Galton–Watson Processes

Ryan McCorvie[*]

October 23rd, 2017

This summarizes some of the basic results of Galton–Watson processe. This basically follows [Dur07] and [LP16]. These are relavent to Hawkes processes since a Hawkes process can be described in terms of a branching process.

1. Definition (Galton–Watson process): The *Galton–Watson* process is a Markov chain branching process with values $Z_n \in \mathbb{Z}_+$ for $n \geq 0$. The quantity Z_n represents the size of the nth generation of a family. Let X be a random variable on $\mathbb{Z}_+$ representing the number of offspring of an individual and let $p_k = P(X = k)$. Let $X_{n,i}$ be i.i.d. copies of X. Starting with $Z_0 = 1$, let

$$Z_{n+1} := \sum_{i=1}^{Z_n} X_{n+1,i}$$

The process Z_n is a model for the number of individuals in the nth generation, each of has offspring independently according to the distribution of X.

When $X \sim \mathsf{Poisson}(m)$ for some $m > 0$, the process is called a Poisson Galton–Watson process. As a warm-up let's calculate the distribution Y for the number of siblings an individual has. Let $m = EX$ be the mean of the offspring distribution.

$$\begin{aligned} P(Y = k) &= P\{ \text{ choose an individual from a family with k+1 children } \} \\ &= \frac{P(\text{ choose individual } \mid X = k+1)}{\sum_{i=0}^{\infty} P(\text{choose individual } \mid X = i+1)} \\ &= \frac{(k+1)p_{k+1}}{\sum_{k=0}^{\infty}(k+1)p_{k+1}} \\ &= m^{-1}(k+1)p_{k+1} \end{aligned}$$

Since the recurrance $mp_k = (k+1)p_{k+1}$ is satisfied only by the Poisson distribution, $X \sim Y$ iff $X \sim \mathsf{Poisson}(m)$.

An event of central importance is $E = \{Z_n = 0 \text{ for some } n\}$ and let $\rho = P(E)$. It follows by induction that once $Z_n = 0$ it remains there forever. The event E is called the *extinction event* and ρ is called the *extinction probability*.

2. Claim *On the event of non-extinction, $Z_n \to \infty$ a.s. provided $p_1 \neq 1$*

Proof. The only non-transient state of the Markov chain Z_n is 0. If $p_0 = 0$ then Z_n is non-decreasing and non-constant, so every finite state $k \geq 1$ is transient. If $p_0 > 0$ and any finite state $k \geq 1$ is non-transient, then Z_n returns to k only if it doesn't immediately go extinct, which is possible with positive probability $p_0^k > 0$. Since the probability of returning to k must be less $1 - p_0^k < 1$, k is transient. ∎

[*]mccorvie@berkeley.edu

1 Probability Generating Function

A key object for studying the Galton–Watson process is the *probability generating function* (p.g.f.)

(1) $$g(s) = E[s^X] = \sum_{k \geq 0} p_k s^k$$

This converges for any $s \in [0,1]$ and uniformly in $[0,r]$ for any $r < 1$ since the sum is dominated by the geometric series $\sum_{k \geq 0} s^k$.

3. Claim *The p.g.f. for Z_n satisfies*

$$g_n(s) = E[s^{Z_n}] = g \circ g \circ \cdots \circ g(s) = g^{(n)}(s)$$

Proof. Calculate

$$E\left[s^{Z_n} \,\middle|\, Z_{n-1}\right] = E\left[s^{\sum_{k=1}^{Z_{n-1}} X_{n,i}} \,\middle|\, Z_{n-1}\right] = \prod_{k=1}^{Z_{n-1}} E\left[s^{X_{n,i}}\right] = g(s)^{Z_{n-1}}$$

Therefore by the tower law

$$g_n(s) = E\left[E\left[s^{Z_n} \,\middle|\, Z_{n-1}\right]\right] = E\left[g(s)^{Z_{n-1}}\right] = g_{n-1}(g(s))$$

Since the base case satisfies $g_1(s) = Es^{X_{1,1}} = Es^X = g(s)$, by induction $g_n(s) = g^{(n)}(s)$ □

