Chapter 7.2. Selected Topics: Galton-Watson Trees

The Galton-Watson process is a stochastic model that describes the evolution of a population where each individual reproduces independently according to the same probability distribution. The model was first introduced by Irénée-Jules Bienaymé in 1845, and later independently developed by Francis Galton and Henry William Watson in 1874 to study the extinction of family surnames.

The original motivation came from Galton's interest in genealogical trees, and whether aristocratic families would persist over time. Watson and Galton derived the probability of surname extinction, finding that family names would eventually disappear. The model has since found applications beyond genealogy, including Biology (population dynamics, cell division), Physics (branching processes in nuclear reactions), Computer Science (analysis of tree data structures), and Economics (branching of firms and organizations).

The population size in the next generation depends only on the current population size and the reproduction law, thus the sequence of population sizes forms a Markov chain.

Definitions and Results

Definition 0.1 (Galton-Watson Process). Let $X_0, X_1, X_2, \dots$ be a sequence of random variables (representing the population sizes in each generation). We say that $(X_t)_{t\geq 0}$ forms a **Galton-Watson process** if:

1. $X_0 = 1$ (we start with a single ancestor).
2. For each $t \geq 1$, $X_{t+1} = \sum_{n=1}^{X_t} Z_{t,n}$, where $Z_{t,n}$ represents the number of offspring of the n-th individual in generation t.
3. The random variables $(Z_{t,n})_{t\geq 0, n\geq 1}$ are i.i.d. on $\mathbb{N}$: $\mathbb{P}(Z_{t,n} = k) = p_k$ with $\sum_{k\geq 0} p_k = 1$.

The probability distribution $(p_k)_{k\geq 0}$ is called the **offspring distribution** of the process.

Remark. The Galton-Watson process is a Markov chain on the state space $S = \mathbb{N}$, where state 0 is absorbing (once the population goes extinct, it vanishes forever). If $X_t = 0$, then $X_{t+1} = X_{t+2} = \cdots = 0$. Given $X_t = n$, X_{t+1} is a sum of n i.i.d. random variables with distribution (p_k). Therefore, defining $(p_k^{(n)})$ as the n-th convolution of (p_k), we have the transition probabilities for X_t:

