

# 1 Introduction

## 1.1 The philosophy of the course

Probability is the logic of science<sup>1</sup>; statistics is a tool for making discoveries from data; and *dynamics* is a language for narrating exact science. In modern mathematics, probability does not necessarily mean stochastic; it is about measure-theoretical representation of a subject.

Newtonian world view is a clockwise machine; complex many-body systems always have “dissipation”. So where is the biological diversity coming from? You should keep this very important question in mind all through this course.

## 1.2 Dynamic models

Experimental biology follows a reductionistic approach in which modular, functional mechanisms are elucidated one piece at a time. But life is a complex phenomenon at every level, from cells to organisms, to populations, due to interactions among multiple, heterogeneous components. Therefore, in all area of biology, mathematical models provide the means for putting the pieces together.

*Dynamic models* describe how a system’s properties, in a simplified representation, change over time. Dynamic models have a unique role in science: It is the only method that is able to definitively provide a *sufficient condition* for an observed phenomenon or phenomena. In modern biology, this is called a *mechanism*. It establishes a causal relation with certainty.

There are two types of models: “data-driven” descriptive models and mechanistic models. One of the best known data-driven descriptive models is perhaps Kepler’s three laws of planetary motion. Most current statistical models obtained from “big data” belong to this category. Even when these models can provide accurate predictions, it does not tell us why the data behave the way they are — a fundamental element of what we call “understanding”. In contrast, a mechanistic model

A dynamic model has two essential components: *state variables* and *dynamic equations*. One should visualize a dynamics as a “point”  $\vec{x} = (x_1, x_2, \dots, x_n)$  moving in a  $n$ -dimensional space as a function of time. One of the most important assumptions in a dynamic model is that the state of the system at time  $t + \Delta t$  is completely determined by the state of the system at time  $t$ :  $\vec{x}(t) \rightarrow \vec{x}(t + \Delta t)$ .

A significant portion of the equations in biology are simply “counting the numbers”, or density. This is discussed in the textbook as “Bathtub models”, or I would like to call it “balance checkbooks”:

$$\frac{dW(t)}{dt} = I(t) - O(t),$$

where  $W(t)$  is the amount of water in the bathtub,  $I(t)$  and  $O(t)$  are the inflow and outflow

---

<sup>1</sup>E. T. Jaynes (2003) *Probability Theory: The Logic of Science*, 1<sup>st</sup> ed., Cambridge Univ. Press.

rates, e.g., the amount of water going into and coming out the bathtub *per unit time*. In the banking language:  $W(t)$  is the amount of money in the account,  $I(t)$  is the rate of deposits, and  $O(t)$  is the rate of expense.

### 1.3 Simple models with a few equations

One important application of mathematical modeling is in population dynamics. This can be about populations of biological organisms, chemical species inside a test tube, or sociological and economical agents. As long as one has the notion of different “individuals”, there is the concept of a “population”.

Just as the bathtub problem, population dynamics usually starts with an equation like this:

$$\text{rate of population increase} = \text{birth rate} - \text{death rate} + \text{immigration rate.} \quad (1)$$

If we use  $x(t)$  to denote the population at time  $t$ , then the above equation becomes

$$\frac{dx}{dt} = x(b(x) - d(x)) + i(x), \quad (2)$$

in which  $b$  and  $d$  are the *per capita* birth and death rates, respectively. Note that one of the most important aspects of birth and death is that if  $x = 0$ , then there will be no possibility of further birth or death. Without immigration, an extinct population will remain extinct. The immigration term  $i(x)$ , however, has a very different feature: It needs not to be zero when  $x = 0$ .

