BRANCHING PROCESSES

1. Galton-Watson Processes

Galton-Watson processes were introduced by Francis Galton in 1889 as a simple mathematical model for the propagation of family names. They were reinvented by Leo Szilard in the late 1930s as models for the proliferation of free neutrons in a nuclear fission reaction. Generalizations of the extinction probability formulas that we shall derive below played a role in the calculation of the critical mass of fissionable material needed for a sustained chain reaction. Galton-Watson processes continue to play a fundamental role in both the theory and applications of stochastic processes.

First, an informal desription: A population of individuals (which may represent people, organisms, free neutrons, etc., depending on the context) evolves in discrete time $n = 0, 1, 2, \dots$ according to the following rules. Each nth generation individual produces a random number (possibly 0) of individuals, called *offspring*, in the $(n+1)$st generation. The offspring counts $\xi_\alpha, \xi_\beta, \xi_\gamma, \dots$ for distinct individuals $\alpha, \beta, \gamma, \dots$ are mutually independent, and also independent of the offspring counts of individuals from earlier generations. Furthermore, they are identically distributed, with common distribution $\{p_k\}_{k\geq 0}$. The *state* Z_n of the Galton-Watson process at time n is the number of individuals in the nth generation.

More formally,

Definition 1. A Galton-Watson process $\{Z_n\}_{n\geq 0}$ with offspring distribution $F = \{p_k\}_{k\geq 0}$ is a discrete-time Markov chain taking values in the set $\mathbb{Z}_+$ of nonnegative integers whose transition probabilities are as follows:

$$P\{Z_{n+1} = k \,|\, Z_n = m\} = p_k^{*m}. \tag{1}$$

Here $\{p_k^{*m}\}$ denotes the m−th convolution power of the distribution $\{p_k\}$. In other words, the conditional distribution of Z_{n+1} given that $Z_n = m$ is the distribution of the sum of m i.i.d. random variables each with distribution $\{p_k\}$. The default initial state is $Z_0 = 1$.

Construction: A Galton-Watson process with offspring distribution $F = \{p_k\}_{k\geq 0}$ can be built on any probability space that supports an infinite sequence of i.i.d. random variables all with distribution F. Assume that these are arranged in a doubly infinite array, as follows:

$$\begin{array}{l} \xi_1^1, \xi_2^1, \xi_3^1, \cdots \\ \xi_1^2, \xi_2^2, \xi_3^2, \cdots \\ \xi_1^3, \xi_2^3, \xi_3^3, \cdots \\ \text{etc.} \end{array}$$

Set $Z_0 = 1$, and inductively define

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{n+1}. \tag{2}$$

The independence of the random variables ξ_i^n guarantees that the sequence $(Z_n)_{n\geq 0}$ has the Markov property, and that the conditional distributions satisfy equation (1).

□

For certain choices of the offspring distribution F, the Galton-Watson process isn't very interesting. For example, if F is the probability distribution that puts mass 1 on the integer 17, then the evolution of the process is purely deterministic:

$$Z_n = (17)^n \quad \text{for every } \; n \geq 0.$$

Another uninteresting case is when F has the form

$$p_0 p_1 > 0 \quad \text{and} \quad p_0 + p_1 = 1.$$

In this case the population remains at its initial size $Z_0 = 1$ for a random number of steps with a geometric distribution, then jumps to 0, after which it remains stuck at 0 forever afterwards. (Observe that for *any* Galton-Watson process, with any offspring distribution, the state 0 is an absorbing state.) To avoid having to consider these uninteresting cases separately in every result to follow, we make the following standing assumption:

Assumption 1. The offspring distribution is not a point mass (that is, there is no $k \geq 0$ such that $p_k = 1$), and it places positive probability on some integer $k \geq 2$. Furthermore, the offspring distribution has finite mean $\mu > 0$ and finite variance $\sigma^2 > 0$.

1.1. **First Moment Calculation.** The inductive definition (2) allows a painless calculation of the means EZ_n. Since the random variables ξ_i^{n+1} are independent of Z_n,

$$\begin{aligned} EZ_{n+1} &= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{m} \xi_i^{n+1}\right) \mathbf{1}\{Z_n = m\} \\ &= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{m} \xi_i^{n+1}\right) P\{Z_n = m\} \\ &= \sum_{k=0}^{\infty} m\mu P\{Z_n = m\} \\ &= \mu E Z_n. \end{aligned}$$

Since $EZ_0 = 1$, it follows that

$$\boxed{EZ_n = \mu^n.} \tag{3}$$

Corollary 1. *If $\mu < 1$ then with probability one the Galton-Watson process dies out eventually, i.e., $Z_n = 0$ for all but finitely many n. Furthermore, if $\tau = \min\{n : Z_n = 0\}$ is the extinction time, then*

$$P\{\tau > n\} \leq \mu^n.$$

Proof. The event $\{\tau > n\}$ coincides with the event $\{Z_n \geq 1\}$. By Markov's inequality,

$$P\{Z_n \geq 1\} \leq EZ_n = \mu^n.$$

□

1.2. **Recursive Structure and Generating Functions.** The Galton-Watson process Z_n has a simple recursive structure that makes it amenable to analysis by generating function methods. Each of the first-generation individuals $\alpha, \beta, \gamma, \dots$ behaves independently of the others; moreover, all of its descendants (the offspring of the offspring, etc.) behaves independently of the descendants of the other first-generation individuals. Thus, each of the first-generation individuals engenders an independent copy of the Galton-Watson process. It follows that a Galton-Watson process is gotten by conjoining to the single individual in the 0th generation Z_1 (conditionally) independent copies of the Galton-Watson process. The recursive structure leads to a simple set of relations among the probability generating functions of the random variables Z_n:

