Part I
Branching Processes

1 17.3.2025

We start with a probability space: $\mathcal{F} = (\Omega, \sigma, \mathbb{P})$. $a \in \Omega$ is a single event in our game, such as the result of a roll of a die. σ is a σ-algebra, over which we define our probability. If we can, we will use the power set of Ω, also known as 2^{Ω}. Otherwise, we will use whatever algebra makes sense. P would be the probability measure.

Usually, we will have $\Omega = \mathbb{N}$, where we can use the power set, and when $\Omega = \mathbb{R}$, we will use the Borel measure.

1.1 Random Variables

We are interested in functions $X : \Omega \to \mathbb{R}$, because we mostly count things, although in general we could replace $\mathbb{R}$ with other spaces that are measurable.

We call X a *random variable* if for any I open in $\mathbb{R}$, $X^{-1}(I) \in \sigma$.

If X, Y are random variables, $X = Y$ will denote functional equivalence, and $X \overset{l}{=} Y$ will denote equivalence *by law*, which is defined as:

$$X \overset{l}{=} Y \Leftrightarrow \forall A \in \sigma \quad \mathbb{P}\left[X\left(A\right) \subset I\right] = \mathbb{P}\left[Y\left(A\right) \subset I\right]$$

1.2 Moments

The *probability distribution function of* X (its PDF), denoted as p_X can be given as:

$$p_X\left(s\right) = \mathbb{P}\left[X = s\right]$$

which for discrete random variables will be p_n as a series, and we'll hide X to make it simpler to write.

For a continuous variable we would use the CDF $\mathbb{P}\left[X \leq s\right]$, and use its differential measure.

The average, or expectation of a random variable X is given by:

$$\bar{X} = \mathbb{E}\left[X\right] = \int p_X\left(s\right) s ds$$

We can have other random variables by setting functions over random variables, and then:

$$\overline{f\left(X\right)} = \mathbb{E}\left[f\left(X\right)\right] = \int f\left(s\right) p_X\left(s\right) ds$$

The *n-th moment* can be similarly written as:

$$M_n\left(X\right) = \mathbb{E}\left[X^n\right]$$

The *n-th central moment* can be written as:

$$C_n(X) = \mathbb{E}\left[(X - \bar{X})^n\right]$$

The first central moment is just 0, and the second is the *variance* $V[X]$, which has to do with how wide the distribution is. The third, called *skewness*, is the degree of how tilted the distribution is around its expected value.

The *n-th factorial moment* can be written as:

$$F_n(X) = \mathbb{E}\left[\prod_{i=0}^{n-1}(X - i)\right]$$

It will come in handy later in this course, which is when we'll discuss it more.

Theorem (which we don't prove here)

If two random variables X, Y have the same moments $\forall n\ M_n(X) = M_n(Y)$, then $X \overset{l}{=} Y$.

If we only had a finite set of equal moments, we would have an infinite number of different distributions that fit, which is a classic interpolation problem, sometimes called the *hamburger problem*.

1.3 Random Processes

Given an index set I with elements i, X_i is a *random process* if for any $i \in I$, X_i is a random variable. The denotation seems confusing, but it's better to keep it simple, because we will have these all over the place.

I will usually be the generation number if we have discrete time, so $\mathbb{N}$, or $\mathbb{R}^+$ if we talk about an index of time for continuous processes.

Distributions are not everything in processes. Often, in random variables, we really only care about equality by law, because things that are similarly distributed are rather the same.

In random processes, we could have dependence, which would ruin this, as the game we play at stage n could very likely depend on the result at earlier stages. In this case, knowing the total distribution at stage n isn't enough, because we care about its dependence with other stages, so conditional probabilities matter.

The *autocorrelation of* X is given by:

$$\varphi_X(s,t) = \mathbb{E}[X_s X_t] - \mathbb{E}[X_s]\,\mathbb{E}[X_t]$$

Usually, we care about how strong the correlation is depending on $s - t$, such that correlations drop off as $s - t$ grows larger.

This has connections with Ergodicity, where taking ensemble averages is the same as taking samples from different far times.

A random process t_n will be called a *time series* if $n \in \mathbb{N}$, $t_n \in \mathbb{R}^+$and $\forall n\ t_n < t_{n+1}$. A time series would matter as the waiting time between events, mostly. We would care about the distribution of *dwell times* of t, $\theta_n = t_{n+1} - t_n$ a lot. A second question would be how much time we must wait until the n-th event, which is simply t_n.

