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:

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

The first central moment is just 0, and the second is the *variance* $V\left[X\right]$, 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:

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

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 $\forall n \; M_n\left(X\right) = M_n\left(Y\right)$, 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:

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

□

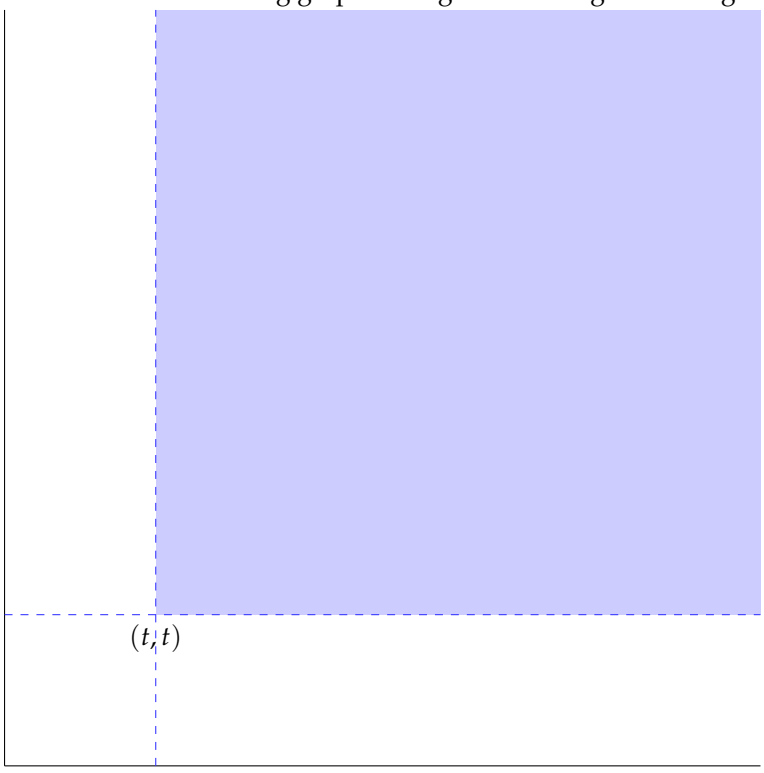
If X is an exponential random variable with parameter λ

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

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:[1]



[1] Thanks to Aviv Barnea for the tikz code for this image

Therefore:

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

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:

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

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

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

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.

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

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 $\nu_n(k)=\mathbb{P}$ [Getting k progeny for any live node in G
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 $\mathbb{P}\left[\xi=n\right]=p_n$. 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:

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

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

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

This behavior is expected: Exponential growth or decay, and a linear growth for critical cases, as the system's expectation meets the deterministic case. The exponential case is also always larger than a population with no immigration.

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

□

4 7.4.2025

4.1 The original Galton-Watson paper

Today we will finish our discussion on discrete branching processes. We will discuss the original Galton-Watson paper.

Let a discrete, stationary branching process X_n, with multiplicity $\{p_\nu\}_{\nu=0}^{\infty}$. The extinction probability, $\mathbb{P}[X_n = 0]$, we denote as $p_n(0)$, just to confuse you with all these ps.

Note that $p_n(0) \leq 1$, as a probability, and with no immigration it is a sink case, so $p_n(0) \leq p_{n+1}(0)$. Therefore as a bounded series that grows monotonically, its limit exists and is finite. We denote $\lim_{n\to\infty} p_n(0) \equiv p_\infty(0)$.

We already saw that the PGF $H_{n+1} = Q(H_n)$. Since $H_n(0) = p_n(0)$, we can get a recursion relation for $H_n(0)$:

$$p_{n+1}(0) = Q(p_n(0))$$

Therefore, since the limit exists:

$$p_\infty(0) = Q(p_\infty(0))$$

We want the roots of $q(x) = Q(x) - x$ in $[0,1]$. Notice that 1 is always a solution to this equation, because Q is a PGF. At $x = 0$ this is always non-negative.

We derive q:

$$\frac{dq}{dx} = \left(\sum_n p_{\nu,n} n x^{n-1}\right) - 1 \leq \left(\sum_n p_{\nu,n} n\right) - 1 = \overline{\nu} - 1$$

This gives us two cases:

1. Supercritical system, $\overline{\nu} > 1$, there can be additional roots in $[0,1]$ except 1. Basically, we don't actually help in this case directly.
2. Critical or subcritical system, $\overline{\nu} \leq 1$, there can only be one root, because the function is strictly monotonically decreasing. Since $p_\infty(0) = 1$ is a root, it is the *only* root.

With a population of size 1, therefore if you are not supercritical, you will go extinct with probability 1 at some point. What happens if you start with a population of size n_0? The PGF is simply the product of all contributions, so the extinction probability is just $p_\infty^{n_0}(0)$. If you are critical, remember that $\mathbb{E}[X_n] = n_0$, and by definition $\lim_{n\to\infty} \mathbb{E}[X_n] = n_0$.

This is confusing, because we don't actually meet the expectation. What really happens is that many chains will die out, but some of them will be very large. The fraction of surviving initial trees will decay, even if the population is actually stable, because the population will be dominated by the few lucky trees who survive.

And that's the Galton-Watson result from their paper.

4.2 Continuous Branching Processes

We are considering a discrete population in continuous time. So far we used the generation system to define our dynamics. We need to move on to dwell times that we previously mentioned.

Now, if we take our Galton-Watson tree, we don't define the next generation for all items simultaneously. Instead, each individual has its own dwell time until it becomes inactive and creates progeny. The problem we are going to encounter is that we want the dwell time and multiplicity of individuals to maybe depend on the current population. Thus, we would need to somehow update the dwell times of individuals and their multiplicity whenever another individual does anything.

Let's try some naive ways to do it:

$$X_{t+dt} = \sum_{i=1}^{X_t} \xi_i$$

The problem is that we need to somehow make sure there are no additional contributions from progeny born in $[t, t+dt]$. This will fail as a definition.

We start with the generation definition of $X_{n+1} = \sum_{i=1}^{X_n} \xi_i$.

We can define $Y_t = X_{n_t}$ where n_t is a counter of a time series. However, we want the rate of events to increase with increasing population size, so while this definition works, it is rather unworkable for most applications. So that's too naive too.

The actual definition we are going to use takes a lot of mathematical proofs to do, and we're not going to do all of them. We are going to wave our hands a bit, bear with me. If you want to know the details, read the book, I will put a link here. Given two random processes, we want to define a stochastic integral. Let X_t, w_t be random processes, and we want $\int_a^t X_s dw_s$.

Let a partition $S = \{t_i\}$ of $[a, t]$. We define $ind(S) = \max\{t_{i+1} - t_i\}$. We take the Riemann sum $I^I(S) = \sum_{i=1}^{n-1} X_{t_i}\left(w_{t_{i+1}} - w_{t_i}\right)$. This is a random variable, and it would also differ if we take $X_{t_{i+1}}$ there, which is weird. The stochastic integral would be:

$$\int_a^t X_s dw_s = \lim_{Ind(S)\to 0} I^I(s)$$

What this requires is that w_t be a Martingale, and that X_t be not too awful. Read the book for the details, because like I mentioned, stochastic integrals are a whole course to themselves.

The $I^I(s)$, like we said, is defined with X_{t_i} in the sum, and is the Ito integral. If we take $X_{\frac{t_{i+1}+t_i}{2}}$, we get the Stratonovich integral. They are not the same. We go with the Ito definition because the resulting process is a Martingale. But again, we don't actually explain any of this, and you need to read the book (or I may take some notes of that book in an appendix to these notes when I do so myself).

