Extinction of Family Name

Student Participants:
Tianzi Wang
Jamie Xu

Faculty Mentor:
Professor Runhuan Feng
The purpose of this project is to give a brief introduction of the Galton-Watson-Bienaymé branching process and develop a numerical example for calculating the extinction probability of American male lines of decent.

Learning Objectives:
- Learn about Galton-Watson Binaymé branching process
- Develop examples of probability generating function
Suppose that one man starts a new family name. Let 

\[ z_n = \text{the number of sons in the n-th generation, } n = 0, 1, 2, \ldots \]

Then \( Z_0 = 1 \). Let \( X \) be a generic random variable standing for the number of sons from a male in the family and \( p_k \) be the probability that a male has \( k \) sons, i.e. \( P(x = k) = p_k \) for every \( k = 0, 1, 2, \ldots \)

The probability generating function (pgf) \( P(s) \) is defined by

\[
P(s) = P_X(s) = E(s^X) = \sum_{k=0}^{\infty} p_k s^k
\]
This family tree demonstrates the example, with green circles represent the sons and orange circles represent the daughters.
We shall write

\[ P_n(s) = \text{the pgf of } Z_n, \quad n = 0, 1, 2, \ldots \]

\[ m = E(X) = P'(1) = \text{the number of sons from a given male} \]

\[ \sigma^2 = V(X) = P''(1) + P'(1) - (P'(1))^2 = \text{the variance of the number of sons from a given male} \]

The stochastic process \( \{Z_n, n \geq 0\} \) is called the Galton-Watson-Bienaymé branching process.

Here are some useful properties of the pgf.

- If \( X \) and \( Y \) are independent, \( P_{X+Y}(S) = P_X(S)P_Y(S) \)

- If \( X_1, X_2, X_3, \ldots, X_n \) are independent random variables, and if \( S_n = X_1 + X_2 + X_3 + \ldots + X_n \)

  \[ P_{S_n}(s) = P_{S_1}(s)P_{S_2}(s) \ldots P_{S_n}(s) \]

- Let \( X_1, X_2, X_3, \ldots, X_n \) be i.i.d. random variables, and \( N \) be a random variable independent of the \( X_i's \). Let the random variable \( S_N = \sum_{k=1}^{N} X_k \), then

  \[ P_{S_N}(s) = P_N(P_X(s)) \]
We carefully examine how the n-th generation carry forward the family name to the n+1 generation. Let us label the males in the n-th generation by 1, 2, 3, ..., $Z_n$ and let $X_i$ be the number of sons from the male with the label i. Then the total number of males in the n+1 generation would be

$$Z_{n+1} = X_1 + X_2 + X_3 + ... + X_{Z_n}$$

By assumption, the $X_i$'s are independent of each other and of $Z_n$. Hence by definition, we have the fundamental equation

$$P_{n+1}(s) = P_n(P(s)) = P\left(P_n(s)\right), \quad n = 0, 1, 2, ...$$
Suppose that the probabilities $p_k'$s are given by
\[ p_k = pq^k, \quad k = 0, 1, 2, \ldots \]
Then we also have
\[ P(s) = \sum_{k=0}^{\infty} pq^k s^k = \frac{p}{1 - qs} \]
\[ P_2(s) = P(P_1(s)) = \frac{p}{1 - \frac{qp}{1 - qs}} = \frac{p(1 - qs)}{1 - qs - pq} \]

We can also use the fundamental equation to calculate moments of $Z_n$ for any $n$.
If $m < \infty$, then $E(Z_n) = m^n$ for all $n \geq 0$
If $\sigma^2 < \infty$, then $V(Z_n) = \begin{cases} \frac{\sigma^2 m^n (m^n - 1)}{m^2 - m}, & \text{if } m \neq 1 \\ \frac{n \sigma^2}{m^2 - m}, & \text{if } m = 1 \end{cases}$
Probability Distribution of the Number of Descendants

In this section, we apply the methodology developed in the previous section to actual data and calculate the probability distribution of descendants in a female line. We will show by example of how to compute the iterative function from the previous section into matrix form.
Define $c_j$ to be the proportion of women who have $j = 0, 1, \ldots$ children. Let $g$ define the proportion of births which are girls. Then, of the $c_1$ mothers with one child, the proportion of daughters is $gc_1$. Of the $c_2$ mothers with two children, we have $g^2c_2$ have two daughters, $2g(1-g)c_2$ with one daughter and $(1-g)^2c_2$ with no daughters. Table 1 shows the proportions that are divided into classes.