Given a time series we can get the dwell times, and given a series of positive random variables we can make them into a time series by adding them up.

A time series (defined by dwell times θ_n, because of the above equivalence) will be called a *renewal process* if $\{\theta_n\}$ are i.i.d.

The *counter of a time series* t_n will be defined by:

$$C_s = \max\{n | t_n \leq s\}$$

It just tells us how many events happened by time t. This is the inverse indexing of knowing when the n-th event happened, which is t_n.

1.4 Markovian and Stationary Processes

A Markovian process is a deep concept which takes its own course. These are processes where we can actually solve things.

A process X_t will be called a *Markov process* if for any $s > t > u$, $P(X_s = x | X_t = y) = P(X_s = x | X_t = y, X_u = z)$.

This means that any knowledge we need is the newest information. After we have new information, we simply don't care what happened before.

A non-Markovian process is for example, a dead time system with paralysis, since we need to know the last time we dropped.

A process X_t will be called a *stationary process* if for any $s > t$ we have $P(X_s = x | X_t = y) = F(x, y, s - t)$. This means that a stationary process is obviously markovian.

Given a time series t_n, and its counter C_t, a function $\lambda : \mathbb{R} \to \mathbb{R}$ will be called the *rate of* C_t if $\mathbb{P}[C_{t+\delta t} = C_t + 1] = \lambda(t)\, dt + o(dt)$, in the sense that:

$$\lim_{\delta t \to 0} \frac{\mathbb{P}[C_{t+\delta t} = C_t + 1]}{\lambda(t)\, dt} = 1$$

A time series will be called *ageless* if its counter's rate is a constant function $\lambda(t) \equiv \lambda$.

Theorem

> The dwell time for the first event in an ageless time series is an exponential random variable.

Let t_n be an ageless time series, and therefore $t_1 = \theta_0$ is the first dwell time as well. Define $F_t(s) \equiv \mathbb{P}[t \leq s]$.

$$\begin{aligned} F_t(s + dt) - F_t &= \mathbb{P}[t \in (s, s + \delta t]] = (1 - F_t(s))\, \mathbb{P}[C_{t+\delta t} = C_t + 1] \\ &= (1 - F_t(s))\, (\lambda \delta t + o(\delta t)) \end{aligned}$$

And therefore:

$$\frac{dF_t}{ds}(s) = \lambda(1 - F_t(s))$$

The solution to which, given that $F_t(0) = 0$ since nothing can happen in no time, exists and is unique (Poincare). It just happens to be $F_t(s) = 1 - e^{-\lambda s}$. You can derive and check for yourself.

Since F is the CDF of the first dwell time, t is an exponential random variable by law. □

1.5 Probability Generating Function

Given a random variable X, with values in $\mathbb{N}$, its *probability generating function* (PGF) is defined by $\mathbb{E}\left[x^X\right]$. We will denote it as $H(X)$.

By definition, $\mathbb{E}\left[x^X\right] = \sum p_n x^n$.

The continuous version would be a transform, kind of like Fourier analysis.

Proposition

The PGF is an analytic function over the unit circle ($x < 1$).

Let $|x| \leq 1$. Thus $|x^n| \leq 1$, and $|x^n p_n| \leq p_n$.

Obviously, $\sum_{n=0}^{k} p_n$ is monotonically increasing and bounded by 1, and thus converges (to 1, as a probability). It is a Majorant for the PGF, and therefore the PGF $\sum p_n x^n$ converges uniformly over the unit circle (Weierstrass M-test). □

The function is analytic over the unit circle, and therefore all its derivatives are defined.

Proposition

If $\lim_{x\to 1} \frac{d^n H}{dx^n} \equiv \frac{d^n H}{dx^n}|_{x=1}$ exists and is finite, then $\frac{d^n H}{dx^n}|_{x=1} = F_n(x)$.

$$\frac{d^n H}{dx^n}|_{x=1} = \lim_{x\to 1} \sum p_n \prod_{k=0}^{n-1} (n-k) x^{n-1} = \sum p_n \prod_{k=0}^{n-1} (n-k) \equiv F_n(X)$$

Where inserting the derivatives into the sum uses our uniform convergence. □

Where doesn't this work? We play until we hit heads, and we get money for 2^n where n is the number of tails we get. The probability to get n is $p_n = 2^{-n}$. Thus:

$$\mathbb{E}[X] = \sum p_n 2^n = \sum 1 = \infty$$

And the PGF goes as:

$$H(X) = \sum 2^{-n} x^n$$

But this has no derivative in $x = 1$.