Proposition 2. *Denote by $\varphi_n(t) = Et^{Z_n}$ the probability generating function of the random variable Z_n, and by $\varphi(t) = \sum_{k=0}^{\infty} p_k t^k$ the probability generating function of the offspring distribution. Then φ_n is the n−fold composition of φ by itself, that is,*

$$\varphi_0(t) = t \qquad \text{and} \tag{4}$$

$$\varphi_{n+1}(t) = \varphi(\varphi_n(t)) = \varphi_n(\varphi(t)) \qquad \forall\, n \geq 0. \tag{5}$$

Proof. There are two ways to proceed, both simple. The first uses the recursive structure directly to deduce that Z_{n+1} is the sum of Z_1 conditionally independent copies of Z_n. Thus,

$$\begin{aligned}\varphi_{n+1}(t) = Et^{Z_{n+1}} &= E\varphi_n(t)^{Z_1} \\ &= \varphi(\varphi_n(t)).\end{aligned}$$

The second argument relies on the fact the generating function of the mth convolution power $\{p_k^{*m}\}$ is the mth power of the generating function $\varphi(t)$ of $\{p_k\}$. Thus,

$$\begin{aligned}\varphi_{n+1}(t) = Et^{Z_{n+1}} &= \sum_{k=0}^{\infty} E(t^{Z_{n+1}} \,|\, Z_n = k)P(Z_n = k) \\ &= \sum_{k=0}^{\infty} \varphi(t)^m P(Z_n = k) \\ &= \varphi_n(\varphi(t)).\end{aligned}$$

By induction on n, this is the $(n+1)$st iterate of the function $\varphi(t)$. □

Problem 1. (A) Show that if the mean offspring number $\mu := \sum_k kp_k < \infty$ then the expected size of the nth generation is $EZ_n = \mu^n$. (B) Show that if the variance $\sigma^2 = \sum_k (k-\mu)^2 p_k < \infty$ then the variance of Z_n is finite, and give a formula for it.

Properties of the Generating Function $\varphi(t)$: Assumption 1 guarantees that $\varphi(t)$ is not a linear function, because the offspring distribution puts mass on some integer $k \geq 2$. Thus, $\varphi(t)$ has the following properties:

(A) $\varphi(t)$ is strictly increasing for $0 \leq t \leq 1$.
(B) $\varphi(t)$ is strictly convex, with strictly increasing first derivative.
(C) $\varphi(1) = 1$.

1.3. **Extinction Probability.** If for some n the population size $Z_n = 0$ then the population size is 0 in all subsequent generations. In such an event, the population is said to be *extinct*. The first time that the population size is 0 (formally, $\tau = \min\{n : Z_n = 0\}$, or $\tau = \infty$ if there is no such n) is called the *extinction time*. The most obvious and natural question concerning the behavior of a Galton-Watson process is: What is the probability $P\{\tau < \infty\}$ of extinction?

Proposition 3. *The probability ζ of extinction is the smallest nonnegative root $t = \zeta$ of the equation*

$$\varphi(t) = t. \tag{6}$$

Proof. The key idea is *recursion*. Consider what must happen in order for the event $\tau < \infty$ of extinction to occur: Either (a) the single individual alive at time 0 has no offspring; or (b) each of its offspring must engender a Galton-Watson process that reaches extinction. Possibility (a) occurs with probability p_0. Conditional on the event that $Z_1 = k$, possibility (b) occurs with probability ζ^k. Therefore,

$$\zeta = p_0 + \sum_{k=1}^{\infty} p^k \zeta^k = \varphi(\zeta),$$

that is, the extinction probability ζ is a root of the Fixed-Point Equation (6).

There is an alternative proof that $\zeta = \varphi(\zeta)$ that uses the iteration formula (5) for the probability generating function of Z_n. Observe that the probability of the event $Z_n = 0$ is easily recovered from the generating function $\varphi_n(t)$:

$$P\{Z_n = 0\} = \varphi_n(0).$$

By the nature of the Galton-Watson process, these probabilities are nondecreasing in n, because if $Z_n = 0$ then $Z_{n+1} = 0$. Therefore, the limit $\xi := \lim_{n\to\infty} \varphi_n(0)$ exists, and its value is the extinction probability for the Galton-Watson process. The limit ξ must be a root of the Fixed-Point Equation, because by the continuity of φ,

$$\begin{aligned}
\varphi(\xi) &= \varphi(\lim_{n\to\infty} \varphi_n(0)) \\
&= \lim_{n\to\infty} \varphi(\varphi_n(0)) \\
&= \lim_{n\to\infty} \varphi_{n+1}(0) \\
&= \xi.
\end{aligned}$$

