

Bio-Mathematics- a Special Reference To Matrices

Prof. Mandira Kar
Department of Mathematics
St. Aloysius College, Jabalpur (M.P.)

Introduction

Mathematics is one of the oldest organized disciplines of human knowledge with a continuous line of development spanning 5000 years or more. It originated from human curiosity and is an endless enterprise.

Mathematics is known as the “Queen of all Sciences’ as it provides solution techniques & increases the potentials of other disciplines like Physics, Chemistry, Biology, etc. In this era of scientific, industrial and I.T revolution, due attention to mathematics is essential for the progress of the world.

Mathematics permeates biology. However mathematics in biology is appreciated only when biologists start reading & doing research. Even today there are only a few mathematicians who are knowledgeable in biology & very few biologists know mathematics.

Mathematical biology is emerging very rapidly for today traditional academic boundaries require interdisciplinary approaches. Biological concepts and models are becoming more quantitative. For a progressive and fruitful research career in biology one must have requisite knowledge of biology, mathematics, and computer science. A conscious effort to learn the necessary mathematics via biological applications is the need of the hour.

Mathematical biology is an interdisciplinary scientific research field with a range of mathematical applications in biology, biotechnology, medicine etc.; Matrices, Linear algebra, Abstract algebra, Calculus, Differential equations, Graph theory, Statistics, Probability, Operations Research are some areas of Mathematics most commonly applied in Biology.

Contemporary biology generates a huge mountain of data & drawing biologically meaningful inferences from these data requires analysis through good mathematical models. Mathematical models are an equation or a set of equations that quantifies a certain phenomena. Presently with the use of various computer s/w the analysis and inference is accelerated.

Research Problem → Model → Analysis & Result → Model
(After data comparison & revision if needed)

Apart from mathematical biology or more commonly known as Biomathematics, Bioinformatics is also the call of the day. Bioinformatics is an interdisciplinary field that draws on knowledge from biology, biochemistry, statistics, mathematics, and computer science. It is mainly the use of computer algorithms to draw inferences from massive sets of biological data.

Some areas of application of math/stat in bioscience are

- ✓ Genomics and proteomics
- ✓ Description of intra- and inter-cellular processes

- ✓ Growth and morphology
- ✓ Exponential decay & growth
- ✓ Epidemiology and population dynamics
- ✓ Neuroscience
- ✓ Mathematical Ecology
- ✓ Environmental/Water Pollution
- ✓ Climate Disturbances
- ✓ Mathematical Physiology (blood/medicine diffusion etc)
- ✓ Flight of birds
- ✓ Swimming of fish in water
- ✓ Population genetics
- ✓ Prey & Predators

need for matrices

Learning matrix algebra is a major step towards biological modeling. Most biological models involve several variables: frequencies of multiple types, numbers of individuals of different ages, states or species and there are countless examples where a number of variables change simultaneously. Multivariate models are much richer and more interesting than models of a single variable, and to analyze such biological models with several variables learning matrix algebra is indispensable.

Matrix Theory

To start analyzing multivariate models, one need to master matrix algebra; and matrix algebra needs some time to learn. I shall make every effort to teach the necessary mathematics for biological applications wherever it is possible.

The main aspects of matrices that one uses in biological modeling are


- ✓ Definition of Matrices
- ✓ Row/column matrices known as vectors
- ✓ Transpose of matrices especially vectors
- ✓ Matrix addition $[A + B]_{ij} = a_{ij} + b_{ij}$
- ✓ Condition for Matrix multiplication

$$\left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] = \left[\begin{array}{c} \circ \\ \circ \\ \circ \end{array} \right]$$

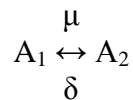
- ✓ Matrix multiplication
- ✓ multiplication of a matrix with a number $a[B]$
- ✓ Non commutativity of matrix multiplication i.e.; $AB \neq BA$
- ✓ Associative property of matrices $(AB)C = A(BC)$
- ✓ Distributive law of matrices $A(B + C) = AB + AC$
- ✓ Systems of linear equations and solutions (unique/infinite)
- ✓ Inconsistent linear equations
- ✓ Graphical representation of a system of two linear equations.