Observation

If $H(x)$ is analytic over the unit circle, written as a power series $H(x) = \sum a_n x^n$ with $a_n > 0$ and $\sum a_n = 1$, then there is some random variable such that H is its PGF, which is defined by the distribution a_n.

Denotation We denote the coefficient of x^n in a McLauren power series of H as $[H(X)]_n = \sum_{\sum n_j = n} \prod_{i=1}^{j} a_{n_i}$

Corollary

If $n \in \mathbb{R}$ and $H(X)$ is a PGF, then $H^n(X)$ is a PGF.

$H^n(X)$ is analytic by composition. $H^n(1) = 1 = \sum b_n$. $b_n > 0$ since it is simply given as multiplications of a_n which are all positive.

$H^n(X) \Rightarrow \sum_{i=1}^{n} X_i$ for X_i which are i.i.d with X.

2 24.3.2025

2.1 Some more preparation

We are almost getting to what a branching process is! We dealt with the PGF last time, and it's going to take us far.

Reminder: If $X : \mathbb{N} \to \mathbb{R}$ is a random variable, we define:

$$H_X(x) = \mathbb{E}\left[x^X\right] = \sum_n p_n x^n$$

Proposition

If X, Y are independent, then

$$H_{X+Y}(x) = H_x(x) H_Y(x)$$

We denote $a_n = \mathbb{P}[X = n]$, $b_n = \mathbb{P}[Y = n]$.

$$\mathbb{P}[x + y = n] = \sum_{k=0}^{n} \mathbb{P}[X = k]\,\mathbb{P}[Y = n - k] = \sum_{k=0}^{n} a_k b_{n-k} = [H_X(x)\, H_Y(x)]_n$$

□

A different proof, which is even shorter, is:

$$H_{X+Y}(x) = \mathbb{E}\left[x^{X+Y}\right] = \mathbb{E}\left[x^X x^Y\right] = \mathbb{E}\left[x^X\right]\mathbb{E}\left[x^Y\right] = H_X(x)\, H_Y(x)$$

□

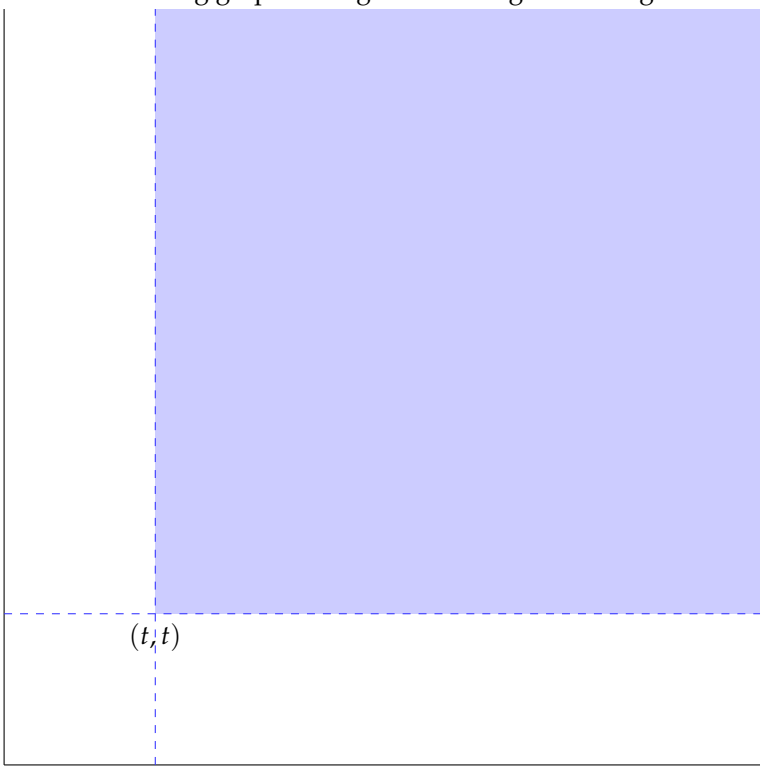
If X is an exponential random variable with parameter λ