4.2.1 Simple Example

Let $w_t = t$ be the deterministic time we are used to, and X_t be discrete. In this case, it is clear why the integral converges, because it is simply the sum of the population so far.

Our goal is to define the branching process, where each individual has a reaction rate $\lambda_\nu(X_t, t)$ to branch in to ν individuals. This is the multiplicity analog from before, where we combine both the number of progeny and the time rate of this type of effect.

Therefore, for each individual, the probability for a reaction that creates ν individuals in an interval $[t, t+dt]$ is $\lambda_\nu(X_t, t)\, dt + o(dt)$.

We want to look at $X_t - X_0$. We take some partition S as before of the time interval $[0, t]$. We can write:

$$X_t - X_0 = \sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right) = \sum_{i=1}^{n}\sum_{\nu=0}^{\infty}(\nu-1)\, Y_\nu\left[X_{t_i}\lambda_\nu(X_{t_i}, t_i)\, dt + o(dt)\right]$$

Where Y_ν are independent Poisson random variables. This is because Y_ν are the counters for the exponential events with rates $X_{t_i}\lambda_\nu(X_{t_i}, t_i)$, so they count the number of events in which the population change is $(\nu-1)$, and we sum over all possible options. The fact that we can do a sum over all possible options might demand something on the rates λ_ν for this to exist, but we're going to ignore that in this course for now. Use finite progeny numbers if you're feeling dirty.

We'll change the order of summation (again, we're not explaining what must happen for this to work):

$$X_t - X_0 = \sum_{\nu=0}^{\infty}(\nu-1)\sum_{i=1}^{n} Y_\nu\left[X_{t_i}\lambda_\nu(X_{t_i}, t_i)\, dt + o(dt)\right] = \sum_{\nu=0}^{\infty}(\nu-1)\, Y_\nu\left[\sum_{i=1}^{n} X_{t_i}\lambda_\nu(X_{t_i}, t_i)\, dt + o(dt)\right]$$

And now we can write:

$$X_t - X_0 = \sum_{\nu=0}^{\infty} (\nu - 1) Y_\nu \left[\int_0^t X_s \lambda_\nu (X_s, s) \, ds \right]$$

So now can define things. X_t will be called a *continuous branching process* if there exist a series of **positive** functions $f_k(x,t)$ such that:

$$X_t = X_0 + \sum_k (k-1) Y_k \left[\int_0^t f_k (X_s, s) \, ds \right]$$

For independent Poisson random variables Y_k.

5 21.4.2025

Last week was Passover, and next week is M&C2025, so we are sparsely having classes for now.

We will also introduce diffusion approximations in this course because 2/3 students are doing this.

We are changing the test at the end of the course to a home test.

And now back to continuous branching processes.

5.1 The Master Equation

We remind that a branching process is given by:

$$X_t = X_0 + \sum_{l=-1}^{N} l Y_l \left[\int_0^t f_l (X_s, s) \, ds \right]$$

And we have this be a finite sum just because we don't want to deal with convergence because that's not the point here.

The branching laws f_l are those who govern the distribution.

We define $\lambda_l (X_t, t)$ as the rate per capita if $f_l (X_t, t) = \lambda_l (X_t, t) X_t + S_l (t)$. This should be taken as the rate per time for an individual in the population to undergo a reaction that yields an l-change in the population, in the short time frame $[t, t+dt]$. Therefore the probability is $\mathbb{P}$ [one individual doing l-interaction in $[t, t+dt]$] $= \lambda_l (X_t, t) \, dt + o(dt)$.

We now define, for any $t_2 > t_1$:

$$H(t_1, t_2, n, k) = \mathbb{P} [X_{t_2} = n | X_{t_1} = k]$$

The backward Master Equation is therefore (The complete probability theorem):

$$\mathbb{P} [X_t = n] = \sum_k \mathbb{P} [X_{t-dt} = k] \, H (t - dt, t, n, k)$$

The forward Master Equation is similarly:

$$\mathbb{P} [X_t = n] = \sum_k \mathbb{P} [X_{dt} = k] \, H (dt, t, n, k)$$

In a stationary system, where there is no direct dependence on the time, $H(t_1, t_2, n, k) = H(t_2 - t_1, n, k)$. In that case, we can write the backward equation as:

$$\mathbb{P} [X_{t+dt} = n] = \sum_k \mathbb{P} [X_t = k] \, H (dt, n, k)$$

and the forward as:

$$\mathbb{P} [X_{t+dt} = n] = \sum_k \mathbb{P} [X_{dt} = k] \, H (t, n, k)$$

We will denote $a_n(t) \equiv \mathbb{P} [X_t = n]$, as usual. The backwards equation can be written as:

$$a_n (t + dt) = \sum_k a_k (t) \, H (dt, n, k)$$

If H can be written as: $H(dt, n, k) = f(n, k) \, dt + o\left(dt^2\right)$, we get:

$$a_n (t + dt) = \sum_k a_k (t) \, f (n, k) \, dt + o\left(dt^2\right)$$

If $f(n, n) = A + o(dt)$, we can make this into an ODE.

5.1.1 Constant per capita rates with no negative immigration

Lets start with the case where the per-capita rates are constant, so $f_l(X_t) = \lambda_l X_t + S_l$, where both λ_l, S_l are constant. We also demand that $S_{-1} = S_0 = 0$ (no non-positive immigration). The case of S_0 is just not very important, because it is a non-event.

In this case:

If $n \neq k$, we can write:

$$H(dt, n, k) = \mathbb{P}[\text{One event takes us from } k \text{ to } n \text{ in } dt] = S_{n-k}dt + \lambda_{n-k}kdt + o(dt)$$

If $n = k$, we can also have no event:

$$H(dt, n, n) = \mathbb{P}[\text{Reaction} \xcancel{\text{of no change}}^{0}] + \mathbb{P}[\text{No event}] = 1 - \left(n\sum_{l=-1}^{N}\lambda_l - \sum_{l=1}^{N}S_l\right)dt$$

We denote $\lambda = \sum_l \lambda_l, S = \sum_l S_l$

To make things simpler, we assumed both $S_0 = \lambda_0 = 0$, which dropped the first term

$$a_n(t+dt) = \sum_{k\neq n} a_k(t)(k\lambda_{n-k} + S_{n-k})dt + a_n(t)(1 - n\lambda - S)dt + o\left(dt^2\right)$$

And thus:

$$a_n(t+dt) - a_n(t) = -(n\lambda + S)a_n(t)dt + \sum_{k\neq n} a_k(t)(k\lambda_{n-k} + S_{n-k})dt$$

And divide by dt and take the limit to get:

$$\frac{da_n}{dt}(t) = -(n\lambda + S)a_n(t) + \sum_{k\neq n} a_k(t)(k\lambda_{n-k} + S_{n-k})$$

Which is a linear set of ODEs, which can be formulated with a matrix that has a sum over columns of 1.

5.1.2 Onto the PGF!

We denote the PGF

$$G(x,t) = \sum_{n=0}^{\infty} a_n(t)x^n$$

We also mark the PGF of the immigration by $r(x) = \sum_{k=1}^{N} x^k S_k$, and $Q(x) = \sum_{k=1}^{N}\lambda_k x^k$.