$$\mathbb{P}(X_{t+1} = k \mid X_t = n) = p_k^{(n)}$$

Notice that $(p^{(n)})$ is determined by

$$p_k^{(1)} = p_k$$
$$p_k^{(n+1)} = \sum_h p_h^{(n)} p_{k-h}$$

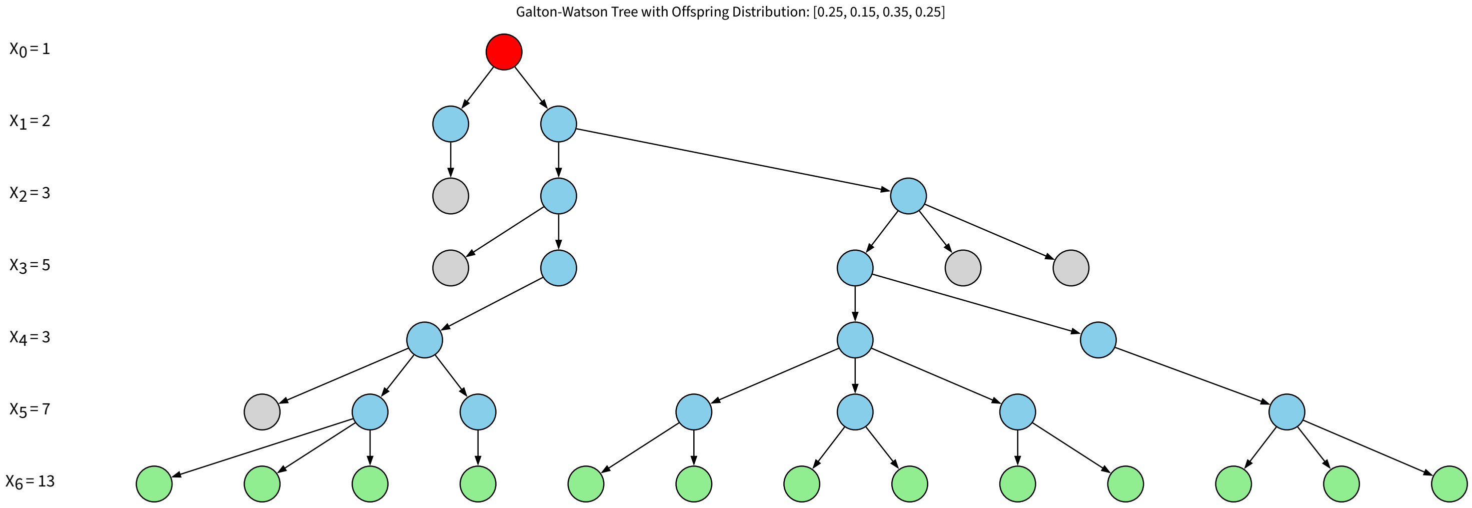

Figure 1

In particular it is important to characterize the probability[1] $\zeta = \mathbb{P}(\tau_0 < \infty)$, the probability that the process will go extinct. While not particularly useful for the computations, we can remark that for the process to go extinct at time t, all the individuals at time $t-1$ must have 0 offspring:

$$\zeta = \sum_{t=1}^{\infty} \mathbb{P}(\tau_0 = t) = \sum_{t=1}^{\infty} \sum_{k=0}^{\infty} \mathbb{P}(\tau_0 = t \mid X_{t-1} = k)\mathbb{P}(X_{t-1} = k) = \sum_{t,k=1}^{\infty} \mathbb{P}(X_{t-1} = k) p_0^k$$

The value of ζ is determined by the values of the offspring distribution (p_k), so we want to understand in particular when $\zeta = 1$ and when $\zeta < 1$. Define the function

$$\phi(z) := \sum_{k=0}^{\infty} p_k z^k, \qquad |z| \leq 1$$

which is the generating function of the offspring distribution. In particular

$$\phi(0) = p_0, \phi(1) = 1, \phi'(1) = \mathbb{E}[Z] = \sum_{k=0}^{\infty} k p_k$$

Lemma 0.1. *Let* $\phi_t(z) := \sum_{k=0}^{\infty} \mathbb{P}(X_t = k) z^k = \mathbb{E}\left[z^{X_t}\right]$*, for* $z \in [0,1]$*. Then it holds*

$$\begin{cases} \phi_0(z) = z \\ \phi_1(z) = \phi(z) \\ \phi_{t+1}(z) = \phi(\phi_t(z)) \end{cases}$$

Moreover ϕ_t *is non-negative, non-decreasing, convex, analytic in* $0 \leq z < 1$.

---

[1] As usual, hereafter $\tau_0 = \inf\{t \geq 0 : X_t = 0\}$ is the hitting time of 0.

Proof. The identities for ϕ_0 and ϕ_1 follow from the definition. Conditioning on Z_1 we have

$$\begin{aligned}\phi_{t+1}(z) &= \sum_{k=0}^{\infty} \mathbb{P}(X_{t+1}=k)z^k = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \mathbb{P}(X_{t+1}=k \mid X_1=j)\mathbb{P}(X_1=j)z^k \\ &= \sum_{j=0}^{\infty} p_j \left(\sum_{k=0}^{\infty} \mathbb{P}(X_t=k)z^k\right)^j = \phi(\phi_t(z))\end{aligned}$$

Alternatively, using conditional expectation, $\mathbb{E}[z^{X_{t+1}}] = \mathbb{E}[\mathbb{E}[z^{X_{t+1}} \mid X_1]] = \mathbb{E}[\mathbb{E}[z^{X_t}]^{X_1}] = \phi(\phi_t(z))$ since, conditional on X_1, X_{t+1} is the sum of X_1 independent Galton-Watson processes.