4. Corollary (Extinction probability) *The extinction probability satisfies* $\rho = \lim_{n\to\infty} g^{(n)}(0)$

Proof. Consider nested events $E_n = \{Z_n = 0\}$ which satisfy $E_{n+1} \supset E_n$ and $E = \bigcup_n E_n$. Since $E_n = g_n(0)$ it must be that $= g^{(n)}(0) \uparrow \rho$. □

5. Corollary (Extinction criterion) *Let $m = EX$ and assume that $p_1 < 1$ (that is, X is not trivially just 1).*

(i) $\rho = 1$ *iff* $m \leq 1$
(ii) The extinction probability is the smallest root $\rho = g(\rho)$ *in* $[0,1]$

Proof. Term-by-term differentiation of the p.g.f. to any order is valid on $[0,1)$. For example, for $r < 1$ the term-by-term derivatives of the p.g.f. converge uniformly on $[0,r]$ since the tail is dominated by $R_n = \sum_{k=n}^{\infty} kr^{k-1}$, which is arbitrarily small for large enough n. Similarly the tail of $\frac{d^k}{dr^k}\frac{1}{1-r}$ bounds the tail of the termwise kth derivative of the p.g.f. on $[0,r]$. Furthermore $g'(s) = E[Xs^{X-1}] \uparrow EX$ as $s \uparrow 1$ by monotone convergence.

Now $g(1) = E1^X = 1$ so $s = 1$ is a solution of $s = g(s)$. Furthermore g is non-decreasing and concave upward since its the convex combination non-decreasing concave upward functions $\{1, s, s^2, \ldots\}$.

Since by assumption $g(s) \neq s$ for some $s \in [0,1]$ (we exclude this case by the assumption X is not almost surely equal to 1), the convex upward expression $g(s) - s$ has at most two roots. If there are two roots, then the slope of $g(s) - s$ given by $g'(s) - 1$ must be positive at the larger root. Hence if $m \leq 1$ then $s = 1$ is the smaller of two roots or there is exactly one root.

Without loss of generality, we can assume $g(0) \in (0,1)$ in which case $g'(s) > 0$ for $s \in [0,1)$. For if $g(0) = 0$ then $\rho = 0$ and $g^{(n)}(0) = 0$ for all n so $g^{(n)}(0) = 0 \to 0$. In this case each individual has at least one descendent and some chance of more than one descendent, so its pretty obvious that exinction never happens. On the other hand if $g(0) = 1$ then $\rho = 1$ and $g_n(0) = 1$ for $n \geq 1$. In this case extinction is certain in the first generation.

Since g is non-decreasing and concave upward $U = [0, r)$ where r is the smallest root of $r = g(r)$. For any $s \in U$, $g(s) > s$ by definition and $g(s) < r$ since g is strictly increasing. Hence $g(U) \subset U$. The points $g^{(n)}(0)$ have a limit point in the compact set $\overline{U}$. Since $g(s) > s$ for any $s < r$, the only possible limit point is r. Hence $\rho = \lim_{n\to\infty} g^{(n)}(0)$ satisfies $\rho = g(\rho)$. ☐

In light of this result, the process is called super-critical, critical or sub-critical according to whether $m > 1$, $m = 1$ or $m < 1$

2 Martingale Techniques and Super-critical Growth

The preceding analyzed Z_n by studying the p.g.f. analytically. Using martingales convergene we can develop the same results and also study the limiting distributions of Z_n.

6. Claim *Let $m = EX$. The process $M_n := Z_n/m^n$ is a martingale.*

Proof. Note that

$$E[Z_{n+1} \mid Z_n] = E\left[\sum_{k=1}^{Z_n} X_{n+1,i} \,\middle|\, Z_n\right] = Z_n E[X] = mZ_n$$

Dividing both sides by m^{n+1} its clear $E[M_{n+1} \mid M_n] = M_n$. Since $M_n \geq 0$, $E|M_n| = EM_n = M_0 = 1 < \infty$. ☐

This immediately gives a simple formula for the average size of each generation, $EZ_n = m^n EM_n = m^n M_0 = m^n$.

7. Corollary *Assume X is not almost surely 1 (so that $p_1 < 1$). When $m \leq 1$ then $\rho = 1$.*

Proof. Note that $\sup E|M_n| = \sup EM_n = 1 < \infty$ so $\lim_{n\to\infty} M_n$ exists almost surely by the martingale convergence theorem. Therefore $\lim_n Z_n/m^n$ exists almost surely. Say the distribution of $\lim_n Z_n/m^n$ is given by a random variable W.