Finally, it remains to show that ξ is the *smallest* nonnegative root ζ of the Fixed-Point Equation. This follows from the monotonicity of the probability generating functions φ_n: Since

$\zeta \geq 0$,

$$\varphi_n(0) \leq \varphi_n(\zeta) = \zeta.$$

Taking the limit of each side as $n \to \infty$ reveals that $\xi \leq \zeta$. □

It now behooves us to find out what we can about the roots of the Fixed-Point Equation (6). First, observe that there is always at least one nonnegative root, to wit, $t = 1$, this because $\varphi(t)$ is a probability generating function. Furthermore, since Assumption 1 guarantees that $\varphi(t)$ is strictly convex, roots of equation 6 must be isolated. The next proposition asserts that there are either one or two roots, depending on whether the mean number of offspring $\mu := \sum_k k p_k$ is *greater* than one.

Definition 2. A Galton-Watson process with mean offspring number μ is said to be *supercritical* if $\mu > 1$, *critical* if $\mu = 1$, or *subcritical* if $\mu < 1$.

Proposition 4. *Unless the offspring distribution is the degenerate distribution that puts mass* 1 *at* $k = 1$*, the Fixed-Point Equation (6) has either one or two roots. In the* supercritical *case, the Fixed-Point Equation has a unique root* $t = \zeta < 1$ *less than one. In the critical and subcritical cases, the only root is* $t = 1$.

Together with Proposition 3 this implies that extinction is certain (that is, has probability one) if and only if the Galton-Watson process is critical or subcritical. If, on the other hand, it is *supercritical* then the probability of extinction is $\zeta < 1$.

Proof. By assumption, the generating function $\varphi(t)$ is *strictly* convex, with strictly increasing first derivative and positive second derivative. Hence, if $\mu = \varphi'(1) \leq 1$ then there cannot be a root $0 \leq \zeta < 1$ of the Fixed-Point Equation $\zeta = \varphi(\zeta)$. This follows from the Mean Value theorem, which implies that if $\zeta = \varphi(\zeta)$ and $1 = \varphi(1)$ then there would be a point $\zeta < \theta < 1$ where $\varphi'(\theta) = 1$.

Next, consider the case $\mu > 1$. If $p_0 = 0$ then the Fixed-Point Equation has roots $t = 0$ and $t = 1$, and because $\varphi(t)$ is strictly convex, there are no other positive roots. So suppose that $p_0 > 0$, so that $\varphi(0) = p_0 > 0$. Since $\varphi'(1) = \mu > 1$, Taylor's formula implies that $\varphi(t) < t$ for values of $t < 1$ sufficiently near 1. Thus, $\varphi(0) - 0 > 0$ and $\varphi(t_*) - t_* < 0$ for some $0 < t_* < 1$. By the Intermediate Value Theorem, there must exist $\zeta \in (0, t_*)$ such that $\varphi(\zeta) - \zeta = 0$. □

1.4. **Tail of the Extinction Time Distribution.** For both critical and subcritical Galton-Watson processes extinction is certain. However, critical and subcritical Galton-Watson processes differ dramatically in certain respects, most notably in the distribution of the time to extinction. This is defined as follows:

$$\tau = \min\{n \geq 1 : Z_n = 0\}. \tag{7}$$

Proposition 5. *Let* $\{Z_n\}_{n \geq 0}$ *be a Galton-Watson process whose offspring distribution* F *has mean* $\mu \leq 1$ *and variance* $\sigma^2 < \infty$*. Denote by* τ *the extinction time. Then*

(A) *If* $\mu < 1$ *then there exists* $C = C_F \in (0, \infty)$ *such that* $P\{\tau > n\} \sim C\mu^n$ *as* $n \to \infty$.
(B) *If* $\mu = 1$ *then* $P\{\tau > n\} \sim 2/(\sigma^2 n)$ *as* $n \to \infty$.

Thus, in the subcritical case, the extinction time has an exponentially decaying tail, and hence finite moments of all orders. On the other hand, in the critical case the extinction time has infinite mean.

Proof. First note that $P\{\tau > n\} = P\{Z_n > 0\}$. Recall from the proof of Proposition 3 that $P\{Z_n = 0\} = \varphi_n(0)$; hence,

$$P\{\tau > n\} = 1 - \varphi_n(0).$$

This shows that the tail of the distribution is determined by the speed at which the sequence $\varphi_n(0)$ approaches 1. In the subcritical case, the graph of the generating function $\varphi(t)$ has slope $\mu < 1$ at $t = 1$, whereas in the critical case the slope is $\mu = 1$. It is this difference that accounts for the drastic difference in the rate of convergence.

