AP3162D: Lecture 1 - Basics of dynamical systems

Hyun Youk

Delft University of Technology
(Dated: February 15, 2018)

.

I. MOTIVATION FOR THIS COURSE

If we were to simply state the goal of our course, it would be: "To understand what properties define a living system and how they arise from the laws of physics." This is an ambitious goal, to be sure, not just for our four months long course but also for any research group as well. In fact, we do not have a comprehensive answer to this question yet. But we have some pieces of the puzzle. In this course, we will learn these pieces and thereby partially fulfil the goal. But we know that there are still many pieces that are missing in the puzzle. In fact, we do not even know how many pieces are missing. There is a lot of work left to be done. In addition to learning what the known pieces of the puzzle are, we will learn both the basic tools and modern techniques that researchers are using to find the missing pieces. Perhaps you will also join the search after taking this course!

II. DYNAMICAL SYSTEMS

A. Motivation for studying dynamical systems

In the most general sense, a **dynamical system** is any system in which there is a rule that describes the motion of a point-particle in an abstract space. "Abstract space" means that the particle is not necessarily a physical particle made of atoms. More specifically, the abstract space is called a phase space and it can have any number of dimensions (e.g., 6-dimensions). A position in the phase space is X (e.g. $X = (x, y, z, \dot{x}, \dot{y}, \dot{z})$ is a position in a 6-dimensional phase space) and for every point X, there is a rule that tells us how the X changes over time. An example of a dynamical system is,

$$\frac{dx}{dt} = ax^2 - b \tag{1}$$

where a and b are positive constants. We can view this as one-dimensional dynamical system, in which the phase space consists of the x-axis. Then Eq. 1 sketches out a **vector field** in one dimension in which the direction of the vector (arrow) indicates in which direction the x changes after an infinitesimal time dt (Fig. 1). A **fixed point** of a dynamical system is a point in the phase space at which the particle can remain at rest over time (i.e., the position of the particle in the phase space does not change over time). A fixed point can be classified as a **stable fixed point** and an **unstable fixed point**. A stable fixed point is a fixed point whereby *any* small perturbation causes the particle to return to the fixed point at a later time. An unstable fixed point is a fixed point whereby *any* small perturbation causes the particle to move further away from the fixed point and not return to it. There are other types of fixed points. We will turn to these in later lectures. In the case of Eq. 1, a fixed point occurs when $dx/dt = 0$. This happens when $ax^2 = b$, which occurs when either $x = \sqrt{b/a}$ or $x = -\sqrt{b/a}$.

Graphically, we can see that $x = \sqrt{b/a}$ is an unstable fixed point and $x = -\sqrt{b/a}$ is a stable fixed point (Fig. 1). We can also see this analytically by approximating Eq. 1 near the fixed points by a Taylor approximation. For example, when $x = -\sqrt{b/a} + \delta x$, where δx is a small value, then Eq. 1 becomes

$$\frac{d(\delta x)}{dt} = a(-\sqrt{b/a} + \delta x)^2 - b \tag{2a}$$

$$= b\left(1 - \delta x\sqrt{\frac{a}{b}}\right)^2 - b \tag{2b}$$

$$= b\left(1 - 2\delta x\sqrt{\frac{a}{b}}\right) - b \tag{2c}$$
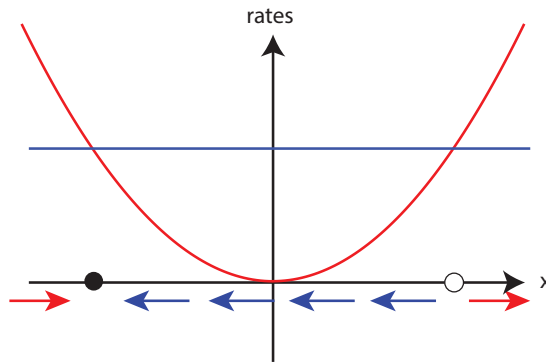
$$= -2\sqrt{ab}\delta x \tag{2d}$$

Figure 1. Example of a dynamical system (Eq. 1). Arrows represent the direction of motion of a particle along the x-axis. The white circle is the unstable fixed point. Black circle is the stable fixed point. To obtain the direction of the arrows, we look at the two rates - ax^2 and b - and determine which one is greater than the other. Then using Eq. 1, we deduce in which direction the x changes over time. The colour of the arrows indicate which of the two rates (b - blue line; ax^2 - red line) is larger.