As an introduction, we start by constructing a few matrix models to illustrate why matrices are indispensable when studying models with several variables starting with a

discrete-time model which is simple enough to be analyzed without matrices, and then generalize the same problem to more variables which makes it necessary to use matrices.

 Change from generation to generation under the force of mutation

Consider a population of two types (allele) A_1 and A_2
 In every generation, each copy of type A_1 mutates into A_2 with probability μ , & A_2 mutates into A_1 with probability δ i.e.;



Assume that population size to be constant, in other words the total population is N in every generation. Assume that N is large enough so that stochastic variation in the next generation or more commonly referred to as genetic drift can be neglected

Denote the frequency of type A_1 in generation t with p_t , such that the population in generation t has $p_t N$ population of A_1 and $(1 - p_t)N$ population of A_2

Given p_t to calculate p_{t+1}

We know that type A_1 in the next generation will be there for two reasons; either it was originally (in generation t) of type A_1 and did not mutate (i.e., remained A_1), or it was originally A_2 and mutated to become A_1 .

The former route produces $p_t N(1 - \mu)$ copies of A_1 , where $p_t N$ is the number of original A_1 and a fraction $1 - \mu$ of those who did not mutate.


The latter route gives $\{(1 - p_t)N\} \delta$ copies of A_1 because of the original $(1 - p_t)N$ of type A_2 , a fraction δ mutates and becomes A_1 .

Hence the number of type A_1 in generation $t+1$ can be written as

$$p_{t+1}N = p_t N (1 - \mu) + \{(1 - p_t)N\} \delta;$$

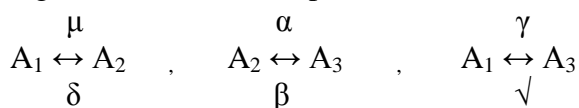
and after dividing with the constant N , we obtain

$$p_{t+1} = p_t (1 - \mu) + (1 - p_t)\delta \quad (1)$$

 Use of Matrices

Now Consider a population of n types A_1, A_2, \dots, A_n . We shall need new notation for the mutation probabilities. As the number of types increases, so does the number of possible mutation probabilities to $n^2 - n$

(eg; for $n = 3$ mutation probabilities are $9 - 3 = 6$, for $n = 4$ it is 12 & so on)



It is not practical to assign a separate letter to each of them. Instead, let q_{ij} denote the probability that A_j mutates into A_i . Notice the order of indices: the second index j shows the original type and the first index i is the new type

q_{ii} (where the two indices are the same) is the probability that that there was no mutation.

We can now write down the frequencies of all n types in the next generation from equation

(1). Denoting the next generation with a prime (say) the equations for ($n = 3$) are

$$\begin{aligned} p'_1 &= q_{11}p_1 + q_{12}p_2 + q_{13}p_3 \\ p'_2 &= q_{21}p_1 + q_{22}p_2 + q_{23}p_3 \\ p'_3 &= q_{31}p_1 + q_{32}p_2 + q_{33}p_3 \end{aligned}$$

Here is the need to introduce matrices

$$P = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \text{ known as Vector or Column Matrix}$$

$$Q = \begin{pmatrix} q_{11} & \cdots & q_{13} \\ \vdots & \ddots & \vdots \\ q_{31} & \cdots & q_{33} \end{pmatrix} \text{ is a } 3 \times 3 \text{ Matrix}$$

$$P' = \begin{pmatrix} p'_1 \\ p'_2 \\ p'_3 \end{pmatrix}$$

$$\begin{pmatrix} q_{11} & \cdots & q_{13} \\ \vdots & \ddots & \vdots \\ q_{31} & \cdots & q_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} p'_1 \\ p'_2 \\ p'_3 \end{pmatrix}$$