Subcritical Case: Consider first the case where $\mu = \varphi'(1) < 1$. Recall from the proof of Proposition 3 that in this case the sequence $\varphi_n(0)$ increases and has limit 1. Thus, for n large, $\varphi_n(0)$ will be near 1, and in this neighborhood the first-order Taylor series will provide a good approximation to φ. Consequently,

$$\begin{aligned} 1 - \varphi_{n+1}(0) &= 1 - \varphi(\varphi_n(0)) \\ &= 1 - \varphi(1 - (1 - \varphi_n(0))) \\ &= 1 - (1 - \varphi'(1)(1 - \varphi_n(0))) + O(1 - \varphi_n(0))^2 \\ &= \mu(1 - \varphi_n(0)) + O(1 - \varphi_n(0))^2. \end{aligned} \tag{8}$$

If not for the remainder term, we would have an exact equality $1 - \varphi_{n+1}(0) = \mu(1 - \varphi_n(0))$, which could be iterated to give

$$1 - \varphi_n(0) = \mu^n(1 - \varphi_0(0)) = \mu^n.$$

This would prove the assertion (A). Unfortunately, the equalities are exact only in the special case where the generating function $\varphi(t)$ is linear. In the general case, the remainder term in the Taylor series expansion (**??**) must be accounted for.

Because the generating function $\varphi(t)$ is convex, with derivative $\varphi'(1) = \mu$, the error in the approximation (8) is negative: in particular, for some constant $0 < C < \infty$,

$$\mu(1 - \varphi_n(0)) - C(1 - \varphi_n(0))^2 \le 1 - \varphi_{n+1}(0) \le \mu(1 - \varphi_n(0)).$$

The upper bound implies that $1 - \varphi_n(0) \le \mu^n$ (repeat the iteration argument above, replacing equalities by inequalities!). Now divide through by $\mu(1 - \varphi_n(0))$ to get

$$\begin{aligned} 1 - C(1 - \varphi_n(0)) \le \frac{\mu^{-n-1}(1 - \varphi_{n+1}(0))}{\mu^{-n}(1 - \varphi_n(0))} \le 1 \quad \Longrightarrow \\ 1 - C\mu^n \le \frac{\mu^{-n-1}(1 - \varphi_{n+1}(0))}{\mu^{-n}(1 - \varphi_n(0))} \le 1. \end{aligned}$$

Thus, successive ratios of the terms $\mu^{-n}(1 - \varphi_n(0))$ are exceedingly close to 1, the error decaying geometrically. Since these errors sum, Weierstrass' Theorem on convergence of products implies that

$$\lim_{n\to\infty} \frac{\mu^{-n}(1 - \varphi_n(0))}{\mu^{-0}(1 - \varphi_0(0))} = \lim_{n\to\infty} \mu^{-n}(1 - \varphi_n(0)) := C_F$$

exists and is positive.

Critical Case: Exercise. (See Problem 4 below.) □

1.5. **Asymptotic Growth Rate for Supercritical Galton-Watson Processes.** It is not hard to see that if a Galton-Watson process Z_n is supercritical (that is, the mean offspring number $\mu > 1$) then either $Z_n = 0$ eventually or $Z_n \to \infty$. Here is an informal argument for the case where $p_0 > 0$: Each time that $Z_n = K$, for some $K \geq 1$, there is chance p_0^K that $Z_{n+1} = 0$. If somehow the process Z_n were to visit the state K infinitely many times, then it would have infinitely many chances to hit an event of probability p_0^K; but once it hits this event, it is absorbed in the state 0 and can never revisit state K. This argument can be made rigorous:

Problem 2. Prove that if a Markov chain has an absorbing state z, and if x is a state such that z is accessible from x, then x is transient.

If Z_n is supercritical, then it follows that with positive probability (=1-probability of extinction) $Z_n \to \infty$. How fast does it grow?

Theorem 6. *There exists a nonnegative random variable W such that*

$$\lim_{n\to\infty} Z_n/\mu^n = W \quad \textit{almost surely.} \tag{9}$$

If the offspring distribution has finite second moment[1] *and $\mu > 1$ then the limit random variable W is positive on the event that $Z_n \to \infty$.*

Given the *Martingale Convergence Theorem*, the convergence (9) is easy; however, (9) is quite difficult to prove without martingales. In section 2 below, I will prove an analogous convergence theorem for a continuous-time branching process.

1.6. **Problems.**

Problem 3. Suppose that the offspring distribution is nondegenerate, with mean $\mu \neq 1$, and let ζ be the smallest positive root of the Fixed-Point Equation. (A) Show that if $\mu \neq 1$ then the root ζ is an *attractive* fixed point of φ, that is, $\varphi'(\zeta) < 1$. (B) Prove that for a suitable positive constant C,

$$\zeta - \varphi_n(0) \sim C\varphi'(\zeta)^n.$$

(Hence the term *attractive* fixed point.)

Problem 4. Suppose that the offspring distribution is nondegenerate, with mean $\mu = 1$. This is called the *critical* case. Suppose also that the offspring distribution has finite variance σ^2. (A) Prove that for a suitable positive constant C,

$$1 - \varphi_n(0) \sim C/n.$$

(B) Use the result of part (A) to conclude that the distribution of the extinction time has the following scaling property: for every $x > 1$,

$$\lim_{n\to\infty} P(\tau > nx \,|\, \tau > n) = C/x.$$

[1] Actually, it is enough that $\sum_{k\geq 3} p_k k \log k < \infty$: this is the *Kesten-Stigum* theorem.