Solving Eq. 2d yields

$$\delta x(t) = \delta x_0 e^{-2\sqrt{ab}t} \tag{3}$$

Thus the δx exponentially decays to zero over time, meaning that any small initial perturbation δx_0 around the fixed point ($x = -\sqrt{b/a}$) returns the particle back to this fixed point. Thus this fixed point is a stable fixed point. By repeating the same procedure for the other fixed point, you will find that any small initial perturbation δx_0 takes the particle further away from the fixed point exponentially over time.

In this course, we are interested in studying basic theory of dynamical systems because biological systems are all about changes that occur over time. That is, *biological systems, involving genes, proteins, and cells, are all about dynamics.* The basic aspects of the theory of dynamical systems thus helps us quantitatively describe biological processes that involve many cells, proteins, and genes. As an example, we next turn to describing a particular type of dynamical system, namely the growth and death of a population of cells.

B.   Cell replication dynamics: Logistic differential equation

We want to describe how the total number of cells (population size) changes over time due to each cell asexually replicating itself (i.e., by mitosis). This may sound simple but there are actually several complications and interesting features even in this seemingly simple task. Let $N(t)t$ be the total number of cells at time t. Let T be the time that it takes to replicate. Then we have

$$N(t) = 2^{t/T} N_0 \tag{4}$$

Note that this has the property that $N(nT) = 2^n N_0$, which is exactly what we should have, but it has the disadvantage that N takes non-integer values. To remedy this, we can divide N by a large volume V. By defining $x(t) = N(t)/V$, we have

$$x(t) = 2^{t/T} x_0 \tag{5}$$

where $x_0 = N_0/V$. Note that x is still discrete since it changes by steps of $1/V$. But by making V to be sufficiently large, we can make x to be nearly continuous. We can rewrite above as

$$x(t) = x_0 exp(tln(2)/T) \tag{6a}$$
$$= x_0 e^{\mu t} \tag{6b}$$

where $\mu = ln(2)/T$ is called the **growth rate**. By taking the derivative of both sides of Eq. 6b, we have

$$\frac{dx}{dt} = \mu x \tag{7}$$

This equation describes an exponential growth of the population. According to Eq. 6b, the population size should approach infinity as time passes. But this would mean that there is an infinite amount of energy in the universe since some energy is required to make a cell (i.e., replicate) as well as an infinite amount of space to keep all those cells. Of course, neither of these is true. Moreover, we assumed that no cell dies, which is also evidently untrue. Let's first include cell death. Let us assume that each cell has some fixed chance of dying. Then we modify Eq. 9 to get

$$\frac{dx}{dt} = \mu x - \gamma x \tag{8}$$

where γ is a constant with a dimension of 1/time. But note that as long as $\mu > \gamma$, we still have the population size going to infinity. To see this, note that the solution to Eq. 9 is

$$x(t) = x_0 exp((\mu - \gamma)t) \tag{9}$$

When $\mu < \gamma$, the population size approaches zero ($x = 0$ is a stable fixed point). Thus neither of the two cases - $\mu > \gamma$ and $\mu < \gamma$ - make sense after a sufficiently long time. We would like a growth dynamics that makes sense for all times, not just for a sufficiently small window of time. The key to a more realistic description of cell growth is the fact that cells have limited space and resources (e.g., sugar - the source of energy). We say that there is a carrying capacity for cells. More specifically, let us define a constant K, which we call the carrying capacity. If the population size is larger than the carrying capacity, then we want the net growth rate to be negative. If the population size is smaller than the carrying capacity, then we want the net growth rate to be positive. A simple way to set-up such a dynamics is the following:

$$\frac{dx}{dt} = \mu_{\text{eff}}x \tag{10a}$$

$$= r\left(1 - \frac{x}{K}\right)x \tag{10b}$$

Note that r is the constant reproduction rate when there is no resource limitation (i.e., when the carrying capacity is infinite). Eq. 10b is called the **logistic equation**. It is one of the most common equations used for modeling population growth. This equation is not easy to analytically solve. But we can find its fixed points and their stability without knowing the solution. A fixed point occurs when $dx/dt = 0$ in Eq. 10b. From this condition, we find two fixed points:

$$x_1 = 0 \qquad\qquad x_2 = K \tag{11}$$

To find their stability, we can use either a Taylor approximation or a graphical method (Fig. 2). Let's use a Taylor approximation to first order in δx. For a small deviation δx about x_1, Eq. 10b becomes

$$\frac{d(\delta x)}{dt} = r\left(1 - \frac{x_1 + \delta x}{K}\right)\left(x_1 + \delta x\right) \tag{12a}$$

$$= r\left(1 - \frac{x_1 + \delta x}{K}\right)x_1 + r\left(1 - \frac{x_1 + \delta x}{K}\right)\delta x \tag{12b}$$

$$\approx r\left(1 - \frac{x_1}{K}\right)x_1 - r\frac{\delta x}{K}x_1 + r\left(1 - \frac{x_1}{K}\right)\delta x \tag{12c}$$

$$= r\left(1 - \frac{2x_1}{K}\right)\delta x \tag{12d}$$

Note that we deliberately kept writing x_1 instead of writing zero in place of it. This is because we can now see that a small perturbation about x_2 yields an equation of the same form:
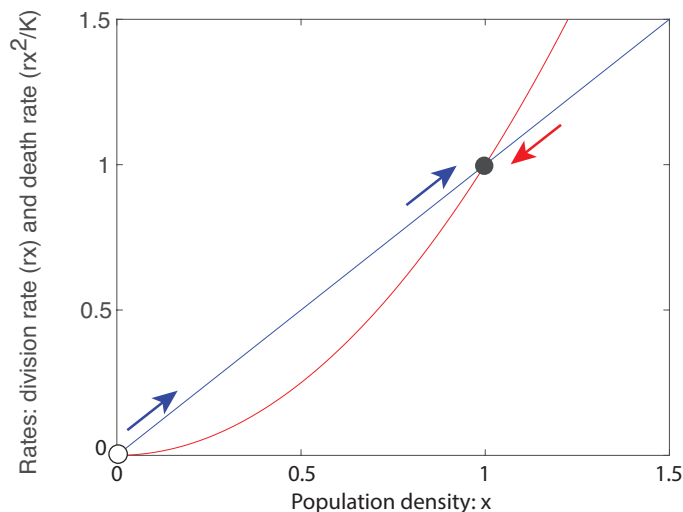
Figure 2. A graphical method for finding the fixed points and analyzing their stability for the logistic differential equation (Eq. 10b). Blue line is the birth rate rx and the red curve is the death rate rx^2/K. Here, we picked $r = 1$ and $K = 1$. The points where the two curves intersect (black and white points) are the fixed points of Eq. 10b. Black point is a stable fixed point and the white point is an unstable fixed point. Blue and red arrows indicate the direction in which the x and the net rate $(rx - rx^2/K)$ change due to a small perturbation in x from its fixed point value.

$$\frac{d(\delta x)}{dt} = r\left(1 - \frac{2x_2}{K}\right)\delta x \tag{13}$$

From Eqs. 12d and 13, we see that for the small perturbation around $x_1 = 0$, we have

$$\delta x(t) = \delta x_0 exp(rt) \tag{14}$$

whereas for the small perturbation around $x_2 = 0$ we have

$$\delta x(t) = \delta x_0 exp(-rt) \tag{15}$$

Thus, a small initial perturbation by an amount δx_0 around the fixed point x_1 leads to an exponentially growing population size to a point that the small perturbation approximation is no longer valid after a long enough time. A small initial perturbation by an amount δx_0 around the fixed point x_2 leads to the population size reducing back to the carrying capacity K. From these, we conclude that x_1 is an *unstable fixed point* and that x_2 is a *stable fixed point*. The advantage of the Taylor approximation method over the graphical method is that we actually obtain exactly *how* fast a small initial perturbation either reduces or grows over time.