Being a power-series with non-negative coefficients, ϕ is non-negative, non-decreasing, convex, and analytic in $|z| < 1$. Thus its powers $\phi \circ \phi \circ \ldots \circ \phi$ and also non-negative, non-decreasing, convex, analytic. ☐

Theorem 0.1. *Let $(X_t)_{t\geq 0}$ be a Galton-Watson process with offspring distribution having probability generating function ϕ and mean $\mu = \mathbb{E}[Z] = \phi'(1) \in [0,\infty]$. Then ζ is the smallest fixed point of ϕ, i.e. the smallest solution to the equation*

$$z = \phi(z), \qquad z \in [0,1]$$

In particular

1. *If $p_0 = 0$, trivially $\zeta = 0$.*
2. *If $p_0 > 0$ and $\mu \leq 1$, then the population eventually goes extinct a.s., that is $\zeta = 1$.*
3. *If $\mu > 1$, then $\zeta < 1$ and*

$$\mathbb{P}(\lim_{t\to\infty} X_t = \infty) = 1 - \zeta$$

Proof. If $X_1 = k$, there will be extinction iff the k (independent) trees descending from each of the k individuals at time 1 will go extinct, thus with probability ζ^k. Thus, conditioning on X_1

$$\zeta = \sum_{k=0}^{\infty} \mathbb{P}(\tau_0 < \infty \mid X_1 = k)\mathbb{P}(X_1 = k) = \sum_{k=0}^{\infty} p_k \zeta^k = \phi(\zeta)$$

Namely ζ is a fixed point of ϕ. We need to check that it is the minimal one.

Let $\zeta_t = \mathbb{P}(X_t = 0) = \phi_t(0)$ be the probability of extinction by generation t. Then ζ_t is non-decreasing and $\zeta = \lim_{t\to\infty} \zeta_t$, by the continuity of probability measures on monotone sequences. Let ξ be any other fixed point $\xi = \phi(\xi)$, then since $\phi_t(z)$ is non-decreasing in z

$$\zeta_t = \phi_t(0) \leq \phi_t(\xi) = \xi$$

Thus taking the limit, $\zeta \leq \xi$.

If $p_0 = 0$ then $\phi(0) = 0$ and thus $\zeta = 0$ is the minimal fixed point, proving point 1..

Notice that $\xi = 1$ is always a fixed point, as $1 = \phi(1)$. However, since ϕ is nondecreasing and convex, assuming $p_0 > 0$ implies that there exists at most one other fixed point. Indeed, by elementary calculus:

- if $\phi'(1) > 1$, there exists exactly one other fixed point $\zeta < 1$.
- if $\phi'(1) \leq 1$, 1 is the unique fixed point, thus $\zeta = \xi = 1$.

We next check that $\mathbb{P}(\lim_{t\to\infty} X_t = \infty) = 1 - \zeta$. For each t such that $X_t \leq k$, there is a probability at least p_0^k of being extinct at the next step. So

$$\mathbb{P}(\tau_0 < \infty \mid |\{t \,:\, X_t \leq k\}| = \infty) = 1$$

Therefore, on $\{\tau_0 = \infty\}$, we have that $X_t \leq k$ only for finitely many times a.s., for any k, and thus X_t diverges a.s. on $\{\tau_0 = \infty\}$. $\square$

