CHAPTER I

INTRODUCTION

We open this chapter with a review of the history of branching processes and some fundamental and basic concepts to be used subsequently in the development of this thesis.

1.1 Historical Background

Until recently it was believed that the theory of branching processes seems first to have begun with the Galton-Watson Criticality Theorem (1873, 1874). However, the research work carried out by C.C. Heyde and E. Seneta (1972) established beyond any doubt the fact that its origin goes back to I. J. Bienaymé.

Irénée Jules Bienaymé was born in Paris on 28 August 1796 and died there on either 19 or 20 October 1878. He joined the civil service in 1820 and was appointed general inspector of finance in 1834. After the revolution of 1848 he retired and devoted all his time to scientific work.

An early work on the extinction of noble families in France, entitled "Mémoire Sur la durée des familles nobles"
de France'', was written by L. F. Benoist de Châteauneuf (1775 – 1856) and first read in two of the meetings of Mémoires de l' Académie royale des sciences morales et politiques de l' Institut de France : Comptes Rendus held on 31 August 1844 and 16 February 1845 (Kendall (1975), p.229). The work gives an account of the decline of some French noble families and of families of celebrities in the arts and sciences.

Undoubtedly, this paper of de Châteauneuf, among others, stimulated J.J. Bienaymé who treated the same problem mathematically in his paper ''De la loi de multiplication et de la durée des familles'' read out on 29 March 1845 (Kendall (1975)). Bienaymé's paper appears as an appendix in Kendall's paper. Both title of the paper and its opening paragraph reveal the similarity in motivation between him and Galton. It also shows that the correct statement of the Criticality Theorem was known to him:

If ..... the mean of the number of male children who replace the number of male of the preceding generation were less than unity, it would be easily realized that families are dying out due to the disappearance of the
numbers of which they are composed. However, the analysis shows further that when this mean is equal to unity families tend to disappear, although less rapidly

...... (quoted in Hayde and Seneta (1977), p.117)

In connection with Bienayme's methods, certain observations can be made. First, it is noticeable that he refers to a difference equation of the first order but of a degree equal to the maximum number of children. This is a reference to what would be written in the form:

\[ q_{n+1} = f(q_n) \]

...(1.1.1)

where \( q_n \) denotes the probability of extinction after \( n \) generation, and

\[ f(s) = \sum_{k=0}^{\infty} p_k s^k, \quad 0 \leq s \leq 1; \quad p_k \geq 0; \quad \sum_{k=0}^{\infty} p_k = 1 \]

is the probability generating function for the number of 'sons' to a 'potential father'.

Second, Bienayme was aware of the facts that \( q_n \) will increase monotonically as \( n \) increases, and that it will converge to a limit \( q \leq 1 \) as \( n \) tends to infinity, which limit will satisfy the equation
 Bienaymé's argument, so far, runs parallel to that of Watson, before a change in noticed.

The clue to the next section of Bienaymé's argument lies in an at first mysterious remark in which he says that when \( m > 1 \), then \( q \) is given by "the root of the equation \((1.1.2)\) which is less than unity" (Kendall (1975),p. 233). Bienaymé's paper, too, states that a population is not allowed by the branching process to remain in a stationary state, contrary to what authors of life tables suppose in their calculation.

In 1873, the Swiss Mathematician, de Candolle, who did not hear of Bienaymé's work, published his work, *Histoire des Sciences et des Savants Depuis Deux Siecles*, in which he pointed to the possibility of a probabilistic interpretation for the phenomenon of the extinction of a large numbers of noble families. In the same year, P. Galton gave the problem a precise formulation as problem 4001, which was published in the *Educational Times*, and in which he says:

A large nation, of whom we will only concern ourselves
with the adult males, \( N \) in number and who each bear separate surnames, colonize a district. Their law of population is such that, in each generation, \( a_0 \) percent of the adult males have no male children who reach adult life; \( a_1 \) have one such male child; \( a_2 \) have two; and so on up to \( a_5 \) who have five. Find (1) what proportion of the surnames will have become extinct after \( r \) generations; and (2) how many instances there will be of the same surname being held by \( n \) persons.