C.   Cell replication dynamics: Logistic difference equation

As mentioned before, the number of cells N and its concentration x, are discrete variables. The fact that $x = N/V$, where we chose V to be some very large volume, mitigates the problem since we could take V to be any large number (it doesn't have to be related to the space in which the cells live in). But if we were interested in the *actual* concentration of cells, then we would have to make the V to be the actual volume of the space (e.g., cell culturing tube) in which the cells live in. In that case, if N is sufficiently small at all times, which would be the case if the carrying capacity is sufficiently low, then the logistic differential equation is unsuitable for describing the population growth. As a remedy, we now study a discrete version of the logistic differential equation. Looking at the logistic differential equation, we can write the following as its discrete counterpart:

$$x_{t+1} - x_t = r\left(1 - \frac{x_t}{K}\right)x_t \tag{16}$$

where t is either zero or a positive integer (i.e., $t = 0, 1, 2, 3....$). Thus the change in time also happens in discrete units of one. Note that x_t is not necessarily an integer since it is a population density. It also changes by discrete units and the smallest discrete unit of change (i.e., $1/V$) approaches zero as the volume V approaches infinity. By rearranging above equation, we obtain

$$x_{t+1} = x_t \left((r+1) - r\frac{x_t}{K} \right) \tag{17}$$

We now make some assumptions to simplify Eq. 17. First, we let $K = 1$. Then we have

$$x_{t+1} = x_t \left((r+1) - rx_t \right) \tag{18}$$

We can analyze above equation but for historical reasons, the actual **logistic difference equation** is

$$x_{t+1} = ax_t \left(1 - x_t \right) \tag{19}$$

where a is a positive constant. Despite the modification from Eq. 18 to Eq. 19, it turns out that both equations share the same main properties. The added advantage of Eq. 19 over Eq. 18 is that the former is simpler to write. We now analyze the properties of the logistic difference equation. First, we note that negative values of x do not make sense for a population density. This also means that we cannot consider values of x that are larger than one for otherwise, x_{t+1} becomes negative in Eq. 19. Thus we conclude that Eq. 19 is valid only when $0 \leq x \leq 1$. Next, we see that $x_t = 0$ is a fixed point for the dynamics, regardless of the value of a. Finally, for modeling population growth, we want a to be a positive constant. It turns out that a cannot be too large though. Specifically, if $a > 4$, we can get x_t to be larger than 1 (you will prove this in problem set 1). Thus, we must have $0 \leq a \leq 4$.

We can obtain the main properties of the logistic difference equation (Eq. 19) without solving it. To do so, we restrict a to certain intervals and then ask how the solution of the equation, x_t, behaves over time.

- Case 1 ($0 < a < 1$): The only stable fixed point is $x = 0$ (you will show this in problem set 1). The reason is that the sequence $\{x_0, x_1, x_2, ...\}$ is a convergent sequence when $a < 1$ (i.e., it is a **Cauchy sequence**). To prove this, we note that any infinite sequence (which is what $\{x_t\}$ dictated by Eq. 19 is) that is non-increasing and bounded below converges. Then noting that zero is the greatest lower bound of our sequence completes the proof. Thus the population dies out no matter what the initial population density is.

- Case 2 ($1 < a < 3$): The *only* stable fixed point is $x = (a-1)/a$ (you will argue that this is true in problem set 1). $x = 0$ is also a fixed point but it is unstable. This means that as long as the initial population density is between zero and one (i.e., $0 < x_0 < 1$), the population size always converges to $x = (a-1)/a$ (note that when $x_0 = 1$, then every cell dies in the next time step (i.e., $x_1 = 0$)). Thus when $a = 2$, we see that half of the carrying capacity (i.e., $1/2$) is the only stable fixed point. In the language of dynamical systems, we also say that $x = (a-1)/a$ is a **global attractor** on the *open* interval $(0, 1)$.

- Case 3 ($3 < a < 4$): Now, $x = (a-1)/a$ is an *unstable* fixed point. As you will show in problem set 1, if the a is only slightly greater than 3, then x_t oscillates over time with a period of 2, meaning that if the population density starts with a value between zero and one (i.e., $0 < x_0 < 1$), then the population density changes to some other value at the next time step, and then it returns to x_0 at $t = 2$. This dynamics then repeats since the population is back to the initial size and Eq. 19 depends only on the current population size, rather than on the population's history. We call this population dynamics a **period-2** oscillation.

