<td>3.522</td>
</tr>
<tr>
<td>.037</td>
<td>13.894*</td>
<td>13.806</td>
<td>*</td>
<td>3.319</td>
</tr>
</tbody>
</table>

For \( \theta t \leq .010 \), 5 terms are all that are needed for convergence.
For \( .010 < \theta t \leq .017 \), 10 terms are needed.
For \( \theta t > .033 \), 40 terms are insufficient.

It is worth noting now as a prelude to Example 1 that it is the
adding and subtracting of moments in any particular term that causes
such rapid convergence. Since \( \lambda_1 = 800, \lambda_2 = 741, \lambda_3 = 684, \lambda_4 = 629, \lambda_5 = 576 \),
the individual components in the coefficient of \((\theta t)^5\) are of the order \(10^{14}\) and \(10^{15}\), yet due to the subtraction of the
components, the actual coefficient is of order \(10^8\).

For Example 2, the \( r \)-th order kinetic model using combinatorial transition rates, three cases are examined: (i) the initial num-
ber of members in the parent population is 10; (ii) the initial number
of members in the parent population is 50; (iii) the initial number of
members in the parent population is 100. These particular cases
were chosen so that comparisons with tables presented by McQuarrie could be made.

Table 5. Example 2: \( r \)-th order kinetic model using combinatorial transition rates: initial population \( N = 10 \), \( r = 2 \), \( \max\{\mathbb{E}(X(0, t))\} = 5 \), \( \max\{\lambda_i\} = \lambda_1 = 45 \).

<table>
<thead>
<tr>
<th>( \theta t )</th>
<th>Approximate ( \mathbb{E}(X(0, t)) )</th>
<th>Exact ( \mathbb{E}(X(0, t)) )</th>
<th>Approximate ( \mathbb{V}(X(0, t)) )</th>
<th>Exact ( \mathbb{V}(X(0, t)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( .01 )</td>
<td>.415</td>
<td>.415</td>
<td>.352</td>
<td>.352</td>
</tr>
<tr>
<td>( .05 )</td>
<td>1.568</td>
<td>1.568</td>
<td>.786</td>
<td>.786</td>
</tr>
<tr>
<td>( .09 )</td>
<td>2.261</td>
<td>2.261</td>
<td>.772</td>
<td>.772</td>
</tr>
<tr>
<td>( .13 )</td>
<td>2.719</td>
<td>2.719</td>
<td>.693*</td>
<td>.692</td>
</tr>
<tr>
<td>( .17 )</td>
<td>3.044</td>
<td>3.044</td>
<td>.612*</td>
<td>.614</td>
</tr>
<tr>
<td>( .21 )</td>
<td>3.285</td>
<td>3.285</td>
<td>*</td>
<td>.547</td>
</tr>
<tr>
<td>( .25 )</td>
<td>3.472</td>
<td>3.472</td>
<td>*</td>
<td>.493</td>
</tr>
<tr>
<td>( .29 )</td>
<td>3.620</td>
<td>3.620</td>
<td>*</td>
<td>.448</td>
</tr>
<tr>
<td>( .33 )</td>
<td>3.728*</td>
<td>3.741</td>
<td>*</td>
<td>.395</td>
</tr>
</tbody>
</table>

For \( \theta t \leq .05 \), 5 terms are adequate for convergence.
For \( .05 < \theta t < .09 \), 10 are adequate for convergence of expectation.
For \( \theta t \geq .33 \), 40 terms are insufficient.

McQuarrie (1967) has tabulated the expectation for the unreacted population for initial \( N = 10 \), \( N = 50 \), \( N = 100 \) at five particular values as a function of \( \theta t \). McQuarrie used the random variable \( Y(0, t) \) equal to the size of the unreacted population at time \( \theta t \), and then tabled \( \frac{\mathbb{E}(Y(0, t))}{N} \) in terms of \( (N)(\theta t) \). Using the transformation \( 10 - 2\mathbb{E}(X(0, t)) = \mathbb{E}(Y(0, t)) \), we get the comparison with the approximate result.
Table 6. Example 2: \( r \)-th order kinetic model using combinatorial transition rates: initial population \( N = 50 \), \( r = 2 \), \( \max \{ E(X(0, t)) \} = 25 \), \( \max \{ \lambda_i \} = \lambda_1 = 1225 \).