HINT for part (A): The Taylor series approximation to $\varphi(t)$ at $\zeta = 1$ leads to the following approximate relationship, valid for large n:

$$1-\varphi_{n+1}(0) \approx 1-\varphi_n(0) - \frac{1}{2}\varphi''(1)(1-\varphi_n(0))^2,$$

which at first does not seem to help, but on further inspection does. The trick is to change variables: if x_n is a sequence of positive numbers that satisfies the recursion

$$x_{n+1} = x_n - b x_n^2$$

then the sequence $y_n := 1/x_n$ satisfies

$$y_{n+1} = y_n + b + b/y_n + \dots.$$

Problem 5. There's a Galton-Watson process in my random walk! Let S_n be the simple nearest-neighbor random walk on the integers started at $S_0 = 1$. Define T to be the time of the first visit to the origin, that is, the smallest $n \geq 1$ such that $S_n = 0$. Define $Z_0 = 1$ and

$$Z_k = \sum_{n=0}^{T-1} \mathbf{1}\{X_n = k \text{ and } X_{n+1} = k+1\}.$$

In words, Z_k is the number of times that the random walk X_n crosses from k to $k+1$ before first visiting 0.

(A) Prove that the sequence $\{Z_k\}_{k\geq 0}$ is a Galton-Watson process, and identify the offspring distribution as a geometric distribution.

(B) Calculate the probability generating function of the offspring distribution, and observe that it is a *linear fractional transformation.* (See Ahlfors, *Complex Analysis,* ch. 1 for the definition and basic theory of LFTs. Alternatively, try the Wikipedia article.)

(C) Use the result of (B) to find out as much as you can about the distribution of Z_k.

(D) Show that $T = \sum_{k\geq 1} Z_k$ is the total number of individuals ever born in the course of the Galton-Watson process, and show that τ (the extinction time of the Galton-Watson process) is the maximum displacement M from 0 attained by the random walk before its first return to the origin. What does the result of problem 4, part (B), tell you about the distribution of M?

2. YULE'S BINARY FISSION PROCESS

2.1. **Definition and Construction.** The *Yule process* is a continuous-time branching model, in which individuals undergo binary fission at random times. It evolves as follows: Each individual, independently of all others and of the past of the process, waits an exponentially distributed time and then splits into two identical particles. (It is useful for the construction below to take the view that at each fission time the fissioning particle survives and creates one new clone of itself.) The exponential waiting times all have mean 1. Because the exponential random variables are mutually independent, the probability that two fissions will occur simultaneously is 0.

A Yule process started by 1 particle at time 0 can be built from independent Poisson processes as follows. Let $\{N_j(t)\}_{j\in\mathbb{N}}$ be a sequence of independent Poisson counting processes. Since the interoccurrence times in a Poisson process are exponential-1, the jump times in the Poisson process $N_j(t)$ can be used as the fission times of the jth particle; at each such fission time, a new particle must be added to the population, and so a new Poisson process $N_k(t)$ must be "activated". Thus, the time T_m at which the mth fission occurs can be defined as follows: set $T_0 = 0$ and

$$T_m = \min\{t > T_{m-1} : \sum_{j=1}^{m}(N_j(t) - N_j(T_{m-1})) = 1\}. \tag{10}$$

Thus, T_m is the first time after T_{m-1} that one of the first m Poisson processes jumps. The size Z_t of the population at time t is then

$$Z_t = m \quad \text{for} \quad T_{m-1} \leq t < T_m. \tag{11}$$

A similar construction can be given for a Yule process starting with $Z_0 = k \geq 2$ particles: just change the definition of the fission times T_m to

$$T_m = \min\{t > T_{m-1} : \sum_{j=1}^{m+k-1}(N_j(t) - N_j(T_{m-1})) = 1\}. \tag{12}$$

Alternatively, a Yule process with $Z_0 = k$ can be gotten by superposing k independent Yule processes Z_t^j all with $Z_0^j = 1$, that is,

$$Z_t = \sum_{j=1}^{k} Z_t^j \tag{13}$$

Problem 6. Show that by suitably indexing the Poisson processes in the first construction (12) one can deduce the superposition representation (13).

Problem 7. Calculate the mean EZ_t and variance $\mathrm{var}(Z_t)$ of the population size in a Yule process. For the mean you should get $EZ_t = e^t$. HINT: Condition on the time of the first fission.

2.2. **Asymptotic Growth.**

Theorem 7. *Let Z_t be the population size at time t in a Yule process with $Z_0 = 1$. Then*

$$Z_t / e^t \xrightarrow{a.s.} W \tag{14}$$

where W has the unit exponential distribution.

The proof has two parts: First, it must be shown that Z_t/e^t converges to something; and second, it must be shown that the limit random variable W is exponentially distributed. The proof of almost sure convergence will be based on a careful analysis of the first passage times T_m defined by (10). Convergence of Z_t/e^t to a positive random variable W is equivalent to convergence of $\log Z_t - t$ to a real-valued limit $\log W$. Since Z_t is a counting process (that is, it is nondecreasing in t and its only discontinuities are jumps of size 1), convergence of $\log Z_t - t$ is equivalent to showing that there exists a finite random variable $Y = -\log W$ such that for any $\varepsilon > 0$,

$$\lim_{m \to \infty} (T_m - \log m) = Y. \tag{15}$$

To accomplish this, we will use the following consequence of the construction (10).