<table>
<thead>
<tr>
<th>Total Children</th>
<th>No. of Daughter</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$c_0$</td>
<td>$c_0$</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>$c_1$</td>
<td>$(1-g)c_1$</td>
<td>$c_1g$</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>$c_2$</td>
<td>$(1-g)^2c_2$</td>
<td>$2g(1-g)c_2$</td>
<td>$g^2c_2$</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>$c_3$</td>
<td>$(1-g)^3c_3$</td>
<td>$3g(1-g)^2c_3$</td>
<td>$3g^2(1-g)c_3$</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>$c_4$</td>
<td>$(1-g)^4c_4$</td>
<td>$4g(1-g)^3c_4$</td>
<td>$6g^2(1-g)^2c_4$</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 1. Distribution of women according to total children and number of daughters.
Table 2 summarizes the total female population in China by number of children. A frequently used measure in demographic studies is the sex ratio at birth (SRB) which is the number of boy infants compared to girl infants who are born within a given period usually represented by the number of boys per 100 girl infant. According to the Chinese 2000 Population Census, the SRB is 119.9.

Table 3 shows the distribution of women according to total number of children and number of daughters using Table 2 and the SRB.

<table>
<thead>
<tr>
<th>Number of Children</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Women</td>
<td>9,080,779</td>
<td>11,519,885</td>
<td>8,548,763</td>
<td>3,187,520</td>
<td>895,589</td>
<td>315,130</td>
</tr>
</tbody>
</table>

Table 2. China Female Population by Number of Children
*Source: United Nations Demographic Yearbook 2000
<table>
<thead>
<tr>
<th>Total Children</th>
<th>$c_i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0.27068</td>
<td>0.27068</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.34339</td>
<td>0.18723</td>
<td>0.15616</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.25482</td>
<td>0.07576</td>
<td>0.12637</td>
<td>0.05270</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.09501</td>
<td>0.01540</td>
<td>0.03854</td>
<td>0.03214</td>
<td>0.00894</td>
<td>0.00000</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.02670</td>
<td>0.00236</td>
<td>0.00787</td>
<td>0.00985</td>
<td>0.00548</td>
<td>0.00114</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.00939</td>
<td>0.00045</td>
<td>0.00189</td>
<td>0.00315</td>
<td>0.00263</td>
<td>0.00110</td>
</tr>
</tbody>
</table>

Table 3. Distribution of women according to total number of children and number of daughters.

*According to the SRB, $g = 100/(100+119.9) = 0.45475$

*$c_i = \frac{\text{number of women with } i \text{ children}}{\text{total number of women}}$
Probability of Eventual Extinction

- $P_0 = 0$
  Every generation has at least one descendant $\rightarrow$ probability of extinction = 0
  - $P_1 = 1$ : one descent every generation
  - $P_1 < 1$ : $Z_n \rightarrow$ infinite
Probability of Eventual distinction

- \( P_0 > 0 \)
  - \{Z_n=0\} is a subset of \{Z_{n+1} = 0 \}
  - \{\text{extinction}\} = \{ Z_n = 0 \text{ for some } n \geq 1 \} =
  - Continuity property

\[ \xi = P(\text{extinction}) = P \left( \lim_{n \to \infty} U \{ Z_k = 0 \} \right) = P \left( \lim_{n \to \infty} \{ Z_n = 0 \} \right) \]
\[ = \lim_{n \to \infty} P \{ Z_n = 0 \} \]
* \( P( Z_n = 0 ) \) increase as \( n \) increases
Probability of Eventual distinction

- \( P_0 > 0 \)
  - \( X_n = P_n(0) = P(Z_n = 0) \) for \( n \geq 0 \)
  - \( X_{n+1} = P_{n+1}(0) = P(P_n(0) = p(x_n)) \)
- Combine the continuity property and the property of \( X_{n+1} \)
  - \( \xi = \lim_{n \to \infty} X_{n+1} = \lim_{n \to \infty} P(X_n) = P(\xi) \)
  - \( \xi \) has to be the smallest root of \( P(S) = S \)