<table>
<thead>
<tr>
<th>( \theta t )</th>
<th>( \text{Approximate} \ E(X(0, t)) )</th>
<th>( \text{Exact} \ E(X(0, t)) )</th>
<th>( \text{Approximate} \ V(X(0, t)) )</th>
<th>( \text{Exact} \ V(X(0, t)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.001</td>
<td>1.168</td>
<td>1.168</td>
<td>1.063</td>
<td>1.063</td>
</tr>
<tr>
<td>.005</td>
<td>4.927</td>
<td>4.927</td>
<td>3.236</td>
<td>3.236</td>
</tr>
<tr>
<td>.009</td>
<td>7.665</td>
<td>7.665</td>
<td>3.863</td>
<td>3.863</td>
</tr>
<tr>
<td>.017</td>
<td>11.382</td>
<td>11.382</td>
<td>3.860*</td>
<td>3.813</td>
</tr>
<tr>
<td>.021</td>
<td>12.700</td>
<td>12.700</td>
<td>*</td>
<td>3.617</td>
</tr>
<tr>
<td>.022</td>
<td>12.981*</td>
<td>12.991</td>
<td>*</td>
<td>3.565</td>
</tr>
<tr>
<td>.025</td>
<td>12.901</td>
<td>13.785</td>
<td>*</td>
<td>3.405</td>
</tr>
</tbody>
</table>

For \( \theta t \leq .004 \), 5 terms are adequate for convergence.
For \( .004 \leq \theta t \leq .01 \), 10 terms are adequate for convergence of expectation.
For \( \theta t \geq .022 \), 40 terms are insufficient.
For a comparison with McQuarrie's table, we have the following:

<table>
<thead>
<tr>
<th>(N)(t)</th>
<th>McQuarrie's Exact Result</th>
<th>Approximation Using at Most 40 Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>.803</td>
<td>.803</td>
</tr>
<tr>
<td>.50</td>
<td>.670</td>
<td>.670</td>
</tr>
<tr>
<td>.75</td>
<td>.576</td>
<td>.576</td>
</tr>
<tr>
<td>1.00</td>
<td>.504</td>
<td>.504</td>
</tr>
<tr>
<td>1.50</td>
<td>.404</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 7. Example 2: r-th order kinetic model using combinatorial transition rates: initial population N = 100, r = 2, \( \max\{E(X(0, t))\} = 50, \max\{\lambda_i\} = \lambda_1 = 4950. 

\( \theta t \)

<table>
<thead>
<tr>
<th>( \theta t )</th>
<th>Approximate ( E(X(0, t)) )</th>
<th>Exact ( E(X(0, t)) )</th>
<th>Approximate ( V(X(0, t)) )</th>
<th>Exact ( V(X(0, t)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0005</td>
<td>2.359</td>
<td>2.359</td>
<td>*</td>
<td>2.144</td>
</tr>
<tr>
<td>.0015</td>
<td>6.468</td>
<td>6.468</td>
<td>*</td>
<td>4.939</td>
</tr>
<tr>
<td>.0025</td>
<td>9.927</td>
<td>9.927</td>
<td>*</td>
<td>6.489</td>
</tr>
<tr>
<td>.0035</td>
<td>12.878</td>
<td>12.878</td>
<td>*</td>
<td>7.320</td>
</tr>
<tr>
<td>.0045</td>
<td>15.424</td>
<td>15.424</td>
<td>*</td>
<td>7.725</td>
</tr>
<tr>
<td>.0055</td>
<td>17.635*</td>
<td>17.644</td>
<td>*</td>
<td>7.873</td>
</tr>
<tr>
<td>.0055</td>
<td>18.493</td>
<td>19.596</td>
<td>*</td>
<td>7.866</td>
</tr>
</tbody>
</table>

For \( \theta t \leq .0015 \), 5 terms are adequate for convergence of expectation.
For \( .0015 < \theta t \leq .0045 \), 10 terms are adequate for convergence of expectation.
For \( \theta t \geq .0055 \), 40 terms are insufficient.
The comparison with McQuarrie's table is

<table>
<thead>
<tr>
<th>(N)((\theta t))</th>
<th>McQuarrie's Exact Result</th>
<th>Approximation Using at Most 40 Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>.801</td>
<td>.801</td>
</tr>
<tr>
<td>.50</td>
<td>.669</td>
<td>.669</td>
</tr>
<tr>
<td>.75</td>
<td>.574</td>
<td>*</td>
</tr>
<tr>
<td>1.00</td>
<td>.502</td>
<td>*</td>
</tr>
<tr>
<td>1.50</td>
<td>.402</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 8. Example 3: epidemic model: initial population N = 10, \(\text{max}\{E(X(0, t))\} = 5, \text{max}\{\lambda_i\} = \lambda_5 = 30.\)

\[
\begin{array}{cccc}
\theta t & \text{Approximate E}(X(0, t)) & \text{Exact E}(X(0, t)) & \text{Approximate V}(X(0, t)) & \text{Exact V}(X(0, t)) \\
.01 & .104 & .104 & .112 & .112 \\
.07 & .900 & .900 & 1.343 & 1.343 \\
.13 & 1.898 & 1.898 & 2.859 & 2.859 \\
.19 & 2.866 & 2.866 & 3.424* & 3.427 \\
.25 & 3.635 & 3.635 & * & 2.986 \\
.31 & 4.171 & 4.171 & * & 2.157 \\
.32 & 4.251* & 4.240 & * & 2.018 \\
\end{array}
\]