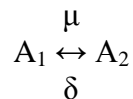
$$= \begin{matrix} q_{11}p_1 + q_{12}p_2 + q_{13}p_3 \\ q_{21}p_1 + q_{22}p_2 + q_{23}p_3 \\ q_{31}p_1 + q_{32}p_2 + q_{33}p_3 \end{matrix}$$

Hence, $P' = QP$ or in general $p_{t+1} = Qp_t$

Conformational states of molecules

A cell membrane has a given number N of trans-membrane channels, which may be either open or closed.

Open channels (denoted by A_1) close at rate μ , whereas closed channels (A_2) open at rate δ . The reaction scheme can be understood as



is similar to the previous example except that this molecular dynamics plays out in continuous time (as opposed to the discrete generations we assumed in the mutation model). Accordingly, μ and δ are rates and not probabilities

Denote the number of open channels by n_1 and the number of closed channels by n_2 .

The dynamics of open and closed channels is given by

$$\frac{dn_1}{dt} = -\mu n_1 + \delta n_2 \quad (2)$$

$$\frac{dn_2}{dt} = \mu n_1 - \delta n_2 \quad (3)$$

The sum of equations (2) & (3) is zero because the total number of membrane channels, $n_1 + n_2 = N$ and it does not change with time. i.e.; $n_2 = N - n_1$.

Using (1) we get Thus, we get $\frac{dn_1}{dt} = -\mu n_1 + \delta(N - n_1)$

If we like, we can also divide this equation with N & put $p = n_1/N$

$$\frac{dp}{dt} = -\mu p + \delta(1 - p) \quad (4)$$

Comparing with the mutation model, the above equation gives the change of p whereas equation (1) gives its value next year. (In continuous time, there is no "next" time step)

Expressing the change of p from equation (1), we obtain $p_{t+1} - p_t = -\mu p_t + \delta(1 - p_t)$, which is directly analogous to equation (4)

Generalization of the continuous-time model to n

Now how do we generalize this continuous-time model to n conformational states of a biomolecule?

Let us take $n = 3$ for illustration, and, as before, let the vector

$P = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ contain the fractions of molecules in the first, second, and third conformational

state. Furthermore, let k_{ij} denote the transition rate at which a molecule in state j switches to state i (*notice again the "backward" notation, the second index marks the initial state*).

The equations describing the dynamics are

$$\frac{dp_1}{dt} = (k_{21} + k_{31})p_1 + k_{12}p_2 + k_{13}p_3$$

$$\frac{dp_2}{dt} = k_{21}p_1 - (k_{12} + k_{32})p_2 + k_{23}p_3$$

$\frac{dp_3}{dt} = k_{31}p_1 + k_{32}p_2 - (k_{13} + k_{23})p_3$ where the sum $(k_{21} + k_{31})$ in the first equation is the rate at which molecules quit state 1 by going either to state 2 or to state 3; the other equations have similar terms.

These equations can be written as

$\frac{dp}{dt} = KP$ where the derivative on the left hand side is the vector of the derivatives,

$$\frac{dp}{dt} = \begin{pmatrix} \frac{dp_1}{dt} \\ \frac{dp_2}{dt} \\ \frac{dp_3}{dt} \end{pmatrix} \text{ and the matrix of the transition rates is}$$

$$K = \begin{pmatrix} (k_{21} + k_{31}) & k_{12} & k_{13} \\ k_{21} & (k_{12} + k_{32}) & k_{23} \\ k_{31} & k_{32} & (k_{13} + k_{23}) \end{pmatrix}$$

Growth of a structured population

It is one of the simplest model of population growth

$$\frac{dN}{dt} = (b - d)N \quad (5)$$

where N is the number of individuals and b and d are respectively the birth and death rates

This equation predicts exponential population growth when the growth rate $(b - d)$ is positive, and exponential decline towards extinction if the growth rate is negative.