We multiply the ODE by x^n:

$$\frac{da_n}{dt}x^n = -(\lambda n + S)a_n(t)x^n + \sum_{k\neq n} x^n a_k(t)(k\lambda_{n-k} + S_{n-k})$$

And sum over all possible values of n to get (since the convergence to G is uniform):

$$\begin{aligned}\frac{\partial G}{\partial t}(x,t) &= -\lambda x\sum_n nx^{n-1}a_n(t) - SG(x,t) + \sum_n\sum_{k\neq n} x^k a_k(t)\left(k\lambda_{n-k}x^{n-k} + S_{n-k}x^{n-k}\right)\\ &= -\lambda x\frac{\partial G}{\partial x}(x,t) - SG(x,t) + \sum_n\sum_k x^k a_k(t)S_{n-k}x^{n-k} + \sum_n\sum_{k\neq n} x^k a_k(t)k\lambda_{n-k}x^{n-k}\end{aligned}$$

The one before last is a multiplication of analytic functions (we added the $k = n$ term since $S_0 = 0$).

$$\frac{\partial G}{\partial t}(x,t) = -\lambda x\frac{\partial G}{\partial x}(x,t) - SG(x,t) + G(x,t)r(x) + x\sum_n\sum_k x^{k-1}a_k(t)k\lambda_{n-k}x^{n-k}$$

We do a similar thing with the last term, except it is the partial derivative of G:

$$\frac{\partial G}{\partial t}(x,t) = -\lambda x\frac{\partial G}{\partial x}(x,t) - SG(x,t) + G(x,t)r(x) + x\frac{\partial G}{\partial x}(x,t)Q(x)$$

To explain this more easily, we can keep the x^n terms outside the convolution, and remember that ka_k is the k-th term in $x\frac{\partial G}{\partial x}$, and thus from the convolution we can tell what the terms are.

It is often written as:

$$\frac{\partial G}{\partial t} = x\frac{\partial G}{\partial x}(x,t)(Q(x) - \lambda) + G(x,t)(r(x) - S)$$

5.1.3 Let's do the equation for the first moment

As usual, $\mathbb{E}\left[\text{Population size}\right] = M_1 = \frac{\partial G}{\partial x}\left(1,t\right)$. We can write an ODE for it:

$$\frac{\partial^2 G}{\partial t \partial x}\left(x,t\right) = \left(Q-\lambda\right)\left(\frac{\partial G}{\partial x} + x\frac{\partial^2 G}{\partial x^2}\right) + x\frac{\partial G}{\partial x}\frac{dQ}{dx} + \frac{\partial G}{\partial x}\left(r-s\right) + G\frac{dr}{dx}$$

We set $x = 1$:

$$\frac{dM_1}{dt} = M_1\overline{\lambda_r} + \overline{S}$$

Which is just the deterministic equation, if we had a constant reaction rate of the mean value and the immigration source was its mean constant.

6 5.5.2025

6.1 Form of the PGF for a branching process with rates and immigration

My thanks to Tomer for supplying the notes for this lesson, which I missed because my airline tried to leave me stranded in London indefinitely.

Define $a_n\left(t\right)$ as the probability of n individuals at time t, assuming one individual at time $t = 0$ and no immigration. Define A_k as the event in which k individuals were born in $[0, dt]$.

We do a backwards derivation for the events in $[0, dt]$ to get:

$$a_n\left(t+dt\right) = \sum_k \mathbb{P}\left[X_t = n | A_k\right]\mathbb{P}\left[A_k\right]$$

From the per-capita definition of the rates, $\mathbb{P}\left[A_k\right] = \lambda_k dt$, and $\mathbb{P}\left[A_0\right] = \left(1-\lambda dt\right)$:

$$a_n\left(t+dt\right) = \sum_k \sum_{\sum_{i=1}^k n_i = n} \prod_{i=1}^k a_{n_i}\left(t\right)\lambda_k dt + \left(1-\lambda dt\right)a_n\left(t\right)$$

$$a_n\left(t+dt\right) - a_n\left(t\right) = -\lambda a_n\left(t\right)dt + \sum_k \sum_{\sum_{i=1}^k n_i = n} \prod_{i=1}^k a_{n_i}\left(t\right)\lambda_k dt$$

So the ODE follows:

$$\frac{da_n}{dt}\left(t\right) = -\lambda a_n\left(t\right) + \sum_k \sum_{\sum_{i=1}^k n_i = n} \prod_{i=1}^k a_{n_i}\left(t\right)\lambda_k dt$$

We can now multiply by x^n and sum over all n to get the PDE for the PGF g:

$$\frac{\partial g}{\partial t}\left(x,t\right) = -\lambda g\left(x,t\right) + \sum_k \lambda_k g^k\left(x,t\right)$$

Assume immigration starts at $t = 0$ according to $\mathbb{P}\left[k \text{ immigrants in } [t, dt]\right] = \gamma_k dt + o\left(dt\right)$, and denote $S = \sum \gamma_k$.

We denote $b_n\left(t\right)$ as the probability of n individuals at time t for this new population. We do a similar derivation with the law of total probability:

$$b_n\left(t+dt\right) = b_n\left(t\right)\left[1 - Sdt\right] + \sum_k \gamma_k dt\left[\sum_{i+m=n} b_i\left(t\right) \prod_{\sum_{j=1}^k m_j = m} a_{m_j}\left(t\right)\right]$$

We used a tricky thing here. If b isn't just immigration but the total, the inner sum that decouples a, b does not work unless we can tell certain individuals have no immigration relating to them. We are, in essence, saying that the first k individuals grow with no immigration and say "Oh, there's some additional thing going on from other contributions".

The ODE for b is given by:

$$\frac{db_n}{dt}\left(t\right) = -Sb_n\left(t\right) + \sum_k \gamma_k\left[\sum_{i+m=n} b_i\left(t\right) \prod_{\sum_{j=1}^k m_j = m} a_{m_j}\left(t\right)\right]$$

If we denote G as the PGF for this case, we get:

$$\frac{\partial G}{\partial t}(x,t) = -SG(x,t) + \sum_n x^n \sum_k \gamma_k \left[\sum_{i+m=n} b_i(t) \prod_{\sum_{j=1}^k m_j = m} a_{m_j}(t) \right]$$

We notice that $\left[g^k(x,t)\right]_m = \prod_{\sum_{j=1}^k m_j = m} a_{m_j}(t)$, and therefore:

$$\sum_{i+m=n} b_i(t) \prod_{\sum_{j=1}^k m_j = m} a_{m_j}(t) = \left[g^k G\right]_n$$

Thus:

$$\frac{\partial G}{\partial t}(x,t) = -SG(x,t) + \sum_k \gamma_k \sum_n x^n \left[g^k G\right]_n = -SG(x,t) + \sum_k \gamma_k G(x,t) g^k(x,t) =$$
$$G(x,t)\left[-S + \sum_k \gamma_k g^k(x,t)\right]$$

And therefore we can solve:

$$G(x,t) = C(x) \exp\left(\int_0^t \left[-S + \sum_k \gamma_k g^k(x,s) \right] ds \right)$$

6.2 Moment Equations

We denote lower case m, f as the uncentered moments and factorial moments of g, and upper case M, F for G.

$$\frac{\partial^2 g}{\partial t \partial x}(x,t) = -\lambda \frac{\partial g}{\partial x}(x,t) + \sum_k \lambda_k k g^{k-1}(x,t) \frac{\partial g}{\partial x}(x,t)$$

Set $x = 1$ and we get:

$$\frac{dm_1}{dt} = -\lambda m_1 + \sum_k \lambda_k k m_1$$

Denote $\alpha = \sum \lambda_k k - \lambda$ to get:

$$\frac{dm_1}{dt} = \alpha m_1$$

and thus

$$m_1 = e^{\alpha t}$$

We know that $f_2 = m_2 - m_1$, and we do the second derivative:

$$\frac{\partial^3 g}{\partial t \partial x^2} = -\lambda \frac{\partial^2 g}{\partial x^2} + \sum_k \lambda_k k \left[(k-1) g^{k-2} \left(\frac{\partial g}{\partial x} \right)^2 + g^{k-1} \frac{\partial^2 g}{\partial x^2} \right]$$