For \(\theta t \leq .05\), 5 terms are adequate for convergence.
For \(.05 < \theta t \leq .07\), 10 terms are adequate for convergence.
For \(\theta t \geq .32\), 40 terms are insufficient.
### Table 9. Example 3: epidemic model: initial population $N = 20$, \[ \max\{E(X(0, t))\} = 10, \max\{\lambda_i\} = \lambda_{10} = 110. \]

<table>
<thead>
<tr>
<th>$\theta t$</th>
<th>Approximate $E(X(0, t))$</th>
<th>Exact $E(X(0, t))$</th>
<th>Approximate $V(X(0, t))$</th>
<th>Exact $V(X(0, t))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.219</td>
<td>.219</td>
<td>.261</td>
<td>.261</td>
</tr>
<tr>
<td>.03</td>
<td>.784</td>
<td>.784</td>
<td>1.277</td>
<td>1.277</td>
</tr>
<tr>
<td>.06</td>
<td>2.010</td>
<td>2.010</td>
<td>4.455*</td>
<td>4.604</td>
</tr>
<tr>
<td>.09</td>
<td>3.654*</td>
<td>3.655</td>
<td>*</td>
<td>9.323</td>
</tr>
<tr>
<td>.12</td>
<td>5.506*</td>
<td>5.429</td>
<td>*</td>
<td>12.148</td>
</tr>
</tbody>
</table>

For $\theta t \leq .04$, 5 terms are adequate for convergence.  
For $.04 < \theta t \leq .06$, 10 terms are adequate for convergence of expectation.  
For $\theta t > .09$, 40 terms are insufficient.

### Table 10. Example 3: epidemic model: initial population $N = 30$, \[ \max\{E(X(0, t))\} = 15, \max\{\lambda_i\} = \lambda_{15} = 240. \]

<table>
<thead>
<tr>
<th>$\theta t$</th>
<th>Approximate $E(X(0, t))$</th>
<th>Exact $E(X(0, t))$</th>
<th>Approximate $V(X(0, t))$</th>
<th>Exact $V(X(0, t))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.345</td>
<td>.345</td>
<td>.453</td>
<td>.453</td>
</tr>
<tr>
<td>.03</td>
<td>1.375</td>
<td>1.375</td>
<td>2.916</td>
<td>2.916</td>
</tr>
<tr>
<td>.06</td>
<td>4.085*</td>
<td>4.057</td>
<td>*</td>
<td>13.714</td>
</tr>
<tr>
<td>.09</td>
<td>8.392</td>
<td>7.816</td>
<td>*</td>
<td>25.702</td>
</tr>
</tbody>
</table>

For $\theta t \leq .02$, 5 terms are adequate for convergence.  
For $.02 < \theta t \leq .03$, 10 terms are adequate for convergence of expectation.  
For $\theta t > .03$, one cannot use any more than 15 terms.
The truncated epidemic model presents interesting problems for the convergence of the Maclaurin power series. We have already commented that it is the summation of positive and negative components in any term which speeds the convergence. The expectation in the truncated epidemic model with $N = 30$ has the form

$$E(X(0, t)) = \sum_{m=1}^{\infty} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{15} G_k^{(m)}(0) \right).$$

From Lemma (3.2) and Theorem (3.1) we see that in $\sum_{k=1}^{15} G_k^{(1)}(0)$ only $G_1^{(1)}(0)$ is non-zero; in $\sum_{k=1}^{15} G_k^{(2)}(0)$, $G_1^{(2)}(0)$ is negative and $G_2^{(2)}(0)$ is positive; in $\sum_{k=1}^{15} G_k^{(3)}(0)$, $G_1^{(3)}(0)$ and $G_3^{(3)}(0)$ are positive and $G_2^{(3)}(0)$ is negative. Therefore as $m$ increases from one to fifteen, a new $G_k^{(m)}(0)$ becomes a non-zero component in the coefficient of $\frac{(\theta t)^m}{m!}$. For $m \geq 16$ each coefficient is comprised of 15 non-zero components.