When $m < 1$ then Z_n decays exponentially in the limit and must tend to 0. In fact the probability of the event $\{Z_n > 0\}$ decays exponentially. Since $Z_n \geq 1$ whenever $Z_n > 0$,

$$P(Z_n > 0) \leq E[Z_n \mid Z_n > 0] = EZ_n = m^n \to 0$$

Another way to see the same thing is to consider the total number of descendents $Z = Z_1 + Z_2 + \ldots$. This has finite expectation $\sum_{k=0}^{\infty} m^k = 1/(1-m)$, which again shows $Z_n = 0$ eventually almost surely.

In critical case $m = 1$, Z_n is a martingale so $\lim_{n\to\infty} Z_n$ exists almost surely. But as claim 2 shows, only 0 is non-transient, so the limit must be 0. Interestingly, though $\rho = 1$, the total number of offsprong $Z = Z_0 + Z_1 + \ldots$ satisfies $EZ = EZ_0 + EZ_1 + \cdots = 1 + 1 + \cdots = \infty$. ☐

Now consider the case $m > 1$. As before, let W be the limiting distribution of M_n. Its possible that $P(W > 0) > 0$, in which case $\rho < 0$ since $Z_n = \mathcal{O}(m^n) \to \infty$ on this set. That is, Z_n grows exponentially whenever the limiting value of M_n is not 0. So, to understand the growth of the supercritical case, the key questions are

1. When is $W > 0$?
2. When $W = 0$ and the process doesn't become extinct, what is the growth rate?

A property is *inherited* if every finite tree has the property, and if whenever a tree has the property, so do the trees rooted at the children of this tree. Inherited properties have a 0-1 law

8. Claim *Every inherited property has conditional probability 0 or 1 given non-extinction*

Proof. Let A be the event that a tree has the property. Every finite tree has the property, so $P(A) \geq \rho$. On the other hand, calculate

$$P(A) = E[P(T \in A \mid Z_1)] \leq E[P(T^{(1)} \in A, \dots, T^{(Z_1)} \in A) \mid Z_1]$$

Since each of the children are i.i.d., this shows $P(A) \leq E[P(A)^{Z_1}] = g(P(A))$. In the regime that $P(A) \geq \rho$ the only way $P(A) \leq g(P(A))$ is if $P(A) = g(P(A))$ and therefore $P(A) \in \{\rho, 1\}$. □

9. Corollary *If $P(W > 0) > 0$ then $P(W = 0) = \rho$ and $\{ W > 0\} = \{Z_n > 0$ for all $\}$ almost surely.*

Proof. The property that $\{W = 0\}$ is inherited, so its probability equals ρ or 1. □

10. Claim *If $EX^2 < \infty$ then $P(W > 0) > 0$*

Proof. Calculate

$$\begin{aligned} E[(M_n - M_{n-1})^2 \mid M_{n-1}] &= m^{-2n} E[(Z_n - mZ_{n-1})^2 \mid Z_{n-1}] \\ &= m^{-2n} E\left[\left(\sum_{k=1}^{Z_{n-1}} (X_{n,i} - m) \right)^2 \,\middle|\, Z_{n-1} \right] \\ &= m^{-2n} Z_{n-1} \sigma^2 \end{aligned}$$

Where $\sigma^2 = \operatorname{var} X = E(X - m)^2$. Therefore the unconditional value of $E(M_n - M_{n-1}) = m^{-2n} m^n \sigma^2 = \sigma^2 / m^n$ Using the usual L^2 decompostiion for martingales

$$EM_n^2 = EM_0^2 + \sum_{k=1}^{n} E(M_k - M_{k-1})^2 = 1 + \sigma^2 \sum_{k=1}^{n} m^{-k}$$

Since the sum on the right converges as $n \to \infty$ we have $\sup_n EM_n^2 < \infty$. Therefore the margingale M_n converges to W in L^2 and hence in L^1. Consequently $EW = \lim_{n\to\infty} EM_n = 1$, and $P(W > 0) > 0$. □

Its actually sufficient that $EX \log^+ X < \infty$.

11. Theorem (Kesten-Stigum) *The following are equivalent when $m \in (1, \infty)$*

(i) $P[W = 0] = \rho$

(ii) $E[W = 0] = 1$

(iii) $E[X \log^+ X] < \infty$

In [AN72] this is proved by analyzing the convergence of a functional equation of the Laplace transform $\phi_n(u) = E[e^{-uM_n}]$. In the limit $\phi(u) = E[e^{-uW}]$ satisfies $\phi(ms) = g(\phi(s))$. In [LP16] the theorem is proved via a probabilistic argument involving size-biased Galton–Watson trees. Here is a less sharp result which shows that growth is almost exponential (even if $E[X \log^+ X] = \infty$).

12. Theorem (Seneta-Heyde) *If $m \in (1, \infty)$ then there exist constants c_n such that*

(i) $\lim Z_n / c_n$ *exists a.s. in* $[0, \infty)$

(ii) $P(\lim Z_n / c_n = 0) = \rho$

(iii) $c_{n+1}/c_n \to m$

Proof. For the proof we find another martingale. Starting with any $s_0 \in (\rho, 1)$ recursively define $s_{n+1} := g^{-1}(s_n)$ so that $s_n \uparrow 1$. The process $\langle s_n^{Z_n} \rangle$ is a margtingale by our choice of s_n. Since its positive and bounded it convertes a.s. and in L^1 to a limit Y with $EY = Es_0^{Z_0} = s_0$.

Rewriting these exponentials $c_n = -1/\log s_n$ gives $s_n^{Z_n} = e^{-Z_n/c_n}$ which shows $\lim Z_n/c_n$ exists a.s. in $[0, \infty]$. To get a handle on these constants, by l'Hôpital's rule

$$. \lim_{s\uparrow 1} \frac{-\log g(s)}{-\log s} = \lim_{s\uparrow 1} \frac{g'(s)s}{g(s)} = m$$

This shows (iii) since $s_n \uparrow 1$. The property $\lim Z_n/c_n = 0$ is inherited so by claim 8 and the fact $E[Y] < 1$, this property must have probability ρ, which shows (ii). Similarly the property $\lim Z_n/c_n < \infty$ is inherited and has probability 1 since $EY > \rho$ which shows (i). □

What happens when a supercritical process dies out? Essentially it behaves like a subcritical processs. Geometrically, the conditional p.g.f. is found by "zooming in" on the graph of g on the square $[0, \rho] \times [0, \rho]$.

13. Claim *A supercritical branching process conditioned to become extinct is a critical branching process. If $X \sim \mathsf{Poisson}(\lambda)$ then the conditional process is Galton–Watson with offspring distribution $\widehat{X} \sim \mathsf{Poisson}(\rho\lambda)$.*

Proof. Let $T = \inf\{n : Z_n = 0\}$ and let $\overline{Z}_n = (Z_n \mid T < \infty)$. The process $\overline{Z}_n$ is also a Markov chain, since the probability of eventual exintiction depends only on the current state, rather than the history. v. More explicitly

$$P(Z_{n+1} = z_{n+1}, T < \infty \mid Z_0 = z_0, Z_1 = z_1, \ldots, Z_n = z_n) = P(Z_n = n, T < \infty \mid Z_n = z_n)$$

Note that $h(k) = P(T < \infty \mid Z_n = k) = \rho^k$ since each individual must independently die out. This function is harmonic in the sense that $E[h(Z_{n+1}) \mid Z_n] = h(Z_n)$, which follows directly from the observation $E[h(Z_{n+1}) \mid Z_n] = g(\rho)^{Z_n} = \rho^{Z_n} = h(Z_n)$.

If $p(x, y)$ is the one-step transition function for Z_n, then Doob's h-transform gives the transition functin for $\overline{Z}_n$

$$\overline{p}(x, y) = \frac{h(y)}{h(x)} p(x, y) = \rho^{y-x} p(x, y)$$

In particular the offspring distribution for $\overline{Z}_n = \rho^{k-1} p_k$ and the p.g.f. is given by $\overline{g}(s) = \sum_{k=0}^{\infty} \rho^{k-1} p_k s^k = g(\rho s)/\rho$. By composition the n-step p.g.f. satisfies $\overline{g}_n(s) = \overline{g}^{(n)}(s) = g_n(\rho s)/\rho$. So the distribution of $\overline{Z}_n$ is exactly the distribution described above, of scaling the p.g.f. vertically and horizonally by ρ.

Finally if Z_n is a Poisson Galton–Watson process then $g(s) = \exp(\lambda(s-1))$ so

$$\overline{g}(s) = \frac{g(\rho s)}{\rho} = \frac{\exp(\lambda(\rho s - 1))}{\exp(\lambda(\rho - 1))} = \exp(\lambda\rho(s-1))$$

This is the p.g.f. of a Poisson Galton–Watson with parameter $\lambda\rho$ □

A similar h-transform conditioning on the event that each individual has an infinite line of descent has the effect of zooming in on the p.g.f in the range $[\rho, 1] \times [\rho, 1]$. In this case the conditional p.g.f. is given by

$$\overline{g}(s) = \frac{g((1-\rho)s + \rho) - \rho}{1 - \rho}$$

3 Limits in the Critical and Sub-Critical Cases

Analogous to the super-critical case, the decay is also exponential in the sub-critical case. This sharpens the previous result that $P(Z_n > 0) \leq m^n$.

14. Theorem (Heathcote, Seneta and Vere-Jones) *For any Galton–Watson process with $m \in (0, \infty)$ the sequence $\langle P(Z > 0)/m^n \rangle$ is decreasng. If $m < 1$ then the following are equivalent*

(i) $\lim_{n\to\infty} P(Z_n > 0)/m^n > 0$
(ii) $\sup E[Z_n \mid Z_n > 0] < \infty$
(iii) $E[X \log^+ X] < \infty$

Here's a "physicist" version with a stronger hypothesis.

15. Claim *Suppose $\sigma^2 = \operatorname{var} X < \infty$ and $m < 1$, then $\lim_{n\to\infty} P(Z_n > 0)/m^n > 0$*

Proof. Using the identity $P(Z_n > 0) = 1 - g_n(0)$, we will study the convergence of $1 - g_n(0) \to 0$ as $n \to \infty$. We know $g''(s) \in (0, \infty)$ for $s \in [0, 1)$ since its concave upward and also $g''(1) < \infty$. The quantity $g''(1) = \lim_{s\uparrow 1} g''(s) = EX(X-1) = \operatorname{var} X + m^2 - m < \infty$. Using the Taylor series with remainder at $s = 1$ and bounding $g''(s)$ by 0 and its maximum C we get a system of inequalities

$$1 - ms \leq g(1-s) \leq 1 - ms + \frac{1}{2}Cs^2$$

Let $\varepsilon_n = 1 - g_n(0)$. Since $\varepsilon_{n+1} = 1 - g(1 - \varepsilon_n)$ we can

$$m\varepsilon_n - \frac{C}{2}\varepsilon_n^2 \leq \varepsilon_{n+1} \leq m\varepsilon_n$$

By induction the right hand side implies $\varepsilon_n \leq m^n$ (an inequality we've already showed using martingales). Dividing through by $m\varepsilon_n$ and using this to get a lower bound for the left side results in

$$1 - \frac{C}{2}m^n \leq \frac{\varepsilon_{n+1}/m^{n+1}}{\varepsilon_n/m^n} \leq 1$$

So the ratios of the terms ε_n/m^n converge to 1, and the error decays exponentially. Since the errors sum, the product of these terms converges to a nonzero quantity, and which proves the limit we are looking for exists

$$\lim_{n\to\infty} \frac{\varepsilon_{n+1}/m^{n+1}}{\varepsilon_0/m^0} = \lim_{n\to\infty} \frac{P(Z_n > 0)}{m^n} = c > 0$$

□

Conditional on non-extinction at time n, there is a limiting distribution of Z_n on $\mathbb{Z}_+$. Contrast this with the pathwise martingale convergence to W, and also to the type of conditioning implicit in the h-transformation when we condition on extinction.

