Galton-Watson Processes

M. Conceição Serra

December 17, 2004

1 Introduction

A Galton-Watson process is a sequence of r.v.'s $\{Z_n, n \in \mathbb{N}_0\}$ defined by

$$\begin{cases} Z_0 = 1 \\ Z_{n+1} = \displaystyle\sum_{i=1}^{Z_n} \xi_{n,i}, \quad n \geq 0, \end{cases} \tag{1}$$

where $\xi_{n,i}, n, i \in \mathbb{N}_0$ are i.i.d. r.v.'s distributed according to a probability function $\{p_k, k \in \mathbb{N}_0\}$. (A sum from one to zero is assumed to be zero.). To avoid trivialities, we assume that $\{p_k, k \in \mathbb{N}_0\}$ is not a degenerated, i.e, that $p_k \neq 1$, for all $k \geq 0$.

The process can be thought of as representing an evolving population of particles. It starts at time 0 with one particle that, after one unit of time, splits into a random number of children according to the reproduction law $\{p_k, k \in \mathbb{N}_0\}$. If we think of Z_1 as the r.v. that represents the size of the first generation, then it will be the total number of particles produced by the initial particle. These particles, after waiting one unit of time, go on to produce a second generation of particles and so on. Therefore, Z_2, the size of generation 2, will be the sum of Z_1 r.v.'s, $\xi_{1,1}, \xi_{1,2}, \ldots \xi_{1,Z_1}$, each of them with probability function $\{p_k, k \in \mathbb{N}_0\}$. The number of "offspring" produced by a single "parent" particle at any given time is assumed to be independent of the history of the process and of the other particles existing at the present. In general we have that Z_{n+1}, the size of generation $n+1$, can be written in form (1), i.e., as the sum of the "offspring" produced by the Z_n particles of the previous generation. Because of the role of $\{p_k, k \in \mathbb{N}_0\}$ in this description, from now on we will refer to it as the reproduction law.

We will prove that $\{Z_n, n \in \mathbb{N}_0\}$ is a Markov chain. If we denote by $\mathscr{B}_n = \sigma(Z_0, Z_1, \ldots, Z_n)$ the σ-algebra generated by $Z_0, Z_1, \ldots, Z_n$, then formulation (1) above gives, for all $k \geq 0$,

$$P[Z_{n+1} = k | \mathscr{B}_n] = P\left[\sum_{i=1}^{Z_n} \xi_{n,i} = k | Z_n\right]$$

proving that the process satisfies the Markov property and, since it runs in discrete time, it is a Markov chain with state space $\mathbb{N}_0$.

To complete the description of the chain we need to compute the transition probabilities. For $j \geq 1, k \geq 0$, we have

$$p_{jk} = P[Z_{n+1} = k | Z_n = j] = \sum_{i_1+i_2+\ldots+i_j=k} p_{i_1} p_{i_2} \ldots p_{i_j} = p_k^{*j}$$

where $\{p_k^{*j}, k \in \mathbb{N}_0\}$ denotes the j-fold convolution of $\{p_k, k \in \mathbb{N}_0\}$. For general values of k and j we have

$$p_{jk} = \begin{cases} p_k^{*j} & if \quad j \geq 1, k \geq 0 \\ 1 & if \quad j = 0, k = 0 \\ 0 & if \quad j = 0, k \geq 1 \end{cases} \tag{2}$$

Since these probabilities do not depend on n, $\{Z_n, n \in \mathbb{N}_0\}$ is a homogeneous Markov chain. If $p_0 > 0$, we see immediately that 0 is an absorbing state and all the other states are transient.

2 Additive property

One of the most important properties of a Galton-Watson process is the additive property. We give no proof but we will use it more or less explicitly.

Additive Property: *Let $\mathscr{B}(\mathbb{Z}_+^\infty)$ be the smallest σ-algebra of subsets of $\mathbb{Z}_+^\infty$ containing all finite dimensional sets and let also $\{Z_n^{(1)}\}, \{Z_n^{(2)}\}, \ldots,$ be independent Galton-Watson processes with the same reproduction law as $\{Z_n\}$. Then, for any $r \in \mathbb{N}$ and for any $A \in \mathscr{B}(\mathbb{Z}_+^\infty)$*

$$P[\{Z_n; n > r\} \in A | Z_r = k] = P\left[\left\{\sum_{j=1}^{k} Z_n^{(j)}; n \geq 1\right\} \in A\right].$$

This property tells us that, given that $Z_r = k$, the distribution of the process after generation r is the same as the distribution of the sum of k independent Galton-Watson processes with the same reproduction law.

In the definition of a Galton-Watson process we usually have $Z_0 = 1$. But this is not very realistic because we may have an initial population with more than one particle. If we assume that these particles reproduce independently of each other and according to the same reproduction law $\{p_k, k \in \mathbb{N}_0\}$, the additive property allows us to make the appropriate adjustments to the case $Z_0 = k$, with $k > 1$.

3 Generating functions and moments

An important tool in the analysis of this processes is the probability generating function (p.g.f.) of the reproduction law

$$f(s) = \sum_{k=0}^{\infty} p_k s^k, \qquad |s| \leq 1,$$

and its iterates

$$f_0(s) = s, \quad f_1(s) = f(s), \quad f_n(s) = \underbrace{(f \circ \ldots \circ f)}_{\text{n times}}(s),$$

where s can be complex in general, but here we assume it to be real. Sometimes we will take derivatives of p.g.f.'s in $s = 1$ but we obviously refer to the left-hand derivatives.

The first thing we will show is that the n-th iterate of f is the p.f.g. of Z_n. In fact, if we denote by ϕ_n the p.g.f. of Z_n, we have the following:

$$\begin{aligned}
\phi_{n+1}(s) &= \sum_{k=0}^{\infty} P[Z_{n+1} = k] s^k \\
&= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P[Z_{n+1} = k | Z_n = j] P[Z_n = j] s^k \\
&= \sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} P[Z_n = j] P[\xi_{n,1} + \ldots + \xi_{n,j} = k] \\
&= \sum_{j=0}^{\infty} P[Z_n = j] \sum_{k=0}^{\infty} P[\xi_{n,1} + \ldots + \xi_{n,j} = k] s^k
\end{aligned}$$

where the $\xi_{n,1}, \ldots, \xi_{n,j}$ are i.i.d. r.v.'s with p.g.f. f. It is well known that the sum of this r.v.'s has p.g.f. equal to $[f(s)]^j$ and it follows that

$$\forall n \geq 0,\ \phi_{n+1}(s) = \sum_{j=0}^{\infty} P[Z_n = j][f(s)]^j = \phi_n(f(s)) \tag{3}$$

Using induction, we will prove that

$$\forall n \geq 0,\ \phi_n(s) = f_n(s). \tag{4}$$

In fact, since $Z_0 = 1$ it is obvious that $\phi_0(s) = f_0(s) = s$. Suppose now that $\phi_n(s) = f_n(s)$ holds, then:

$$\phi_{n+1}(s) = \phi_n(f(s)) = f_n(f(s)) = f_{n+1}(s)$$

where the first equality follows from (3).

Using (3) and (4) we will compute the expectation and the variance of Z_n. We adopt the following notation:

$$m = E[Z_1] = f'(1) \quad \text{and} \quad \sigma^2 = Var[Z_1] = f''(1) + m^2 - m$$

and assume that these quantities are finite.

Differentiating (3) and setting $s = 1$ yields, for $n \geq 1$,

$$\phi'_n(1) = f'(1)\, \phi'_{n-1}(f(1)) = m\, \phi'_{n-1}(1).$$

Iteration produces

$$\phi'_n(1) = m^{n-1}\, \phi'_1(1).$$

Thus, $E[Z_n] = m^n$, for all $n \geq 0$.

To compute the variance, observe that

$$Var[Z_n] = \phi''_n(1) + \phi'_n(1) - [\phi'_n(1)]^2 = \phi''_n(1) + m^n - m^{2n}.$$

From (4), we have that $\phi_n(s) = f(\phi_{n-1}(s))$. Taking the second derivative of ϕ_n in point $s = 1$, we get the following relation:

$$\phi''_n(1) = f''(1)\, [\phi'_{n-1}(1)]^2 + f'(1)\, \phi''_{n-1}(1) = f''(1)\, m^{2(n-1)} + m\, \phi''_{n-1}(1)$$

Iteration produces

$$\begin{aligned} \phi''_n(1) &= f''(1)\{m^{2n-2} + m^{2n-3}\} + m^2\, \phi''_{n-2}(1) \\ &= f''(1)\, \{m^{2n-2} + m^{2n-3} + m^{2n-4}\} + m^3 \phi''_{n-3}(1) \\ &\vdots \\ &= f''(1)\, \{m^{2n-2} + m^{2n-3} + \ldots + m^{n-1}\} \end{aligned}$$

Thus,

$$\begin{aligned} Var[Z_n] &= (\sigma^2 + m^2 - m)\{m^{2n-2} + m^{2n-3} + \ldots + m^{n-1}\} + m(1 - m^n) \\ &= \begin{cases} \sigma^2 m^{n-1} \frac{m^n - 1}{m-1} & if \quad m \neq 1 \\ n\sigma^2 & if \quad m = 1 \end{cases} \end{aligned}$$

4 Probability of extinction

We now consider the problem that originated the study of Galton-Watson processes, the probability of extinction. This problem, posed by Francis Galton in 1873 and first studied by the Reverend Henry Watson, was in fact concerned with the extinction of family names in the British peerage.

The event *extinction* is defined in the following way

$$Q = \bigcup_{k=1}^{\infty} [Z_k = 0].$$

Since $[Z_k = 0]$, $k \geq 0$, is an increasing sequence of events, we have

$$\begin{aligned} P[Q] &= P\left\{\bigcup_{k=1}^{\infty}[Z_k=0]\right\} = \lim_{n\to\infty} P\left\{\bigcup_{k=1}^{n}[Z_k=0]\right\} \\ &= \lim_{n\to\infty} P[Z_n=0] = \lim_{n\to\infty} f_n(0) \end{aligned}$$

If we denote by q the probability of extinction, we have just proved that

$$q = \lim_{n\to\infty} f_n(0). \tag{5}$$

Our first theorem is the classical result on extinction of Galton-Watson processes and shows that the value of q depends on the value of m, the mean value of the reproduction law.

Theorem 4.1 *If $m \leq 1$ then $q = 1$. If $m > 1$, then q is the unique solution in interval $[0, 1)$ of equation*

$$s = f(s). \tag{6}$$

Proof:

The first thing we need to notice is that, from the definition of f as a power series with nonnegative coefficients $\{p_k\}$, we have that f is a strictly convex and increasing function in $[0, 1]$, except when $p_0 + p_1 = 1$. Therefore, unless stated otherwise, we will assume that $p_0 + p_1 < 1$.

In the deduction of (5), we have already seen that $\{f_n(0)\}_n$ is a non-decreasing sequence in n. Due to the continuity of f we also have

$$q = \lim_{n\to\infty} f_{n+1}(0) = \lim_{n\to\infty} f(f_n(0)) = f(\lim_{n\to\infty} f_n(0)) = f(q)$$

and this shows that q is always a solution of equation (6) in interval $[0, 1]$.

If $m \leq 1$, then $f'(s) < 1$ for $0 \leq s < 1$, and we can conclude that there exists $c \in (s, 1)$ such that $f(1) - f(s) = f'(c)(1 - s) < 1 - s$. Thus, $f(s) > s$ for $s \in [0, 1)$ and this shows that, if $m \leq 1$, the only possibility is $q = 1$.

If $m = f'(1) > 1$, then $f(s) < s$ in some left neighborhood of 1 since $f(1) = 1$. But, since $f(0) \geq 0$, there exists at least one solution of equation (6) in interval $[0, 1)$. We want to prove that this solution is unique. In fact, if there were two solutions, say $0 \leq s_0 < s_1 < 1$, then Rolle's theorem applied to the function $g(s) = f(s) - s$ would imply the existence of θ and η, $s_0 < \theta < s_1 < \eta < 1$, such that $g'(\theta) = 0 = g'(\eta)$ which is the same $f'(\theta) = 1 = f'(\eta)$. But this is impossible because f is strictly convex. To finish the proof we still need to show that q cannot be 1. In fact, if $q = \lim_{n\to\infty} f_n(0) = 1$ then $f_{n+1}(0) = f(f_n(0)) < f_n(0)$ if $f_n(0)$ were in a left neighborhood of 1, which contradicts the fact that $\{f_n(0)\}_n$ is a non-decreasing sequence.

To conclude, observe that if $p_0 + p_1 = 1$ then $f(s) = p_0 + (1 - p_0)s$ and, since we exclude the degenerated case, $m < 1$. In this case, the only solution of equation (6) is 1 and therefore $q = 1$. ■

In Theorem 4.1 we can see how the value of m plays an important role in the behavior of the process. Therefore there is a good reason to classify a Galton-Watson process as *subcritical* if $m < 1$, *critical* if $m = 1$ and *supercritical* if $m > 1$.

Instability is another interesting characteristic of Galton-Watson processes. The sequence $\{Z_n, n \in \mathbb{N}_0\}$ either goes to ∞ or goes to 0; it does not remains positive and bounded, even in critical case, as the next theorem shows.

Theorem 4.2 *Provided that* $p_1 < 1$, $P[Z_n \to \infty] = 1 - q$.

Proof:

If $q = 1$ then $[Z_n \to 0]$ a.s. and obviously $P[Z_n \to \infty] = 0$.

Using induction, it is possible to prove that, for $0 \leq q < 1$, $f'_n(q) = [f'(q)]^n$, $n \geq 1$. It obviously holds for $n = 1$ and, if $f'_n(q) = [f'(q)]^n$ then

$$f'_{n+1}(q) = \{f_n(f(q))\}' = f'(q)f'_n(f(q)) = f'(q)f'_n(q) = [f'(q)]^{n+1}.$$

If $0 < q < 1$, then for all $n, k \in \mathbb{N}$, we have

$$P[1 \leq Z_n \leq k] = \sum_{j=1}^{k} P[Z_n = j] \leq \sum_{j=1}^{k} jP[Z_n = j]\frac{q^{j-1}}{q^k} \leq \frac{f'_n(q)}{q^k} = \frac{[f'_n(q)]^n}{q^k}$$

Since $f'(q) < 1$ the series $\sum_{n=1}^{\infty}[f'(q)]^n$ converges and so does the series $\sum_{n=1}^{\infty} P[1 \leq Z_n \leq k]$. Using Borel-Cantelli lemma, it follows that the process cannot visit the interval $[1, k]$ infinitely often. Therefore the process can only visit state 0 infinitely often (happens with probability q) or it grows indefinitely, i.e.,$[Z_n \to \infty]$, and these are disjoint events.

If $q = 0$ then $p_0 = 0$ and, since $p_1 < 1$, the process is non-decreasing and so $Z_n \to \infty$ a.s. since $Z_n \to \infty$ in probability. ■

5 Stationary distributions

Theorem 4.2 indicates that this is not the kind of processes that can have a stationary distribution. In fact, we prove that, if $p_1 < 1$, the only stationary distribution is the trivial one $(1, 0, 0, \ldots)$.

Suppose there exists a non trivial stationary distribution $\pi = (\pi_0, \pi_1, \ldots)$. The numbers π_i, $i \geq 0$, satisfy the following conditions:

$$\pi_i \geq 0, \qquad \sum_{i=0}^{\infty} \pi_i = 1 \quad \text{and} \quad \pi_i = \sum_{j=0}^{\infty} \pi_j p_{ij} \tag{7}$$

where the p_{ij} are the transition probabilities introduced in (2). We split the proof in two cases:

(a) Suppose $p_0 > 0$. Then $p_{j0} > 0,\ j = 0, 1, 2, \ldots$. Since

$$\pi_0 = \pi_0 p_{00} + \pi_1 p_{10} + \ldots$$

and $p_{00} = 1$, we have that

$$\pi_1 p_{10} + \pi_2 p_{20} + \ldots = 0$$

and therefore the unique solution is $\pi_i = 0,\ i \geq 1$. From (7) it follows that $\pi_0 = 1$.

(b) Suppose $p_0 = 0$. Then, for $i < j,\ p_{ij} = 0$ and (7) yields

$$\pi_i = \pi_1 p_{1i} + \ldots + \pi_i p_{ii},\ i = 1, 2, \ldots.$$

Let π_n be the first of the numbers $\pi_1, \pi_2, \ldots$ that is positive. Then equation above is just $\pi_n = \pi_n p_{nn}$. But, due to the additive property, p_{nn} is the coefficient of s^n in the expansion of $[f(s)]^n$ and then $p_{nn} < 1$ (or we would be in the case $p_1 = 1$). Therefore we have $\pi_n = 0$ which contradicts the hypothesis. Hence $\pi_1, \pi_2, \ldots$ must all be zero and again $\pi_0 = 1$.

6 Some limit theorems

In this section we give a short (very short indeed...) presentation of some of the limiting theorems of Galton-Watson processes. We give no proofs of the theorems but we try to convince the reader of the importance of the results.

We start with the critical case. We have seen before that, in this case, for all $n \geq 0$, $E[Z_n] = 1$ and $Var[Z_n] = n\sigma^2 \to \infty$, as $n \to \infty$. Besides, Theorem 4.2 states that, if $p_1 < 1$, $Z_n \to 0$ a.s.. Therefore, although extinction is an almost sure event, we might expect the process to achieve very high values before extinction. The theorem we state below was first proved by Kolmogorov and gives the rate of convergence to 0 of the sequence $\{P[Z_n > 0]\}_n$.

Theorem 6.1 *If $m = 1$ and $\sigma^2 < \infty$ then*

$$\lim_{n\to\infty} nP[Z_n > 0] = \frac{2}{\sigma^2}.$$

Since

$$1 = E[Z_n] = E[Z_n | Z_n > 0]P[Z_n > 0] + 0P[Z_n = 0]$$

we get

$$E[Z_n | Z_n > 0] = \frac{1}{P[Z_n > 0]}$$

and Theorem 6.1 allows us to conclude that

$$E[Z_n|Z_n > 0] \backsim \frac{n\sigma^2}{2}$$

Since the mean of the conditioned process is growing at rate n, it is reasonable to divide the process by factor n. This leads to the next theorem that was first proved by Yaglom.

Theorem 6.2 *If $m = 1$ and $\sigma^2 < \infty$ then*

$$\lim_{n\to\infty} P\left\{\frac{Z_n}{n} > z|Z_n > 0\right\} = \exp\left\{-\frac{2z}{\sigma^2}\right\},\ z \geq 0.$$

For obvious reasons this result is also known as the exponential limit law for the critical process.

The subcritical processes have the least interesting behavior of the 3 kinds of processes. We state just one theorem that again concerns the rate of convergence to 0 of sequence $\{P[Z_n > 0]\}_n$.

Theorem 6.3 *If $m < 1$, then*

$$\lim_{n\to\infty} \frac{P[Z_n > 0]}{m^n} = \begin{cases} 0 & if \quad \sum k\, p_k \log k = \infty \quad or \quad p_0 = 1 \\ C & if \qquad otherwise \end{cases}$$

where C is a positive constant.

For the supercritical processes we will concentrate mainly in the study of the sequence $\left\{W_n = \frac{Z_n}{m^n}, n \in \mathbb{N}_0\right\}$. This sequence is a martingale with respect to $B_n = \sigma(Z_0, Z_1, \ldots, Z_n)$ and, since $W_n \geq 0$, there exists a r.v. W such that

$$\lim_{n\to\infty} W_n = W\ a.s. \tag{8}$$

In fact, this is true for $0 < m < \infty$. This is obviously a nice result but it does not tell us anything about W. Using Fatou's lemma we can conclude that

$$E[W] \leq \liminf E[W_n] = E[Z_0]$$

but this does not exclude the possibility that $W \equiv 0$, in which case we can only conclude that m^n is too big a normalizing factor. However, if $\sigma^2 < \infty$ then we can assert that W is not degenerated, as the next theorem shows. It also shows that convergence in (8) holds also in mean square.

Theorem 6.4 *If $m > 1$ and $\sigma^2 < \infty$, then:*

(i) $\lim_{n\to\infty} E[(W_n - W)^2] = 0$;

(ii) $E[W] = 1$ *and* $Var[W] = \frac{\sigma^2}{m^2-m}$;

(iii) $P[W = 0] = q$.

Now observe the following:

$$\begin{aligned} Z_n - m^n W &= Z_n - m^n \lim_{k\to\infty} \frac{Z_{n+k}}{m^{n+k}} \\ &= Z_n - \lim_{k\to\infty} \frac{Z_{n+k}}{m^k} =^d Z_n - \lim_{k\to\infty} \frac{1}{m^k} \sum_{j=1}^{Z_n} Z_k^{(j)} \\ &= \lim_{k\to\infty} \sum_{j=1}^{Z_n} \left\{ 1 - \frac{Z_k^{(j)}}{m^k} \right\} \end{aligned}$$

where the $\{Z_k^{(j)}\}$ are independent Galton-Watson processes with the same reproduction law as $\{Z_n\}$. Denoting by $W^{(j)} = \lim_{k\to\infty} \frac{Z_k^{(j)}}{m^k}$, we get

$$\frac{Z_n - m^n W}{m^n} =^d \frac{1}{m^n} \sum_{j=1}^{Z_n} (1 - W^{(j)}) \Leftrightarrow m^n (W_n - W) =^d \sum_{j=1}^{Z_n} (1 - W^{(j)})$$

and

$$\frac{m^n (W_n - W)}{\sqrt{Z_n}} =^d \frac{1}{\sqrt{Z_n}} \sum_{j=1}^{Z_n} (1 - W^{(j)}).$$

These calculations together with the variance computed in last theorem indicate that, conditionally upon $[Z_n \nrightarrow 0]$,

$$\frac{m^n (W_n - W)}{\sqrt{Z_n}} \to^d N(0, \frac{\sigma^2}{m^2 - m}).$$

And this finishes the sequence of results exposed here.

Bibliography:

Athreya, K.B. and Ney, P. (1972), *Branching Processes*, Springer-Verlag, Berlin

Harris, T.E. (1963), *The Theory of Branching Processes*, Springer-Verlag, Berlin

Jagers, P. (1975), *Branching Processes with Biological Applications*, Wiley, New-York.