We will use the expression "in control" to mean that

$$\sum_{k=1}^{15} \frac{(G_k^{(m)}(0))}{m!} \text{ is of at most order } 10^{m+1};$$

so that

$$\sum_{k=1}^{15} \frac{(G_k^{(m)}(0))}{m!} \text{ is at most of order } 10^{-m} \text{ is at most of order } 10.$$ Otherwise we say that

$$\sum_{k=1}^{15} \frac{(G_k^{(m)}(0))}{m!} \text{ is "out of control."}$$
Now because of the summation of positive and negative components in the coefficients \( \sum_{k=1}^{15} \frac{(G^{(m)}_k(0))}{m!}, \ 1 \leq m \leq 15 \), the coefficients \( \sum_{k=1}^{15} \frac{(G^{(m)}_k(0))}{m!} \) remain in control even though the individual components might be of much larger order. As an example we will look at \( \sum_{k=1}^{15} \frac{(G^{(5)}_k(0))}{5!} \) for the truncated epidemic model with \( N = 30 \). \( \sum_{k=1}^{15} G^{(5)}_k(0) \) is composed of the following five components:

\[
\frac{1}{5!} \lambda_1^4 (\lambda_4^4) \approx (2.025)(10)^5 \\
- \frac{1}{5!} \lambda_1 \lambda_2 (\lambda_1^3 + \lambda_2^3 + \lambda_2^2 + \lambda_1^2 \lambda_2) \approx (-5.44)(10)^6 \\
\frac{1}{5!} \lambda_1 \lambda_2 \lambda_3 (\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3) \approx (2.49)(10)^7 \\
- \frac{1}{5!} \lambda_1 \lambda_2 \lambda_3 \lambda_4 (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) \approx (-3.69)(10)^7 \\
\lambda_1 \lambda_2 \lambda_3 \lambda_4 \approx (1.71)(10)^7
\]

The sum of the coefficients, however, is approximately \((-5.94)(10)^4\) and so the coefficient of \((\theta t)^5\) is in control.

Now the coefficient for \((\theta t)^{15}\), the last coefficient in which a new component is introduced is approximately \((3.48)(10)^{13}\), also in control. The coefficient for \((\theta t)^{16}\), however is out of control, i.e., approximately equal to \((-3.042648329277622)(10)^{21}\).
What has happened is that there is no new component added into the sum to reduce the size of the coefficient. That this is the case can be seen by the fact that \( \prod_{j=1}^{15} \frac{\lambda_j}{15!} = (2.028432003401158)(10)^{20} \). If there were a \( \lambda_{16} \) approximately equal to \( \lambda_{15} = 240 \), the coefficient would be back in control. Let \( \lambda_{16} = \lambda_{15} + \Delta \) for some \( \Delta \).

\[
G_{16}'(0) = \prod_{j=1}^{16} \frac{\lambda_j}{16!} = (3.042648601737) + \frac{\Delta}{16} \prod_{j=1}^{15} \frac{\lambda_j}{15!};
\]

if we add this to the known coefficient

\[
\sum_{k=1}^{15} G_k'(0) = (-3.042648329277622)10^{21}
\]

we get approximately \((2.71459)(10)^{14} + \frac{\Delta}{16} \prod_{j=1}^{15} \frac{\lambda_j}{15!}\). If \( \Delta \) is small, the coefficient is back in control.

We can use this argument to justify taking only the partial sum of the first 15 terms in the Maclaurin expansion when the partial sum of the first 14 terms and the partial sum of the first 15 terms are approximately equal.

Let \( \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{15} \) be defined as in the process; let the auxiliary transition parameters \( \lambda_{16}, \lambda_{17}, \lambda_{18}, \ldots \) be introduced to keep the coefficients in control, i.e.,
\[ \lambda_{16} = \lambda_{15} + \Delta_{16} \]
\[ \lambda_{17} = \lambda_{16} + \Delta_{17} \]
\[ \lambda_{18} = \lambda_{17} + \Delta_{18} \]
\[ \vdots \]

and \( \lambda_i \neq \lambda_j \) for all \( i \) and \( j \). Call the c. d. f. for the new sum of the waiting times \( H_k(\theta t) \). Then for small \( \theta t \) such that

\[
\sum_{m=1}^{15} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{15} H_k^{(m)}(0) \right)
\]

is approximately equal to

\[
\sum_{m=1}^{\infty} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{15} H_k^{(m)}(0) \right)
\]

we note that only the first 15 \( \lambda_i \)'s are used in the sum on the left-handside which is to say that the new \( E(X(0, t)) \) depends only on

\[
\sum_{k=1}^{15} G_k(\theta t) = \sum_{k=1}^{15} \Pr(T_k \leq t)
\]

and hence

\[
E(X(0, t)) = \sum_{m=1}^{\infty} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{15} H_k^{(m)}(0) \right)
\]
is approximately equal to \[ \sum_{m=1}^{15} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{15} G_k^{(m)}(0) \right) \] and so we can ignore all terms in the expansion after \[ \sum_{k=1}^{15} \frac{(\theta t)^{15}}{15!} G_k^{(15)}(0). \]

In practice what was done was to observe the partial sums of the first 14 terms and the partial sum of the first 15 terms. When

\[ \epsilon < \left| \sum_{m=1}^{14} \left( \sum_{k=1}^{15} G_k^{(m)}(0) \frac{(\theta t)^m}{m!} \right) - \sum_{m=1}^{15} \left( \sum_{k=1}^{15} G_k^{(m)}(0) \frac{(\theta t)^m}{m!} \right) \right| \]

for some \( \epsilon \) and some \( \theta t \), we realized that we had only a poor approximation to \( E(X(0,t)) \).