  Interestingly, as we increase the a further above 3, mathematicians have found that the population density still oscillates but with a different period. In particular, they have observed a **period doubling** - the population density oscillates as a period-4 oscillator, and then if we increase the value of a further, it becomes a period-8 oscillator. If we increase the value of a even further, then the period-8 oscillator becomes a period-16 oscillator and so on. This occurs all the way until we hit $a = 3.57$. When $3.57 \leq a < 3.6787$, mathematicians have observed that the population density can have an infinitely many even-numbered periods. Which one of the even periods the population density has depends on the initial population density x_0. Furthermore, at $a = 3.6786$, mathematicians have observed that odd-numbered periods begin to appear. Finally, when $3.82 < a \leq 4$,

mathematicians have found that all-integer values of periods, even or odd, can occur. Which one is the actual period again depends on the initial population density. Although we would like to prove these statements, doing so is beyond the scope of our course and in fact involves advanced mathematics. It is no surprise, then, that the logistic difference equation (Eq. 19) is and has been one of the most frequently studied equations in dynamical systems.

- Case 4 (a=4): Here, the population size changes *chaotically* over time. This means that if we have two populations that begin with very slightly different initial densities (e.g., they are different by only 0.0000001), then these two populations may change very differently over time - so much so that we would not be able to believe that the two populations really came from such a slightly different initial densities. Note that Eq. 19 is a deterministic equation. Thus, if we have two populations that have the same initial density, then these two populations will have *exactly* the same population density in subsequent times. Chaos does *not* mean stochastic - we will turn to stochasticity later. **Chaotic dynamics** simply means that two different initial values that can be made arbitrarily close to each other can still yield two very distinct dynamics that do not closely follow each other. Since Eq. 19 is a **deterministic equation**, we say that it produces a **deterministic chaos** when $a = 4$. You will computationally explore this chaotic dynamics in problem set 1.

In summary, the logistic difference equation (Eq. 19) shows that a seemingly simple-looking, deterministic equation that models biological systems can have rich properties (which have occupied mathematicians for a long time) and even generate highly non-trivial, unpredictable dynamics such as chaotic dynamics.

D. Cell replication dynamics: Stochastic growth and death

So far, we have used deterministic equations to model cell division and death. We saw that defining the effective growth rate μ_{eff}, which can be a positive or a negative number, and having this depend on the population size can produce a population that does not infinitely grow. But a shortcoming so far is that all the equations that we used above are deterministic equations. In other words, they were equations for which if we know the population density at a given time, then the equation tells us what the population density is at all subsequent times and previous times. Nothing is left undetermined. This "clockwork" or "mechanical" view of population dynamics is unrealistic. As we will see in this course, cellular processes such as gene expression, cell division, and cell death, are **stochastic**. This means that if we were to watch a single cell, we cannot predict with absolute certainty when the cell will divide or die. Why is this? This question leads us to the topic of **stochastic dynamics**.

A typical reason that is often given in the literature for cellular processes, including cell growth and division, being stochastic is that there are very few copies of each molecule inside a cell that govern each cellular process. As an example, a cell may contain between 1 to 100 copies of an mRNA that is transcribed from a particular gene. Dividing 1 to 100 mRNA molecules by the volume V of a cell yields a discrete series of concentrations: $1/V$, $2/V$,?, and $100/V$. Thus, ordinary differential equations like the logistic differential equation (Eq. 10b) fail to capture these discrete changes in the quantity that they model (e.g., population density x in the case of the Eq. 10b). But intuition tells you that if $1/V$ is "small enough", then we can "smooth out" the discreteness. As we will see, in the case of the logistic equation, this means that $1/V$ is much smaller than $< x >/V$, where $< x >$ is the average population density. But note that it is not easy to see from this reasoning, that number of molecules or cells involved in a process change by discrete units or that they are small, why we cannot use the logistic difference equation which does change the population density x in discrete units. In fact, there is a deeper reason behind why cellular processes are stochastic.

