Chapter 5

Branching Processes

In this relatively short chapter we will discuss a particular type of infinite Markov chain called a branching process and study some of its properties. The type of branching process we study is sometimes called a “Galton-Watson process”. Sir Francis Galton was a very famous statistician who was interested in the probability that a given family name would go extinct (the sort of thing that statisticians of the day cared about). This stochastic process since then has found many applications in genetics and even in particle physics.

5.1 Basic Setup

Consider a population which consists of individuals that can produce offspring. Let \( X_n \) denote the number of members in the \( n \)th generation. Let \( Z_{in} \) equal the number of offspring produced by the \( i \)th member of the \( n \)th generation. Thus,

\[
X_{n+1} = \begin{cases} 
\sum_{i=1}^{X_n} Z_{in}, & \text{if } X_n \geq 1, \\
0, & \text{if } X_n = 0.
\end{cases}
\]

We assume that the \( Z_{in} \) are independent, and identically distributed with mean \( \mu \) and variance \( \sigma^2 > 0 \). It is often assumed that \( X_0 = 1 \). Denote the distribution of \( Z_{in} \) by:

\[ P_j = P(Z_{in} = j). \]

**Fact 5.1.** \( \{X_n : n \geq 0\} \) is a Markov chain with infinite state space.

In general, if \( P_j > 0 \) for \( j \geq 2 \) then it is clear that the Markov chain has an infinite state space. To see that it is Markov chain, let us consider the distribution of \( X_{n+1}|X_n = j \). This is the same as the distribution of,

\[
\sum_{i=1}^{j} Z_{in},
\]

where \( Z_{in} \) are all i.i.d. This does not depend in any way on the past \( \{X_{n-1}, \ldots, X_0\} \)
**Question 5.1.** Recall, \( P_0 = P(Z_{in} = 0) \). Assume \( P_0 > 0 \). Classify the states of the chain as recurrent or transient.

We know that the state \( \{0\} \) is absorbing and so is in its own class. For the rest of the states \( \{2,3,4,\ldots\} \) we note that they are all communicating and transient, since there is a non-zero probability that we will hit the absorbing state 0 and have a return probability \(< 1\).

### 5.2 Mean and Variance: Short/Medium term Behavior

A natural question is to characterize the distribution of the population after \( n \) time steps, i.e. after \( n \) generations. This turns out to be a bit difficult but computing the mean and variance is relatively straightforward.

**Fact 5.2.**

\[
E[X_n] = \mu^n.
\]

We will prove this via induction.

**Base Case:** Since we always take \( X_0 = 1 \), the base case is straightforward.

**Induction:** Observe that,

\[
E[X_n] = \sum_j E[X_n|X_{n-1} = j]P(X_{n-1} = j)
= \sum_j E\left[ \sum_{i=1}^j Z_{in} \right] P(X_{n-1} = j)
= \sum_j j\mu P(X_{n-1} = j) = \mu E[X_{n-1}] = \mu \times \mu^{n-1} = \mu^n.
\]

We can also calculate the variance:

**Fact 5.3.**

\[
\text{Var}(X_n) = \begin{cases} 
  n\sigma^2 & \text{if } \mu = 1, \\
  \sigma^2 \mu^{n-1} \left( \frac{1-\mu}{1-\mu} \right) & \text{if } \mu \neq 1.
\end{cases}
\]

Analogous to the law of total expectation, there is a law of total variance that tells us how to calculate the variance of a random variable by conditioning on another random variable, i.e.

\[
\text{Var}(X_n) = E[\text{Var}(X_n|X_{n-1})] + \text{Var}[E(X_n|X_{n-1})].
\]

Now, conditioning on \( X_{n-1} \) the mean of \( X_n \) is \( X_{n-1}\mu \) and the variance is \( X_{n-1}\sigma^2 \), so we obtain

\[
\text{Var}(X_n) = \sigma^2 E[X_{n-1}] + \mu^2 \text{Var}(X_{n-1})
= \sigma^2 \mu^{n-1} + \mu^2 \text{Var}(X_{n-1})
= \sigma^2 \left[ \mu^{n-1} + \mu^n \right] + \mu^4 \text{Var}(X_{n-2}).
\]

Continuing to unroll this recursion, and noting \( \text{Var}(X_0) = 0 \) we obtain that,

\[
\text{Var}(X_n) = \sigma^2 \left[ \mu^{n-1} + \mu^n + \ldots + \mu^{2n-2} \right],
\]

which is precisely the stated fact.
5.3 Probability of Dying Out: Long-term Behavior

Of interest is the probability that the population will eventually die out, if \( X_0 = 1 \). Denote this probability \( \pi_0 \). As one would expect, \( \pi_0 \) depends on the distribution \( Z \). Remarkably, however, whether or not \( \pi_0 = 1 \), depends only on the mean of \( Z \), denoted \( \mu \). In particular, \( \pi_0 = 1 \) if and only if \( \mu \leq 1 \).

**Fact 5.4.** If \( \mu < 1 \), then \( \pi_0 = 1 \).

To see this we note that,

\[
\mathbb{P}(X_n > 0) = \sum_{j=1}^{\infty} \mathbb{P}(X_n = j) \\
\leq \sum_{j=1}^{\infty} j \mathbb{P}(X_n = j) \\
= \mathbb{E}[X_n] = \mu^n.
\]

So the probability that \( X_n > 0 \) tends to 0 as \( n \to \infty \), so we conclude that the \( \pi_0 \to 1 \).

**Fact 5.5.** One key fact about a branching process is that if \( P_0 > 0 \) then either the population will die out or \( \to \infty \). Why?

Recall that we showed earlier that the state \( \{0\} \) is absorbing while the rest are transient. Let us focus our attention on states \( \{1, \ldots, k\} \) for some finite \( k \). Then since the states \( \{1, \ldots, k\} \) are transient we must eventually not return to them, i.e. there are two possibilities we get absorbed into 0 (i.e. the population dies out) or the population reaches \( k + 1 \). Since \( k \) was arbitrary this means that for any finite \( k \), the population dies out or grows beyond \( k \) (i.e. grows to \( \infty \)).
More generally, there are three regimes of interest:

- $\mu < 1$, population will eventually die out, i.e. $P(X_n = 0) \to 1$.

- $\mu = 1$, same as above, but we won’t prove this. Intuitively, the mean population size is 1 but the variance will keep growing so reasonable chance at some point the population will die off.

- $\mu > 1$, in this case $P(X_n = 0) < 1$, i.e. there is some chance the population size will converge to infinity.

**Fact 5.6.** In the case when $\mu > 1$, verify that the limiting probability $\lim_{n \to \infty} P(X_n = 0) = \pi_0$ must satisfy:

$$
\pi_0 = \sum_{j=0}^{\infty} \pi_0^j P_j.
$$

We rely on the usual trick of conditioning on the first step. So we have that,

$$
\pi_0 = \mathbb{P}(\text{ever dying out}|X_0 = 1) = \sum_j \mathbb{P}(\text{ever dying out}|X_0 = 1, X_1 = j) \mathbb{P}(X_1 = j|X_0 = 1)
$$

$$
= \sum_j \mathbb{P}(\text{ever dying out}|X_0 = 1, X_1 = j) P_j
$$

$$
= \sum_j \pi_0^j P_j,
$$

since if there are $j$ different individuals, their branches die out with probability $\pi_0$ independently, and so the whole population dies out with probability $\pi_0^j$.

**Fact 5.7.** Of course, $\pi_0 = 1$ always solves this equation. It turns out that when $\mu > 1$ there is always a value $\pi_0 < 1$ that satisfies this equation. The smallest (strictly) positive solution of this equation is the $\pi_0$ we are interested in.

You will prove this fact in your HW.

**Example 5.1.** If $P_0 = \frac{1}{2}$, $P_1 = P_2 = \frac{1}{4}$ then determine $\pi_0$.

We go through two steps, first we calculate the mean:

$$
\mu = 1 \times P_1 + 2 \times P_2 = \frac{3}{4} < 1,
$$

so we know that $\pi_0 = 1$.

**Example 5.2.** If $P_0 = \frac{1}{4}$, $P_1 = \frac{1}{4}$ and $P_2 = \frac{1}{2}$ then determine $\pi_0$. 

67
We again calculate the mean, to obtain:

\[ \mu = 1 \times \frac{1}{4} + 2 \times \frac{1}{2} = \frac{5}{4} > 1, \]

so we know that \( \pi_0 < 1 \), we write down the equation above:

\[ \pi_0 = \frac{1}{4} + \frac{1}{4} \pi_0 + \frac{1}{2} \pi_0, \]

which has two solutions \( \pi_0 = \{1/2, 1\} \). So we take the smallest positive solution: in this case it is 1/2.

Example 5.3. In each of the previous examples, if instead we had an initial population of size \( N \) (where \( N \) is some fixed number), what is the probability that the population would die out? i.e. what is the new \( \pi_0 \)?

In the first case, we would still have \( \pi_0 = 1 \), and in the second case we would need each of the \( N \) populations to die off, i.e. we obtain that \( \pi_0 = \frac{1}{2^N} \).