The point of the comment is that a computer routine which gives only the sum of \( n \) terms and not the partial sums up to and including the first \( n \) terms will give misleading results. The problem is that for large \( \lambda_i \)'s the number of terms needed for convergence seems to be quite large, but for small \( \theta t \), the first \( N \) terms will converge and will converge to \( E(X(0,t)) \).
BIBLIOGRAPHY


APPENDICES
APPENDIX I

The purpose of this appendix is to trace the development of the asymptotic estimator 
\[ \hat{\theta}_t = \left( \frac{\ln(k+1)}{N} + \frac{(k-2)+2}{k \ln(k+1)} \right) \] 
for the truncated epidemic model with initial number of infectives, \( y \), equal to one and the initial number of susceptibles equal to \( N \).

We assume in this model that since for all \( i \) the \( \lambda_i \) increase as \( N \) increases, \( \hat{\theta}_t \) is in a small neighborhood of zero for very large \( N \). Further noting that the exact estimate for \( k = 1 \) is
\[ \hat{\theta}_t = \frac{1}{\lambda_1 - \lambda_2} \ln \frac{\lambda_1}{\lambda_2} = \frac{1}{N-2} \ln \left( \frac{2N-2}{N} \right), \] 
we further assume that for all \( k \geq 1 \) \( \hat{\theta}_t \) can be written in the form
\[ \hat{\theta}_t = \sum_{i=1}^{\infty} \frac{A_i(k)}{N} \] 
and that
\[ \lim_{N \to \infty} \frac{1}{N} = 0 \] 
for all \( i \).

Our problem then is to find a general form for \( A_1(k) \) and \( A_2(k) \) and use \( \hat{\theta}_t = \frac{A_1(k)}{N} + \frac{A_2(k)}{N^2} \) as an approximation to \( \hat{\theta}_t \). We will examine the calculations done for \( k = 2 \). The likelihood equation is
\[ \frac{N}{(N-2)(2N-6)} e^{-N\theta t} - \frac{(2N-2)}{(N-2)(N-4)} e^{-(2N-2)\theta t} \]
\[ + \frac{(3N-6)}{(2N-6)(N-4)} e^{-(3N-6)\theta t} = 0 \]
which can be simplified by multiplication by the least common denominator and division by \( e^{-N\theta t} \) to
\[ (N^2 - 4N) - (4N^2 - 16N + 12)e^{-(N-2)\theta t} + (3N^2 - 12N + 12)e^{-(2N-6)\theta t} = 0 \]

Assuming that \( \hat{\theta} t \) can be written as \( \sum_{i=1}^{\infty} \frac{A_i(k)}{N^i} \) and writing \( e^x \) in its Maclaurin expansion form, we get

\[ (A.1) \quad (N^2 - 4N) - (4N^2 - 16N + 12) \left( \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} (\sum_{i=1}^{\infty} \frac{A_i(k)}{N^i})^j \right) \]

\[ + (3N^2 - 12N + 12) \left( \sum_{j=0}^{\infty} \frac{(-2)^j}{j!} (\sum_{i=1}^{\infty} \frac{A_i(k)}{N^i})^j \right) = 0. \]

Treating the lefthandside of this equation as a polynomial in \( N \), we can write the equation as

\[ a_1 N^2 + a_2 N^1 + a_3 N^0 + a_4 N^{-1} + \ldots + a_j N^{-(j-3)} + \ldots = 0 \]

where \( a_i = \phi_i(A_1(k), A_2(k), A_3(k), \ldots) \). We may write this equation in the form \( a_1 N^2 + a_2 N^1 + a_3 + O(N) = 0 \). Considering the first two terms as an approximation to the lefthandside of the equation, we write \( a_1 N^2 + a_2 N + a_3 = 0 \) as an approximate likelihood equation. Since the equation is true for all \( N \), we argue that \( a_1 = a_2 = a_3 = 0 \).

Therefore expanding Equation (A.1) and collecting the coefficients of \( N \), we get
\[ 0 = a_1 = 1 + (-4)(1 - \frac{A_1(k)}{1!} + \frac{A_1^2(k)}{2!} - \frac{A_1^3(k)}{3!} + \ldots) \]
\[ + 3(1 - \frac{2A_1(k)}{1!} + \frac{(2A_1(k))^2}{2!} - \frac{(2A_1(k))^3}{3!} + \ldots) \]
\[ = 1 - 4 \left( \sum_{j=0}^{\infty} \frac{(-A_1(k))^j}{j!} \right) + 3 \left( \sum_{j=0}^{\infty} \frac{(-2A_1(k))^j}{j!} \right) \]
\[ = 1 - 4e^{A_1(k)} - 2A_1(k) \]
which implies \( A_1 = 0 \) or \( A_1 = \ln 3 \), the \( A_1(k) = 0 \) being part of the trivial solution \( \hat{a}_1 = 0 \) which implies \( A_i(k) = 0 \) for all \( i \).