Consider a population with many subpopulations  $\vec{x} = (x_1, x_2, \dots, x_n)$ , all  $x_i \geq 0$ . In the absence of immigration, if we denote  $r_i(\vec{x}) = b_i(\vec{x}) - d_i(\vec{x})$ , then

$$\frac{dx_i}{dt} = x_i r_i(\vec{x}), \quad (3)$$

and the per capita growth rate for the entire population, which is also the mean per capita growth rate,

$$\bar{r} = \frac{\sum_{i=1}^n \frac{dx_i}{dt}}{\sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n x_i r_i(\vec{x})}{\sum_{i=1}^n x_i}, \quad x_i \geq 0. \quad (4)$$

Then,

$$\frac{d\bar{r}(\vec{x})}{dt} = \left[ \frac{\sum_{i=1}^n x_i r_i^2}{\sum_{i=1}^n x_i} - \left( \frac{\sum_{i=1}^n x_i r_i}{\sum_{i=1}^n x_i} \right)^2 \right] + \frac{\sum_{i,j=1}^n x_i x_j r_j \left( \frac{\partial r_i}{\partial x_j} \right)}{\sum_{i=1}^n x_i} \quad (5)$$

We note that the term inside  $[\dots]$  on the right-hand-side is never negative:

$$\frac{\sum_{i=1}^n x_i r_i^2}{\sum_{i=1}^n x_i} - \left( \frac{\sum_{i=1}^n x_i r_i}{\sum_{i=1}^n x_i} \right)^2 = \frac{\sum_{i=1}^n x_i (r_i - \bar{r})^2}{\sum_{i=1}^n x_i} \geq 0, \quad (6)$$

In fact, it is actually the variance of  $r_i$  among the different subpopulations. Therefore, it is always positive if there are variations among  $r_i$ . This mathematical result is a part of the ideas of both Adam Smith, on economics, and Charles Darwin, on the natural selection. In fact, the term  $[\dots]$  in Eq. (5) has been identified by R. A. Fisher, the British statistician and evolutionary biologist, as the “growth of fitness due to natural selection”.<sup>2</sup> And here is a quote from Smith’s *magnum opus* “An Inquiry into the Nature and Causes of the Wealth of Nations” (1776):

“As every individual, therefore, endeavours as much as he can both to employ his capital in the support of domestic industry, and so to direct that industry that its produce may be of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public good. It is an affectation, indeed, not very common among merchants, and very few words need be employed in dissuading them from it.”

Eq. 5 can be written as:

$$\frac{d}{dt} \left( \frac{\sum_{i=1}^n x_i r_i}{\sum_{i=1}^n x_i} \right) - \frac{\sum_{i=1}^n x_i \frac{dr_i(\vec{x})}{dt}}{\sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n x_i (r_i - \bar{r})^2}{\sum_{i=1}^n x_i}. \quad (7)$$

This equation can be phrased as “the change in the per capita growth rate of an entire population is never less than the average change in per capita growth rate of the subpopulations”.<sup>3</sup> Eq. 5 also shows that  $d\bar{r}/dt$  could be negative if the last term on the right-hand-side is large and negative. Therefore, it is interesting to investigate under what circumstances it is positive or negative.

First, we note that if all  $r_i$  are constant, independent of  $\vec{x}$ , then this last term is zero since  $(\partial r_i / \partial x_j) = 0$ .

Second, if  $r_i$  is a linear function of  $\vec{x}$ :  $r_i(\vec{x}) = \sum_{k=1}^n w_{ik} x_k$ . Furthermore, one can always

<sup>2</sup>Edwards, A.W.F (1994) The fundamental theorem of natural selection. *Biol. Rev.* 69, 443–474.

<sup>3</sup>Price, G.R. (1972) Fishers fundamental theorem made clear. *Ann. Hum. Genet. Lond.* 36, 129–140.

decompose a matrix into a symmetric and an anti-symmetric parts:  $w_{ij} = w_{ij}^S + w_{ij}^A$ . Then

$$\begin{aligned} \frac{\sum_{i,j=1}^n x_i x_j r_j \left( \frac{\partial r_i}{\partial x_j} \right)}{\sum_{i=1}^n x_i} &= \frac{\sum_{i,j,k=1}^n x_i w_{ij} x_j w_{jk} x_k}{\sum_{i=1}^n x_i} \\ &= \frac{\sum_{i,j,k=1}^n x_i w_{ij}^S x_j w_{jk}^S x_k}{\sum_{i=1}^n x_i} + \frac{\sum_{i,j,k=1}^n x_i w_{ij}^A x_j w_{jk}^A x_k}{\sum_{i=1}^n x_i} \\ &= \frac{\sum_{j=1}^n x_j \left( \sum_{i=1}^n x_i w_{ij}^S \right)^2}{\sum_{i=1}^n x_i} - \frac{\sum_{j=1}^n x_j \left( \sum_{i=1}^n x_i w_{ij}^A \right)^2}{\sum_{i=1}^n x_i}. \end{aligned} \quad (8)$$

Hence, a symmetric interaction between subpopulations  $i$  and  $j$  increases the  $\bar{r}$ , and an anti-symmetric interaction between subpopulations  $i$  and  $j$  decreases the  $\bar{r}$ . Competition and symbiosis are the former type, and predator and prey are the latter type.

## 1.4 Complex dynamics such as a single protein in water

<http://www.youtube.com/watch?v=iaHHgEoa2c8>

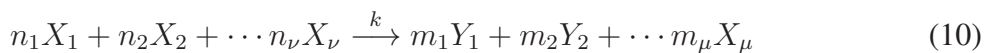
<http://www.youtube.com/watch?v=gFcp2Xpd29I>

<http://www.youtube.com/watch?v=Y79Xl0LfYI4>

## 1.5 Michaelis-Menten enzyme kinetics



The *law of mass action* from chemical reaction theory states that a chemical reaction like



has a rate constant  $k$ , and the rate of reaction  $J$ , e.g., number of chemical reaction (10) per unit time:

$$J = k x_1^{n_1} x_2^{n_2} \cdots x_\nu^{n_\nu},$$

where  $x_k$  is the concentration of chemical species  $X_k$  among the reactants. Then

$$\frac{dx_k}{dt} = -n_k J, \quad k = 1, 2, \cdots, \nu;$$

and

$$\frac{dy_k}{dt} = m_k J, \quad k = 1, 2, \cdots, \mu,$$

where  $y_k$  is the concentration of chemical species  $Y_k$  among the products.

Applying the law of mass action to Eq. (9), we have

$$\frac{ds}{dt} = k_{-1}c - k_1es, \quad (11a)$$

$$\frac{de}{dt} = (k_{-1} + k_2)c - k_1es, \quad (11b)$$

$$\frac{dc}{dt} = k_1es - (k_{-1} + k_2)c, \quad (11c)$$

$$\frac{dp}{dt} = k_2c. \quad (11d)$$

The initial conditions are

$$s(0) = s_0, \quad e(0) = e_0, \quad c(0) = p(0) = 0. \quad (11e)$$

We observe that  $\frac{dc}{dt} + \frac{de}{dt} = 0$  and  $\frac{ds}{dt} + \frac{dc}{dt} + \frac{dp}{dt} = 0$ . This can be understood by going through the biochemical reaction “mechanism” and recognize that the total enzyme  $e_0$  and total substrates  $s_0$  are conserved. Substituting equations

$$c + e = e_0, \quad s + c + p = s_0$$

into Eq. (11), and eliminating  $e$  and  $p$ , we have

$$\frac{ds}{dt} = k_{-1}c - k_1e_0s + k_1cs, \quad (12a)$$

$$\frac{dc}{dt} = k_1e_0s - k_1cs - (k_{-1} + k_2)c \quad (12b)$$

$$s(0) = s_0, \quad c(0) = 0. \quad (12c)$$

*Non-dimensionalization.* The two equations in (14) are not yet ready to be analyzed computationally. Note that since a computation has to have all the parameters in the equations assigned with numerical values, explore the general behavior of a system differential equations involves many calculations for different parameter values. Thus the fewer the parameter, the better. The system (14) seems to have four parameters:  $k_1, k_{-1}, k_2, e_0$ , and  $s_0$ . But actually, it has less.

Note that the one does not have to use the standard unit, such as molar, for the concentrations  $s$  and  $c$ , nor standard unit for time, such as second, for  $t$ . Rather, one can try to use some “internal units”. First, we note that  $k_{-1}$  must have “dimension” of  $[\text{time}]^{-1}$  since  $k_{-1}c \sim \frac{ds}{dt}$  which is  $[\text{concentration}][\text{time}]^{-1}$ . Similarly,  $k_1$  has a dimension of  $[\text{concentration}]^{-1}[\text{time}]^{-1}$ , thus  $k_1e_0$  has a dimension of  $[\text{time}]^{-1}$ . Now let us introduce “non-dimensionalized variables”

$$u = \frac{s}{s_0}, \quad v = \frac{c}{e_0}, \quad \text{and} \quad \tau = k_1e_0t \quad (13)$$

Then, (14) becomes

$$\frac{du}{d\tau} = \left( \frac{k_{-1}}{k_1 s_0} \right) v - u + uv, \quad (14a)$$

$$\left( \frac{e_0}{s_0} \right) \frac{dv}{d\tau} = u - uv - \left( \frac{k_{-1} + k_2}{k_1 s_0} \right) v, \quad (14b)$$

$$u(0) = 1, v(0) = 0. \quad (14c)$$

in which combined parameters  $e_0/s_0 = \epsilon$ ,  $k_2/(k_1 s_0) = \lambda$  and  $(k_{-1} + k_2)/(k_1 s_0) = K$  are all dimensionless. We finally arrive at

$$\frac{du}{d\tau} = -u + (u + K - \lambda)v, \quad (15a)$$

$$\epsilon \frac{dv}{d\tau} = u - (u + K)v, \quad (15b)$$

$$u(0) = 1, v(0) = 0. \quad (15c)$$

It has only three parameters!

One of the important features of enzyme reaction systems inside a cell is that  $e_0 \ll s_0$ . That is  $\epsilon \ll 1$ .

## 2 Radioactive decay and exponential random time

### 2.1 Random variables, probability density function, etc.

A random variable  $X$  taking a real value has a probability density function (pdf)  $f_X(x)$ :

$$\int_{-\infty}^{\infty} f_X(x) dx = 1. \quad (16)$$

The meaning of the  $f_X(x)$  is this

$$\Pr\{x < X \leq x + dx\} = f_X(x) dx. \quad (17)$$

Then, the cumulative distribution of  $X$ :

$$F_X(x) = \Pr\{X \leq x\} = \int_{-\infty}^x f_X(z) dz, \text{ and } f_X(x) = \frac{dF_X(x)}{dx}. \quad (18)$$

The mean (or expected value) and variance of  $X$  are

$$\langle X \rangle = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (19)$$

$$\text{Var}[X] = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx. \quad (20)$$

in which we have denoted  $E[X]$  by  $\mu$ . Two most important examples of random variables taking real values are “exponential” and “normal”, also called Gaussian.

Learn to use `rnorm()`, `rexp()`, `hist()`, and `nls(log(hdata[,2]) ~ a-b*hdata[,1]^2,` in which `hdata` contains the density function obtained from `hist`.

**The pdf of a function of a random variable  $X$ .** Let us have a random variable  $X$  with pdf  $f_X(x)$ . Now consider a differentiable, monotonic increasing function  $g(\cdot)$  and let  $Y = g(X)$ . So  $Y$  is also a random variable. What is the distribution of  $Y$ ? We note that

$$\Pr\{Y < y\} = \Pr\{X < g^{-1}(y)\}, \text{ i.e., } F_Y(y) = F_X[g^{-1}(y)]. \quad (21)$$

Therefore,

$$f_Y(y) = \frac{d}{dy} \Pr\{Y < y\} = \frac{d}{dy} \int_{-\infty}^{g^{-1}(y)} f_X(x) dx = f_X(g^{-1}(y)) \frac{d}{dy} (g^{-1}(y)). \quad (22)$$

Eq. (22) should be remembered as

$$f_Y(y) dy = f_X(x) dx, \text{ in which } x = g^{-1}(y) \text{ or } y = g(x). \quad (23)$$

There is a clear graphical interpretation of the formulae (21) and (23).

## 2.2 Exponential distribution

The simplest linear ordinary differential equation

$$\frac{dx}{dt} = -rx \quad (24)$$

is widely taught as a model for radioactive decay problem. More precisely, consider a block of radioactive material, the  $x(t)$  is the remaining radioactive material at time  $t$ :

$$x(t) = x(0)e^{-rt}. \quad (25)$$

The parameter  $r$  is the “rate of decay” per atom.

If all the atoms in the block are identical and independent, then  $x(t)$  can also be interpreted as the probability of a single atom in the population still not decayed at time  $t$ :

$$p(t) = e^{-rt}. \quad (26)$$

Sometime, this is called “survival probability” in the population dynamics.

However, a more careful inspection of the decays of individual atoms, one realizes that the occurrence of the “event”, i.e., a click in a Geiger counter, is random. The time when an atom decay,  $T$  is a random variable with a probability density function  $f_T(t)$ :

$$f_T(t)dt = \Pr\{t < T \leq t + dt\}, \quad (t \geq 0) \quad (27)$$

which reads “ $f_T(t)dt$  is the probability of random time  $T$  being in the interval  $(t, t + dt]$ . Then, at time  $t$ , the probability the atom is still no decayed, i.e.,  $T > t$ , is the survival probability:

$$p(t) = \Pr\{T > t\} = \int_t^\infty f_T(s)ds. \quad (28)$$

We therefore have

$$f_T(t) = -\frac{dp(t)}{dt} = re^{-rt}. \quad (29)$$

The random time  $T$  has an exponential distribution. Its mean value, also called expected value, is

$$\langle T \rangle = \int_0^\infty t f_T(t)dt = \frac{1}{r}. \quad (30)$$

In fact, there is a variance in the random time  $T$ :

$$Var[T] = \langle T^2 \rangle - \langle T \rangle^2 = \left(\frac{1}{r}\right)^2. \quad (31)$$



### 2.3 The minimum of $n$ identical, independent distribution

Why is the exponential distribution so prevalent in nature? To answer this question, let us consider the following problem:  $T_1$  and  $T_2$  are two independent distributions for two random times  $T_1$  and  $T_2$ . We are interested in the

$$T^* = \min \{T_1, T_2\}. \quad (32)$$

And we have

$$\Pr\{T^* > t\} = \Pr\{T_1 > t, T_2 > t\} = \Pr\{T_1 > t\} \Pr\{T_2 > t\}. \quad (33)$$

This is because the *multiplication rule* of two independent random events: The joint probability is the product of the probabilities. Therefore, if one has  $n$  identical and independently distributed random times  $T_1, T_2, \dots, T_n$ , then their minimum  $T^*$  has a distribution

$$\Pr\{T^* > t\} = \Pr\{T_1 > t\} \cdots \Pr\{T_n > t\} = \left(\varphi_T(t)\right)^n, \quad (34)$$

in which  $\varphi_T(t) = \Pr\{T > t\}$  is a monotonically decreasing function with  $\varphi_T(0) = 1$  and  $\varphi_T(\infty) = 0$ . Therefore, if  $\varphi'_T(0) = r \neq 0$  and  $n$  is very large, we have

$$\lim_{n \rightarrow \infty} \left[ \varphi_T \left( \frac{t}{n} \right) \right]^n = \lim_{n \rightarrow \infty} \left[ 1 + \varphi'_T(0) \left( \frac{t}{n} \right) \right]^n = e^{-rt}. \quad (35)$$

Why is there a  $1/n$  on the left-hand-side of Eq. (35)? This is because with larger and larger  $n$ , the mean time for  $T^*$  is getting smaller and smaller. In fact, it scales as  $1/n$ . If we had not introduced the  $1/n$ , the limit of  $(\varphi_T(t))^n$  would be 0 for all  $t > 0$ .

***n exponential iid.*** In statistics, “iid” stands for “identical and independently distributed”. If we consider  $n$  identical, independent atoms, each with an exponential waiting time  $e^{-rt}$ , then the time for the first decay,  $T^* = \min\{T_i, 1 \leq i \leq n\}$  follows the distribution

$$\Pr\{T^* > t\} = \Pr\{T_1 > t, \dots, T_n > t\} = \left( \Pr\{T > t\} \right)^n = e^{-nrt}. \quad (36)$$

Note we have used the fact that all  $T_i$  are independent. Therefore, the rate for one decay from  $n$  atoms is  $nr$ .

***Exponential time is memoryless.*** Two measurements of  $T$ , one starts at  $t = 0$ , another starts at  $t = t_0$ , will give identical result:

$$\frac{\Pr\{T > t_0 + t\}}{\Pr\{T > t_0\}} = \frac{e^{-r(t_0+t)}}{e^{-rt_0}} = e^{-rt}. \quad (37)$$

## 2.4 Dynamics of a decreasing population

We can now re-interpret the equation in (24):

$$dp(t) = -rp(t)dt. \quad (38)$$

In an infinitesimal time interval  $(t, t + dt]$ , the change in the survival probability of a single atom is  $rp(t)dt$ .

Now consider a population of identical, independently distributed (iid) atoms. Let  $p_n(t)$  be the probability of having  $n$  radioactive atoms. There are two events that change the  $p_n(t)$ :

(a) A decay of one of  $n + 1$  radioactive atoms. This increases  $p_n(t)$  while decreases  $p_{n+1}(t)$ ; the rate is  $(n + 1)r$ .

(b) A decay of one of  $n$  radioactive atoms. This decreases  $p_n(t)$  while increases  $p_{n-1}(t)$ . The rate is  $nr$ .

Therefore, considering each event can occur in the infinitesimal time interval  $(t, t + dt]$ , we have

$$dp_n(t) = (n + 1)rp_{n+1}(t)dt - nrp_n(t)dt. \quad (39)$$

We now consider a population with  $N$  total individuals at  $t = 0$ . The individuals are identical and independent, with individual “death rate”, i.e., death rate per capita,  $r$ .

To characterize the dynamics of population,  $X(t)$ ,  $X$  takes values  $0, 1, 2, \dots, N$ , one no longer can say that at time  $t$ , the  $X(t)$  is such and such. However, one can predict at time  $t$ , the probability of  $X(t) = n$ :

$$p_n(t) = \Pr\{X(t) = n\}. \quad (40)$$

The  $p_n(t)$  satisfies the system of differential equations

$$\frac{d}{dt}p_n(t) = r(n + 1)p_{n+1}(t) - nrp_n(t). \quad (41)$$

## 2.5 Mean value of the population dynamics

If the population  $X_n(t)$  is random with distribution  $p_n(t)$ , then its mean value is

$$\langle X(t) \rangle = \sum_{i=0}^N i \Pr\{X(t) = i\} = \sum_{i=0}^N ip_i(t). \quad (42)$$

Then we have

$$\begin{aligned}\frac{d}{dt}\langle X(t) \rangle &= \sum_{n=0}^{\infty} n \frac{dp_n(t)}{dt} \\ &= \sum_{n=0}^{\infty} n (r(n+1)p_{n+1}(t) - rnp_n(t)) \\ &= r \sum_{n=0}^{\infty} n(n+1)p_{n+1}(t) - r \sum_{n=0}^{\infty} n^2 p_n(t) \\ &= r \sum_{n=0}^{\infty} (n+1)^2 p_{n+1}(t) - r \sum_{n=0}^{\infty} (n+1)p_{n+1}(t) - r \sum_{n=0}^{\infty} n^2 p_n(t) \\ &= -r \sum_{n=0}^{\infty} (n+1)p_{n+1}(t) \\ &= -r \langle X(t) \rangle.\end{aligned}$$

This is the true meaning of equation (24).