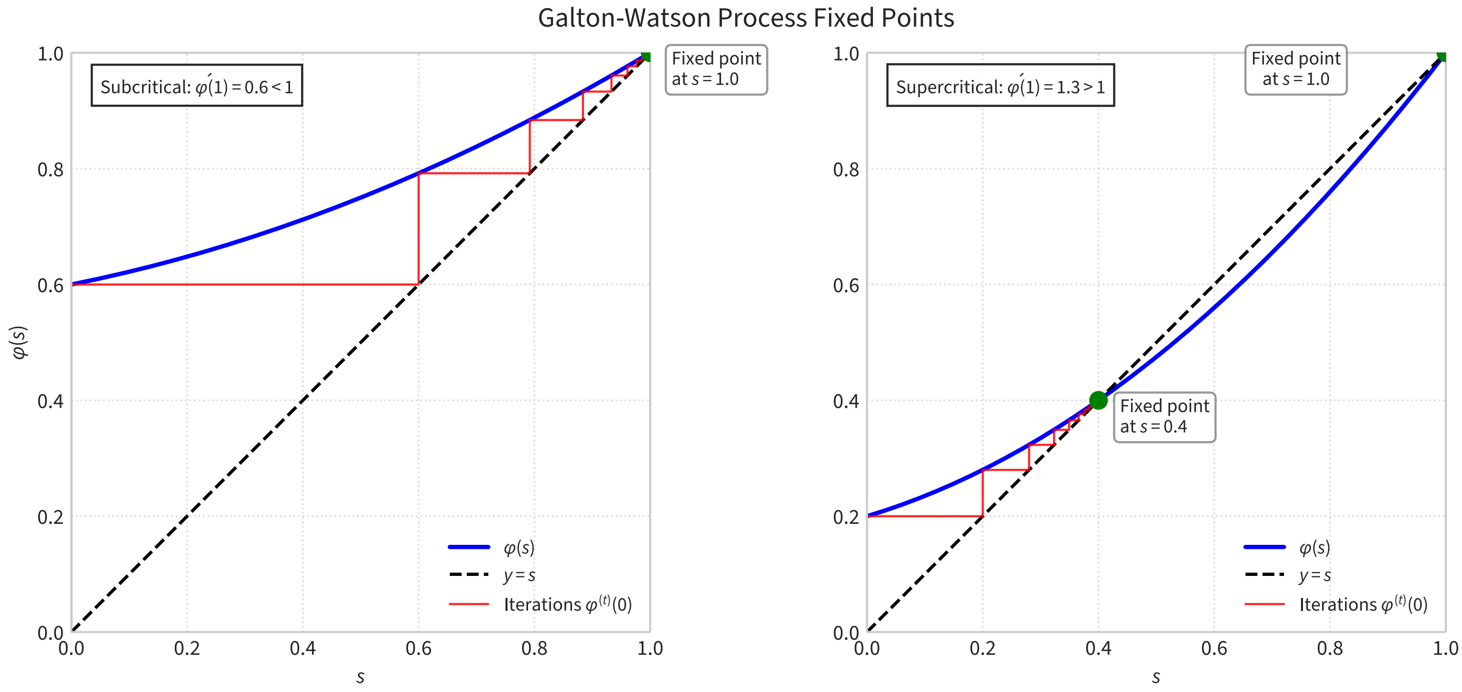

Figure 2: The extinction probability is the smallest fixed point of $\phi(z)$. Except in trivial cases ($p_1 = 1$, which gives $\phi(z) \equiv z$ and $\zeta = 0$), there are only two possibilities. If $\phi'(1) \leq 1$, there is only one such fixed point, $\zeta = 1$. If $\phi'(1) > 1$, there exist two fixed points: 1 and $\zeta < 1$.

Basic Examples

Example 0.1 (Binary splitting)**.** Consider a bacterium having probability $p_0 \in (0,1)$ to die before replication, probability $p_2 \in (0,1)$ to replicate into two exact copies (mitosis), and probability $p_1 = 1 - p_0 - p_2$ to undergo a mitosis where one of the two copies dies immediately (so the net offspring has size 1). The offspring distribution has $p_k = 0$ for $k \geq 3$ and $\mu = 1 - p_0 + p_2$. To find the probability ζ of extinction, we need to solve

$$z = \phi(z) \equiv p_0 + p_1 z + p_2 z^2 \equiv p_0 + (1 - p_0 - p_2) z + p_2 z^2$$

The solutions are $z = 1$ and $z = p_0/p_2$. Therefore

$$\zeta = \begin{cases} p_0/p_2 & \text{if } p_0 < p_2 \\ 1 & \text{if } p_0 \geq p_2 \end{cases}$$

and indeed $\zeta = 1$ iff $\mu \leq 1$.

Example 0.2 (Poisson offspring distribution)**.** Suppose the number of offspring follows a Poisson distribution with parameter $\lambda > 0$:

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N}$$

The probability generating function is:

$$\phi(z) = \sum_{k=0}^{\infty} z^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{\lambda(z-1)}$$

The mean is $\mu = \lambda$, so:

- If $\lambda \leq 1$, extinction occurs with probability 1.
- If $\lambda > 1$, the extinction probability ζ is the smallest solution to $e^{\lambda(z-1)} = z$.