Hence \( A_1 = \ln 3 \).

Continuing by collecting the coefficients of the terms in \( N^1 \) in the expansion of Equation (A.1), we get
\[ a_2 = -4 - 4(-A_2) \sum_{j=0}^{\infty} \frac{A_1^j(k)}{j!} + 2A_1(k) \sum_{j=0}^{\infty} \frac{A_1^j(k)}{j!} + 4 \sum_{j=0}^{\infty} \frac{A_1^j(k)}{j!} \]
\[ - 6(A_2(k)) \sum_{j=0}^{\infty} \frac{2A_1^j(k)}{j!} - 3A_1(k) \sum_{j=0}^{\infty} \frac{2A_1^j(k)}{j!} + 2 \sum_{j=0}^{\infty} \frac{2A_1^j(k)}{j!} \]
\[ = -4 + 4A_2(k)e^{-A_1(k)} - 8A_1(k)e^{-A_1(k)} + 16e^{-A_1(k)} - 6A_2(k)e^{-2A_1(k)} \]
\[ - 2A_1(k) - 2A_1(k) + 18A_1(k)e^{-2A_1(k)} - 12e^{-2A_1(k)} \]
Since $A_1(k) = \ln 3$, the equation simplifies to

$$a_2 = -4 + \frac{4}{3} A_2(k) - \frac{8}{3} \ln 3 + \frac{16}{3} - \frac{2}{3} A_2(k) + 2\ln 3 - \frac{4}{3}$$

or

$$\frac{2}{3} A_2(k) = \frac{2}{3} \ln 3$$

which implies

$$A_2(k) = \ln 3.$$

Continuing in this fashion for $k = 2, 3, 4, 5, 6$, we get the following results:

<table>
<thead>
<tr>
<th>k</th>
<th>$A_1(k)$</th>
<th>$A_2(k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\ln 3$</td>
<td>$\ln 3$</td>
</tr>
<tr>
<td>3</td>
<td>$\ln 4$</td>
<td>$1 + 2/3 \ln 4$</td>
</tr>
<tr>
<td>4</td>
<td>$\ln 5$</td>
<td>$2 + 2/4 \ln 5$</td>
</tr>
<tr>
<td>5</td>
<td>$\ln 6$</td>
<td>$3 + 2/5 \ln 6$</td>
</tr>
<tr>
<td>6</td>
<td>$\ln 7$</td>
<td>$4 + 2/6 \ln 7$</td>
</tr>
</tbody>
</table>

We note that these forms can be written for $k = 2, 3, 4, 5, 6$ as

$$A_1(k) = \ln (k+1) \quad \text{and} \quad A_2(k) = (k-2) + 2/k \ln (k+1).$$

Using the fact, as shown on page 17, that any function $h(k,N)$ is an asymptotic solution to the maximum likelihood Equation (1.1) if $\lim_{N \to \infty} h(k,N) = 0$, we form the asymptotic estimator.
\[ \tilde{\Theta}_t = \left( \frac{\ln(k+1)}{N} + \frac{(k-2) + 2/k \ln(k+1)}{N^2} \right) \]

after noting that for fixed \( k \) and \( t \)

\[
\lim_{N \to \infty} \tilde{\Theta}_t = 0.
\]
The purpose of this appendix is to make parallelisms between the pure birth process with distinct transition parameters and the pure birth process with a common transition parameter, that is the Poisson process. We will also demonstrate how the probabilities in the Yule process may be generated using \((1-e^{-\lambda \theta t})^k\) as the form for the distribution function of the sum of \(k\) waiting times.

As one might imagine all the main results of the thesis hold for the case where the waiting times are independent, identically distributed exponentially with common parameter \(\lambda \theta\). The main results as applied to the Poisson process are (1) the maximum likelihood estimate for \(t\) when \(k\) events are observed in \((0, t)\) is the mode of the density for the distribution of the sum of \((k+1)\) waiting times parameterized as a function of \(\theta t\); (2) if \(G_k(\theta t)\) is the c.d.f. for the transformed sum of \(k\) waiting times as a function of \(\theta t\), then for \(k \geq 1\)

\[
G_k^{(m)}(0) = 0 \quad \text{if} \quad m \leq k-1 \\
= \lambda^k \quad \text{if} \quad m = k \\
= \frac{(-1)^{m-k} \lambda^k}{(m-k)!} \mu'_{(k, m-k)} \quad \text{if} \quad m \geq k+1
\]

where \(\mu'_{(k, m-k)}\) is the \((m-k)\)-th raw moment of a gamma distribution with \(\alpha = k, \beta = \lambda\) (a comparable form exists, of course,
for the c.d.f. \( F_k(0) \) of the untransformed sum of the waiting times);

(3) the Maclaurin power series expansion of the moments can be used to give a closed form for the moments of the Poisson distribution.