1. Why are cellular processes stochastic?: A "deeper" answer

Although a few copies of molecules are involved in a cellular process, it does not logically follow that the process should be non-deterministic (i.e., stochastic). For instance, if we have five billiard balls elastically bouncing off each other and off the sides of a pool table, their motion may seem completely random over time. This collection of balls may even be a chaotic dynamical system. Nonetheless, the laws are completely deterministic. Our inability to calculate the precise trajectory of each ball is due to our limitations in mathematical or technical inability to precisely measure the exact position and velocity of each ball (this precision matters for dynamics occurring over a long time scale). But at the end of the day, the Newton's deterministic laws of motion, completely dictates how each ball moves in the future and how it moved in the past. This is not a stochastic system despite consisting of only five balls. So just saying that a cellular process like cell division involves a discrete number of cells or that a few copies of mRNA and proteins inside a cell control the cell division, is insufficient to explain why cellular processes

are stochastic. The deeper and more subtle reason, ironically, is that many molecules are actually involved in cellular processes including cell division. These molecules include water molecules that bombard the few mRNA molecules inside a cell. Bombardment by these molecules cause each mRNA molecule to take a random walk (collectively, this is called a **diffusion**) inside a cell. We give this process a stochastic description because we do not know the position and velocity of every water molecule inside the cell. Moreover, we do not know the position and velocity of every other type of molecules inside the cell that are diffusing around and bombarding the few mRNA and proteins inside the cell that control cell division or death. In short, the randomness is apparent - it is only due to our ignorance of the state of all the particles inside and outside the cells. That is why we cannot describe exactly where each transcription factor will be inside a cell. Thus we cannot describe exactly when each transcription factor or ribosome or RNA polymerase will bind at a particular location on a DNA or a RNA. If we view molecules diffusing inside a cell, simply from a classical physics perspective, then the randomness is really due to our ignorance of what every single molecule is doing. This is precisely what happens in systems of interest in statistical physics, such as a box of ideal gas particles. When it comes to cell division or cell death, we are also ignorant of all the factors that affect them. In many cases, we are not even sure what and how many factors affect cell division and death. Thus, *our lack of knowledge about all the factors involved in cell division and death necessitates a stochastic description of the population dynamics.*

Now that we know why cellular processes, including cell division and death, are stochastic, we now describe one of two main approaches for describing stochastic processes called the **Master equation** approach. The other approach, called the "Fokker Planck approach", will be covered when we describe stochastic gene expression in a later lecture.

2. Master equation description of stochastic cell division and death

Since we lack all the information that is necessary to predict exactly when the cell will divide or die, the best that we can do is find the probability that a cell will divide or die in within some time interval. Let $P_n(t)$ be the probability that there are n cells in the population at time t. To experimentally measure this probability, we could try two things. Suppose that we start with 1000 populations of cells that each start with 10 cells. Moreover, suppose that all cells are genetically identical. If we take a microscope camera and take a snapshot of the 1000 populations at a later time t and then count the number of cells in each population, we can plot the histogram of the number of cells in each population. This histogram estimates the $P_n(t)$ for every n (this estimate becomes exact as the number of populations that we measure becomes infinite, which is experimentally impossible). Another method is, instead of taking a snapshot in time, we can just focus on a single cell and make a time-lapse movie of that cell. We will then record when the cell divides, potentially several time, before it dies. Then after a sufficiently long enough time, we can plot the histogram of the *intervals* between adjacent cell divisions. If the systems is ergodic, then the histograms obtained by both methods would be equal. In most cases, we assume that indeed both histograms would be equal. Mathematically proving that a system is indeed ergodic is difficult and it actually depends on time scale that we restrict ourselves to. For example, if we wait long enough, mutations would occur in the cell's genome, which might in turn cause the cell to divide with a different average division time. Evidently, according to this reasoning, cellular processes are not ergodic over a sufficiently long time scale. But for the purpose of most experiments, we are not interested in how a circuit behaves in evolutionary time scales. Thus we assume ergodicity. Then consider the "flow chart" of events shown in Fig. 3.
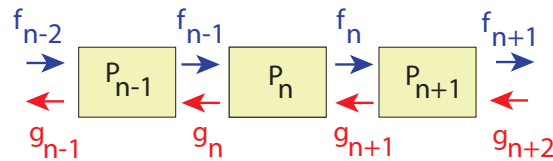

Figure 3. A **Markov chain**: Visualisation of stochastic birth and death of a cell. n is the number of cells in a population. $f_n dt$ to be the probability that, in a population of n cells, there is one cell division within an infinitesimal time interval dt. $g_n dt$ is the probability that, in a population of n cells, there is one cell dying within an infinitesimal time interval dt.