And setting $x = 1$ again we get:

$$\frac{df_2}{dt} = -\lambda f_2 + \sum_k \lambda_k k \left[(k-1) e^{2\alpha t} + f_2 \right] = \alpha f_2 + e^{2\alpha t} \sum \lambda_k k (k-1)$$

We denote $\sum \lambda_k k (k-1) \equiv D_2$ to get:

$$\frac{df_2}{dt} = \alpha f_2 + D_2 e^{2\alpha t}$$

We solve the homogeneous and particular solutions:

$$f_{2h} = Ce^{\alpha t}, f_{2p} = Ae^{2\alpha t}$$

and solve for A, C to get:

$$f_2(t) = \frac{D_2}{\alpha} \left(e^{2\alpha t} - e^{\alpha t} \right)$$

Using the form of G and denoting $Q(x,t) = \int_0^t \left[-S + \sum_k \gamma_k g^k(x,s)\right] ds$, we can get (We ignore $c(x)$ because our initial population is assumed 0):

$$\frac{\partial G}{\partial x} = \frac{\partial Q}{\partial x} e^{Q(x,t)} = \left(\int_0^t ds \sum_k \gamma_k k g^{k-1} \frac{\partial g}{\partial x} \right) e^Q$$

Set $x = 1$:

$$\frac{dM_1}{dt} = \left(\int_0^t \sum_k \gamma_k k e^{\alpha s} ds \right) 1 = \frac{\left(e^{\alpha t} - 1\right)}{\alpha} \sum_k \gamma_k k$$

7 12.5.2025

Today we deal with the Feynman-α formula.

We have a time interval $[0,t]$. We introduce a source at 0, and we stop looking at time t, and we start a detector in the range $[t-T,t]$. We have a branching process for the population running over the interval. Our question is what is the distribution of the number of counts in the detector $x(t,T)$.

We denote λ_a as the rate of individual death with no counts (per capita). λ_f is the rate of fissions (per capita), i.e. to branch. λ_d is the rate of individual detection (a form of death where we count it). ν is the multiplicity with a multiplicity vector $\{p_\nu\}$. S is the source rate (immigration), with a multiplicity of $p_{\gamma,i} = \delta_{1i}$.

To solve this problem, we should start with a simpler problem. We'll flip our axis so that we move backwards in time. Now time will flow from $s = 0$ at t, to T and then to $s = t$.

We denote:

$$a_n(t,T) = \mathbb{P}\left[\text{an individual born } t \text{ before the window closes creates } n \text{ counts assuming no source}\right]$$

We will use a backwards derivation with the law of total probability, to ask what did that individual do when it was born during a short period dt, in $[t+dt,t]$ (We extend the time range by dt such that the individual has slightly more time to run before the detector starts). We also denote $\lambda = \lambda_f + \lambda_a + \lambda_d$. Notice that detections can happen outside our window, we just won't count them in a_n! This model assumes the detector is in the system and affects it, but does not record. We will denote $\chi_I(t) = \begin{cases} 1 & t \in I \\ 0 & t \notin I \end{cases}$.

$$a_n(t+dt,T) = a_n(t)(1-\lambda)dt + \lambda_a dt \delta_{0n} + \lambda_d dt \left[\left(1 - \chi_{[0,T]}(t)\right) \delta_{0n} + \chi_{[0,T]}(t) \delta_{1n} \right] + \lambda_f dt \sum_\nu p_\nu \sum_{\sum_i n_i = \nu} \prod_{i=1}^{\nu} a_{n_i}(t,T) + o(dt)$$

From which we get the PDE:

$$\frac{\partial a_n}{\partial t} = -\lambda a_n + \lambda_a \delta_{0n} + \lambda_d \left[\left(1 - \chi_{[0,T]}(t)\right) \delta_{0n} + \chi_{[0,T]}(t) \delta_{1n} \right] + \lambda_f \sum_\nu p_\nu \sum_{\sum_i n_i = \nu} \prod_{i=1}^{\nu} a_{n_i}(t,T)$$

Multiply by x^n and sum over n to get the PGF g of it:

$$\frac{\partial g(x,t,T)}{\partial t} = -\lambda g(x,t,T) + (\lambda_a + \lambda_d) + \lambda_d \chi_{[0,T]}(t)(x-1) + \lambda_f \sum_\nu p_\nu g^\nu(x,t,T)$$

And the last term is $r(g)$ where r is the characteristic polynomial of the multiplicity (its PGF).

Assume that before the window closes, we introduce a Poisson source (like we mentioned above, that it has multiplicity 1 and a constant rate with no memory).

We denote

$$b_n(t,T) = \mathbb{P}\left[\text{a source that starts } t \text{ before the window of length } T \text{ closes generates } n \text{ counts}\right]$$

Like in the previous lesson, we get G for b using g for a, and we assume an initial, empty population. We split over whether the source happens in $[t,t+dt]$.

$$b_n(t+dt,T) = (1 - Sdt) b_n(t,T) + Sdt \sum_{k=0}^{n} a_k(t,T) b_{n-k}(t,T)$$

And we get the PDE:

$$\frac{\partial b_n}{\partial t}(t,T) = -Sb_n(t,T) + S\sum_{k=0}^{n} a_k(t,T)\, b_{n-k}(t,T)$$

We multiply and sum to get the PGF G :

$$\frac{\partial G}{\partial t}(x,t,T) = -SG(x,t,T) + SG(x,t,T)\, g(x,t,T) = SG(x,tT)\,(g(x,t,T)-1)$$

Which is often called *Bartlett's Equation*.
And therefore:

$$G(x,t,T) = \exp\left(-S\int_0^t (1-g(x,\tau,T))\, d\tau\right) = e^{Q(x,t,T)}$$

Where we don't need a $c(x)$ to multiply it since at $t=0$ we know we have 0 population.

$$M_1 = \frac{\partial G}{\partial x}(1,t,T) = G(1,t,T)\frac{\partial Q}{\partial x}(1,t,T) = \frac{\partial Q}{\partial x}(1,t,T)$$
$$F_2 = \frac{\partial^2 G}{\partial x^2}(1,t,T) = G(1,t,T)\left(\frac{\partial Q}{\partial x}\right)^2(1,t,T) + G(1,t,T)\frac{\partial^2 Q}{\partial x^2}(1,t,T) = \left(\frac{\partial Q}{\partial x}\right)^2(1,t,T) + \frac{\partial^2 Q}{\partial x^2}(1,t,T)$$

And we can derive Q:

$$\frac{\partial Q}{\partial x}(1,t,T) = S\int_0^t \frac{\partial g}{\partial x}(1,\tau,T)\, d\tau = S\int_0^t m_1(\tau,T)\, d\tau$$
$$\frac{\partial Q}{\partial x}(1,t,T) = S\int_0^t \frac{\partial^2 g}{\partial x^2}(1,\tau,T)\, d\tau = S\int_0^t f_2(\tau,T)\, d\tau$$

And now we need to get m_1, f_2.

$$\frac{\partial^2 g}{\partial t \partial x}(x,t,T) = -\lambda\frac{\partial g}{\partial x}(x,t,T) + \lambda_d\chi_{[0,T]} + \lambda_f\sum_\nu p_\nu \nu g^{\nu-1}(x,t,T)\frac{\partial g}{\partial x}(x,t,T)$$

We denote $\alpha = \lambda - \lambda_f\overline{\nu}$ and set $x=1$.

$$\frac{\partial m_1}{\partial t}(t,T) = \frac{\partial^2 g}{\partial t\partial x}(1,t,T) = -\lambda m_1(t,T) + \lambda_d\chi_{[0,T]}(t) + \lambda_f\sum_\nu p_\nu \nu m_1(t,T) = -\alpha m_1 + \lambda_d\chi_{[0,T]}(t)$$
$$m_1(0) = 0$$

Onto the next one:

$$\frac{\partial^3 g}{\partial t\partial x^2}(x,t,T) = -\lambda\frac{\partial^2 g}{\partial x^2}(x,t,T) + \lambda_f\sum_\nu p_\nu\nu\left[(\nu-1)\, g^{\nu-2}(x,t,T)\left(\frac{\partial g}{\partial x}\right)^2(x,t,T) + g^{\nu-1}\frac{\partial^2 g}{\partial x^2}(x,t,T)\right]$$

And we set $x=1$:

$$\frac{\partial f_2}{\partial t}(t,T) = \frac{\partial^3 g}{\partial t\partial x^2}(1,t,T) = -\lambda f_2 + \lambda_f\left[\overline{\nu} f_2(t,T) + m_1^2(t,T)\,\overline{\nu(\nu-1)}\right] = -\alpha f_2(t,T) + \lambda_f\overline{\nu(\nu-1)}m_1^2(t,T)$$

And now to solve these equations. We assume $\alpha \neq 0$, otherwise we need a different, linear solution.
We use the solution for a linear ODE, where if $\frac{dy}{dt} = ay + g, y(0) = y_0$ then:

$$y(t) = e^{at}\left[\int_0^t e^{-as} g(s)\, ds + y_0\right]$$

And therefore:

$$m_1(t) = e^{-\alpha t}\left[\int_0^t e^{\alpha s}\lambda_d\chi_{[0,T]}(s)\, ds\right] = \frac{\lambda_d}{\alpha}\begin{cases} -e^{-\alpha(t-T)} + e^{-\alpha t} & t > T \\ \left(1-e^{-\alpha t}\right) & t \leq T \end{cases}$$

Where the top comes from the integral being cut off at T and the bottom where we set $\chi = 1$ in the whole range.

We can rewrite this as:

$$m_1(t) = \frac{\lambda_d}{\alpha}\left[\left(1-e^{-\alpha t}\right)\chi_{[0,T]}(t) - \left(e^{-\alpha(t-T)} - e^{-\alpha t}\right)\Theta(t-T)\right]$$

Where Θ is the heavyside function, which turns from 0 to 1 when the argument hits 0.
We can get M_1now. When $t \in [0,T]$:

$$M_1(t,T) = \frac{S\lambda_d}{\alpha}\int_0^t \left(1-e^{-\alpha s}\right)ds = \frac{S\lambda_d}{\alpha}\left(t + \frac{1}{\alpha}\left(1-e^{-\alpha t}\right)\right)$$

If $t > T$:

$$\begin{aligned} M_1(t,T) &= \frac{S\lambda_d}{\alpha}\left[\int_0^T m_1(s)\,ds + \int_T^t m_1(s)\,ds\right] = \frac{S\lambda_d}{\alpha}\left[\left(T + \frac{1}{\alpha}\left(1-e^{-\alpha T}\right)\right) - \int_T^t \left(e^{-\alpha(s-T)} - e^{-\alpha s}\right)ds\right] \\ &= \frac{S\lambda_d}{\alpha}\left[\left(T + \frac{1}{\alpha}\left(1-e^{-\alpha T}\right)\right) + \frac{1}{\alpha}\left(e^{-\alpha(t-T)} - e^{-\alpha t} - 1 + e^{-\alpha T}\right)\right] \\ &= \frac{S\lambda_d}{\alpha}\left[T + \frac{1}{\alpha}e^{-\alpha t}\left(e^{\alpha T} - 1\right)\right] \end{aligned}$$

So given together:

$$M_1(t,T) = \frac{S\lambda_d}{\alpha}\begin{cases} t + \frac{1}{\alpha}\left(1-e^{-\alpha t}\right) & 0 \le t \le T \\ T + \frac{1}{\alpha}e^{-\alpha t}\left(e^{\alpha T} - 1\right) & t > T \end{cases}$$

If $\alpha > 0$ (subcritical system), and we take $t \to \infty$, this will be just:

$$\lim_{t\to\infty} M_1(t,T) = \frac{S}{\alpha}\lambda_d T = \frac{ST\lambda_d}{\lambda - \lambda_f \overline{\nu}} = ST\frac{\lambda_d}{\lambda}\frac{1}{1-\frac{\lambda_f}{\lambda}\overline{\nu}} = \frac{STP_d}{1-k}$$

Which is actually what we should have expected! The population size stabilizes at its $\frac{S}{\alpha}$ asymptotic value, and we measure it with a detection efficiency P_d for a time length T.

Basically, we waited long enough for the population to stabilize. If the system is supercritical, this would have exploded.

For homework we will have to solve for F_2.

8 19.5.2025

We have about 5 more lectures, but we'll reorder the class because we're all so good.

We will skip all the legwork and give the solutions for m_2:

$$\frac{\partial f_2}{\partial t}(t,T) = -\alpha f_2(t,T) + \lambda_f \overline{\nu(\nu-1)} m_1^2(t,T)$$

And so:

$$f_2(t,T) = \frac{\overline{\nu(\nu-1)}\lambda_d^2\lambda_f}{\alpha^3}\begin{cases} 2\left(\sinh(\alpha t) - \alpha t\right) & 0 \le t \le T \\ e^{-2\alpha t}\left(2e^{\alpha t} - e^{2\alpha t} + 2e^{\alpha(t+T)} - 2e^{\alpha t}(1+\alpha t) - 1\right) & t > T \end{cases}$$

This m_2 is just for the population starting with one individual that branches. We still need to get to F_2 and M_2 for the case with a Poisson source. We had M_1 above and the generating function G, and we saw:

$$M_2 - M_1 = F_2 = \frac{\partial^2 G}{\partial x^2}(1,t,T) = \left(\frac{\partial Q}{\partial x}\right)^2(1,t,T) + \frac{\partial^2 Q}{\partial x^2}(1,t,T)$$

And therefore:

$$M_2 - M_1 = M_1^2(t,T) + S\int_0^t f_2(\tau,T)\,d\tau$$

And if we denote $X(t,T)$ as the number of counts:

$$\mathbb{V}[X(t,T)] = M_1 + S\int_0^t f_2(\tau,T)\,d\tau$$

Which is ugly A.F., but in the limit it is somewhat nicer:

$$\mathbb{V}\left[\lim_{t\to\infty} X(t,T)\right] = \mathbb{V}\left[X(T)\right] = \mathbb{E}\left[X(T)\right] + S\frac{\overline{\nu(\nu-1)}\lambda_d^2\lambda_f}{\alpha^4}\left[e^{-\alpha T}+\alpha T-1\right]$$

And now we can define the Feynman-Y function:

$$Y(T) = \lim_{t\to\infty}\frac{\mathbb{V}\left[X(t.T)\right]}{\mathbb{E}\left[X(t,T)\right]} - 1 = \frac{S\frac{\overline{\nu(\nu-1)}\lambda_d^2\lambda_f}{\alpha^4}\left[e^{-\alpha T}-\alpha T-1\right]}{\frac{\lambda_d}{\alpha}ST} = \frac{\overline{\nu(\nu-1)}\lambda_d\lambda_f}{\alpha^2}\left[1-\frac{1-e^{-\alpha T}}{\alpha T}\right] = Y_\infty\left(1-\frac{1-e^{-\alpha T}}{\alpha T}\right)$$

8.1 Logistic Model

Our system so far has been dependent on the criticality of the system, and in a critical system, there is no stable state. The limit depends only on the initial condition, but any known drift will also drift our estimate for the limit.