Proposition 8. *Let T_m be the fission times in a Yule process Z_t with $Z_0 = k$. Then the interoccurrence times $\tau_m := T_m - T_{m-1}$ are independent, exponentially distributed random variables with expectations $E\tau_m = 1/(m+k-1)$.*

Proof (Sketch). The random variable T_m is the first time after T_{m-1} at which one of the Poisson processes $N_j(t)$, for $1 \leq j \leq m+k-1$, has a jump. Times between jumps in a Poisson process are exponentially distributed with mean 1, and jump times in independent Poisson processes are independent. Thus, the time until the next jump in m independent Poisson processes is the minimum of m independent exponentials, which is exponentially distributed with mean $1/m$.

This is not quite a complete argument, because the "start" times T_m are random. However, it is not difficult (exercise!) to turn the preceding into a rigorous argument by integrating out over the possible values of T_m and the possible choices for which Poisson processes jump at which times. □

The family of exponential distributions is closed under scale transformations: In particular, if Y is exponentially distributed with mean 1 and $\alpha > 0$ is a scalar, then αY is exponentially distributed with mean α. Since the variance $\mathrm{var}(Y)$ of a unit exponential is 1, it follows that the variance $\mathrm{var}(\alpha Y)$ of an exponential with mean α is α^2. Consequently, if $\tau_m = T_m - T_{m-1}$ is the time between the $(m-1)$th and the mth fission times in a Yule process with $Z_0 = 1$, then

$$E\tau_{m+1} = m^{-1} \quad \text{and} \quad \mathrm{var}(\tau_{m+1}) = m^{-2}, \tag{16}$$

and so

$$E\,T_{m+1} = \sum_{k=1}^{m} k^{-1} \sim \log m \quad \text{and} \quad \mathrm{var}(T_{m+1}) = \sum_{k=1}^{m} k^{-2} \to \zeta(2) < \infty \tag{17}$$

as $m \to \infty$, where $\zeta(2) = \sum_{k=1}^{\infty} k^{-2}$. In particular, the variance of T_m remains bounded as $m \to \infty$, and so the distribution of T_m remains concentrated around $\log m$. In fact, $T_m - \log m$ converges, to a possibly random limit, by the following general result about random series of independent random variables:

Theorem 9. *Let X_j be independent random variables with mean $EX_j = 0$ and finite variances $\mathrm{var}(X_j) = \sigma_j^2$. Then*

$$\sum_{j=1}^{\infty} \sigma_j^2 := \sigma^2 < \infty \quad \Longrightarrow \quad \lim_{n\to\infty} \sum_{j=1}^{n} X_j := S \tag{18}$$

exists and is finite with probability one, and the limit random variable S has mean zero and variance σ^2.

A proof of Theorem 9, based on Wald's Second Identity, is given in section 3 below. Modulo this, we have proved (15), and hence that $W = \lim_{t\to\infty} Z_t/e^t$ exists and is finite and strictly positive with probability 1.

2.3. **Characterization of the Exponential Distributions.** It remains to show that the limit random variable W is exponentially distributed with mean 1. For this, we appeal to self-similarity. Let $T = T_1$ be the time of the first fission. At this instant, two identical offspring particles are produced, each of which engenders its own Yule process. Thus,

$$\begin{aligned} Z_t &= 1 \quad \text{if } t < T \quad \text{and} \\ Z_t &= Z'_{t-T} + Z''_{t-T} \quad \text{if } t \geq T \end{aligned} \tag{19}$$

where Z'_s and Z''_s are independent Yule processes – and independent of the fission time T – each started with $Z'_0 = Z''_0 = 1$ particle. Divide each side by e^t and let $t \to \infty$ to get

$$W = e^{-T}(W' + W'') = U(W' + W'') \tag{20}$$

where T is a unit exponential and W', W'' are independent replicas of W, both independent of T. Note that $U = e^{-T}$ is uniformly distributed on the unit interval.

Proposition 10. *If W is a positive random variable that satisfies the distributional equation* (20) *then W has an exponential distribution. Conversely, there exist (on some probability space) independent unit exponential random variables T, W', W'' such that the random variable W defined by* (20) *also has the unit exponential distribution.*

Proof. The converse half is easy, given what we know about Poisson processes: Take a unit-intensity Poisson process N_t and let ν be the time of the *second* occurrence. Then ν is the sum of two independent unit exponentials. Furthermore, we know that the time of the *first* occurrence is, conditional on ν, uniformly distributed on the interval $[0, \nu]$. Thus, if we multiply ν by an independent uniform-[0,1], we obtain a random variable whose distribution coincides with that of the time of the first occurrence in a Poisson process. (Note: The random variable $U\nu$ so obtained is not the same as the time of first occurrence in N_t, but its distribution must be the same.)