As shown in Fig. 3, we define $f_n dt$ to be the probability that, in a population of n cells, there is one cell division within an infinitesimal time interval dt. We define $g_n dt$ to be the probability that, in a population of n cells, there is one cell dying within an infinitesimal time interval dt. By keeping the time interval dt to be infinitesimally small, we ensure that at most one cell is created or destroyed in this time interval. Once again, $P_n(t)$ is the probability that there are n cells in the population at time t. In an experiment involving, say 10,000 populations that all started with

the same number of cells, $P_n(t)$ would be the fraction of those 10,000 populations that have n cells at time t. The diagram shown in Fig. 3 is called the **Markov chain** because it shows a chain of events and the probabilities f_n and g_n do *not* depend on how P_n was changing over time but rather on its value *now*. Such a stochastic process is called a **Markov process**. Note that the Markov chain in Fig. 3 is a dynamical system because a current position n, would change over time by the stochastic rules dictated by the f_n's and g_n's. Specifically, looking at the Markov chain (Fig. 3), we see that the dynamical system is dictated by

$$\frac{dP_n}{dt} = f_{n-1}P_{n-1} + g_{n+1}P_{n+1} - g_nP_n - f_nP_n \tag{20}$$

Eq. 20 is called the **Master equation**. As we will see in this course, it applies not just to cell division and death but to all stochastic processes, whether they are biological or not, including gene expression and random-walk of particles. The Master equation holds for every value of n. There are infinitely many equations (one for each P_n) that are coupled to each other because P_n depends on P_{n-1} and P_{n+1}. We do not know how to analytically (without computers) solve the Master equation for arbitrary functions f_n and g_n. We will see in this course that there are computational algorithms that simulate how P_n changes over time (e.g., **Gillespie algorithm**). Analytically, we can find the steady-state values of P_n (i.e., P_n no longer depends on time) if we restrict f_n and g_n to be particular functions of n. Let us now turn to one such situation.

3. Master equation - Constant probabilities for cell division and death

We now analyze Eq. 20 when the *probability density* for cell division f_n is kn, where k is a positive constant, and the *probability density* for cell death g_n is γn, where γ is a positive constant. Here, we are simply saying that each cell has the same constant probability of giving birth (kdt) and a constant probability of dying (γdt) in a time interval dt. Here we ask how the average number of cells $<n>$ changes over time. By the definition of averages, we have

$$\frac{d<n>}{dt} = \frac{d}{dt}\sum_{n=0}^{\infty} nP_n \tag{21a}$$

$$= \sum_{n=0}^{\infty} n\frac{dP_n}{dt} \tag{21b}$$

Substituting the master equation (Eq. 20) into the dP_n/dt in Eq. 21b leads to

$$\frac{d<n>}{dt} = -(k+\gamma)\sum_{n=0}^{\infty} n^2 P_n + k\sum_{n=0}^{\infty} n(n-1)P_{n-1} + \gamma\sum_{n=0}^{\infty} n(n+1)P_{n+1} \tag{22}$$

We can simplify Eq. 22 by using a mathematical trick: $n = (n+1) - 1$ and $n = (n-1) + 1$. We also use the fact that $P_{-1} = 0$ at all times (since we cannot have $n < 0$) and that the sum of all the P_n's equals to one at all times (since each P_n is a probability). Furthermore, using the fact that

$$<n>= \sum_{n=0}^{\infty} nP_n, \tag{23}$$

Eq. 22 becomes

$$\frac{d<n>}{dt} = (k-\gamma)<n> \tag{24}$$

Eq. 24 has the same form as the deterministic equation for population size (Eq. 9) when the birth and death rates are held constant. The only difference is in the meaning: Eq. 9 assumes that the population density "x", and thus $n = xV$, are deterministic where as Eq. 24 does not treat the number of cells n as a deterministic quantity. In other words, according to Eq. 24 assumes that two populations with the same number of cells at the same time will keep on having the same number of cells at every instance in the future and that they had the same number of cells in

the past. On the other hand, Eq. 24 is a deterministic equation that governs how the average number of cells $< n >$ changes over time that we derived by starting from a stochastic description, in which two populations that have the same number n of cells at the same time, can change in their size differently in the future. In deriving Eq. 24, we also see where the rate constants μ and γ in Eq. 24 come from: They are simply the probability densities for division and death respectively.