$$\mathbb{P}[x \leq t] = \int_0^t \lambda e^{-\lambda s} ds$$

and Y is an independent exponential random variable with parameter Λ, then:

$$\min(X, Y) \sim Exp(\lambda + \Lambda)$$

This can be proven graphically. Take a square of in an X, Y plane, where they are both less than some t. The space that excludes the opposite square in that half place across from (t, t) is where the minimum is under t. This can be seen in the following graphic image of the integration range:



Therefore:

$$\mathbb{P}[\min(X, Y) \leq t] = 1 - \iint_{\mathbb{R}^{2+} \setminus D} \lambda \Lambda e^{-\lambda s} e^{-\Lambda \omega} ds d\omega = \int_t^\infty ds \int_t^\infty \lambda \Lambda e^{-\lambda s} e^{-\Lambda \omega} d\omega = 1 - e^{-(\lambda + \Lambda)t}$$

And therefore its CDF is that of a random exponential variable with parameter $\lambda + \Lambda$.□

Let X, Y be random variables. The law of dependent probability is just:

$$\mathbb{E}[X|Y=y] = \sum \mathbb{P}[X=n|Y=y]\, n$$

Mishpat Hahachlaka (the law of total expectation) says that:

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

This is a form of Fubini's theorem, obviously, that we can take the integral in whichever order we like.

The law of total variance says that:

We need of course for the expectation and variance to both exist, for this to make sense.

$$V[X] = \mathbb{E}[V[X|Y]] + V[\mathbb{E}[X|Y]]$$

And NOW, we can finally get to branching processes, but we'll start with Galton-Watson Trees, even though we won't actually use them much as a formalism.

2.2 Galton-Watson Trees

They were an economist and a mathmatician who tried to figure out why the number of different noble family names was dropping off.

We follow a series G_n of trees in the graph theory sense $G_n = (P_n, V_n)$, and a color function $f_n : P_n \to \{0,1\}$. 1 marks a *living* node, and 0 a dead node We always take $G_1 = (\{p_0\}, \phi)$, $f_1(p_0) = 1$. At any stage in the series we sample the number of new nodes and edges we will introduce for each node in $f_n^{-1}(1)$, and we add those as leafs. We set $f_n^{-1}(1) = \{p | p \in P_n \backslash P_{n-1}\}$.

We still need to define how we sample the new progeny, of course, for this to be complete.

The series G_n will be defined by the series of distributions $\{\nu_n(k)\}$ where $\nu_n(k) = \mathbb{P}[\text{Getting } k \text{ progeny for any live node in } G_n]$. We can have this distribution depend on the size of $f_n^{-1}(1)$. We won't consider non-Markovian cases in this course.

We move on to a modern definition that is more common.

2.3 Branching Processes (Finally!)

A random process X_n will be called a *Discrete Branching Process* if there exist random variables $\xi_{i,n}$ that give values in $\mathbb{N}$, which are all i.i.d, such that:

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{i,n}$$

A random process X_n will be called a *Branching Process with Immigration* if it follows:

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{i,n} + I_n$$

where I_n is some random variable with values in $\mathbb{Z}$, but limited to not causing a negative X_{n+1}, to avoid summing over negative populations.

A discrete branching process will be called *stationary* if $\xi_{i,n}$ are i.i.d for all n, so they can be written as $\xi_{i,n} \sim \xi$. Specifically, they do not depend on the population size X_n, nor n.

The dynamics is then dependent only on a single law ξ, with a probability vector $\mathbb{P}[\xi = n] = p_n$. If we have immigration, we also demand $I_n \sim I$ be i.i.d.

The Microbe Problem

We start with an initial population $n_0 = 1$ (i.e. $X_0 = 1$ with probability 1). With probability p it will mitosis (become 2), with $1 - p$ it will die.

What is the distribution of the number of microbes in generation n?

We define $Q(z) = (1-p) + pz^2$, which is the PGF for the number of progeny of a single microbe. We denote the compunding of a function Q on itself k times as $Q^{[k]}$.