Equation (5) assumes that all individuals are equivalent, i.e., they all have the same birth and death rates; such an unstructured population can be described by a single variable N , the number of equivalent individuals. Suppose now that the population contains juveniles and

adults, and let respectively N_1 and N_2 denote their numbers. Adults produce juveniles at rate b and die at rate d . Juveniles mature into adults at rate m and die at rate δ

The dynamics of this structured population is thus given by

$$\frac{dN_1}{dt} = bN_2 - mN_1 - \delta N_1$$

$$\frac{dN_2}{dt} = mN_1 - dN_2$$

- ✓ bN_2 is the only positive term in dN_1/dt is, because the only way to get new juveniles added to N_1 is the reproduction of adults.
- ✓ The term mN_1 is negative in the first equation but a positive term in the second equation juveniles cease to be juveniles and are removed from N_1 ; they however do not disappear from the population but become adults and are added to N_2 .
- ✓ Death affects each stage separately.

We can define a mathematical model for this with the help of matrices as

$$\begin{pmatrix} \frac{dN_1}{dt} \\ \frac{dN_2}{dt} \end{pmatrix} = \begin{pmatrix} -m + \delta & b \\ m & -d \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$$

The question arises for which parameter values will this structured population grow? Some intuition helps us to appreciate the significance of the juvenile stage. If m is very large, then juveniles mature almost instantaneously after birth; in this case, the model is very close to equation (5) and the approximate condition for population growth is again $b - d > 0$. If however m is very small, then the offspring spend an enormously long time as juveniles and (assuming $\delta > 0$) almost all of them die before they would mature; hence the population goes extinct (unless b is enormously large to compensate for juvenile mortality).

A metapopulation model

A metapopulation is a collection of local populations connected by dispersal (migration). For a simple example, consider a metapopulation of only two local populations with discrete-time population dynamics.

At the beginning of a year, each member of population i (where i is 1 or 2) produces F_i offspring on average. Adults survive till the next year with probability P_i . Only the offspring disperse (juvenile dispersal) such that an offspring born in population i moves to population j with probability m_{ji} (recall the order of indices); m_{ii} is for no dispersion

With N_1 and N_2 denoting respectively the number of individuals in populations 1 and 2 the model assumptions lead to the equations

$$\begin{aligned} N_1' &= m_{11}F_1N_1 + m_{12}F_2N_2 + P_1N_1 \\ N_2' &= m_{21}F_1N_1 + m_{22}F_2N_2 + P_2N_2 \end{aligned}$$

where prime denotes the next generation. The $m_{11}F_1N_1$ term of the first equation gives the number of offspring produced in population 1 that also stayed in population 1. The term $m_{12}F_2N_2$ is the number of offspring produced in population 2 that moved over to population 1. P_1N_1 is the number of adults of population 1 that are still alive in the next year. Because adults do not move, only local adults contribute to the population size.

The second equation is constructed in a similar manner. The **Mathematical model** can be given as $N' = AN$ where

$$N' = \begin{pmatrix} N1' \\ N2' \end{pmatrix}, A = \begin{pmatrix} (m11F1 + P1) & m12F2 \\ m21F1 & (m22F2 + P) \end{pmatrix} \& N = \begin{pmatrix} N1 \\ N2 \end{pmatrix}$$

Frequencies/Probability of genotypes

Consider a diploid population of individuals, where two variants A_1 & A_2 of a gene are segregating. Assuming that each individual can be genotyped i.e; every individual may be homozygote (A_1A_1), homozygote, (A_2A_2) or heterozygote (A_1A_2).