The direct half is harder. I will show that if W is a positive random variable that satisfies (20) then its *Laplace transform*

$$\varphi(\theta) := E e^{-\theta W} \tag{21}$$

must coincide with the Laplace transform of the exponential distribution with mean α, for some value of $\alpha > 0$. By the Uniqueness Theorem for Laplace transforms, this will imply that W has an exponential distribution. The strategy will be to take the Laplace transform of both sides of (20), and to split the expectation on the right side into two, one for the event $\{U < 1 - \varepsilon\}$ and the other for $\{U \geq 1 - \varepsilon\}$. Letting $\varepsilon \to 0$ will then lead to a first-order differential equation for $\varphi(\theta)$ whose only solutions coincide with Laplace transforms of exponential distributions. The sordid details: equation (20) and the independence of U, W', W'' imply that for any $\varepsilon > 0$

and every $\theta > 0$,

$$\begin{aligned}\varphi(\theta) = Ee^{-\theta W} &= \int_0^1 Ee^{-u\theta(W'+W'')}\,du \\ &= \int_0^{1-\varepsilon} Ee^{-u\theta(W'+W'')}\,du + \int_{1-\varepsilon}^1 Ee^{-u\theta(W'+W'')}\,du \\ &= \int_0^1 Ee^{-\theta(1-\varepsilon)u(W'+W'')}\,du(1-\varepsilon) + \int_{1-\varepsilon}^1 Ee^{-u\theta W'}Ee^{-u\theta W''}\,du \\ &= (1-\varepsilon)\varphi(\theta-\theta\varepsilon) + \int_{1-\varepsilon}^1 \varphi(u\theta)^2\,du\end{aligned}$$

Subtract $\varphi(\theta(1-\varepsilon))$ from both sides and divide by ε to get

$$\frac{\varphi(\theta)-\varphi(\theta-\theta\varepsilon)}{\varepsilon} = -\varphi(\theta-\theta\varepsilon) + \frac{1}{\varepsilon}\int_{1-\varepsilon}^1 \varphi(u\theta)^2\,du.$$

Now take $\varepsilon \to 0$ and use the continuity and boundedness of $\varphi(\theta)$ together with the Fundamental Theorem of Calculus to conclude that

$$-\theta\varphi'(\theta) = -\varphi(\theta) + \varphi(\theta)^2. \tag{22}$$

It is easily checked that for any $\alpha > 0$ the Laplace transform $\varphi_\alpha(\theta) = \alpha/(\alpha+\theta)$ of the exponential distribution with mean $1/\alpha > 0$ is a solution of the differential equation (22). This gives a one-parameter family of solutions; by the uniqueness theorem for first-order ordinary differential equations, it follows that these are the *only* solutions. □

3. Convergence of Random Series

This section is devoted to the proof of Theorem 9. Assume that $X_1, X_2, \dots$ are independent random variables with means $EX_j = 0$ and finite variances $\sigma_j^2 = EX_j^2$, and for each $n = 0, 1, 2, \dots$ set

$$S_n = \sum_{j=1}^n X_j. \tag{23}$$

Wald's Second Identity . *For any* bounded *stopping time* T*,*

$$ES_T^2 = E\sum_{j=1}^T \sigma_j^2. \tag{24}$$

Proof. Since T is a stopping time, for any integer $k \geq 1$ the event $\{T \geq k\} = \{T > k-1\}$ depends only on the random variables X_i for $i < k$, and hence is independent of X_k. In particular, if $j < k$ then $EX_jX_k\mathbf{1}\{T \geq k\} = EX_j\mathbf{1}\{T \geq k\}EX_k = 0$. Now suppose that T is a *bounded* stopping time;

then $T \le m$ almost surely for some integer $m \ge 1$. Thus,

$$\begin{aligned}
ES_T^2 &= E\left(\sum_{k=1}^{m} X_k \mathbf{1}\{T \ge k\}\right)^2 \\
&= \sum_{k=1}^{m} EX_k^2 \mathbf{1}\{T \ge k\} + 2 \sum_{1 \le j < k \le m} EX_j X_k \mathbf{1}\{T \ge k\} \\
&= \sum_{k=1}^{m} EX_k^2 \mathbf{1}\{T \ge k\} \\
&= \sum_{k=1}^{m} EX_k^2 E\mathbf{1}\{T \ge k\} \\
&= \sum_{k=1}^{m} \sigma_k^2 E\mathbf{1}\{T \ge k\} \\
&= E\left(\sum_{k=1}^{T} \sigma_k^2\right).
\end{aligned}$$

□

Corollary 11. *(L^2 Maximal Inequality) Assume that the total variance $\sigma^2 := \sum_{j=1}^{\infty} \sigma_j^2 < \infty$. Then for any $\alpha > 0$,*

$$P\{\sup_{n \ge 1} |S_n| \ge \alpha\} \le \frac{\sigma^2}{\alpha^2}. \tag{25}$$

Proof. Define T to be the first n such that $|S_n| \ge \alpha$, or $+\infty$ if there is no such n. The event of interest, that $\sup|S_n| \ge \alpha$, coincides with the event $\{T < \infty\}$. This in turn is the increasing limit of the events $\{T \le m\}$ as $m \to \infty$. Now for each finite m the random variable $T \wedge m$ is a bounded stopping time, so Wald's Identity implies

$$ES_{T\wedge m}^2 = E\left(\sum_{j=1}^{T\wedge m} \sigma_j^2\right) \le \sigma^2.$$