We denote $H_n(x)$ as the PGF for the n-th generation. $H_0(x) = x$.

$$H_n(x) = \sum a_{n,k} x^k$$

By definition.

We recursively know that, by the law of total probability based on what happened in generation 1:

$$\begin{aligned} a_{n,k} &= \mathbb{P}[k \text{ in generation } n | \text{mitosis}]\, p + \mathbb{P}[k \text{ in generation } n | death]\,(1-p) = \\ &= p \sum_{j=0}^{k} a_{n-1,j} a_{n-1,k-j} + \delta_{0,k}(1-p) \\ &= p\left[H_{n-1}^2(x)\right]_k + + \delta_{0,k}(1-p) \end{aligned}$$

We multiply by x^k and sum over them:

$$H_n(x) = pH_{n-1}^2(x) + (1-p) = Q(H_{n-1}(x))$$

since in the first generation we either die off to 0 and stay there, or we have two individuals that need to compound in $n-1$ generations.

$$H_n(x) = Q^{[n]}(x)$$

And therefore:

$$\begin{aligned} H_0 &= x \\ H_1 &= px^2 + 1 - p \\ H_2 &= p^3 x^4 + 2p^2(1-p)x^2 + p(1-p)^2 + (1-p) \end{aligned}$$

And so on and so forth.
What is the expectation and variance?

$$M_1(n) = \mathbb{E}[X_n] = \frac{dH_n}{dx}|_1 = \frac{dH_n}{dH_{n-1}}\frac{dH_{n-1}}{dx}|_1 = 2pH_{n-1}|_1\mathbb{E}[H_{n-1}] = 2pM_1(n-1) = (2p)^n 1 = (2p)^n$$

We can divide this into 3 cases, based on the ordering of $2p$ and 1, into subcritical, critical and supercritical regimes $M_1(n) \underset{n\to\infty}{\to} \begin{cases} 0 & 2p < 1 \\ n_0 & 2p = 1 \\ \infty & 2p > 1 \end{cases}$

Similarly, we can do the variance:

$$F_2(n) = V(X_n) = \frac{d^2H_n}{dx^2}|_1 = \left[Q''(H_{n-1})\left(H'_{n-1}\right)^2 + Q'(H_{n-1})H''_{n-1}\right]_1 = 2pM_1^2(n-1) + 2pF_2(n-1)$$
$$= 2p(2p)^{2(n-1)} + 2pF_2(n-1) = (2p)^{2n-1} + 2pF_2(n-1) = 2pF_2(n-1) + \frac{1}{2p}\left((2p)^2\right)^n$$

This is an inhomogeneous recursion relation of the form:

$$F_2(n) = F_{2,h}(n) + F_{2,p}(n)$$
$$F_{2,h}(n) = 2pF_{2,h}(n-1) \Rightarrow F_{2,h} = C(2p)^n$$

Reminder, generally the particular solution of the equation:

$$a_n = qa_{n-1} + \gamma^n A$$

is of the form $a_{n,p} = B\gamma^n$ unless $\gamma = q$, in which case $a_{n,p} = Bn\gamma^n$. Thus, for $2p \neq 1$, we get:

$$F_{2,p}(n) = B\left[(2p)^2\right]^n$$

When $2p = 1$, $2p = (2p)^2$ and we need the other particular solution, and we get:

$$F_{2,p}(n) = Bn$$

Which results in a particular solution that is solved to:

$$Bn = B(n-1) + 1 \Rightarrow B = 1$$

And therefore:

$$F_2(n) = C + n$$

where C comes from boundary conditions, so for our case $C = 0$, because $F_2(0) = \mathbb{E}[X_n(X_n - 1)] = \mathbb{E}\left[X_n^2\right] - 1 = 0$.

$$F_2(n) = n$$

The variance is simply:

$$V(X_n) = \mathbb{E}\left[X_n^2\right] - \mathbb{E}^2[X_n] = n$$

This is the irregular case where the mean does not actually describe a likely case of the population. There will be many populations that grow and multiple populations that die off, and they cancel our **on average**, but the behavior of the population is a cone that starts at n_0 and moves up or down with a random walk.

2.4 Generalized Microbe Problem

Now, when we mitosis, we have a probability distribution for our progeny, which we will name *multiplicity*, and mark it $\{p_\nu(n)\}_{n=0}^{\infty}$. We assume that all its moments exist and are finite. Death is now a part of life, such that mitosis to 0 is death.

We still have, in the exact same way:

$$a_n(k) = \delta_{k,0} p_\nu(0) + \sum_{l=0}^{\infty} p_\nu(l) \sum_{\sum_{j=1}^{l} k_j = k} \prod_{j=0}^{l} a_{n-1}(k_j)$$

$$a_n(k) = \delta_{k,0} p_\nu(0) + \sum_{l=0}^{\infty} p_\nu(l) \left[H_{n-1}^l\right]_k$$

$$H_n(x) = Q^{[n]}(x)$$

where Q is the new multiplicity PGF.

We still get, because we never used anything comples:

$$M_1(n) = \left(\sum \nu p_\nu\right) M_1(n-1) \equiv \bar{\nu} M_1(n-1) = \overline{\nu}^n n_0$$

So we still get the same criticality behavior, based on the sign of $\overline{\nu} - 1$.

Definition We call a stationary branching process *subcritical, critical and supercritical* based on $\overline{\nu}$ being less than, equal or greater than 1.

3 31.3.2025

> I did not attend this lecture, but rather wrote this off of Tomer's notes. Thanks, Tomer!

Last time we studied stationary branching processes, which are uniquely defined by their multiplicity $\{p_\nu(n)\}_{n=0}^{\infty}$. We saw that for an initial population $n_0 = 1$, the PGF for the population at generation n was given recursively by:

$$g_{n+1}(x) = Q(g_n(x))$$

where Q is the PGF for the multiplicity.
We also saw that $\mathbb{E}[X_n] = \overline{\nu}^n$.
Let us now compute the variance. For the first moment we had:

$$\frac{dg_{n+1}}{dx}(x) = \frac{d}{dx}\left[\sum p_{\nu,k} g_n^k(x)\right] = Q'(g_n(x))\, g_n'(x)$$

Thus

$$\frac{d^2 g_{n+1}}{dx^2}(x) = \frac{d}{dx}\left[Q'(g_n(x))\, g_n'(x)\right] = Q''(g_n(x))\left[g_n'(x)\right]^2 + Q'(g_n(x))\, g_n''(x)$$

Setting $x = 1$ we get:

$$F_2(n+1) = \overline{\nu(\nu-1)} M_1^2(n) + \overline{\nu} F_2(n) = \overline{\nu(\nu-1)}\overline{\nu}^{2n} + \overline{\nu} F_2(n)$$

For $\overline{\nu} \neq 1$ we solve the homogeneous and particular solutions:

$$a_h(n) = \overline{\nu}^n, \quad a_p(n) = A\overline{\nu}^{2n}$$

Solving for A we get:

$$A\overline{\nu}^{2n+2} = A\overline{\nu}^{2n} + \overline{\nu(\nu-1)}\overline{\nu}^{2n}$$

And thus:

$$A\overline{\nu}^2 = A + \overline{\nu(\nu-1)}$$

$$A = \frac{\overline{\nu(\nu-1)}}{\overline{\nu}^2 - 1}$$

For $\overline{\nu} = 1$, $M_1(n) = 1$ and so:

$$F_2(n+1) = F_2(n) + \overline{\nu(\nu-1)}$$

Again:

$$a_h(n) = 1, \quad a_p(n) = An$$

which is solved for A similarly to be:

$$A = \overline{\nu(\nu-1)}$$

Which results in a simple form for F_2:

$$F_2(n) = C + n\overline{\nu(\nu-1)}$$

Which, when we translate to a variance would be:

$$V[X_n] = \mathbb{E}\left[X_n^2\right] - \mathbb{E}^2[X_n] = \mathbb{E}\left[X_n^2\right] - 1$$

$$F_2(n) = \mathbb{E}[X_n(X_n - 1)] = \mathbb{E}\left[X_n^2\right] - \mathbb{E}[X_n] = \mathbb{E}\left[X_n^2\right] - 1 = V[X_n]$$

And thus, since

$$V[X_0] = 0 \Rightarrow C = 0$$

So finally

$$V[X_n] = n\overline{\nu(\nu-1)} = n\left(\overline{\nu^2} - 1\right) = n\sigma^2$$

where σ^2 is the variance of the multiplicity.

3.1 Example

Let $p_{\nu,i} = \frac{1}{3}\sum_{i=0}^{2}\delta_{ij}p_{\nu,j}$. What is the criticality of the system? What is g_2 if $n_0 = 1$? What is g_2 assuming $n_0 = 1000$?

3.1.1 Criticality

$$\overline{\nu} = \frac{1}{3} + 2\frac{1}{3} = 1$$

So it is critical.

3.1.2 $g_2, n_0 = 1$

$$g_0(x) = x$$

$$g_1(x) = Q(g_0(x)) = Q(x) = \frac{1}{3}\left(1 + x + x^2\right)$$

$$g_2(x) = \frac{1}{3}\left(1 + g_1(x) + g_1^2(x)\right) = \frac{1}{3}\left(1 + \frac{1}{3}\left(1 + x + x^2\right) + \frac{1}{9}\left(1 + x + x^2\right)^2\right)$$

Which you can simplify, if you must.

3.1.3 $g_2, n_0 = 1000$

We know that if X, Y are i.i.d. random variables, $g_{X+Y} = g_X g_Y$, and so $g_{\sum X_i} = \prod g_{X_i}$ in general, if $\{X_i\}$ are i.i.d. random variables. More specifically:

$$g_{\sum X_i} = g_{X_1}^{n_0}$$

And thus:

$$g_{n_0=1000} = \left[\frac{1}{3}\left(1 + \frac{1}{3}\left(1 + x + x^2\right) + \frac{1}{9}\left(1 + x + x^2\right)^2\right)\right]^{1000}$$

3.2 Stationary branching processes with immigration

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{i,n} + I_n$$

We denote the multiplicity of ξ as p_ν, as before, and that of I as p_i. The PGFs are denoted as Q, I respectively. We denote the moments of I with symbols such as $\overline{\nu}_i$ for the first.

Theorem

Denote g_n as the PGF of the n-th generation, given $X_0 = 1$. Therefore:

$$g_{n+1}(x) = g_n(Q(x))\, I(x)$$

Proof By definition:

$$g_{n+1}(x) = \mathbb{E}\left[x^{X_{n+1}}\right] = \mathbb{E}\left[x^{\sum_{i=1}^{X_n}\xi_{i,n}+I_n}\right] = \mathbb{E}\left[x^{\sum_{i=1}^{X_n}\xi_{i,n}} x^{I_n}\right]$$

using the law of total probability:

$$\begin{aligned} g_{n+1}(x) &= \mathbb{E}\left[\mathbb{E}\left[x^{\sum_{i=1}^{X_n}\xi_{i,n}} x^{I_n} | X_n\right]\right] = \mathbb{E}\left[\mathbb{E}\left[x^{\sum_{i=1}^{X_n}\xi_{i,n}}\right] | X_n\right] \mathbb{E}\left[x^{I_n}\right] \\ &= \mathbb{E}\left[Q^{X_n}(x)\right] \mathbb{E}\left[x^{I_n}\right] = g_n(Q(x))\, I(x) \end{aligned}$$

□

Corollary

$$M_1(n) = \begin{cases} \left(1 - \frac{\overline{\nu_i}}{1-\overline{\nu}}\right)\overline{\nu}^n + \frac{\overline{\nu_i}}{1-\overline{\nu}} & \overline{\nu} \neq 1 \\ 1 + n\overline{\nu_i} & \textit{otherwise} \end{cases}$$

Proof

$$\frac{dg_{n+1}}{dx} = g_n'(Q)\, Q' I + g_n(Q)\, I'$$

Set $x = 1$ and we get:

$$M_1(n+1) = M_1(n)\,\overline{\nu} + \overline{\nu_i}$$

If $\overline{\nu} \neq 1$, we solve the homogeneous and inhomogeneous parts. The homogeneous is still $C\overline{\nu}^n$. The particular solution we assume constant at A, which solves for $A = \frac{\overline{\nu_i}}{1-\overline{\nu}}$. We use the initial condition to be 1, to get the solution above.

If $\overline{v} = 1$, $M_1(n+1) = M_1(n) + \overline{v_i}$, so the homogeneous solution is constant this time around. The particular solution we assume to be An, which solves as $A = \overline{v_i}$. With the initial condition, we get the solution above.

□