16. Theorem (Yaglom) *For each $k = 1, 2, \ldots$, $q_k = \lim_{n\to\infty} P(Z_n = k \mid Z_n > 0)$ exists and is a probability distribution. Moreover the p.g.f. of q_k given by $h(s) = \sum_{k=1}^{\infty} q_k s^k$ satisifes*

$$h(g(s)) = mh(s) + 1 - m$$

Proof. Let $h_n(s)$ be the p.g.f. for $P(Z_n = k \mid Z_n > 0)$. Then a moment's consideration shows that

$$h_n(s) = \frac{g_n(s) - g_n(0)}{1 - g_n(0)} = 1 - \frac{1 - g_n(s)}{1 - g_n(0)}$$

Because it is more amenable to analysis, let define $\widehat{h}_n := 1 - h_n(s) = \frac{1-g_n(s)}{1-g_n(0)}$. Also, for reasons evident in the next calculation, define $\eta(s) := \frac{1-g(s)}{1-s}$.

$$\frac{\widehat{h}_{n+1}(s)}{\widehat{h}_n(s)} = \frac{1-g_{n+1}(s)}{1-g_{n+1}(0)} \bigg/ \frac{1-g_n(s)}{1-g_n(0)} = \frac{1-g(g_n(s))}{1-g_n(s)} \bigg/ \frac{1-g(g_n(0))}{1-g_n(0)} = \frac{\eta(g_n(s))}{\eta(g_n(0))}$$

Now $\eta(s)$ is an increasing function of s because g is monotonic and concave upward, and η represents the slope of the secant line from $(s, g(s))$ to $(1,1)$. By induction because g is increasing and $s > 0$ we have $g_n(s) > g_n(0)$, so $\eta(g_n(s)) > \eta(g_n(0))$. This means that $\widehat{h}_n(s)$ is increasing in n and converges pointwise in $[0,1]$ to some function $\widehat{h}(s)$. Next calculate,

$$\widehat{h}_n(g(s)) = \frac{1-g_n(g(s))}{1-g_n(0)} = \frac{1-g_{n+1}(s)}{1-g_{n+1}(0)} \frac{1-g(g_n(0))}{1-g_n(0)} = \widehat{h}_{n+1}(s)\eta(g_n(0))$$

Now $\eta(s) \uparrow m < 1$ as $s \uparrow 1$ since it approaches the tangent $g'(1)$, so $\eta(g_n(0)) \to m$ as $n \to \infty$. We deduce the following functional equations

$$\widehat{h}(g(s)) = m\,\widehat{h}(s) \qquad \Longleftrightarrow \qquad h(g(s)) = mh(s) + 1 - m$$

Its immediate that $h(1) = h(g(1)) = 1 - m + mh(1)$, or $h(1) = 1$ which shows the limits of the probabilities sum to 1 and therefore h is a probability generating function.

I guess we have neglected to show that the limit of the generating functions is again a generating function. Its not true in general that the uniform limit of real analytic functions is real analytic. I think the key is the functional equation, which certainly has a unique (up to multiplicative) solution as a generating function. $\square$

The critical case is arguably the most interesting, but we'll satisfy ourselves with

17. Theorem (Kesten, Ney and Spitzer) *Suppose $m = 1$ and let $\sigma^2 = \operatorname{var} X$. Then*

(i) ***Kolmogorov's estimate*** $\lim_{n\to\infty} nP(Z_n > 0) = 2/\sigma^2$

(ii) ***Yaglom's limit law*** *If $\sigma < \infty$ then the conditional distribution of Z_n/n given $Z_n > 0$ converges as $n \to \infty$ to an exponential law with mean $\sigma^2/2$*

Proof. As noted before $\lim_{s\uparrow 1} g''(s) = \sigma^2 + m^2 - m$ which in this case is just σ^2. So using the Taylor expansion $g(1-\varepsilon) = 1 - \varepsilon + \frac{1}{2}\sigma^2\varepsilon^2 + \mathrm{o}(\varepsilon^2)$ we calculate for $s = 1 - \varepsilon$

$$\Delta_1(s) = \frac{1}{1-g(s)} - \frac{1}{1-s} = \frac{g(s)-s}{(1-g(s))(1-s)} \approx \frac{\frac{1}{2}\sigma^2\varepsilon^2 + \mathrm{o}(\varepsilon^2)}{(\varepsilon + \mathcal{O}(\varepsilon^2))\varepsilon} \to \frac{1}{2}\sigma^2 \text{ as } \varepsilon \to 0$$

Now let

$$\Delta_n(s) := \frac{1}{1-g_n(s)} - \frac{1}{1-g_{n-1}(s)} = \frac{1}{1-g(g^{(n-1)}(s))} - \frac{1}{1-g^{(n-1)}(s)} = \Delta_1(g^{(n)}(s))$$

Now since $g^{(n)}(s) \uparrow 1$ as $s \uparrow 1$ clearly $\lim_{s\uparrow 1} \Delta_n(s) = \lim_{s\uparrow 1} \Delta_1(s) = \sigma^2/2$. Moreover since $g^{(n)}(s) > g^{(n-1)}(s) > \ldots g(s) > s$, the convergence is uniform for all n. That is, the error term $|\Delta_1(s) - \sigma^2/2|$ controls all of the error terms $|\Delta_n(s) - \sigma^2/2|$. This means given a sequence $s_n \uparrow 1$ the following limit exists

$$\lim_{n\to\infty} \frac{1}{n(1-g_n(s_n))} - \frac{1}{n(1-s_n)} = \lim_{n\to\infty} \frac{\Delta_n(s_n) + \cdots + \Delta_1(s_n)}{n} = \frac{\sigma^2}{2}$$

If we write $\alpha := \lim n(1-s_n)$ (which can take any value in $[0,\infty]$) then we can reformulate this as

$$(*) \qquad \lim_{n\to\infty} n(1-g_n(s_n)) = \frac{1}{\sigma^2/2+\alpha^{-1}}$$

For part (i) consider

$$l = \lim_{n\to\infty} nP(Z_n > 0) = \lim_{n\to\infty} n(1-g_n(0))$$

But if we define $s_n = g_n(0)$, noting $2n \to \infty$ as $n \to \infty$ we can rewrite this equation

$$l = \lim_{n\to\infty} 2n(1-g^{(2n)}(0)) = \lim_{n\to\infty} 2n(1-g^{(n)}(g^{(n)}(0))) = 2\lim_{n\to\infty} n(1-g_n(s_n))$$

Comparing with $(*)$ we see the limit satisfies the following equation

$$l = \frac{2}{\frac{\sigma^2}{2}+l^{-1}} \qquad \Longrightarrow \qquad l = \frac{2}{\sigma^2}$$

For part (ii) consider the conditional Laplace transform of Z_n/n given by

$$\mathcal{L}_n(u) = E[e^{-uZ_n/n} \mid Z_n > 0] = \frac{g_n(e^{-u/n})-g_n(0)}{1-g_n(0)} = 1 - \frac{n(1-g_n(e^{-u/n}))}{n(1-g_n(0))}$$

By $(*)$ as $n \to \infty$ the numerator tends to $1/(\sigma^2/2+u^{-1})$ and by (i) the denominator tends to $2/\sigma^2$. Hence

$$\lim_{n\to\infty} \mathcal{L}_n(u) = 1 - \frac{\sigma^2/2}{\sigma^2/2+u^{-1}} = \frac{1}{u\sigma^2/2+1}$$

This is the Laplace transform of a continuous exponential distribution with mean $2/\sigma^2$ $\square$

4 Multi-type Galton–Watson

This section briefly states the analogous results for multi-type Galton–Watson processes. For proofs see [Har63]. In this process there are k types of individuals, and the one-generation transition probabilities $p^{(i)}(j_1,\dots,j_k)$ specify the probability that an individual i has j_1 offspring of type 1, j_2 of type 2, etc. For $\boldsymbol{j} \in \mathbb{Z}_+^k$, using the multi-index notation $\boldsymbol{s}^{\boldsymbol{j}} = s_1^{j_1}\dots s_k^{j_k}$ and the vector versions of $\boldsymbol{g}(\boldsymbol{s}) = (g^{(1)}(\boldsymbol{s}),\dots,g^{(k)}(\boldsymbol{s}))$ and $\boldsymbol{p}(\boldsymbol{j}) = (p^{(1)}(\boldsymbol{j}),\dots,p^{(k)}(\boldsymbol{j}))$, we can write the multitype p.g.f. as

$$\boldsymbol{g}(\boldsymbol{s}) = \sum_{\boldsymbol{j}\in\mathbb{Z}_+^k} \boldsymbol{p}(\boldsymbol{j})\boldsymbol{s}^{\boldsymbol{j}}$$

18. Definition (Multi-type Galton–Watson): A k-type Galton–Watson process is a Markov chain $\{\mathbf{Z_n} : \mathbf{n}=\mathbf{0},\mathbf{1},\dots\}$ on $\mathbb{Z}_+^k$ with transition function

$$P(\boldsymbol{i},\boldsymbol{j}) = P(\boldsymbol{Z}_{n+1} = \boldsymbol{j} \mid \boldsymbol{Z}_n = \boldsymbol{i}) = \text{ coefficient of } \boldsymbol{s}^{\boldsymbol{j}} \text{ in } [\boldsymbol{g}(\boldsymbol{s})]^{\boldsymbol{i}}$$

To emphasize the dependence on the initial distribution of particles $\boldsymbol{Z}_0 = \boldsymbol{i}$ sometimes we write $\boldsymbol{Z}_n^{(\boldsymbol{i})}$

The generating function satisfies the usual recursion $\boldsymbol{f}_{n+n'}(\boldsymbol{s}) = \boldsymbol{f}_n(\boldsymbol{f}_{n'}(\boldsymbol{s}))$. If $\boldsymbol{f}(\boldsymbol{s}) = \boldsymbol{M}\boldsymbol{s}$ where $\boldsymbol{M}$ is a $k\times k$ matrix, then the process is said to be singular. Each particle has a single offspring, and the process is equivalent to an ordinary finite Markov chain where the particle type corresponds to the Markov chain state. We'll assume the process is non-singular.

Let $m_{ij} = E\boldsymbol{Z}_{1,j}^{(i)}$ be the expected number of type j offspring starting with a single individual of type i. We assume the m_{ij} exist and define the *mean matrix* $\boldsymbol{M} = (m_{ij})_{i,j=1,\ldots,k}$. Analogous with the single-type case, have the following relations

$$m_{ij} = \frac{\partial g^{(i)}}{\partial s_j}(\mathbf{1}) \qquad EZ_{n,j}^{(i)} = \frac{\partial g_n^{(i)}}{\partial s_j}(\mathbf{1}) \qquad E[\boldsymbol{Z}_n \mid \boldsymbol{Z}_0] = \boldsymbol{Z}_0 \boldsymbol{M}^n$$

If there is an n such that $\boldsymbol{M}^n$ is strictly positive (every entry in the matrix is strictly positive) then the process $\boldsymbol{Z}_n$ is *positive regular*.

The role of the critical parameter m is now played by the spectral radius (maximum eigenvalue) $\lambda = ||\boldsymbol{M}||$.

19. Theorem *If $\{\boldsymbol{Z}_n\}$ is positive regular and nonsingular then*

$$P(\boldsymbol{Z}_n = \boldsymbol{j} \text{ infinitely often}) = 0$$

for any $\boldsymbol{j} \neq 0$

In other words, just as in the single-type case, 0 is the only non-transient state. This means that the process either tends to an infinite number of indifiduals for some type, or the process goes extinct. Let $\rho^{(i)}$ be the probability of eventual extinction of the process whose initial state is a single individual of type i. Then let $\boldsymbol{\rho} = (\rho^{(1)}, \ldots, \rho^{(k)})$.

20. Theorem *Assume $\{\boldsymbol{Z}_n\}$ is positive regular and non-singular. Let $\lambda = ||\boldsymbol{M}||$ be the maximum eigenvalue of $\boldsymbol{M}$.*

(i) If $\lambda \leq 1$ then $\boldsymbol{\rho} = \mathbf{1}$. If $\lambda > 1$ then $\boldsymbol{\rho} < \mathbf{1}$ (the vectorial inequality holds componentwise)

(ii) $\lim_{n\to\infty} \boldsymbol{f}_n(\boldsymbol{s}) = \boldsymbol{\rho}$

(iii) The only solution of $\boldsymbol{s} = \boldsymbol{f}(\boldsymbol{s})$ in $[0,1]^k$ is $\boldsymbol{\rho}$

Thus the process is supercritical, critical or subcritical according to whether $\lambda > 1$, $\lambda = 1$ or $\lambda < 1$

[AN72] Krishna B Athreya and Peter E Ney. *Branching Processes*. Springer Verlag, 1972.

[Dur07] Richard Durrett. *Random Graph Dynamics*. Cambridge University Press, 2007.

[Har63] Theodore E. Harris. *The Theory of Branching Processes*. Springer-Verlag, 1963.

[LP16] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, 2016.