We want to generalize our model to the Logistic ODE model (named by Verhult, the name is terrible and you should look that French guy up and complain to his long forgotten grave).

$$\frac{dy}{dt}(t) = \alpha y(t) - \epsilon^2 y^2(t)$$

This equation is solvable, but we don't actually care about that. What we care about is that there are two stable points. If $y(t)=0$, it will remain there forever. The other is $y(t)=\frac{\alpha}{\epsilon^2}$.

If $\alpha<0$, the other solution is non-physical and it is moot. So we will only consider supercritical systems, with $\alpha>0$.

The second derivative around 0 changes from negative to positive, so 0 is not an absorbing state, since if $y(0)\in\left(0,\frac{\alpha}{\epsilon^2}\right)$, $y(t)$ will increase always towards $\frac{\alpha}{\epsilon^2}$. Therefore, for any $y(0)>0$, $\frac{\alpha}{\epsilon^2}$ is an attracting state for the ODE in that region, and $y(t)$ will monotonically converge to it from any positive starting point. We call $\frac{\alpha}{\epsilon^2}$ the *carrying capacity* of the system.

We want to discuss the stochastic logistic model, but what we just talked about is fully deterministic.

We will model the stochastic model by introducing feedback into the death rate.

For any $k\geq 0$, λ_k is a constant. For $k=-1$ (death), $\lambda_{-1}(X)=\lambda_{-1,0}+\epsilon^2 X$. These, we remind you, are the per capita rates of reaction with k progeny.

We start with the law of total probability, as always, for population of size n, denoted as $a_n(t)$, which started as $a_n(0)=\delta_{n1}$.

$$a_n(t+dt) = (1-\lambda(1)\,dt)\,a_n + \sum_{k=-1}^{L}\lambda_k dt????$$

And we can't use the backward derivation of looking at $[0,dt]$, because what would we use for combining the effects of different populations when death rates actually depend on them? We must, instead, use the forward derivation.

$$a_n(t+dt) = a_n(t)\left(1-[\lambda(n)\,dt]^n\right) + \sum_{k=-1}^{L} a_{n-k}(t)\,(n-k)\,\lambda_k(n-k)\,dt + o(dt)$$

Which the first term in first order can be simplified to $1-n\lambda(n)\,dt$:

$$a_n(t+dt) = a_n(t)\left(1-n\lambda(n)\,dt\right) + \sum_{k=-1}^{L} a_{n-k}(t)\,(n-k)\,\lambda_k(n-k)\,dt + o(dt)$$

$$\frac{da_n}{dt}(t) = -\left(\lambda_{-1,0}n+\epsilon^2 n^2 + n\sum_{k=0}^{L}\lambda_k\right)a_n(t) + \left[\lambda_{-1,0}+\epsilon^2(n+1)\right](n+1)\,a_{n+1}(t) + \sum_{k=0}^{L}\lambda_k(n-k)\,a_{n-k}(t)$$

We can break this into the regular and feedback terms:

$$\frac{da_n}{dt}(t) = -\left(\sum_{k=0}^{L}\lambda_k+\lambda_{-1,0}\right)na_n(t) - \epsilon^2 n^2 a_n(t) + \left(\lambda_{-1,0}(n+1)\,a_{n+1}(t) + \sum_{k=0}^{L}\lambda_k(n-k)\,a_{n-k}\right) + \epsilon^2(n+1)^2\,a_{n+1}(t)$$

We can multiply by x^n and sum over all options to go for the PGF G:

$$\begin{aligned}\frac{\partial G}{\partial t}(x,t) = &-\left(\sum_{k=0}^{L}\lambda_k + \lambda_{-1,0}\right)\sum na_n x^n - \epsilon^2\sum_n n^2 a_n x^n \\ &+ \sum_{k=0}^{L}\lambda_k \sum_n x^n (n-k)\, a_{n-k} + \lambda_{-1,0}\sum_n a_{n+1}(n+1)\, x^n \\ &+ \epsilon^2\sum_n (n+1)^2 a_{n+1}x^n\end{aligned}$$

We denote $\Lambda = \sum_k \lambda_k + \lambda_{-1,0}$ for simplicity:

$$\frac{\partial G}{\partial t} = -\Lambda x\frac{\partial G}{\partial x} + \lambda_{-1,0}\frac{\partial G}{\partial x} - \epsilon^2 x\left(\frac{\partial}{\partial x}\left(x\frac{\partial G}{\partial x}\right)\right) + \frac{\partial G}{\partial x}\sum_{k=0}^{L}\lambda_k x^{k+1} + \epsilon^2\left(\frac{\partial}{\partial x}\left(x\frac{\partial G}{\partial x}\right)\right)$$

Which is simplified to:

$$\frac{\partial G}{\partial t} = \left(-\Lambda x + \lambda_{-1,0} + \sum_{k=0}^{L}\lambda_k x^{k+1}\right)\frac{\partial G}{\partial x} + \epsilon^2(1-x)\left(\frac{\partial G}{\partial x} + x\frac{\partial^2 G}{\partial x^2}\right)$$

We notice that $q(x) = \lambda_{-1,0} + \sum_{k=0}^{L}\lambda_k x^{k+1}$ as the characteristic polynominal of the multiplicity rates (not a PGF!!!).

$$\frac{\partial G}{\partial t} = (q(x) - \Lambda x)\frac{\partial G}{\partial x} + \epsilon^2(1-x)\left(\frac{\partial G}{\partial x} + x\frac{\partial^2 G}{\partial x^2}\right)$$

9 26.5.2025

It's moments time, and there are some complications which are great.

$$\frac{\partial^2 G}{\partial t\partial x} = (q-\Lambda x)\frac{\partial^2 G}{\partial x^2} + \left(\frac{dq}{dx} - \Lambda\right)\frac{\partial G}{\partial x} + \epsilon^2(1-x)\left(2\frac{\partial^2 G}{\partial x^2} + x\frac{\partial^3 G}{\partial x^3}\right) - \epsilon^2\left(\frac{\partial G}{\partial x} + x\frac{\partial^2 G}{\partial x^2}\right)$$

$$\frac{\partial^3 G}{\partial t\partial x^2} = (q-\Lambda x)\frac{\partial^3 G}{\partial x^3} + 2\left(\frac{dq}{dx} - \Lambda\right)\frac{\partial^2 G}{\partial x^2} + \frac{d^2 q}{dx^2}\frac{\partial G}{\partial x} + \epsilon^2(1-x)\left(3\frac{\partial^3 G}{\partial x^3} + x\frac{\partial^4 G}{\partial x^4}\right) - 2\epsilon^2\left(2\frac{\partial^2 G}{\partial x^2} + x\frac{\partial^3 G}{\partial x^3}\right)$$

And we set $x = 1$:

$$\frac{dM_1}{dt} = \alpha M_1 - \epsilon^2(M_1 + F_2) = \alpha M_1 - \epsilon^2 M_2$$

Where $\alpha = \frac{dq}{dx}(1) - \Lambda = \sum_{k=1}^{L} k\lambda_k - \Lambda = \Lambda\left(\sum_{k=1}^{L} kp_k - 1\right) = \Lambda(\overline{\nu} - 1)$

$$\frac{dF_2}{dt} = 2\alpha F_2 + D_2 M_1 - 2(2F_2 + F_3)$$

Where $D_2 = \frac{d^2 q}{dx^2}(1) = \sum_{k=1}^{L} k(k-1)\lambda_k$
And if we look for $M_2 = F_2 - M_1$, this would be something like:

$$\frac{dM_2}{dt} = 2\alpha(M_2 - M_1) + M_1 D_2 - 2(2F_2 + F_3) - \alpha M_1 + \epsilon^2 M_2$$

And we are going to have an infinite set of coupled moment equations. Blah.
What's interesting, is that the equilibrium point isn't going to be at $\frac{\alpha}{\epsilon^2}$ anymore! Having the same equilibrium point would require:

$$M_{1,\infty} = \frac{\epsilon^2}{\alpha}M_{2,\infty}$$

And we only get the same equilibrium point as the deterministic one ($\frac{\alpha}{\epsilon^2}$) if $M_{2,\infty} = \left(\frac{\alpha}{\epsilon^2}\right)^2$, which means $M_{2,\infty} = M_{1,\infty}^2$, which only happens if the Variance is 0, since $\mathbb{V}[x] = M_2 - M_1^2$. This never happens for a stochastic process unless they collapse to a deterministic point, which it won't. So what we get is that the stochastic process has a different equilibrium expectation than the deterministic value.

Having an infinite set of coupled equations means that we're not going to be able to get a closed form solution easily for this system. We can tell that the set point is lower, since $M_{2,\infty}$ is strictly positive. We don't know how to analytically solve the stochastic carrying capacity, and there may not even be a closed form. We can introduce a closure and solve the set, but that's going to introduce its own bias in the estimate.

Let us try a little. We will denote $H = \frac{\partial G}{\partial x}$ for simplicity and try to solve the differential equation for the limit of G:

$$0 = H\left(q - \Lambda x\right) + \epsilon^2\left(1 - x\right)\left(H + x\frac{dH}{dx}\right)$$

$$\frac{dH}{dx} = H\frac{\left(q - \Lambda x\right) - \epsilon^2\left(x - 1\right)}{\epsilon^2 x\left(x - 1\right)}$$

Since $x = 1$ is a root of both the numerator and denominator and q is a polynomial in x, we can write this as:

$$\frac{dH}{dx} = H\left(x\right)C\left(\frac{1}{x} + P\left(x\right)\right)$$

And therefore:

$$H\left(x\right) = C_1 e^{\int_0^x C\left(\frac{1}{s} + P(s)\right)ds}$$

And therefore:

$$G\left(x\right) = C_2 + C_1\int_0^x e^{C\int_0^y\left(\frac{1}{s} + P(s)\right)ds}dy$$

This has two degrees of freedom, C_1, C_2, which are determined by the following conditions:

- $G\left(1\right) = 1$
- G is analytic over the unit circle

This will require that $G_\infty\left(x\right) = 1$, when $C < 0$, since that function will have a pole otherwise. Oh dear god, that means the numerical simulations that fully allow us to sample the asymptotic behavior isn't actually an asymptotic behavior! It's a quasi-stable state and not a stable state. The actual stable state is a dead population!

What we need is the conditional probability of survival, to take away the quasi-stability!

$$b_n\left(t\right) = \mathbb{P}\left[X\left(t\right) = n | X\left(t\right) \neq 0\right] = \frac{\mathbb{P}\left[X\left(t\right) = n\right]}{\mathbb{P}\left[X\left(t\right) \neq 0\right]} = \begin{cases} \frac{a_n}{1 - a_0} & n \neq 0 \\ 0 & n = 0 \end{cases}$$

And therefore:

$$\tilde{G} = \sum b_n x^n = \frac{G\left(x,t\right) - a_0\left(t\right)}{1 - a_0\left(t\right)}$$

And therefore:

$$\tilde{G}_{M.S}\left(x\right) = \lim_{t\to\infty}\frac{G\left(x,t\right) - a_0\left(t\right)}{1 - a_0\left(t\right)}$$

This is a formal definition, but we don't know how to actually solve this to get the meta-stable condition. Can we use this to get a differential equation for $\tilde{G}$? The extinction problem to know a_0 is terrible and we can't solve it, but we know the PDE for G, and thus:

$$\tilde{G}_{M.S}\left(x,t\right) = \frac{G\left(x,t\right) - G\left(0,t\right)}{1 - G\left(0,t\right)}$$

10 9.6.2025

We are going to deal with the diffusion approximation today.

We start with the Langevin equation:

$$\frac{dx}{dt} = F\left(x,t\right) + \eta$$

Where η is some stochastic process. What they wanted to do is introduce random noise into a differential equation. The subtext is that the contribution of the stochastic term is smaller than the deterministic term. We would have wanted η to have a bound variance and so on.

The problem we run into is that the differential term isn't well defined with this jumping stochastic process. Ito defined this problem with Ito calculus, which is where we are going today. We tried to do proofs and so on when necessary so far, but Ito calculus is its own course so we're going to hand waive quite a bunch.

10.1 Brownian Motion

We are calling it BM, but its mathematical name is usually a Wiener process. A process X will be called BM if:

1. $\forall t, t_0 > 0 \quad \mathbb{E}\left[X\left(t_0 + t\right) | X\left(t_0\right)\right] = X\left(t_0\right)$
2. $\forall t, s > 0 \quad X\left(t+s\right) - X\left(t\right) \sim N\left(0, \sqrt{s}\right)$

Does such a process exist? Yes. There are multiple constructions, the known one is Donsker's. Look up Donsker's Theorem, which does an acceleration to a drunk walk.

$$W_t = \lim_{n\to\infty} \frac{X\left(\underline{nt}\right)}{\sqrt{n}}$$

Is proven to exist and have the BM conditions.

Now we define an Ito integral (a stochastic integral), using BM variables. Let X_t, W_t be random processes and $f : \mathbb{R} \to \mathbb{R}$. Let S be some partition of $[0, T]$. We define:

$$I\left(S\right) = \sum_{i=1}^{N} f\left(x_{t_i}\right)\left(W_{t_{i+1}} - W_{t_i}\right)$$

This is a random variable, which also depends on S. Denote $i\left(S\right)$ as the refinement parameter of S, i.e. $i\left(S\right) = \max_i\left\{t_{i+1} - t_i\right\}$. We define:

$$\int_o^T f\left(X_s\right) dW_s = \lim_{i(S)\to 0} I\left(S\right)$$

If it exists.

1. Does it exist for some specific series of partitions S_i? Under which conditions and what type of convergence this is for a random variable?
2. Can we get a limit in the general sense written above, in terms of not having to specify the series of S?

Ito's Theorem

If W_t is a BM (Wiener process), X_t has a bounded variation and f is locally square-integrable and $f\left(X_s\right)$ is adapted to the filtration of W_t, then the Ito integral is well defined for any series of partitions. It also works when dW_t is a δ distribution.

We can then write:

$$X_t = X_0 + \int_0^t f\left(X_s, s\right) ds + \int_0^t B\left(X_s, s\right) dW_s$$

This is a random process, well defined, where the initial condition is some random distribution, the first Ito integral is with a δ distribution and the second is an Ito integral with a Wiener process. Now we have an integral equation for X_t, where we look for a process to satisfy this.

We denote this integral equation with a differential system for ease:

$$dX_t = f\left(X_t, t\right) dt + B\left(X_t, t\right) dW_t$$

This equation gives us the intuition for when this SDE is a useful model. When in every small time scale we can approximate with a deterministic behavior on $X\left(t\right)$ that is proportional to dt, and a normal variable with mean 0 and its variance is proportional to dt.

The deterministic term is called the *drift term*, and the second term is called the *noise term* or the *diffusion term*. When there is no diffusion, we have a regular ODE.

Classic Example

We already discussed renewal processes back in this course. A renewal process is a time series where the dwell times are i.i.d. If t_n is a renewal process, the counter of t_n is defined by $R\left(t\right) = \sum_i \chi_{t_i<t} = \max\left\{n | t_n < t\right\}$.

If we denote $\mu = \mathbb{E}\left[\theta\right]$, then $\lim_{t\to\infty} \frac{\mathbb{E}[R(t)]}{t} = \frac{1}{\mu}$. This is because if we have enough measurements, we expect them to be the rate (1 over the period) times the total time.

The stronger version of this is if we denote $\sigma^2 = \mathbb{V}[\theta]$, then:

$$\lim_{t\to\infty} \frac{R(t) - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} \sim N(0,1)$$

.

In an engineering sense, this means that $R(T) \sim \frac{T}{\mu} + \sqrt{\frac{\sigma^2}{\mu^3}}\sqrt{T}N(0,1) = N\left(\frac{T}{\mu}, \sigma^2\frac{T}{\mu^3}\right)$, whenever $T \gg \mu$.

If we assume that we look at a time interval $\Delta t \gg \mu$, we can write:

$$R(\Delta t) - R(0) \approx \frac{1}{\mu}\Delta t + \frac{\sigma^2}{\mu^3}\Delta W_{\Delta t}$$

This yells at us to use an SDE, because we have a drift and a proper noise term. If we have time scales such that the macroscopic solution time $t \gg \mu$, we can approximate $R(t)$ with the SDE:

$$dR(t) = \frac{1}{\mu}dt + \sqrt{\frac{\sigma^2}{\mu^3}}dW_t$$

An extension to this example

If the dwell time somewhat depends on the count rate so far, such that this is no longer a renewal process. This means $\mu(R(t))$, then if the change in μ is still slow because R changes slowly enough and the dependence is slow enough, but we can still assume that μ is constant in a time step Δt, we can still write:

$$dR = \frac{1}{\mu(R(t))}dt + \sqrt{\frac{\sigma^2(R(t))}{\mu^3(R(t))}}dW_t$$

This happens when R is large and a single count barely changes it, and then if μ depends on large changes in R, this will work well.

10.2 Diffusion approximation to a branching process

Let X_t be a branching process with reaction rates λ_l (per capita rates for l progeny, $l \geq 0$, and 0 is death).

We look at the change rate in some small time dt, i.e. $X(t+dt) - X(t)$. We denote x_i as the progeny individual i has in that time. We know that this is just:

$$X(t+dt) - X(t) = \sum_{i=1}^{X(t)} x_i - X(t)$$

We know that:

$$\begin{aligned}
\mathbb{E}[x_i] &= \sum_l P(x_i = l)(l-1) = \sum_l l\lambda_l dt + (1 - \lambda dt) + o(dt) \\
\mathbb{V}[x_i] &= \mathbb{E}\left[x_i^2\right] - \mathbb{E}^2[x_i] = \sum_l l^2\lambda_l dt + \cancel{(1 - \lambda dt)} - \cancel{1} - 2\sum_l l\lambda_l dt + 2\lambda dt + o(dt) \\
&= \sum_l l^2\lambda_l dt - 2\sum_l l\lambda_l dt + \lambda dt + o(dt) = \sum \lambda_l (l-1)^2 dt + o(dt)
\end{aligned}$$

The C.L.T says that we can approximate:

$$\frac{X(t+dt) - X(t) - X(t)\left(\sum_l l\lambda_l dt - \lambda dt\right) + o(dt)}{\sqrt{X(t)\left(\sum \lambda_l (l-1)^2\right)dt + o(dt)}} \sim N(0,1)$$

So we can write something like:

$$dX_t = X_t\left(\sum_l l\lambda_l - \lambda\right)dt + \sqrt{X(t)\left(\sum_l \lambda_l (l-1)^2\right)}dW_t$$

Which is the diffusion approximation to the branching process.

Next lesson we'll add detection.

11 30.6.2025 (After the Iran war)

This is an extra lesson, since we lost 2 with the war. We are considering the SIR model (Susceptible, Infected and Recovered/Removed) in the stochastic case.

We assume $S + I + R = N$ which is a time constant, so the population has an initial capacity that stays the same, and things just move ahead.

11.1 Deterministic Model

The deterministic model is rather simple:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta}{N}SI \\ \frac{dI}{dt} &= \frac{\beta}{N}SI - \mu I \\ \frac{dR}{dt} &= \mu I\end{aligned}$$

The time consistency arises from the equations themselves, and if we don't care about R we could just drop the whole thing. A few things to immediately note:

1. $S(t)$ monotonically decreases, $R(t)$ monotonically increases to N in most cases
2. $S(t) = N$ means that $\frac{dI}{dt} = I(\beta - \mu)$. It is slightly meaningless in that specific case since $I \equiv 0$, but meh.
3. If $N, S \gg I$, we can make the above assumption anyway, because $S \sim N$.

The base infection coefficient $R_0 = \frac{\beta}{\mu}$.

Since $S(t)$ monotonically decreases, for all t we can use Gronwall's inequality ($\frac{S}{N}\beta - \mu \leq \beta - \mu$) to get $I(t) \leq Ce^{(\beta-\mu)t} = Ce^{\mu(R_0-1)t}$. If $\beta < \mu$, there is really no epidemic, since the infected population is vanishingly small. To make something an epidemic, we only care if $R_0 > 1$.

If we look at $I(t)$, $\frac{dI}{dt} = I\left(\frac{S}{N}\beta - \mu\right)$, we will reach a maximum of infected when $S = \frac{\mu}{\beta}N$, which is called "herd immunity", because so many were infected that we can no longer support an epidemic. This can be rewritten for I, R as $(N - (I + R)) = \frac{N}{R_0}$. This can be solved for I to get:

$$I = N\left(1 - \frac{1}{R_0}\right) - R$$

If R is large, herd immunity is reached with a smaller maximal infected number, which is important because they are putting a burden on the healthcare system.

11.2 Stochastic model

Now onto the stochastic problem. This is a more complex issue but very interesting, because it is much closer to the actual mathematics of this problem.

We have S_t, I_t, R_t and we look at $[t, t + dt]$:

$$\begin{aligned}\mathbb{P}[S_t \to S_t - 1, I_t \to I_t + 1, R_t] &= \frac{S_t I_t}{N}\beta dt + o(dt) \\ \mathbb{P}[S_t, I_t \to I_t - 1, R_t + 1] &= \mu I_t dt + o(dt) \\ \mathbb{P}[S_t, I_t, R_t] &= 1 - \frac{S_t I_t}{N}\beta dt - \mu I_t dt + o(dt)\end{aligned}$$

We will look at the generating function $G(x, y, z) = \mathbb{E}\left[x^S y^I z^R\right]$. To get a term for this, our usual methods will work. When we then derive by x, y, z to get the mean behavior:

$$\begin{aligned}\frac{d\mathbb{E}[S]}{dt} &= -\frac{\mathbb{E}[SI]}{N}\beta \\ \frac{d\mathbb{E}[I]}{dt} &= \frac{\mathbb{E}[SI]}{N}\beta - \mu\mathbb{E}[I] \\ \frac{d\mathbb{E}[R]}{dt} &= \mu\mathbb{E}[I]\end{aligned}$$

Just like in the logistic model, the stochastic behavior changes the mean behavior! We only get the deterministic model if we implicitly assume that $\mathbb{E}[SI] = \mathbb{E}[S]\mathbb{E}[I]$, which isn't ever true!

The natural way to do this is to compute $cov(S,I) = \mathbb{E}[SI] - \mathbb{E}[S]\mathbb{E}[I]$ from the real system once, and then we can look at $\rho(S,I) = \frac{cov(S,I)}{\sigma(S)\sigma(I)}$. The problem is that this is expensive, and if we run many populations then we might as well have run the whole thing and know the full populations...

11.3 Test

The test will be a home test, at 17.7. We will have our materials open. Dubi will be abroad. He will give us a model, and if we understand the logistic model we can probably handle it.

We will need to Chapman-Kolmogorov that model, and it will have some trick to it.

It may be a home work rather than a test.