Galton, then, turned to his friend, W. Watson, who transformed the problem into one of iteration of generating functions
\[
f(s) = \sum_{j=0}^{\infty} p_j s^j, \quad p_j = a_j/100, \quad \text{i.e., the probability of a father begetting } j \text{ male children reaching adult life,}
\]
then Watson, in 1874, defines a sequence recursively by
\[
f_r = f_{r-1} \circ f. \quad \text{His conclusion is that the answer to the first question is the term independent of } s \text{ in } f_r(s) \text{ and gives the number of surnames with } k \text{ representatives in the } r \text{th generation as the coefficient of } s^k \text{ in } f_r(s) \text{ multiplied by } N.
\]
However, Watson’s solution contains an algebraic oversight and he incorrectly concludes that each family will eventually
die out with probability one. If Watson had read Schröder's work on functional iteration in *Mathematische Annalen* (1971), as it was suggested by Kendall (1966), he would have experimented with linear fractional generating functions and discovered that his last conclusion was wrong and lacked due deliberation.

Consequently, another half a century passed before the correct extinction probability was known, (for Bienaymé's work came to light only in 1972).

The Galton-Watson process seems to have been neglected for a long time. Only in 1922 did R. A. Fisher touch upon the topic in a genetical context and followed it up in 1930 to study random variations in frequencies of genes. By that time, and in 1927, J. B. S. Haldane had applied the model to genetics and roughly sketched a correct answer of the Criticality Theorem, namely that essentially the extinction probability is one exactly when the mean \( m = f'(1) \leq 1 \).

In 1929, the same problem was independently treated by the Danish Erlang in the *Matematisk Tidskrift*. Erlang's treatment of the problem shows that he realized as Watson had not, that
equation (1.1.2) can have two roots in the relevant interval 
[0,1], and that in fact there will be one root in [0,1] in 
addition to the root \( q = 1 \) if and only if the expected number 
of sons per parent, \( m \) is greater than unity. From some remarks 
of Steffensen, included in his "Deux problèmes du calcul des 
probabilités" (1933), it is reasonable to believe that Erlang, 
before his death, conjectured what is in fact the basic theorem 
of the subject: it is always the smallest root of (1.1.2) which 
is the appropriate one; thus extinction is almost certain for 
subcritical populations with \( m < 1 \) and for critical populations 
with \( m = 1 \), but there is always a positive chance of survival 
for supercritical populations with \( 1 < m \leq \infty \).

A clear and detailed proof of this theorem was made by 
J. F. Steffensen in 1930 and 1933. Commenting on Steffensen's 
effort, W. P. Elderton remarked that the probabilities \( p_k \) 
might in practice prove to be in geometric progression. Steffensen, 
in his turn, quickly realized that if we put

\[
p_0 = a, \quad p_k = (1-a) (1-\beta) \beta^{k-1} \quad (k = 1, 2, \ldots)
\]

then \( f \) will be a linear fractional function and the iterations
can be made explicit (Kendall (1966), p.389).

After reading Steffensen’s article in the *Matematisk Tidskrift*, A.J. Lotka (1931) applied the branching process theorem to the data contained in the 1920 United States census of white males obtaining $q = 0.88$ as the probability of the termination of the male line of descent from a newborn male.

Due mainly to the efforts of D. Hawkins and S. Ulam, T.E. Harris, and A.M. Yaglom, the final solution to the Galton-Watson process was successfully evolved between 1944 and 1950.

More details concerning the historical development of branching processes can be found in Harris (1963), Kendall (1966, 1975), Jagers (1975), and Heyde and Seneta (1972, 1977).

1.2 Markov Chain

The stochastic process $\{Z_n, n = 0, 1, 2, \ldots\}$ is called a Markov chain if, for $j, k, j_1, j_2, \ldots, j_{n-1} \in \mathbb{N}$ (or any subset of the set of all integers $\mathbb{N}$),

$$P(Z_n = k | Z_{n-1} = j, Z_{n-2} = j_{n-1}, \ldots, Z_1 = j_2, Z_0 = j_1)$$

$$= P(Z_n = k | Z_{n-1} = j) = P(j_k \text{ (say)})$$
Whenever the first member is defined.

The probability of $Z_n$ being in state $k$ given that $Z_{n-1}$ is in state $j$ is called one-step transition probability and denoted by $p_{jk}$.

The transition probability may or may not be independent of the time $n$. If $p_{jk}$ is independent of $n$, the Markov chain is said to be homogeneous (or to have stationary transition probabilities).

1.3 Bienaymé-Galton-Watson Branching Process:

Let the random variables $Z_0$, $Z_1$, $Z_2$, ... denote the size of (or the number of objects in) the 0th, 1st, 2nd, ... generations respectively. Let the probability that an object (irrespective of the generation to which it belongs) generates $k$ similar objects be denoted by $p_k$, where $p_k > 0$; $k = 0, 1, 2, ...$; $\sum_{k=0}^{\infty} p_k = 1$.

The sequence $\{ Z_n : n = 0, 1, 2, ... \}$ constitutes a Bienaymé-Galton-Watson (or simply BG) branching process with offspring distribution $\{ p_k \}$. 
Formally \( \{Z_n, n = 0, 1, 2, \ldots \} \) is a time homogeneous Markov chain with state space \( N \) and with transition probabilities

\[
P_{ij} = P(Z_{n+1} = j | Z_n = i) = \begin{cases} p_j^{i+1} & \text{if } i \geq 1, j \geq 0 \\ \\
\delta_{ij} & \text{if } i = 0, j \geq 0 \\ \end{cases}
\]

\( \delta_{ij} \) being the Kronecker delta and \( \{p_j^{i+1}, j = 0, 1, 2, \ldots \} \) being the \( i \)-fold convolution of \( \{p_j, j = 0, 1, 2, \ldots \} \). The transition probabilities satisfy

\[
P_{00} = 1 \text{ and } \sum_{j=0}^{\infty} P_{ij} s^j = (\sum_{j=0}^{\infty} P_{ij} s^j)^i, i \geq 1; 0 \leq s \leq 1.
\]

From the definition of \( \{Z_n\} \) as a Markov chain with a given transition function, we know from general considerations (the Kolmogorov theorem) that there is a probability space \( (\Omega, F, P) \) on which \( \{Z_n(w); n \geq 0\} \) are defined, and have the distributions determined by (1.3.1).

Throughout we shall assume that

(1) the process start with a single ancestor, i.e., \( Z_0 = 1 \);
(2) \( F \) is non-degenerate, i.e., \( p_k < 1 \) for all \( k \), and that \( P(Z_0 = o) < 1 \);
(3) \( p_0 + p_1 < 1 \).
It is clear from (1.3.1) that if \( Z_n = 0 \), then \( Z_{n+k} = 0 \) for all \( k \geq 0 \). Thus 0 is an absorbing state, and reaching 0 is the same as the process being extinct. All other states 1, 2, ... are transient, that is, \( Z_n \to \infty \) a.s. on \( B^c \) where

\[ E : = \{ Z_n = 0 \text{ eventually} \} \] is the set of extinction.

1.4 Probability Generating Function and Extinction Probability

An important and useful tool in deriving properties of the BGW branching process and of more sophisticated branching processes is the probability generating function (p.g.f.), and it will be the main object of attention in this thesis.

As we have mentioned in section 1.1, Watson noticed the very important fact that the p.g.f. for \( Z_n \) is the nth functional iterate of the p.g.f. for \( Z_1 \). That is, if

\[ f(s) = \sum_{k=0}^{\infty} p_k s^k, \quad |s| < 1 \] \hspace{1cm} (1.4.1)

is the p.g.f. for \( Z_1 \), and

\[ f_n(s) = \sum_{k=0}^{\infty} p(Z_n = k) s^k \] \hspace{1cm} (1.4.2)

is the p.g.f. for \( Z_n, n = 0, 1, 2, ... \), with \( f_0(s) = s \) and \( f_1(s) = f(s) \), then
\[ f_n(s) = f(f(\ldots(f(s))\ldots)), \]

Further

\[ f_n(s) = f(f_{n-1}(s)) = f_{n-1}(f(s)), \quad n=1,2,\ldots \quad (1.4.3) \]

In particular, setting \( s = o \) in (1.4.2), \( P(Z_n = o) = f_n(o) \).

Also if \( m = f'(l) = E(Z_1) < \infty \), then \( f_n(1) = E(Z_n) = m^n \), and

\[ \sigma^2 = f''(1) + f'(1)^2 = \text{Var} (Z_1) < \infty, \quad \text{then} \]

\[ \text{Var} (Z_n) = \begin{cases} 
\sigma^2 m^n (m^n - 1)/(m^n - m) & \text{if } m \neq 1 \\
\sigma^2 n & \text{if } m = 1
\end{cases} \]

Thus the variance of \( Z_n \) increases or decreases geometrically if \( m > 1 \) or \( m < 1 \) and linearly if \( m = 1 \).

The process is called subcritical, critical, supercritical or explosive, depending on whether \( m < 1, \ m = 1, \ 1 < m < \infty \), or \( m = \infty \).

Let \( s \) be real. From the definition of \( f \) as a power series with non-negative coefficients \( \{p_k\} \) adding to 1, and with \( p_0 + p_1 < 1 \), we see at once that:

(1) \( f \) is strictly convex and increasing in \([0,1]\).
(2) \( f(0) = p_0, f(1) = 1 \);

(3) if \( m \leq 1 \) then \( f(s) > s \) for \( s \in [0,1] \);

(4) if \( m > 1 \) then \( f(s) = s \) has a unique root in \([0,1]\).

Now, let \( q \) denote the probability of eventual extinction of the BGW branching process. Then

\[
q = P(\lim_{n \to \infty} Z_n = 0) = \lim_{n \to \infty} P(Z_n = 0) = \lim_{n \to \infty} f_n(0).
\]

Let \( q \) be the smallest root of \( f(s) = s \) for \( s \in [0,1] \). Then (1) - (4) imply that there is such a root, and furthermore,

**Lemma 1.4.1.** If \( m \leq 1 \) then \( q = 1, P(Z_n \to \infty) = 0 \) while if \( m > 1 \) then \( q < 1, P(Z_n \to \infty) > 0 \).

1.5 **Some Basic Theorems on BGW Branching Process**

Our purpose of this section is to give some useful limit theorems about \( Z_n \), and study the behaviour of \( Z_n \) when \( n \) is large. It has already been seen that the sequence \( \{Z_n\} \) either goes to \( \infty \) or goes to \( 0 \), it does not remain positive and bounded, even in case \( m = 1 \) (for the proofs of these theorems we refer to Athreya and Ney (1972) Ch. I, and Asmussen and Hering (1983), Ch. III).
It is imperative at this stage to define various modes of convergence of a sequence \( \{x_n\} \), \( n = 1, 2, \ldots \), to the random variable \( X \).

1. Convergence in probability means for each \( \epsilon > 0 \), we have
\[
\lim_{n \to \infty} P \left( |x_n - X| > \epsilon \right) = 0;
\]

2. Convergence in mean square means
\[
\lim_{n \to \infty} E \left| x_n - X \right|^2 = 0;
\]

3. Convergence with probability 1 means with probability 1 the \( \lim_{n \to \infty} x_n \) exists and is equal to \( X \), i.e.,
\[
\lim_{n \to \infty} x_n = X \text{ a.s.}
\]

For subcritical process, we have

**Theorem 1.5.1.** If \( o < m < 1 \), then
\[
E(s^{z_n} | z_n > 0) = \frac{f_n(s) - f_n(o)}{1 - f_n(o)} = k(s) \quad \text{as } n \to \infty,
\]

\( s \in [0, 1] \), where \( k(s) \) is the unique p.g.f. solution of
\[
l - k(f(s)) = n \ (l - k(s)); \quad k(o) = o.
\]

\[
(1.5.1)
\]
The mean of this distribution is \( k'(1) = \lim_{n \to \infty} \frac{m^n}{1 - f_n(c)} = \mu(f) \)

ew \( k'(1) > 1 \).

Note that \( P(Z_n > c) = 1 - f_n(c) \sim \frac{m^n}{\mu(f)} \).

If \( f''(1) < \infty \), then \( k'(1) \) and \( k''(1) \) are finite, we can differentiate (1.5.1) twice as \( s = 1 \) to get

\[
k''(1) = \frac{\mu(f) f''(1)}{m(1-m)}.
\]

For critical branching process, we have

**Lemma 1.5.1.** If \( m = 1 \) and \( \sigma^2 = \text{Var} Z_1 < \infty \), then

\[
\lim_{n \to \infty} \frac{1}{n} \left[ \frac{1}{1 - f_n(c)} - \frac{1}{1 - s} \right] = \frac{\sigma^2}{2},
\]

uniformly for \( 0 \leq s < 1 \).

Since \( P(Z_n > c) = 1 - f_n(c) \), thus we obtain (setting \( s = c \) in Lemma 1.5.1) the following estimate of the rate of convergence to \( o( \) first proved by Kolmogorov (1938)).

**Theorem 1.5.2.** If \( m = 1 \) and \( \sigma^2 < \infty \), then as \( n \to \infty \)

\[
1 - f_n(c) = P(Z_n > c) \sim \frac{2}{n \sigma^2}.
\]

Since every non-zero state is transient, then either
\( Z_n \to 0 \) or \( Z_n \to \infty \). In critical case \( Z_n \to 0 \) w.p.1. On the other hand the limit probabilities of the sequence of conditional distributions of \( \left\{ Z_n \mid Z_n > 0 \right\} \) are zero, and hence this process is divergent to \( \infty \). An idea as to the rate of divergence is given by a simple moment calculation:

\[
1 = E Z_n = E(Z_n \mid Z_n > 0), P(Z_n > 0) > 0, P(Z_n = 0)
\]
implying that

\[
E(Z_n \mid Z_n > 0) = \frac{1}{P(Z_n > 0)} \sim \frac{n f''(1)}{2} \quad \text{(by Theorem 1.5.2)}
\]
i.e., the mean of the conditional process is growing at the rate \( n \). Therefore, it is reasonable to cut the process down by a factor \( n \). This leads to the following:

**Theorem 1.5.3.** If \( p_1 \neq 1 \), \( m = 1 \) and \( \mu = \frac{1}{2} f''(1) < \infty \), then

\[
\lim_{n \to \infty} P \left( \frac{Z_n}{n} > \lambda \mid Z_0 = 1, Z_n > 0 \right) = e^{-\lambda/\mu}, \lambda > 0.
\]

This theorem was originally proved by Yaglom (1947) under a third moment assumption.

**Corollary 1.5.1.** Under the same conditions of Theorem 1.5.3
we have
\[
\lim_{n \to \infty} E \left( \frac{Z_n^2}{n} \mid Z_0 = 1, Z_n > 0 \right) = \mu.
\]

For supercritical branching processes, letting \( \mathcal{W}_n = \frac{Z_n}{n^\alpha} \),
n = 0, 1, 2, ..., \( \{ \mathcal{W}_n \} \) forms a Markov chain. We have
\[
E(\mathcal{W}_n^2) = \frac{1}{n^{2\alpha}} E(Z_n^2) = 1 + \frac{\sigma^2 (1 - \alpha \beta - \alpha \gamma)}{\alpha^2 \beta - \alpha \beta}.
\]

The branching processes satisfy the additive property,
namely, given \( Z_n = 1 \), the process \( \{ Z_{n+k} \mid k = 0, 1, 2, \ldots \} \)
is the sum of \( k \) independent copies of the process \( \{ Z_0 = 1, Z_1, Z_2, \ldots \} \). Using the Markov property this immediately yields
\[
E(Z_{n+k} \mid Z_n = i_n, Z_{n-1} = i_{n-1}, \ldots, Z_1 = i_1, Z_0 = i_0) = E(Z_k \mid Z_0 = 1) = i_n \cdot m^k, k=0,1,2,\ldots
\]
then we have
\[
E(\mathcal{W}_{n+k} \mid \mathcal{W}_n, \ldots, \mathcal{W}_1, \mathcal{W}_0) = E(\mathcal{W}_{n+k} \mid \mathcal{W}_n) = \mathcal{W}_n \text{ a.s.}
\]
\( k, n = 0, 1, 2, \ldots, \ldots \) \((1.5.2)\)
The first equality of \((1.5.2)\) holding because \( \mathcal{W}_0, \mathcal{W}_1, \ldots \) is
a Markov chain. Thus we get the following result:
Theorem 1.5.4. If $0 < m < \infty$, then $\left\{ W_n ; \mathcal{F}_n ; n = 0, 1, 2, \ldots \right\}$, where $\mathcal{F}_n$ is a $\sigma$-algebra generated by $Z_0, Z_1, \ldots, Z_n$, form a martingale. Furthermore, since $W_n \geq 0$, there exists a random variable $W$ such that

$$\lim_{n \to \infty} W_n = W \text{ a.s.} \quad \ldots \quad (1.5.3)$$

From (1.5.3) we see that $Z_n \sim m^n W$, this means that the population increases at a geometric rate, in accordance with the Malthusian law of growth. Although, Theorem 1.5.4 gives an interesting result under a very weak hypothesis it tell us nothing about $W$. It could be meaningful, if at all, only when $m > 1$ and moreover when $\sigma^2 < \infty$. We can then assert that $W$ is non-degenerate.

Theorem 1.5.5. If $m > 1$, $\sigma^2 < \infty$, and $Z_0 = 1$, then

1. $$\lim_{n \to \infty} \mathbb{E} (W_n - W)^2 = 0;$$
2. $$\mathbb{E} W = 1, \text{ Var } W = \sigma^2 / (m^2 - m);$$
3. $$P(W = 0) = q = P(Z_n = 0 \text{ for some } n).$$

The mean square convergence of $W_n$ to $W$ in the above theorem was first established by Harris (1948).
An application of Laplace-Stieltjes transform reveals something more about the random variable $W$.

Let $\phi_n(u) = \mathbb{E}(e^{-uW_n})$ and $\phi(u) = \mathbb{E}(e^{-uW})$, $u \geq 0$ be the Laplace-Stieltjes transform of the distribution functions of $W_n$ and $W$ respectively. Then writing $t_n = e^{-u/n}$,

$$
\phi_n(u) = \mathbb{E}(e^{-uW_n}) = \mathbb{E}(t_n^Z_n) = f_n(t_n)
$$

and

$$
\phi_{n+1}(mu) = \mathbb{E}(e^{-muZ_{n+1}/\sqrt{n+1}}) = \mathbb{E}(t_n^Z_{n+1})
$$

$$
= f_{n+1}(t_n) = f(f_n(t_n)) \text{ by (1.4.3)}
$$

i.e. $\phi_{n+1}(mu) = f(\phi_n(u))$.

Since the random variables $W_n$ converge in probability to $W$, their distributions converge to that of $W$ and $\phi_n(u) \to \phi(u)$ when $u \geq 0$. Therefore letting $n \to \infty$ in the last equation we find that the Laplace-Stieltjes transform of $W$ satisfies the fundamental equation

$$
\phi(mu) = f(\phi(u)), \quad u \geq 0
$$

with $\phi(0^+) = 1$ \quad \ldots (1.5.4)

The solution of (1.5.4) is unique up to a scale factor,
that is, if $\phi_1$ and $\phi_2$ are two solutions, then there is a constant $c$ such that $c < c < c^*$ and $\phi_1(u) = \phi_2(cu)$.


The distribution function of $W$ has a probability of mass $q$ at the origin and is absolutely continuous on $(0, \infty)$ with continuous positive density if $a > 1$ and $c^2 < c^*$ (Harris (1963), p. 16).

1.6 Ballman-Harris Branching Process

So far we have been considering branching processes $\{Z_n: n \geq 0\}$ in discrete time: an object after one unit of time produces, similar objects according to offspring distribution $\{P_k\}$. Now we proceed to consider a generalization such that the lifetimes of objects are i.i.d. random variables. Instead of the process $\{Z_n: n \geq 0\}$ we shall consider the process $\{Z(t); t \geq 0\}$, where $Z(t)$ equals the number of objects (or particles, individuals, organisms) at time $t$. The process $\{Z(t); t \geq 0\}$ may or may not be Markovian. If the lifetimes of objects are exponential random variables, then the process $\{Z(t); t \geq 0\}$ is a Markovian process. In this section however, we consider the general case where the lifetimes
of objects do not necessarily have exponential distributions.

Suppose that an object (ancestor) at time \( t = 0 \) initiates the process. At the end of its lifetime, it produces a random number of descendants according to the offspring distribution \( \{P_k\} \) (with p.g.f. \( h(s) \)). We assume that these descendants act independently of each other and that at the end of its lifetime, each one produces its own offspring with the same distribution \( \{P_k\} \), and that the process continues as long as objects are present. The lifetimes \( T \) are independent random variables with distribution function \( G(t) = P(T \leq t) \); object production is independent of the present state or past history of the process; and the lifetimes and object production variables are independent.

The stochastic process \( \{Z(t), t \geq 0\} \) is known as an age-dependent or general time branching process. This process is sometimes also referred to as the Bellman-Harris process, after Bellman and Harris first considered it in 1948.

As before, generating function of the process will continue to be a key to our analysis, and the present work will center around an integral equation satisfied by the generating function.
of $Z(t)$, where

$$F(s,t) = \sum_{k=0}^{\infty} P[Z(t) = k] s^k$$  \hspace{1cm} \text{...(1.6.1)}

To find $P[Z(t) = k]$, we shall condition on the lifetime $T$ at which the ancestor dies bearing $i$ offspring. We have

$$P[Z(t) = k] = \int_0^\infty P[Z(t) = k \mid T = u] G(u)$$

$$= \int_0^t P[Z(t) = k \mid T = u] G(u) + \int_t^\infty P[Z(t) = k \mid T = u] G(u)$$

In case of the second term, $u > t$. Given that $T = u$, the number of objects at time $t$ is then still 1 (the ancestor), and the second term yields $\delta_{kk}[1-G(t)]$, where $\delta_{jk}$ is the Kronecker delta. In case of the first term $u \leq t$, the ancestor dies at time $u \leq t$, leaving $i$ direct descendants; the probability of this is $p_i G(u)$, and further these $i$ descendants (who independently initiate processes at time $u$) leave $k$ objects in the remaining time $t-u$; the probability of this is

$$\sum_{i=0}^{\infty} p_i P_{i}^{*1}[Z(t-u) = k]$$

where $P_{i}^{*1}$ is the 1-fold convolution of $P$. Thus

$$P(Z(t) = k) = [1-G(t)] \delta_{kk} + \int_0^t G(t) \sum_{i=0}^{\infty} p_i P_{i}^{*1}[Z(t-u) = k].$$
Now multiplying through by $s^k$, summing over $k$, then the g.f. $F(s,t)$ of Bellman-Harris process satisfies the integral equation

$$F(s,t) = s \left[ 1-G(t) \right] + \int_0^t S(h \left[ F(s,t-u) \right], dG(u), |s| \leq 1$$

This integral equation cannot easily be solved in the general case. However, in particular when $G' (t) = b e^{-bt}$, we can see that this integral equation reduces to

$$F(s,t) = e^{-bt} + b t e^{-bt} \int_0^t S(h \left[ F(s,u) \right], e^{bu}, du.$$ 

whence

$$\frac{dF}{dt} = b \left[ h \left[ F(s,t) \right] - F(s,t) \right].$$

We conclude this section with the following observation about the expectation $M(t) = E(Z(t))$. Its asymptotic behaviour will, however, be studied in chapter IV.

The expectation $M(t)$ of a Bellman-Harris process $Z(t)$, $t \geq 0$, $Z_0 = 1$ satisfies the integral equation

$$M(t) = [1-G(t)] + m \int_0^t M(t-u) \cdot dG(u) \quad \ldots (1.6.2)$$

where $m = h'(1)$ is the mean of the offspring distribution. Further, if $m = 1$, then $M(t) = 1$ is a solution of (1.6.2).
To prove (1.6.2) we shall condition on the lifetime \( T \) of the ancestor, to get

\[
M(t) = E[Z(t)] = \int_0^\infty E[Z(t)|T=u] \, dG(u) = \int_0^t E[Z(t)|T=u] \, dG(u) \int_t^\infty E[Z(t)|T=u] \, dG(u)
\]

If \( u > t \), then the number of objects at time \( t \) is still 1, then \( E[Z(t)|T=u] = E(Z_0) = 1 \).

If \( u \leq t \), then the ancestor lives for time \( u \) at the end of which, it leaves 1 offspring with probability \( p_1 \), each of these offspring initiates a process, the number of objects of such a process having the same distribution as \( Z(t-u) \). Thus for \( u \leq t \)

\[
E[Z(t)|T=u] = \sum_{i=0}^{\infty} p_1 E[Z(t-u)] = m \cdot i(t-u).
\]

Then

\[
M(t) = m \int_0^t M(t-u) \, dG(u) + \int_t^\infty dG(u)
\]

Thus (1.6.2) follows, when \( m = 1 \), then the solution of (1.6.2) is obviously \( M(t) = 1 \).

1.7 Definition of Lattice Variable

A discrete random variable \( X \) is said to be a lattice
variable (or to have a lattice distribution) if it takes values \( c + nd \) \((n = 0, 1, 2, \ldots)\), where \(c\) and \(d\) are positive constants. The largest \(d\) is said to be the period of the distribution.

When \(c = 0, d = 1\), a lattice variable becomes an integer valued variable. The classification of random variables is given below:

```
Random Variable

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Lattice</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Integer valued</td>
</tr>
</tbody>
</table>
```

```
| Continuous               |
|                          |
| Non-lattice              |
|                          |
| Non-integer valued       |
```

The distribution is non-lattice when the corresponding random variable is other than lattice, that is, it is either discrete non-lattice or continuous.
1.8 Compound Distributions

In chapter II, we shall use distributions which are formed by "mixture" of discrete distributions. For discrete distributions "compounding" is commonly used in place of "mixing" and the resultant distributions are called compound distributions.

If \( \{ F_j(x_1, \ldots, x_n) \} \) \((j = \ldots -1, 0, 1, 2, \ldots)\) represent different (proper) cumulative distribution functions and if \( a_j \geq 0, \sum_{j=-\infty}^{\infty} a_j = 1 \), then

\[
P(x_1, \ldots, x_n) = \sum_{j=-\infty}^{\infty} a_j F_j(x_1, \ldots, x_n)
\]

also is a proper cumulative distribution function. This is called a mixture of the distributions \( \{ F_j \} \).

If \( F_j(x_1, \ldots, x_n) \) are of the form \( F(x_1, \ldots, x_n, \theta_1, \ldots, \theta_n) \).
The \( a_j \)'s may then be regarded as probabilities in a (discrete) joint distribution of the \( \theta \)'s. The idea may then be extended to suppose that the \( \theta \)'s have a joint continuous distribution with probability density function \( p(\theta_1, \ldots, \theta_s) \). Then the cumulative distribution function of the mixture is

\[
P(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta_1, \ldots, \theta_s) F(x_1, \ldots, x_n, \theta_1, \ldots, \theta_s) d\theta_1 \ldots d\theta_s.
\]
In either case (discrete or continuous distribution of θ's) we can write

\[ P(x_1, \ldots, x_n) = E_{\theta_1, \ldots, \theta_s} P(x_1, \ldots, x_n; \Theta_1, \ldots, \Theta_s). \]

Where the subscripts to \( E \) indicate that expectation is taken with respect to the variables \( \Theta_1, \ldots, \Theta_s \). Note that \( \Theta_1, \ldots, \Theta_s \) do not appear in the mixture distribution. Of course there may be other parameters in the \( F_j \)'s which are not eliminated by being integrated or summed out.

If the cumulative distribution function of a random variable is \( P(x \mid \Theta_1, \ldots, \Theta_k) \), depending on the \( k \) parameters \( \Theta_1, \ldots, \Theta_k \), then a 'compound' (or 'mixture') distribution is constructed by ascribing to some, or all, of the \( \Theta \)'s, a probability distribution. The new distribution has the cumulative distribution function \( E \left[ P(x \mid \Theta_1, \ldots, \Theta_k) \right] \), the expectation being taken with respect to the joint distribution of the \( \Theta \)'s.

It will be convenient to denote compound distributions in the symbolic form

\[ F_1 \wedge F_2 \]

\( \theta \)
where \( F_1 \) represents the origin distribution, \( \theta \) the varying parameters and \( F_2 \) the compounding distribution. Thus, for example

\[
\text{Poisson} (\theta) \wedge \text{Gamma} (\alpha, \beta)
\]

means a compound Poisson distribution by associating the gamma distribution with probability density function

\[
P_\theta(x) = \left[ \frac{\beta^\alpha}{\Gamma(\alpha)} \right] -1 x^{\alpha-1} e^{-\theta/x} \cdot \left( \frac{x}{\beta} \right)^{-\alpha} \quad \left( 0 < \theta, \alpha, \beta > 0 \right)
\]

to the expected value \( \theta \), of a Poisson distribution.