Exercise 0.1. Consider a Galton-Watson process where each individual has a random number of offspring according to the geometric distribution with parameter $p \in (0,1)$, i.e., $p_k = p(1-p)^k$ for $k \geq 0$. Calculate the extinction probability and the expected value of the offspring distribution, in terms of p.

Solution

We have $\phi(z) = \sum_{k=0}^{\infty} p(1-p)^k z^k = \frac{p}{1-(1-p)z}$ and $\mu = \phi'(1) = (1-p)/p$. Solving $\phi(z) = z$ leads to the roots are 1 and $\frac{p}{1-p}$. Thus

$$\zeta = \begin{cases} 1 & \text{if } p \geq 1/2 \\ \frac{p}{1-p} & \text{if } p < 1/2 \end{cases}$$

consistently with the value of μ.

Exercise 0.2. Consider a random walk Y_t on $\mathbb{Z}$ with $p_{x,x+1} = 1 - p_{x,x-1} = p \in (0, 1/2]$, conditioned to start positive, that is we assume $Y_0 = 0$, $Y_1 = +1$ for the sake of simplicity. Let τ_0^+ be, as usual, the time of the first return to 0, so it is finite a.s. since $p \leq 1/2$, and define

$$X_n := |\{0 \leq t < \tau_0^+ \ : \ Y_t = n, Y_{t+1} = n+1\}|$$

as the number of times that Y_t crosses the edge $(n, n+1)$, before returning to 0.

Prove that (X_n) is a Galton-Watson process, find the offspring distribution and the extinction probability ζ.

Solution

Fix $n \geq 1$ and let σ_l be the time of the l-th crossing of $(n, n+1)$, thus $Y_{\sigma_l - 1} = n$ and $Y_{\sigma_l} = n+1$. The stopping time

$$\xi_l := \inf\{t > \sigma_l : Y_t = n\} \in \mathbb{N}$$

is the first return to level n after the k-th crossing. The number of crossings of $(n+1, n+2)$ that occur in the open interval (σ_k, ξ_k) is

$$Z_{n,l} := |\{\sigma_l < t < \xi_l : Y_t = n+1, Y_{t+1} = n+2\}|$$

By the strong Markov property, the $Z_{n,l}$ are independent and thus the conditions of Definition 0.1 hold. We interpret the crossing of $(n, n+1)$ as a "parent". Starting at $n+1$, the random walk must either step to $n+2$ (creating an offspring) or step to n (dying). Since $p_{x,x+1} = p$, the probability of stepping to $n+2$ before n is simply p (a single step right), and the probability of stepping to n before $n+2$ is $1-p$ (a single step left). If the walk hits $n+2$, it returns to $n+1$ a.s., allowing for further "offspring" attempts. Thus, $Z_{n,l}$ follows a geometric distribution with success parameter p:

$$\mathbb{P}(Z_{n,l} = k) = p^k(1-p), \quad k \geq 0$$

We thus have $\phi(z) = (1-p)/(1-pz)$, $\mu = \phi'(1) = p/(1-p) \leq 1$ since $p \leq 1/2$. Therefore $\zeta = 1$ for all $p \leq 1/2$ (if $p > 1/2$ then the argument above is not correct, since the return to 0 from $Y_1 = 1$ is not a.s.).

Moments and Asymptotic Behavior

We next examine the long-time behavior, in particular in the supercritical case $\mu > 1$.

Proposition 0.1. *Let $(X_t)_{t \geq 0}$ be a Galton-Watson process with offspring distribution (p_k).*

1. *For $\mu = \mathbb{E}[Z] = \sum_k kp_k$, it holds $\mathbb{E}[X_t] = \mu^t$ for $t \geq 0$.*
2. *Assuming $\sigma^2 := \mathrm{Var}(Z) = \sum_k (k-\mu)^2 p_k < \infty$, it holds for $t \geq 0$*

$$\mathrm{Var}(X_t) = \begin{cases} \sigma^2 \mu^{t-1} \frac{\mu^t - 1}{\mu - 1} & \text{if } \mu \neq 1 \\ t\sigma^2 & \text{if } \mu = 1 \end{cases}$$

Proof.

1. $\mathbb{E}[X_t] = \phi_t'(1)$. So by induction, using Lemma 0.1, $\phi_1'(1) = \phi'(1) = \mu$. $\phi_{t+1}'(1) = \phi'(\phi_t(1))\phi_t'(1)$. But $\phi_t(1) = 1$ and $\phi_t'(1) = \mu^t$ by the induction hypothesis.
2. For the variance, we reason similarly using $\mathrm{Var}(X_t) = \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2 = \phi_t''(1) + \phi_t'(1) - \phi_t'(1)^2$, and reasoning by induction.

$\square$

Theorem 0.2. *There exists a random variable $M \geq 0$ with $\mathbb{E}[M] < \infty$ and $\mathbb{P}(M = 0) \geq \zeta$, such that the limit*

$$\lim_{t \to \infty} X_t \mu^{-t} = M$$

holds a.s.

Proof. By the definition of X_t

$$\mathbb{E}[X_t \mid (X_{t-1}, X_{t-2}, \dots, X_0)] = \mu X_{t-1}$$

Therefore $M_t := X_t \mu^{-t}$ is a non-negative martingale. By Doob's convergence theorem, the limit M of M_t exists a.s. and is integrable. Whenever $\tau_0 < \infty$, X_t vanishes eventually, so $M = 0$ on $\{\tau_0 < \infty\}$. $\square$

Remark. The previous theorem establishes that $X_t \mu^{-t}$ stabilizes a.s. (possibly at 0) as $t \to \infty$:

$$X_t = \mu^{-t}(M + o_t(1))$$

as $t \to \infty$. This asymptotic behavior can be visualized through numerical simulatio: each time we sample a Galton-Watson process, we get a possibly (if $\mu > 1$) different limit M (i.e. the limit M is random). So on each simulation ω, we build the random variable $M(\omega)$, just taking the limit. If we want to estimate the distribution of M, we can just estimate numerically $\mathbb{P}(M \in A)$ using a Monte-Carlo method: we count the fraction of times (for T large enough), X_T/μ^T fell into the set A.

The numerical validity of Theorem 0.2 is immediate to grasp[2].

When $\mu = 1$ (the critical case) and extinction occurs with probability 1, one can study the behavior of the process conditioned on non-extinction. This result can be proved analytically using Lemma 0.1.

---

[2] Feel free to edit the code to include different sampling methods, resizing the editor.

Theorem 0.3 (Conditional Limit Theorem). *Consider a critical Galton-Watson process ($\mu = 1$, $\sigma^2 < \infty$). Then:*

$$\mathbb{P}(X_t > 0) \sim \frac{2}{\sigma^2 t} \quad \text{as } t \to \infty$$

Furthermore, conditional on $\{\tau_0 > t\}$, the normalized population size X_t/t converges in distribution to an exponential distribution:

$$(X_t/t \mid \tau_0 > t) \xrightarrow{d} \text{exp}\left(\frac{2}{\sigma^2}\right)$$

Extensions and Modern Applications

Generalizations

Galton-Watson processes serve as fundamental models with numerous generalizations, e.g.:

- **Multi-type Galton-Watson processes:** Each individual can be one of several types, with type-dependent offspring distributions. The process is now a vector-valued Markov chain.
- **Continuous-time branching processes:** The reproduction occurs at random times according to a Poisson process, leading to continuous-time Markov chains.
- **Spatial processes:** Each individual is assigned a location on space, and the offspring distribution depends on the location.

While the historical roots of the Galton-Watson process are in genealogy, the model describes effectively any system where "items" produce "copies" of themselves independently.

Epidemiology and Viral Spreading

In modern epidemiology, the offspring distribution represents the number of secondary infections caused by a single infected individual. The mean μ corresponds to the basic reproduction number, R_0. Unlike standard SIR models which assume deterministic mixing, branching processes capture the *stochasticity of early outbreaks*. They are particularly useful for modeling "super-spreading" events, where the offspring distribution (p_k) is heavy-tailed (e.g., p_0 is high, but p_k for large k is non-negligible). This helps calculate the probability that a virus introduced into a new population will die out naturally or cause an epidemic.