Using Theorem (1.1) we write

\[
\Pr[(X(0,t)=k] = \Pr[T_k \leq t] - \Pr[T_{k+1} \leq t] = F_k(t) - F_{k+1}(t)
\]

where \( F_k(t) \) is the c.d.f. of the gamma distribution with parameters \( \alpha = k, \beta = 1/\lambda \). We note that

\[
\Pr[X(0,t)=k] = \left(\frac{\lambda \theta t}{k!}\right)^k e^{-\lambda \theta t} = \frac{1}{\lambda \theta} \frac{F^{(1)}(k)}{F_{k+1}(t)}
\]

and hence

\[
\Pr[(X(0,t)=k] = \frac{1}{\lambda \theta} F^{(1)}_{k+1}(t) = F_k(t) - F_{k+1}(t).
\]

If we reparameterize by means of \( T = \theta T \) and let \( G_k(\theta t) \) be the new c.d.f., we get

\[
\Pr[X(0,t)=k] = \frac{1}{\lambda} G^{(1)}_{k+1}(\theta t) = G_k(\theta t) - G_{k+1}(\theta t).
\]

Our first result then is that the \( \theta t \) which maximizes the likelihood function is the modal point of \( G_{k+1}(\theta t) \), namely \( \hat{\theta} t = \frac{k}{\lambda} \).

To show that for \( k \geq 1 \)
\[ G_k^{(m)}(0) = 0 \quad \text{if} \quad m \leq k - 1 \]
\[ = \lambda^k \quad \text{if} \quad m = k \]
\[ = \frac{(-1)^{m-k}}{(m-k)!} \lambda^k \mu_{(k,m-k)}' \quad \text{if} \quad m \geq k+1 \]

where \( \mu_{(k,m-k)}' \) is the \((m-k)\)-th raw moment of a gamma distribution with \( \alpha = k, \ \beta = \lambda \), we will first note that \( \mu_{(k,i)}' \) the \( i \)-th raw moment can be written as

\[ \lambda^i (k)(k+1)...(k+i-1) = \lambda^i \binom{k+i-1}{i} i! \]

For \( m \leq k \) we employ Lemma (3.2). For \( m \geq k+1 \), we employ an interesting proof suggested by Prof. H.D. Brunk. If one expands

\[ G_k^{(1)}(\theta t) = \frac{\lambda^k}{(k-1)!} (\theta t)^{k-1} e^{-\lambda \theta t} \]

in a Maclaurin series and then expands \( e^{-\lambda \theta t} \), one gets

\[ G_k^{(1)}(\theta t) = \sum_{r=0}^{\infty} G_k^{(r+1)}(0) \frac{(\theta t)^r}{r!} = \frac{\lambda^k}{(k-1)!} (\theta t)^{k-1} \sum_{n=0}^{\infty} \frac{(-1)^n (\lambda \theta t)^n}{n!} \]

\[ = \frac{\lambda^k}{(k-1)!} \sum_{n=0}^{\infty} \frac{(-1)^n (\theta t)^{n+k-1}}{n!} \lambda^n \]

For \( r \geq k \), comparing coefficients one gets
\[
G_k^{(k+1)}(0) = \lambda^k \frac{k!}{(k-1)!} (-1)^k
\]
\[
G_k^{(k+2)}(0) = \lambda^k \frac{(k+1)!}{(k-1)!} (-1)^2 \lambda^2
\]
and in general
\[
G_k^{(k+i)}(0) = \lambda^k \frac{(k+i-1)!}{(k-1)!} (-1)^i (\lambda)^i.
\]

While the cases for \( m \leq k \) could have been handled in the same manner, we wish to emphasize the parallelism between Chiang's lemma and the case in which the transition parameters are equal.

We note that for a fixed \( m \)

\[
\sum_{k=1}^{\infty} G_k^{(m)}(0) = \lambda^m (-1)^{n-1} \left( \binom{m-1}{m-1} - \binom{m-1}{m-2} + \binom{m-1}{m-3} + \ldots + \binom{m-1}{0} \right)
\]

If we note that by the expansion of the characteristic function of the Poisson distribution,

\[
\phi_{X(0,t)}(s) = \exp(\lambda \theta t e^{is} - 1)
\]

that the highest power of \((\lambda \theta t)\) in the \(r\)-th moment is \(r\) and if we note the expansion of \(E(X(0,t))^n\) in its Maclaurin series

\[
E(X(0,t))^n = \sum_{m=1}^{\infty} \frac{(\theta t)^m}{m!} \left( \sum_{k=1}^{\infty} (k^n-(k-1)^n)G_k^{(m)}(0) \right)
\]
then the implications are

\[ E(\mathbf{X}(0, t))^n = \sum_{i=1}^{n} (-1)^{i-1}(\lambda t)^i \left( \sum_{k=0}^{i-1} (-1)^k (k+1)(n-k^n)\right) \]

and

\[ (-1)^{m-1} \sum_{k=1}^{m-1} (-1)^{k-1}(k^n-(k-1)^n) = 0 \quad \text{for} \quad (m-1) \geq n > 0. \]