Hence,

$$\alpha^2 P\{T \le m\} \le ES_{T\wedge m}^2 \mathbf{1}\{T \le m\} \le ES_{T\wedge m}^2 \le \sigma^2.$$

□

Convergence of Random Sequences: Strategy. Let $\{s_n\}_{n\ge 1}$ be a sequence of real (or complex) numbers. To show that the sequence s_n converges, it suffices to prove that it is Cauchy; and for this, it suffices to show that for every $k \ge 1$ (or for all sufficiently large k) there exists an integer n_k such that

$$|s_{n_k} - s_n| \le 2^{-k} \quad \text{for all } n \ge n_k. \tag{26}$$

Now suppose that the sequence S_n is *random*. To prove that this sequence converges with probability one, it suffices to exhibit a sequence of integers n_k such that the *complements* G_k^c of the events

$$G_k := \{|S_{n_k} - S_n| \le 2^{-k} \ \forall\, n \ge n_k\} \tag{27}$$

occur only finitely many times, with probability one. For this, it is enough to prove that

$$\sum_{k=1}^{\infty} P(G_k^c) = E \sum_{k=1}^{\infty} \mathbf{1}_{G_k^c} < \infty, \tag{28}$$

because if the expectation is finite then the random count itself must be finite, with probability one. This is the *Borel-Cantelli* criterion for convergence of a random series.

Proof of Theorem 9. Assume then that the random variables S_n are the partial sums (23) of independent random variables X_j with means $EX_j = 0$ and variances $EX_j^2 = \sigma_j^2$ such that the total variance $\sigma^2 = \sum_j \sigma_j^2 < \infty$. Then for every $k \ge 1$ there exists $n_k < \infty$ such that

$$\sum_{j=n_k}^{\infty} \sigma_j^2 \le 8^{-k}.$$

By the Maximal Inequality,

$$P(G_k^c) = P\{\sup_{n \ge n_k} |S_n - S_{n_k}| \ge 2^{-k}\} \le 8^{-k}/4^{-k} = 2^{-k}.$$

Since $\sum_k 2^{-k} < \infty$, the Borel-Cantelli criterion is satisfied, and so the sequence S_n is, almost surely, Cauchy, and therefore has a finite limit S. Exercise: If you know the basics of measure theory, prove that $ES = 0$ and $ES^2 = \sigma^2$. Hint: First show that $S_n \to S$ in L^2, and conclude that the sequence S_n is uniformly integrable.

□

4. THE POLYA URN

4.1. **Rules of the game.** The *Polya urn* is the simplest stochastic model of *self-reinforcing behavior*, in which repetition a particular act makes it more likely that the same act will be repeated in the future. Suppose that every afternoon you visit a video-game arcade with two games: MS. PAC-MAN and SPACE INVADERS. On the first day, not having played either game before, you choose one at random. With each play, you develop a bit more skill at the game you choose, and your preference for it increases, making it more likely that you will choose it next time: In particular, if after n visits you have played MS. PAC-MAN R_n times and SPACE INVADERS $B_n = n - R_n$ times, then on the $(n+1)$st day the chance that you decide to put your quarter in SPACE INVADERS is

$$\theta_n := \frac{B_n + 1}{n + 2}.$$

It is natural to ask if after a while your relative preferences for the two games will begin to stabilize, and if so to what?

It is traditional to re-formulate this model as an *urn model*. At each step, a ball is chosen at random from among the collection of all balls (each colored either RED or BLUE) in the urn, and is then replaced, together with a new ball of the same color. More formally:

Definition 3. The *Polya urn* is a Markov chain (R_n, B_n) on the space $\mathbb{N} \times \mathbb{N}$ of positive integer pairs (r, b) with transition probabilities

$$p((r,b),(r+1,b)) = r/(r+b),$$
$$p((r,b),(r,b+1)) = b/(r+b). \tag{29}$$

The default initial state is $(1,1)$. The associated sampling process is the sequence X_n of Bernoulli random variables defined by $X_n = 1$ if $R_{n+1} = R_n + 1$ and $X_n = 0$ if $R_{n+1} = R_n$.

4.2. **The Polya urn and the Yule process.** Hidden within the Yule binary fission process is a Polya urn. Here's how it works: Start two independent Yule processes Y_t^R and Y_t^B, each having one particle at time 0 (thus, $Y_0^R = Y_0^B = 1$). Mark the particles of the process Y_t^R "red", and those of the process Y_t^B "blue". Set

$$Y_t = Y_t^R + Y_t^B;$$

then Y_t is itself a Yule process, with initial state $Y_0 = 2$.

Start the Yule process with two particles, one RED, the other BLUE. (Or, more generally, start it with r_0 red and b_0 blue.) Recall that at the time T_m of the mth fission, one particle is chosen at random from the particles in existence and cloned. This creates a new particle of the same color as its parent. Thus, the mechanism for duplicating particles in the Yule process works exactly the same way as the replication of balls in the Polya urn: in particular, the sequence of draws (RED or BLACK) made at times $T_1, T_2, \ldots$ has the same law as the sampling process associated with the Polya urn.

(To be continued.)