Polymerase Chain Reaction (PCR)

PCR is a molecular biology technique used to amplify a single copy of a DNA segment into millions of copies. In an ideal scenario, this is a binary Galton-Watson process with $p_2 = 1$ (deterministic doubling). In reality, the replication efficiency is $p < 1$. The process is modeled as a branching process where each DNA strand replicates with probability p or fails with probability $1 - p$. Branching process theory is used to estimate the variance of the final copy number, which is crucial for quantitative PCR (qPCR) accuracy. The basic Galton-Watson model must be refined to account for two physical realities:

- **Saturation:** The branching cannot grow exponentially forever. As the population X_t approaches the carrying capacity K (depletion of reagents), the reproduction mean $\mu(X_t)$ drops below 1.
- **Parameter Uncertainty:** The replication efficiency p is not known exactly, and may depend on the equipment conditions.

We model this by placing a prior distribution on the efficiency, for example $p \sim \text{Beta}(\alpha, \beta)$, and assuming a density-dependent offspring distribution. For a population size X_t, the effective efficiency becomes $p_{eff}(X_t) = p \cdot (1 - X_t/K)$.

If we run T cycles, for a successful detection, we need that the final amount of strands X_T passes a certain value $\bar{x}$. An important question is: *How many cycles T must we run to ensure detection?* In other words, we are interested in the hitting time $\tau := \inf\{t : X_t \geq \bar{x}\}$. Even if at larger values of X_t a more deterministic behavior is in place (due to the law of large numbers), the early branching phase and the uncertainty in p contribute to the variance of the random variable τ.

If $\pi = \text{Beta}(\alpha, \beta)$ is the prior on p, the probability of failure (false negative) at cycle T is:

$$\mathbb{P}(\text{False Negative}) = \int_0^1 \mathbb{P}(\tau > T \mid p)\pi(p)dp$$

As an example, we can we estimate the Cumulative Distribution Function (CDF) of τ. We sample p according to π, and for each sampled value we run a simulation of the Galton-Watson process with offspring distribution depending on X_t. For a fixed accepted risk of false negative $\varepsilon > 0$, we then select the run-time T^* such that $\mathbb{P}(\tau \leq T^*) \geq 1 - \epsilon$ (e.g., 99%).

Note that the variance in the “exponential phase” (seen as the horizontal spread of the curves) directly translates to the uncertainty in the quantitative estimate. This horizontal shift corresponds to the random variable M in the martingale limit $X_t \approx M\mu^t$. In the saturation regime, determining M allows us to “back-calculate” the initial viral load.

It is meaningful here to plot make a logaritmic plot:

- As long as X_t is small compared to the carrier capacity $X_t \ll K$, we are in a classical Galton-Watson regime (offspring distribution constant at each generation). Thus we have

  $$\log X_t \simeq \log M + t \log \mu$$

  We see this clearly in the log plot below, where we see over different simulations the “width” of $\log M$, while (after the very first iterations), a linear (for $\log X_t$, exponential for X_t) growth takes place with a coefficient independent of the sample.
- As X_t grows, we see the growth plateauing since we get closer to saturation and the offspring distribution loses efficiency.

Ideally, the prior distribution (the value of the parameters α and β) is provided by the manufacturer of the replication equipment. We can see the expected behavior emerge quite clearly in the Galton-Watson simulation. Actually we get values very similar to those used in laboratories for detection of low viral charges (as we are starting with just one)[3].

Algorithmic Analysis

In computer science, Galton-Watson processes appear in the probabilistic analysis of algorithms and data structures. For instance, the *height of random binary search trees* or the recursion depth of the Quicksort algorithm can be analyzed by embedding these discrete structures into continuous-time branching processes. The “offspring” here correspond to the recursive sub-problems generated by a partition step.

---

[3] Feel free to edit the code resizing the editor box below.

Neutron Transport Theory

Although an older application, it remains vital for nuclear safety. In a nuclear reactor, a neutron colliding with a nucleus may induce fission, releasing more neutrons. This is a Galton-Watson process where μ is the criticality factor. - $\mu < 1$: Subcritical (reaction dies out). - $\mu = 1$: Critical (steady power output). - $\mu > 1$: Supercritical (power increases exponentially, potentially dangerous). Safety calculations rely on computing the probability of extinction (fission chain termination) versus exponential runaway.