Feller (1968) makes the remark

\[ \sum_{k=0}^{j} (-1)^{j-k}{j \choose k}k^n = 0 \quad \text{if} \quad j > n \]

\[ = j! \quad \text{if} \quad j = n \]

and makes the suggestion of expanding \((1-\mathbf{e}^t)^n\) and considering the \(j\)-th derivative evaluated at \(t = 0\) in order to prove the remark.

From the remark it is an easy matter to see that the second part of the implication above is an application of the remark; the suggestion leads to a simpler form for the moments, i.e.,

\[ E(\mathbf{X}(0, t))^n = \sum_{i=1}^{n} (-1)^i \frac{(\lambda t)^i}{i!} \left( \sum_{k=0}^{i} (-1)^k {i \choose k}k^n \right) \]

The Yule Process with \(\lambda_1 = 1\lambda\)

We will use the Feller suggestion as motivation and since
\(D^n(1-e^{-t})^j = (-1)^n D^n(1-e^{-t})^j,\) we will assume for the moment that

\[D^n(1-e^{-t})^j = 0 \quad \text{if} \quad j < n\]
\[= j! \quad \text{if} \quad j = n\]

where \(D\) is the differential operator. Since this is the basic property of the distribution function of the sum of \(n\) waiting times for a particular set of transition parameters, we will consider

\[F_1(t) = (1 - e^{-\lambda t})\]

as the c.d.f. for \(T_1 = W_1\) where \(\lambda\) is the transition parameter for \(W_1\) and \(G_1(\theta t)\) is the c.d.f for the reparameterized \(T_1 = \theta T_1\). Then letting \(\{X(0, t) = k\}, \quad k \geq 0,\) be the event that \(k\) undefined occurrences have occurred in \((0, t),\) we define the probabilities for the events as

\[Pr[X(0, t) = 0] = 1 - F_1(t) = 1 - G_1(\theta t)\]

and

\[Pr[X(0, t) = k] = (F_1(t))^k - (F_1(t))^{k+1}\]
\[= (G_1(\theta t))^k - (G_1(\theta t))^{k+1}\]

for \(k \geq 1.\) We further define \((F_1(t))^k\) to be \(F_k(t)\) and \((G_1(\theta t))^k\) to be \(G_k(\theta t)\).

We note that...
\[
\Pr[X(0, t) = k] = (1 - G_1(\theta t)G_1(\theta t))^k, \quad k \geq 1
\]

\[
= (G^{(1)}_1(\theta t)G_1(\theta t))^k
\]

\[
= \frac{1}{(k+1)!} G^{(1)}_{k+1}(\theta t).
\]

If we expand \( G^{(1)}_1(\theta t)G_1(\theta t) \) we get

\[
\Pr[X(0, t) = k] = e^{-\lambda_0 t} \sum_{i=0}^{k} \binom{k}{i} e^{-\lambda_0 t} + \binom{k}{2} e^{-2\lambda_0 t} - \cdots + (-1)^{k} \binom{k}{k} e^{-k\lambda_0 t}
\]

\[
= \frac{k!}{0!k!} e^{-\lambda_0 t} - \frac{k!}{1!(k-1)!} e^{-2\lambda_0 t} + \cdots
\]

\[+ (-1)^{k} \frac{k!}{k!0!} e^{-(k+1)\lambda_0 t}
\]

If we write \( \lambda_1 = 1\lambda, \lambda_2 = 2\lambda, \ldots, \lambda_i = i\lambda, \ldots \) and note that

\[
\prod_{j \neq i} (\lambda_j - \lambda_i) = (k+1-i)!(i-1)! \lambda^k
\]

we can write

\[
\Pr[X(0, t) = k] = \sum_{i=0}^{k+1} \frac{\lambda_{1}\lambda_{2}\lambda_{3}\cdots\lambda_{i}}{k+1} e^{-\lambda_1 \theta t}
\]

We then recognize this as \( 1/((k+1)\lambda) \) times the density function of a general Erlang distribution with transition parameters \( \lambda_i = i\lambda \).

From our earlier work we know that
\[ G_k^{(m)}(0) = 0 \quad \text{if} \quad m \leq k - 1 \]

\[ = \lambda_k^k \quad \text{if} \quad m = k \]

and we see that Feller's remark is a particular example of Chiang's lemma with \( \lambda_i = i \).