Let the number of A_1A_1 homozygotes be N_{11} ,
 the number of A_2A_2 homozygotes be N_{22} and
 the number of A_1A_2 heterozygotes be N_{12} .
 Let the sample size be $N = N_{11} + N_{12} + N_{22}$

Let D, H and R denote respectively the frequencies/probabilities (i.e., the number of individuals having a given genotype per the total number) of genotypes A_1A_1 , A_1A_2 and A_2A_2 ,

$$\text{Then, } D = N_{11}/N, H = N_{12}/N, R = N_{22}/N$$

Obviously, we have $D + H + R = (N_{11} + N_{12} + N_{22})/N = N/N = 1$ (total probability) i.e., the frequencies add up to 1 (or 100%)

Each diploid individual has two genes. There are $2N$ genes in N individuals.
 What is the frequency /probability of A_1 ?

For this, 1st count the number of A_1 variants.

Each A_1A_1 homozygote individual harbors two A_1 gene therefore there are total of $2N_{11}$ genes in homozygotes; and each A_1A_2 heterozygote harbors one A_1 allele, which makes a total of N_{12} alleles in heterozygotes. A_2A_2 homozygotes have no A_1 at all.

Hence, the total number of A_1 genes is thus $2N_{11} + N_{12}$. Dividing the number of A_1 with the total number $2N$ we get the frequency of A_1 in the population

$$p = \frac{2N_{11} + N_{12}}{2N} = \frac{N_{11}}{N} + \frac{N_{12}}{2N} = D + \frac{H}{2}$$

Because the frequencies must also add up to 1, the frequency of A_2 is $q = 1 - p$.

Random mating

Given the parent population with genotypic frequencies D, H and R.

We now find the frequencies of genotypes among their off-spring.

We assume that off -springs are formed via random mating. That is each parent has the same chance to reproduce & each of the two genes of the parent has equal chance to get into the off - spring

What is the fraction of off -spring who inherits allele A_1 from both parents?

First choose a mother randomly from the population. With probability D, the mother has genotype A_1A_1 and therefore all her eggs carry A_1 ; with probability H, the mother is heterozygote (A_1A_2) and only half her eggs are A_1 ; and there is no other way

Summing up these two possibilities, the fraction $D + H/2$ of the eggs have gene A_1 .
This is exactly the frequency of A_1 alleles in the parents, $p = D + H/2$.

By the same logic, we can say that a fraction p of these eggs received an A_1 sperm (from father).

Hence p^2 of all off -spring inherits gene A_1 from both parents.

We can thus say that the probability of A_1A_1 homozygotes among the off -spring is given by **$D' = p^2$ where D' denotes the frequency of A_1A_1 homozygotes in the next generation** whereas, **p is the frequency of gene A_1 in the initial generation**

Similarly the frequency of A_2A_2 homozygote off -spring is given by **$R' = q^2$**

Next, the frequency of heterozygote off -springs.

In this case the egg is A_1 and the sperm is A_2 , or vice-versa

The probability of the first is $p \times q$, & the second probability is $q \times p$, which

Summing up the frequency of heterozygote off- spring is given by **$H' = pq + qp = 2pq$** and
 $D' + R' + H' = p^2 + q^2 + 2pq = (p+q)^2 = 1$

The frequency does not change from generation to generation. Indeed, in the off spring generation the frequency of gene A_1 is given by $p' = D' + \frac{H'}{2} = p$ which is the same as the frequency of allele A_1 in the initial population.

Bibliography

1. Mathematical Methods in Biology; Eva Kisdi
2. Mathematical Methods in Biology- Vectors and matrices; Eva Kisdi
3. Mathematical Methods in Biology; J. David Logan, William Wolessky
4. Mathematical Modelling- Teaching & Assessment in a technology rich world; P. Galbraith, W Blum, G Booker & Ian Huntley

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:

<http://www.iiste.org>

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: <http://www.iiste.org/journals/> All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: <http://www.iiste.org/book/>

Academic conference: <http://www.iiste.org/conference/upcoming-conferences-call-for-paper/>

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar

