CHAPTER 4

# MODEL BUILDING

[*This chapter is based on the lectures of Dr. A.M. Mathai of CMS Palai, Kerala, India.*]

## 4.0. Introduction

There are various types of models that one can construct for a given set of data. The types of model that is chosen depends upon the type of data for which the model is to be constructed. If the data are coming from a deterministic situation then there may be already an underlying mathematical formula such as a physical law. Perhaps the law may not be known yet. Hence one may try to fit a model to the data collected about the situation. When the physical law is known then there is no need to fit a model but for verification purposes one may substitute data points into the speculated physical law.

## 4.1. Deterministic Situations

For example, a simple physical law for gases says that the pressure $P$ multiplied by volume $V$ is a constant under a constant temperature. Then the physical law that is available is

$$PV = c$$

where $c$ is a constant. When it is a mathematical relationship then all pairs of observations on $P$ and $V$ must satisfy the equation $PV = c$. If $V_1$ is one observation on $V$ and if the corresponding observation is $P_1$ for $P$ then $P_1V_1 = c$ for the same constant $c$. Similarly, other pairs of observations $(P_2, V_2)$, ... will satisfy the equation $PV = c$. If there are observational errors in observing $P$ and $V$ then the equation may not be satisfied exactly by a given observational pair. If the model proposed $PV = c$ is not true exactly then of course the observational pairs $(P_i, V_i)$ for some specific $i$ may not satisfy the equation $PV = c$.

### 4.1.1. Differential equations

Another way of describing a deterministic situation is through the help of differential equations. We may be able to place some plausible assumptions on the rate of change of one variable with respect to another variable. For example, consider the growth of a bacterial colony over time. The rate of change of the population size may be proportional to the actual size. For example if new cells are formed by cell division then the number of new cells will depend upon the number of cells already there. Growth of human population in a community is likely to be proportional to the number of fecundable women in the community. In all these situations the rate of change of the actual size, denoted by $y$, over time $t$, will be a constant multiple of $y$ itself. That is,

$$\frac{\mathrm{d}y}{\mathrm{d}t} = ky \tag{4.1.1}$$

where $k$ is a constant. The solution is naturally

$$y = m\mathrm{e}^{kt}$$

where $m$ is an arbitrary constant. This constant $m$ is available if we have an initial condition such as when $t = 0$, $y_0 = m\mathrm{e}^0 = m$ or $m = y_0$ or the model is

$$y = y_0\mathrm{e}^{kt},\ t \geq 0. \tag{4.1.2}$$

In this model there is only one parameter $k$ (unknown quantity sitting in the model) and this $k$ is the constant of proportionality of the rate of change of the population size $y$ over time $t$.

**Example 4.1.1.** Coconut oil ($u$ = demand in kilograms) and palm oil ($v$ = demand in kilograms) are two competing vegetable cooking oil that are sold in a supermarket. The shopkeeper found that if either coconut oil or palm oil is absent then the demand sores exponentially. the rate of change of demand over time is found to be the following:

$$\begin{aligned}
&\frac{\mathrm{d}u}{\mathrm{d}t} = 2u - v \text{ and } \frac{\mathrm{d}v}{\mathrm{d}t} = -u + v \\
&\quad \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \\
&\quad \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}W = AW,\ W = \begin{pmatrix} u \\ v \end{pmatrix},\ A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}.
\end{aligned}$$

Suppose that the starting stocks at $t = 0$ are $u = 100$ kilos and $v = 200$ kilos, time being measured in days. What will be the demands (1) on the 10th day, (2) eventually.

**Solution 4.1.1.** Note that individually the demands vary exponentially, that is, when $v = 0$, $\frac{\mathrm{d}u}{\mathrm{d}t} = 2u \Rightarrow u = \mathrm{e}^{2t}$ and when $u = 0$, $\frac{\mathrm{d}v}{\mathrm{d}t} = v \Rightarrow v = \mathrm{e}^{t}$. Hence we may look for a general solution of the type

$$u = \mathrm{e}^{\lambda t}x_1 \text{ and } v = \mathrm{e}^{\lambda t}x_2$$

where $\lambda, x_1, x_2$ are unknown quantities. Then

$$W = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{\lambda t}x_1 \\ \mathrm{e}^{\lambda t}x_2 \end{pmatrix} = \mathrm{e}^{\lambda t}X, X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Pre-multiplying both sides by $A$ we have

$$\begin{aligned}
&\frac{\mathrm{d}u}{\mathrm{d}t} = \lambda\mathrm{e}^{\lambda t}x_1,\ \frac{\mathrm{d}v}{\mathrm{d}t} = \lambda\mathrm{e}^{\lambda t}x_2 \\
&\frac{\mathrm{d}}{\mathrm{d}t}W = AW = \mathrm{e}^{\lambda t}AX \Rightarrow \\
&\lambda\mathrm{e}^{\lambda t}x_1 = 2\mathrm{e}^{\lambda t}x_1 - \mathrm{e}^{\lambda t}x_2 \\
&\lambda\mathrm{e}^{\lambda t}x_2 = -\mathrm{e}^{\lambda t}x_1 + \mathrm{e}^{\lambda t}x_2.
\end{aligned}$$

That is,

$$AX = \lambda X,\ A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix},\ X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

This means $\lambda$ is an eigenvalue of $A$.

$$\begin{aligned} |A - \lambda I| = 0 &\Rightarrow \left| \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0 \\ &\Rightarrow \begin{vmatrix} 2-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} = 0 \\ &\Rightarrow \lambda^2 - 3\lambda + 1 = 0 \Rightarrow \lambda_1 = \frac{3+\sqrt{5}}{2},\ \lambda_2 = \frac{3-\sqrt{5}}{2}. \end{aligned}$$

What are the eigenvectors? Corresponding to $\lambda_1 = \frac{3+\sqrt{5}}{2}$ we have the second equation from $AX = \lambda_1 X$

$$-x_1 + \left(1 - \frac{3+\sqrt{5}}{2}\right) x_2 = 0 \Rightarrow x_1 = \frac{1-\sqrt{5}}{2}$$

for $x_2 = 1$ or one solution is

$$W_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix}.$$

For $\lambda_2 = \frac{3-\sqrt{5}}{2}$, substituting in $AX = \lambda_2 X$ then from the first equation

$$\left(2 - \frac{3-\sqrt{5}}{2}\right) x_1 - x_2 = 0 \Rightarrow x_2 = \frac{1+\sqrt{5}}{2}$$

for $x_1 = 1$. Thus an eigenvector is

$$W_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix}.$$

Then the general solution is

$$\begin{aligned} W &= c_1 W_1 + c_2 W_2 \\ &= c_1 e^{\lambda_1 t} \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix} + c_2 e^{\lambda_2 t} \begin{bmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{bmatrix} \end{aligned}$$

where $\lambda_1$ and $\lambda_2$ are given above. From the initial conditions, by putting $t = 0$

$$W_0 = \begin{bmatrix} 100 \\ 200 \end{bmatrix} = c_1 \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{bmatrix} \Rightarrow$$

$$\begin{aligned} 100 &= \frac{(1-\sqrt{5})}{2} c_1 + c_2 \\ 200 &= c_1 + \frac{(1+\sqrt{5})}{2} c_2. \end{aligned}$$

This gives $c_1 = 75 - 25\sqrt{5}$ and $c_2 = 50\sqrt{5}$. Thus the answers are the following: (1): For $t = 10$

$$u = (75 - 25\sqrt{5})e^{(\frac{3+\sqrt{5}}{2})(10)}(\frac{1-\sqrt{5}}{2}) + 50\sqrt{5}e^{(\frac{3-\sqrt{5}}{2})(10)} \text{ and}$$

$$v = (75 - 25\sqrt{5})e^{(\frac{3+\sqrt{5}}{2})(10)} + 50\sqrt{5}e^{(\frac{3-\sqrt{5}}{2})(10)}(\frac{1+\sqrt{5}}{2}).$$

Both these tend to $+\infty$ when $t \to \infty$ and hence the demand explodes eventually.

### 4.1.2. Algebraic models

Another category of models for a deterministic situation is algebraic in nature. Suppose that we want to construct models for mass, pressure, energy generation and luminosity in the Sun then we have to speculate about the nature of matter density distribution in the Sun. The Sun can be taken as a gaseous ball in hydrostatic equilibrium. The matter density can be speculated to be more and more concentrated towards the center or core of the Sun and more and more thin when moving outward from the core to the surface. This density distribution is deterministic in nature. But due to lack of knowledge about the precise mechanism working inside the Sun one has to speculate about the matter density distribution. If $r$ is an arbitrary distance from the center of the Sun then one can start with a simple linear model for the matter density $f(r)$ at a distance $r$ from the center. Then our assumption is that

$$f(r) = a(1 - x),\ a > 0,\ x = \frac{r}{R_o} \tag{4.1.3}$$

where $R_o$ is the radius of the Sun. This model indicates a straight line decrease from the center of the Sun, going outward. From observational data we can see that (4.1.3) is not a good model for the density distribution in the Sun. Another speculative solar model (model for the Sun) is of the form

$$f(r) = f_0(1 - x^{\delta})^{\rho} \tag{4.1.4}$$

where $\delta$ and $\rho$ are two parameters. It can be shown that we have specific choices of $\delta$ and $\rho$ so that the model in (4.1.4) can be taken as a good model for the density distribution in the Sun. Once a model such as the one in (4.1.4) is proposed then simple mathematical formulation from (4.1.4) will give models for mass $M(r)$, pressure $P(r)$ and luminosity $L(r)$, at arbitrary point $r$. Such models are called analytical models. Various such analytical solar and stellar models may be seen from Mathai and Haubold (1988: *Modern Problems in Nuclear and Neutrino Astrophysics*) and later papers by the same authors and others.

### 4.1.3. Computer models

Problems such as energy generation in the Sun or in a star depend on many factors. One such factor is the distribution of matter density in the star. A model based on just one variable such as matter density will not cover all possibilities. Hence what scientists often do is to take into account all possible factors and all possible forms in which the factors enter into the scenario. This may result in dozens of differential equations to solve. Naturally, an analytical solution or exact mathematical formulation will not be possible. Then the next best thing to do is to come up with some numerical solutions with the help of computers. Such models are called computer models. The standard solar model, which is being widely used now, is a computer model. One major drawback of a computer model is that the computer will give out some numbers as numerical solutions but very often it is not feasible to check and see whether the numbers given out by the computer are correct or not.

### 4.1.4. Power function models

There are a number of physical situations where the rate of change of an item is in the form of a power of the original item. In certain decaying situations such as certain radiation activity, energy production, reactions, diffusion processes, etc the rate of change over time is in the form of the power. If $y$ is the original amount and if it is a decaying situation then the law governing the process may be of the form

$$\frac{dy}{dt} = -y^{\alpha}, \text{ (decaying)} \tag{4.1.5}$$

$$\frac{dy}{dt} = y^{\alpha}, \text{ (growing).} \tag{4.1.6}$$

Then a solution of the decaying model in (4.1.5) is of the form

$$y = [1 - a(1-\alpha)t]^{\frac{1}{1-\alpha}} \Rightarrow \frac{dy}{dt} = -ay^{\alpha}, \; a > 0 \tag{4.1.7}$$

and the growth model in (4.1.6) has the solution

$$y = [1 + a(1-\alpha)t]^{-\frac{1}{\alpha-1}} \Rightarrow \frac{dy}{dt} = ay^{\alpha}, \; a > 0. \tag{4.1.8}$$

In (4.1.7) the rate of decay of $y$ is proportional to a power of $y$. For example if $\alpha = \frac{1}{2}$ then the rate of change of $y$ over time is proportional to the square root of $y$. The current hot topics of Tsallis statistics and $q$-exponential functions are associated with the differential equations in (4.1.5) and (4.1.6) with resulting solutions in (4.1.7) and (4.1.8) respectively. The power parameter $\alpha$ may be determined from some physical interpretation of the system or one may estimate from data. The behavior in (4.1.5) and (4.1.6) is also associated with some generalized entropies in Information Theory. Some details may be seen from Tsallis (1988) and Mathai and Haubold (2007, 2007a, 2007b).

### 4.1.5. Input-output models

There are a large number of situations in different disciplines where what is observed is the residual effect of an input activity and an output activity. Storage of water in a dam or storage of grain in a silo at any given time is the residual effect of the total amount of water flowed into the dam and the total amount of water taken out for irrigation, power generation etc. Grain is stored in a silo during harvest seasons and grain is sold out or consumed from time to time and thus what is in the storage silo is the residual effect of the input quantity and output quantity. A chemical called Melatonin is produced in every human being's body. The production starts in the evening, it attains the maximum peak by around 1 am and then it comes down and by morning the level is back to normal. At any time in the night the amount of Melatonin seen in the body is what is produced, say $x$, minus what is consumed or converted or spent by the body, say $y$. Then what is seen in the body is $z = x - y, x \geq y$. It is speculated that by monitoring $z$ one can determine the body age of a human being, because it is seen that the peak value of $z$ grows as the person grows in age and it attains a maximum during what we will call the middle age of the person and the peak value of $z$ starts coming down when the person crosses the middle age. Thus it is of great importance in modeling this $z$ and lot of studies are there on $z$.

Another similar situation is the production of certain particles in the Sun called solar neutrinos. These particles are capable of passing through almost all media and practically inert, not reacting with other particles. There are experiments being conducted in USA, Europe and Japan to capture these neutrinos. There is discrepancy between what is theoretically predicted from physical laws and what is captured through these experiments. This is known as the solar neutrino problem. Modeling the amount of captured neutrinos is a current activity to see whether there is any cyclic pattern, whether there is time variation and so on. What is seen mathematically is an input-output type behavior. There are also several input-output type situations in economics, social sciences, industrial production, commercial activities and so on. An analysis of the general input-output structure is given in Mathai (1993) and some studies of the neutrino problem may be seen from Haubold and Mathai (1994,1995). In input-output modeling, the basic idea is to model

$$z = x - y$$

by imposing assumptions on the behaviors or types of $x$ and $y$ and assumption about whether $x$ and $y$ are independently varying or not.

### 4.1.6. Pathway model

Another idea that was introduced recently, see Mathai (2005), is to come up with a model which can switch around, assume different functional forms, or move from one functional form to another, thus creating a pathway to different functional forms. Such a model will be useful in describing transitional stages in the evolution of any item or describing chaos or chaotic neighborhoods where there is a stable solution. The pathway model is created on the space of rectangular matrices. For the very special case of real scalar positive variable $x$ the model is of the following form:

$$f_1(x) = c_1 x^{\gamma}[1 - a(1-\alpha)x^{\delta}]^{\frac{1}{1-\alpha}}, \; 1 - a(1-\alpha)x^{\delta} > 0, \; a > 0 \tag{4.1.9}$$

where $\gamma, \delta \geq 0, a > 0$ are parameters and $c_1$ can behave as a normalizing constant if the total integral over $f(x)$ is to be made equal to 1 so that with $f(x) \geq 0$ for $x \geq 0$ this $f(x)$ can act as a density for a positive random variable $x$ also. The parameter $\alpha$ is the pathway parameter. Note that for $\alpha < 1$ and $1 - a(1-\alpha)x^{\delta} > 0$ we have (4.1.9) belonging to the generalized type-1 beta family of functions. When $\alpha > 1$

writing $1-\alpha = -(\alpha-1)$ we have

$$f_2(x) = c_1 x^{\gamma}[1 + a(\alpha-1)x^{\delta}]^{-\frac{1}{\alpha-1}}, \ \alpha > 1, \ 0 \le x < \infty. \tag{4.1.10}$$

The function in (4.1.10) belongs to the generalized type-2 beta family of functions. But when $\alpha \to 1$ both the forms in (4.1.9) and (4.1.10) go to the exponential form

$$f_3(x) = c_3 x^{\gamma} e^{-ax^{\delta}}, \ \alpha \to 1, a > 0, \delta > 0, x \ge 0. \tag{4.1.11}$$

This $f_3(x)$ belongs to the generalized gamma family of functions. Thus the basic function $f_1(x)$ in (4.1.9) can move to three different families of functions through the parameter $\alpha$ as $\alpha$ goes from $-\infty$ to 1, then from 1 to $\infty$ and finally the limiting case of $\alpha \to 1$. If the limiting form in (4.1.11) is the stable solution to describe a system then unstable neighborhoods of this stable solution are described by (4.1.9) and (4.1.10). All the transitional stages can be modeled with the help of the parameter $\alpha$. For example, as $\alpha$ moves from the left to 1 this function $f_1(x)$ moves closer and closer and finally to $f_3(x)$. Similarly as $\alpha$ moves to 1 from the right, $f_2(x)$ moves closer and closer and finally to $f_3(x)$.

The same model can create a pathway for certain classes of differential equations, a pathway for classes of statistical densities and a pathway for certain classes of generalized entropies. Thus essentially there are three pathways in the model in (4.1.9), the entropic pathway, the differential pathway and the distributional pathway. More details may be seen from Mathai and Haubold (2007a,b)

### 4.1.7. Fibonacci sequence model

Consider the growth of a certain mechanism in the following fashion. It takes one unit of time to grow and one unit of time to reproduce. Reproduction means producing one more item or splitting into two items. For example consider the growth of rabbit population in the certain locality. Suppose that we count only female rabbits. Suppose that every female rabbit gives birth to one female rabbit at every six month's period. After six months of giving birth the mother rabbit can give birth again, one unit of time. But the daughter rabbit needs six months time to grow and then six months to reproduce, or two units of time. Then at each six months the population size will be of the form $1, 1, 2, 3, 5, 8, 13, ...$ or the sum of the two previous numbers is the next number. This sequence of numbers is called the Fibonacci sequence of numbers. Thus any growth mechanism which takes one unit of time to grow and one unit of time to reproduce can be modeled by a Fibonacci sequence. Some application of these types of models in explaining many types of patterns that are seen in nature may be seen from Mathai and Davis (1974).

**Example 4.1.2.** When a population is growing according to a Finonacci sequence and if $F_k$ denotes the population size at the $k$-th generation, with the initial conditions $F_0 = 0, F_1 = 1$ then the difference equation governing the growth process is $F_{k+2} = F_{k+1} + F_k$. Evaluate $F_k$ for a general $k$.

**Solution 4.1.2.** This can be solved by a simple property of matrices. To this end write one more equation that is an identity $F_{k+1} = F_{k+1}$ and write the system of equations

$$\begin{aligned} F_{k+2} &= F_{k+1} + F_k \\ F_{k+1} &= F_{k+1} \end{aligned} \qquad (a_1)$$

Take

$$V_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \qquad (a_2)$$

Then the equation in $(a_1)$ can be written as

$$V_{k+1} = AV_k, \ A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}. \qquad (a_3)$$

But from the initial condition

$$V_0 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

From $(a_3)$

$$F_1 = AV_0 \Rightarrow F_2 = AV_1 = A^2V_0 \Rightarrow F_k = A^kV_0. \qquad (a_4)$$

Therefore we need to compute only $A^k$ because $V_0$ is already known. For example if $k = 100$ then we need to multiply $A$ by itself 100 times, which is not an easy process. Hence we will make use of some properties of eigenvalues. Let us compute the eigenvalues of $A$.

$$\begin{aligned} |A - \lambda I| = 0 &\Rightarrow \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = 0 \\ &\Rightarrow \lambda^2 - \lambda - 1 = 0 \\ &\Rightarrow \lambda_1 = \frac{1+\sqrt{5}}{2} \text{ and } \lambda_2 = \frac{1-\sqrt{5}}{2}. \end{aligned}$$

Therefore the eigenvalues of $A^k$ are $\lambda_1^k$ and $\lambda_2^k$. Let us compute the eigenvectors. Let $X = \binom{x_1}{x_2}$ be the eigenvector corresponding to $\lambda_1$. Then

$$\begin{aligned} (A - \lambda_1 I)X = O &\Rightarrow \begin{bmatrix} 1 - \frac{(1+\sqrt{5})}{2} & 1 \\ 1 & -\frac{(1+\sqrt{5})}{2} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\Rightarrow X_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix}. \end{aligned}$$

Similarly the eigenvector corresponding to $\lambda_2$ is

$$X_2 = \begin{pmatrix} \lambda_2 \\ 1 \end{pmatrix}.$$

Then the matrix of eigenvectors is given by

$$Q = (X_1, X_2) = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \Rightarrow Q^{-1} = \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix}.$$

Then

$$A = QDQ^{-1} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \left[ \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{pmatrix} \right].$$

Since $\lambda$ and $\lambda^k$ share the same eigenvectors,

$$A^k = \frac{1}{\lambda_1 - \lambda_2}\begin{bmatrix}\lambda_1 & \lambda_2\\ 1 & 1\end{bmatrix}\begin{bmatrix}\lambda_1^k & 0\\ 0 & \lambda_2^k\end{bmatrix}\begin{bmatrix}1 & -\lambda_2\\ -1 & \lambda_1\end{bmatrix}$$
$$= \frac{1}{\lambda_1 - \lambda_2}\begin{bmatrix}\lambda_1^{k+1} - \lambda_2^{k+1} & -\lambda_2\lambda_1^{k+1} + \lambda_1\lambda_2^{k+1}\\ \lambda_1^k - \lambda_2^k & -\lambda_2\lambda_1^k\lambda_1\lambda_2^k\end{bmatrix}. \qquad (a_5)$$

But

$$V_k = A^k V_0 = A^k\begin{pmatrix}1\\0\end{pmatrix} \Rightarrow$$
$$V_k = \frac{1}{\lambda_1 - \lambda_2}\begin{bmatrix}\lambda_1^{k+1} - \lambda_1^{k+1}\\ \lambda_1^k - \lambda_2^k\end{bmatrix}.$$

That is,

$$F_k = \frac{1}{\lambda_1 - \lambda_2}(\lambda_1^k - \lambda_2^k)$$
$$= \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k\right\}. \qquad (a_6)$$

Then $F_k$ can be computed for any $k$ by using $(a_6)$. We can make some interesting observations. Since $\left|\frac{1-\sqrt{5}}{2}\right| < 1$, $\lambda_k \to 0$ as $k \to \infty$. Hence for large values of $k$,

$$F_k \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k \Rightarrow \lim_{k\to\infty}\frac{F_{k+1}}{F_k} = \frac{1+\sqrt{5}}{2}. \qquad (4.1.12)$$

This $\frac{1+\sqrt{5}}{2}$ is known as the golden ratio which appears at many places in nature

# Exercises

**4.1.1.** Do Example 4.1 if the matrix $A$ is the following:

$$A = \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}.$$

**4.1.2.** A biologist has found that the owl population $u$ and the mice population $v$ in a particular region are governed by the following system of differential equations, where $t$ = time being measured in months.

$$\frac{du}{dt} = 2u + 2v$$
$$\frac{dv}{dt} = u + 2v.$$

Initially there are two owls and 100 mice. Solve the system. What will be the populations of owl and mice after 5 months? What will be populations eventually?

**4.1.3.** Suppose that 2 falcons are also in the region in Exercise 4.1.2. Thus the initial populations of falcon $f$, owl $u$ and mince $v$ are $f_0 = 2, u_0 = 2, v_0 = 100$.

Suppose that the following are the differential equations:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = f + u + v$$
$$\frac{\mathrm{d}u}{\mathrm{d}t} = -f + u + 2v$$
$$\frac{\mathrm{d}v}{\mathrm{d}t} = -f - u + 3v.$$

Solve for $f, u, v$ and answer the same questions in Exercise 4.1.2, including falcons

**4.1.4.** For the model of density in equation (4.1.3) evaluate the mass of the Sun at an arbitrary value $r$ if the law is

$$\frac{\mathrm{d}M(r)}{\mathrm{d}r} = 4\pi r^2 f(r)$$

where $f(r)$ is the density at a distance $r$ from the center.

**4.1.5.** If pressure in the Sun is given by the formula

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{G\ M(r) f(r)}{r^2}$$

where $G$ is a constant and $f(r)$ is the density, evaluate the pressure at arbitrary distance $r$ from the center.

**4.1.6.** For the pathway model in (4.1.9) evaluate $c_1$ if $f_1(x)$ is a density for $\alpha < 1$ and $1 - a(1-\alpha)x^{\delta} > 0, a > 0, \delta > 0$.

**4.1.7.** For the pathway model in (4.1.10) evaluate $c_2$ if $f_2(x)$ is a density for $x \geq 0, a > 0, \delta > 0, \alpha > 1$.

**4.1.8.** For the pathway model in (4.1.11) evaluate $c_3$.

**4.1.9.** For the models in (4.1.9), (4.1.10) and (4.1.11) compute the Mellin transform of $f_i(x)$ or $\int_x x^{s-1} f_i(x)\mathrm{d}x, i = 1, 1, 3$ and state the conditions for their existence.

**4.1.10.** Draw the graphs of the models in (4.1.9),(4.1.10) and (4.1.11) for various values of the pathway parameter $\alpha$ and for some fixed values of the other parameters and check the movement of the function.

**4.1.11.** Suppose that a system is growing in the following fashion. The first stage size plus 2 times the second stage plus 2 times the third stage equals the fourth stage. That is,

$$F_k + 2F_{k+1} + 2F_{k+2} = F_{k+3}.$$

Suppose that the initial conditions are $F_0 = 0, F_1 = 1, F_2 = 1$. Compute $F_k$ for $k = 100$.

**4.1.12.** Suppose that a system is growing in the following fashion: The first stage size plus 3 times the second stage size plus the third stage size equals the fourth stage or

$$F_k + 3F_{k+1} + F_{k+} = F_{k+3}.$$

Compute $F_k$ for $k = 50$ if $F_0 = 0, F_1 = 1, F_2 = 1$.

**4.1.13.** In a savannah there are some lions and some deers. Lions eat out deers but deers do not eat lions. There are small games like rabbits in the area. If no deer is left the lions cannot survive for long hunting small games. If no lion is there then deer population will explode. It is seen that the lion population at the $i$-th stage, $L_i$, and the deer population at the $i$-th stage, $D_i$, are related by the system

$$\begin{aligned} L_{i+1} &= 0.1L_i + 0.4D_i \\ D_{i+1} &= 1.2D_i - pL_i \end{aligned}$$

where $p$ is some number. The initial population at $i = 0$ are $L_0 = 10, D_0 = 100$. Compute the population of lion and deer for $i = 1, 2, 3, 4$. For what value of $p$, ($i$) the deer population will be extinct eventually, ($ii$) the lion population will be extinct eventually, ($iii$) lion and deer population will explode. When a number falls below 1 it should be counted as extinction.

**4.1.14.** Show that every third Fibonacci number is odd, starting with $1, 1$.

**4.1.15.** Consider the sequence $F_k + F_{k+1} = F_{k+2}$ with $F_0 = 0, F_1 = a > 0$. Show that

$$\lim_{k\to\infty} \frac{F_{k+1}}{F_k} = \frac{1+\sqrt{5}}{2} = \text{ golden ratio.}$$

# 4.2. Some Non-deterministic Situations

So far we were dealing with deterministic situations where there was no presence of random variables or chance variation. Since a lot of practical situations are random or non-deterministic in nature, when we talk about model building, people naturally think that we are trying to describe a random or non-deterministic situation by mathematical modeling. Attempts to describe random situations have given birth to different branches of science. Stochastic process is one such area where we study a collection of random variables. Time series analysis is an area where we study a collection of random variables over time. Regression is an area where we try to describe random situations by analyzing conditional expectations. As can be seen, even to give a basic description of all the areas and disciplines where we build models it will take hundreds of pages. Hence what we will do here is to pick a few selected topics and give basic introduction to these topics.

## 4.2.1. Random walk model

Consider a simple practical situation of a drunkard coming out of the liquor shop, denoted by $S$ in Figure 4.1.

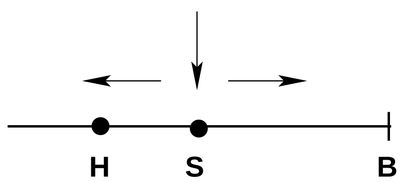


Figure 4. 1

He tries to walk home. Since he is fully drunk assume that he walks in the following manner. At every minute he will either take a step to the right or to the left. Let us assume a straight line path.

Suppose that he covers 1 foot (about a third of a meter) at each step. He came out from the shop as indicated by the arrow. Then if his first step is to the right then he is one foot closer to home, whereas if his first step is to the left then he is farther way from home by one foot. At the next minute he takes the next step either to his right or to his left. If he had taken the second step also to the left then now he is farther away from home by two feet. One can associate a chance or probability for taking a step to the left or right at each stage. If the probabilities are $\frac{1}{2}$ each then at each step there is 50% chance that the step will be to the left and 50% chance that the step will be to the right. If the probabilities of going to the left or right are $0.6$ and $0.4$ respectively then there is a 60% chance that his first step will be to the left.

Some interesting questions to ask in this case are the following: What is the chance that eventually he will reach home? What is the chance that eventually he will get lost and walk away from home to infinity? Where is his position after $n$ steps? There can be several types of modifications to the simple random walk on a line. In Figure 4.1 a point is marked as $B$. This $B$ may be a barrier. This barrier may be such that once he hits the barrier he falls down dead or the walk is finished, or the barrier may be such that if he hits the barrier there is a certain positive chance of bouncing back to the previous position so that the random walk can continue and there may be a certain chance that the walk is finished, and so on.

An example of a 2-dimensional random walk is the case of Mexican jumping beans. There is a certain variety of Mexican beans (lentils). If you place a dried bean on the table top then after a few seconds it jumps by itself in a random direction to another point on the table. This is due to an insect making the dry bean as its home and the insect moves around by jumping. This is an example of a two-dimensional random walk. We an also consider random walk in space and random walk in higher dimensions.

## 4.2.2. Branching process model

In nature there are many species which behave in the following manner. There is a mother and the mother gives birth to $n$ offsprings once in her life-time. After giving birth the mother dies. The number of offsprings could be $0, 1, 2, ..., k$ where $k$ is a fixed number, not infinitely large. A typical example is the banana plant. One banana plant gives only one bunch of bananas. You cut the bunch and the mother plant is destroyed. the next generation offsprings are the new shoots coming from the bottom. The number of shoots could be $0, 1, 2, 3, 4, 5$, usually a maximum of 5 shoots. These shoots are the next generation plants. Each shoot,when mature, can produce one bunch of bananas each. Usually after cutting the mother banana plant the farmer will separate the shoots and plant all shoots, except one, elsewhere so that all have good chances of growing up into healthy banana plants.

Another example is pineapple. Again one pineapple plant gives only one pineapple. The pineapple itself will have one shoot of plant at the top of the fruit and other shoots will be coming from the bottom. Again the mother plant is destroyed when the pineapple is plucked. Another example is certain species of spiders. The mother carries the sack of eggs around and dies after the new offsprings are born. Another example is salmon fish. From the wide ocean the mother salmon enters into fresh water river, goes to the birthplace of the river, overcoming all types of obstacles on the way, and lays one bunch of eggs and then promptly dies. Young salmons come out of these eggs and they find their way down river to the ocean. The life cycle is continued.

These examples are examples for branching processes. Interesting questions to ask are the following: what is the chance that the species will be extinct eventually? This can happen if there is a positive probability of having no offspring in a given birth. What is the expected population size after $n$ generations? Branching process is a subject matter by itself and it is a special case of a general process known as birth and death process.

### 4.2.3. Birth and death process model

This can be explained with a simple example. Suppose that there is a good pool area in a river, a good fishing spot for fishermen. Fish move in and move out of the pool area. If $N(t)$ is the number of fish in the pool area at time $t$ and if one fish moved out at the next unit of time then $N(t+1) = N(t) - 1$. On the other hand if one fish moved into the pool area at the next time unit then $N(t+1) = N(t) + 1$. When one addition is there then one can say that there is one birth and when one deletion is there we can say that there is one death. Thus if we are modeling such a process where there is possibility of birth and death then we call it a birth and death model.

### 4.2.4. Time series models

Suppose that we are monitoring the flood level at Meenachil River at the Pastoral Institute. If $x(t)$ denotes the flood level on the $t$-th day, time $= t$ being measured in days, then at the zeroth day or starting of the observation period the flood level is $x(0)$, the next day the flood level is $x(1)$ and so on. We are observing a phenomenon namely flood level over time. In this example, time is measured in discrete time units. Details of various types of processes and details of the various techniques available for time series modeling of data are available in SERC School Notes of 2005, 2006, 2007. Since these are available to the students the material will not be elaborated here.

### 4.2.5. Regression type models

When people talk abut modeling they are usually talking about regression type models. The idea stems from the following considerations. Suppose that we want to come up with the ‘best-fitting’ curve to a given data, ‘best-fitting’ in some sense. As an example let $y$ denote the increase in milk yield by experimental cows under a certain diet $x$. The experimenter has some observations on $(x, y)$ such as $(0, 2), (1, 3), (2, 5), (3, 7)$. This means that when no diet is administered ($x = 0$), the milk yield is 2 units, when one unit of the diet is administered ($x = 1$) the yield is 3 units and so on. Taking $y$ as a function of $x$, say $y = f(x)$, what is the ‘best-fitting’

$f(x)$ for the data, 'best-fitting' in some sense. Any arbitrary function $f(x)$ can be taken as a model for $y$ then the value predicted by $f(x)$ at $x = x_i$ is $f(x_i)$ and this $f(x_i)$ may be far away from the observed $y_i$. Then the error is $e_i = y_i - f(x_i) =$ observed - predicted by the model. Then

$$e_i^2 = [y_i - f(x_i)]^2, \; i = 1, ..., n$$

if there are $n$ observations. If the variable $y$ is predicted by $f(x)$ then the error is

$$e = y - f(x) \Rightarrow e^2 = [y - f(x)]^2.$$

Our aim is to minimize the distance between $y$ and $f(x)$ so that the predicted value and the true value will be closer to each other and thus select the particular function $f$ which will minimize the distance between $y$ and $f(x)$. One such measure of squared distance is $E[y - f(x)]^2$ where $E$ denotes the expected value when $y$ is a random variable. We may want to use the model $f$ to predict $y$ at a given value of $x$. Then $y$ is the only variable and $x$ is a preassigned number, such as what is the predicted $y$ when $x = 2.8$ units of the diet? Our problem reduces to minimizing $E[y - f(x)]^2$ observing that minimizing the distance $= \sqrt{E(y - f(x))^2}$ is equivalent to minimizing $E(y - f(x))^2$.

$$\min_f E[y - f(x)]^2 \Rightarrow f = ? \tag{4.2.1}$$

This is already available from elementary algebra or calculus. We know that for an arbitrary $a$, $E[y - a]^2$ is a minimum when $a = E(y)$.

$$\begin{aligned} E[y - a]^2 &= E[y - E(y) + E(y) - a]^2 = E[y - E(y)]^2 + E[E(y) - a]^2 \\ &\quad + 2E[y - E(y)][E(y) - a] \\ &= \text{Var}(y) + [E(y) - a]^2 \end{aligned} \tag{4.2.2}$$

because

$$E[y - E(y)][E(y) - a] = [E(y) - a]E[y - E(y)] = E[y - a][E(y) - E(y)] = 0$$

and

$$E[E(y) - a]^2 = [E(y) - a]^2$$

because $E(y) - a$ is a constant, where $\text{Var}(\cdot)$ denotes the variance of $(\cdot)$. In (4.2.2) $\text{Var}(y)$ does not contain $a$. Hence the right side is a minimum when

$$[E(y) - a]^2 = 0 \Rightarrow a = E(y). \tag{4.2.3}$$

since we are dealing with real quantities. Then when $f(x)$ at a given $x$ is used

$$\min_f E[y - f(x)]^2 \Rightarrow f(x) = E(y|x) \tag{4.2.4}$$

or the conditional expectation of $y$ given $x$. This conditional expectation, $E(y|x)$, is the best predictor, best in the sense of minimizing the expected squared error. This 'best' predictor is defined as the regression of $y$ on $x$. That is,

Regression of $y$ on $x = E(y|x) =$ is the best predictor of $y$ at a given value of $x$.

## 4.2.6. Linear predictors

If the 'best' predictor is a linear function in the conditioned variable $x$ then the predictor is of the form

$$E(y|x) = \alpha + \beta x \tag{4.2.5}$$

for all values of $x$, where $\alpha$ and $\beta$ are constants. Standard procedure of statistical distribution theory provides the following values for $\alpha$ and $\beta$. Taking expectation on both sides of (4.2.5) we have

$$E[E(y|x)] = E(y) \tag{4.2.6}$$

and from the right side we have $\alpha + \beta E(x)$. Therefore

$$E(y|x) - E(y) = \beta[x - E(x)]. \tag{4.2.7}$$

Multiplying both sides of (4.2.7) by $x - E(x)$ and then taking expected values with respect to $x$ on both sides we obtain

$$\beta = \frac{\mathrm{Cov}(y, x)}{\mathrm{Var}(x)}. \tag{4.2.8}$$

Substituting this back we have

$$\alpha = E(y) - \beta E(x) \tag{4.2.9}$$

where $\beta$ is available from (4.2.8). Thus the best predictor is

$$f(x) = E(y|x) = \alpha + \beta x \tag{4.2.10}$$

where the values of $\beta$ and $\alpha$ are available from (4.2.8) and (4.2.9) respectively. This is the case when the conditional expectation of $y$ given $x$ is linear in $x$ for all $x$.

Observe that when more than one variable is involved as the independent variables or the variables to be preassigned, the procedure remains the same. The regression of $y$ on $x_1, ..., x_k$ is the conditional expectation of $y$ given $x_1, ..., x_k$, that is, $E(y|x_1, ..., x_k)$. This function may be linear in $x_1, ..., x_k$ or not depending upon the conditional distribution of $y$ at given $x_1, ..., x_k$. When the conditional distribution is not available then one has to go for estimating the parameters $\alpha$ and $\beta$ in the linear regression or the various parameters involved in $E(y|x_1, ..., x_k)$, linear or not. Usually the method of least squares is used in estimating the parameters in such situations. If the functions involved are not linear in the unknowns then techniques from non-linear least squares are to be used. Since these techniques are available in standard textbooks, and statistical packages are also available, more details of model building will not be given here.

# Exercises

**4.2.1**. If the regression of $y$ on $x_1$ and $x_2$ is of the form

$$E(y|x_1, x_2) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$

evaluate $\alpha_0, \alpha_1, \alpha_2$ in terms of expected values, variances and covariances of $y, x_1, x_2$.

**4.2.2**. Show that

$$f(x,y) = \frac{1}{x^2\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(y-3-2x)^2},$$

is a density for $-\infty < y < \infty, x \geq 1$ and zero elsewhere. Then compute the regression of $y$ on $x$ in this case.

**4.2.3**. If the joint density of $x, y$ is given by

$$f(x,y) = 2, x \leq y \leq 1$$

and zero elsewhere show that the regression of $y$ on $x$ as well as the regression of $x$ on $y$ are not linear in the conditioned variable.

**4.2.4**. Suppose that it is known that the regression of $y$ on $x$ is linear in $x$ but neither the joint distribution of $x$ and $y$ nor the conditional distribution of $y$ given $x$ is available. But we have $n$ data points $(x_1, y_1), ..., (x_n, y_n)$ where $x_i$'s are preassigned numbers and the $y_i$'s are the corresponding observations. By minimizing $\sum_{i=1}^{m}(y_i - a - bx_i)^2$ with the help of calculus, obtain the estimates $\hat{a}$ and $\hat{b}$ of $a$ and $b$ respectively.

**4.2.5** For the Exercise 4.2.4 obtain the estimates of $a$ and $b$ by minimizing the error sum of squares as done in Exercise 4.2.4 but by not using calculus and show that the estimates are

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$
$$\hat{b} = \frac{\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{n}(x_i - \bar{x})^2}$$

where $\bar{x}$ and $\bar{y}$ are the averages of the values on $x$ and $y$ respectively.

**4.2.6**. Repeat Exercise 4.2.4 if the model to be fitted is $y_i = a + bx_i + cx_i^2 + e_i$, $i = 1, ..., n$ where $e_i$ is the error in the observation. This means to minimize $\sum_{j=1}^{n}(y_j - a - bx_j - cx_j^2)^2$ by using calculus and estimate $a, b, c$.

**4.2.7.** Fit the model in Exercise 4.2.4 on the data points $(x, y) = (0, 1), (1, 3), (2, 6), (3, 7), (4, 10)$ and estimate $y$, if possible, for $(i) : x = 2.8$; $(2) : x = 4.1$, $(3) : x = 15$.

**4.2.8.** Construct two examples of conditional density where the regression of $y$ on $x$ is $2 + 3x - x^2$.

**4.2.9.** Construct two examples of conditional density which is normal or Gaussian but the regression of $y$ on $x$ is not linear in $x$.

**4.2.10.** Suppose that the conditional density of $y$ given $x$ is Gaussian and the regression of $y$ on $x$ is linear in $x$. What What is the general class of joint densities of $x$ and $y$.

# 4.3. Some Multivariate Data Analysis Techniques

## 4.3.0. Introduction

.

There are a lot of procedures and techniques in the literature of data analysis involving data on many scalar variables, vector variables and matrix variables. There are a number of statistical packages available also, detailing some of these techniques. It is not feasible to exhaust all the techniques available in the literature. Hence I have selected a few techniques and a brief outline on those will be given here. More of them may be seen from books on multivariate statistical analysis. I have selected the ones which can be put in terms of simple mathematical framework so that they can be easily understood without having much knowledge about Probability and Statistics.

## 4.3.1. Principal components analysis

This is basically a variable-selection procedure. Suppose that a social scientist wants to compare individuals belonging to two different tribes. She does not really know what are the important variables on which she should take measurements or observations so that a meaningful comparison can be done. Hence she had taken measurements on all possible items of variation such as lengths of legs, arms, thigh bone, cheek bone, breadth width of skull, many measurements on dimensions of eyes, ears, nose, other facial characteristics and so on. Suppose that she has made measurements on 1000 such characteristics. Now it is practically impossible to analyze the data on these 1000 variables and come up with some meaningful way to compare the two tribes. Hence there should be a way of selecting the most relevant (relevant for comparison purposes) variables out of these 1000 variables so that she can study these subset of variables and analysis of the data on these variables may become feasible.

One criterion that she can use while looking for 'important' variables is the following: If both tribes have five fingers in the hand then the number of fingers is not a variable to be taken. The length of the thumb of adult males in the two tribes are more or less the same then that is not a variable to be taken into account. Suppose that in one tribe the thumb is 5cm in length and the variation within the tribe, measured in terms of standard deviation, is $0.005$cm or the length is $5 \pm 3(0.005)$ if we take a three standard deviation limit from the mean value 5. In the other tribe suppose that the length of the thumb is again 5cm with standard deviation $0.00501$ then we say that practically the two expected lengths are more or less the same and hence this is not an important variable for comparison purposes. Suppose that the length of thigh bone in one tribe is with average length 0.6 meters with a wide variability of 0.2 meters. Here the variable $x$ = length of thigh bone is not concentrated at $0.6$ meters. It is widely varying and hence with almost certainty we cannot say that the expected length is 0.6 meters. This is a variable that one should take into account. Thus one criterion is that the variable having more variability or where the variance (square of standard deviation) is large is an important variable to study.

The joint variability of two scalar variables is measured by the covariance between the two variables and the configuration of variability in a set of variables is measured by the covariance matrix $V = (v_{ij})$ where $v_{ij}$ is the covariance between $x_i$ and $x_j$ if $x_1, ..., x_k$ are the individual variables, and then $V$ will be a $k \times k$ positive definite or positive semi-definite matrix when we deal with real scalar random variables $x_1, ..., x_k$. Since linear functions also include individual variables, instead of taking variables one at a time we may consider all possible linear functions of $x_1, ..., x_k$ and study the variance in these linear combinations. Let

$$u = A'X = a_1x_1 + ... + a_kx_k = X'A,$$

where $A' = (a_1, ..., a_k)$, $X' = (x_1, ..., x_k)$ and a prime denotes the transpose, be an arbitrary linear function of $x_1, ..., x_k$ where $a_1, ..., a_k$ are constants. Then from elementary algebra

$$\mathrm{Var}(A'X) = A'VA$$

where $V$ is the covariance matrix of the random vector $X$. We can select that linear function for which $A'VA$ is a maximum for an arbitrary $A$. Note that since $V$ is at least positive semi-definite $A'VA$ can increase beyond bound and hence we may restrict $A$ within a hypersphere or impose the condition $A'A = 1$. Then the problem is to maximize

$$\phi = A'VA - \lambda(A'A - 1).$$

where $\lambda$ is a Lagrangian multiplier. From elementary algebra, differentiating $\phi$ with respect to $A$ and equating to a null vector we have

$$2VA - 2\lambda A = O \Rightarrow (V - \lambda I)Z = O.$$

But for a non-null solution the coefficient matrix $V - \lambda I$ has to be singular or its determinant zero. That is,

$$|V - \lambda I| = 0$$

or $\lambda$ is an eigenvalue of $V$ and $A$ is the corresponding eigenvector. Since $A'A = 1$ we have

$$VA = \lambda A \Rightarrow A'VA = \lambda A'A = \lambda$$

or the largest eigenvalue of $V$ gives a local maximum for the variance of the linear function. The best choice of the variable or the best linear function of the variables, for which the variance is a maximum, is $A'X$, say $A_1X$, where $A_1$ is the eigenvector corresponding to the largest eigenvalue of $V$. This $A_1'X$ is called the first principal component or the most important linear function. The second principal component is $A_2X$ where $A_2$ the eigenvector corresponding to the second largest eigenvalue of $V$ and so on. This variable selection process is known as the principal components analysis in statistical literature.

**Example 4.3.1.** Show that the following $V$ can represent a covariance matrix and compute the principal component where

$$V = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

**Solution 4.3.1.** If $V$ is a covariance matrix then $V$ has to be at least positive semi-definite. Let us check the definiteness. The matrix is symmetric.

$$2 > 0, \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0, |V| = 4 > 0.$$

Hence $V$ is positive definite. Let us compute the eigenvalues of $V$.

$$|V - \lambda I| = 0 \Rightarrow \begin{vmatrix} 2-\lambda & 0 & 1 \\ 0 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (2-\lambda)(\lambda^2 - 4\lambda + 2) = 0$$

$$\Rightarrow \lambda_1 = 2 + \sqrt{2}, \lambda_2 = 2, \ \lambda_3 = 2 = \sqrt{2}$$

are the three eigenvalues. The largest one is $\lambda_1 = 2 + \sqrt{2}$ is given by

$$(A - \lambda_1 I)X = O \Rightarrow \begin{bmatrix} -\sqrt{2} & 0 & 1 \\ 0 & -\sqrt{2} & 1 \\ 1 & 1 & -\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow X_1 = \begin{bmatrix} 1 \\ 1 \\ \sqrt{2} \end{bmatrix}.$$

Normalizing $X_1$ we get

$$\alpha_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \ \alpha_1'\alpha_1 = 1.$$

Here $\alpha_1$ is the normalized coefficient vector and hence the first principal component is

$$u_1 = \alpha_1' X = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{\sqrt{2}}{2}x_3.$$

Now going through the same procedure we see that the normalized eigenvector corresponding to the second largest eigenvalue $\lambda_2 = 2$ is $\alpha_2' = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)$ and therefore the second principal component is

$$u_2 = \frac{1}{\sqrt{2}}x_1 - \frac{1}{\sqrt{2}}.$$

The third principal component can be seen to be

$$u_3 = \frac{1}{2}x_1 + \frac{1}{2} - \frac{\sqrt{2}}{2}x_3.$$

We can also observe the general properties that $\alpha_i'\alpha_i = 1, i = 1, 2, 3$ and $\alpha_i'\alpha_j = 0, i \neq j$, $\text{Var}(u_1) = \lambda_1 > \text{Var}(u_2) = \lambda_2 > \text{Var}(u_3) = \lambda_3$.

### 4.3.2. Canonical correlation analysis

We have already considered the problem of prediction involving one dependent variable or the variable to be predicted and one or more independent variables or variables which are conditioned or at whose preassigned values the dependent variable is predicted. Now we will generalize this idea. Suppose that we have several dependent variables or variables to be predicted and several independent variables. We can consider a problem of predicting a linear function of variables by using other linear functions of variables. Let us look at a vector of $p_1 + p_2$ variables and let

$$X = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix}, \; X_{(1)} = \begin{bmatrix} x_1 \\ \vdots \\ x_{p_1} \end{bmatrix}, \; X_{(2)} = \begin{bmatrix} x_{p_1+1} \\ \vdots \\ x_{p_1+p_2} \end{bmatrix}, \; p = p_1 + p_2.$$

Let us assume that $p_1 \leq p_2$ for convenience. Now, let us consider arbitrary linear functions of $X_{(1)}$ and $X_{(2)}$. Let the covariance matrix of $X$ be denoted by $V$ and let it be partitioned conformally to take care of the first $p_1$ components and the next $p_2$ components of $X$. Let

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

where $V_{11}$ is $p_1 \times p_1$, $V_{22}$ is $p_2 \times p_2$, $V_{12}$ is $p_1 \times p_2$ and $V_{21} = V_{12}'$. Let the arbitrary functions of $X_{(1)}$ and $X_{(2)}$ be $u = \alpha' X_{(1)}$ and $w = \beta' X_{(2)}$ where $\alpha' = (\alpha_1, ..., \alpha_{p_1})$ and $\beta' = (\beta_{p_1+1}, ..., \beta_{p_1+p_2})$ are arbitrary constants. Since linear correlation is invariant under location and scale parameters, without loss of generality we assume that $\text{Var}(u) = 1$ and $\text{Var}(w) = 1$. That is,

$$1 = \text{Var}(u) = \alpha' V_{11} \alpha, \; 1 = \text{Var}(w) = \beta' V_{22} \beta.$$

The linear correlation between $u$ and $w$ is given by

$$\frac{\text{Cov}(u, w)}{[\text{Var}(u)\text{Var}(w)]^{\frac{1}{2}}} = \text{Cov}(u, w) = \alpha' V_{12} \beta = \beta' V_{21} \alpha$$

since $\text{Var}(u) = 1$ and $\text{Var}(w) = 1$. Then the principle of maximizing the correlation between $u$ and $w$ reduces to maximizing $\alpha' V_{12} \beta$ subject to the conditions $\alpha' V_{11} \alpha = 1$ and $\beta' V_{22} \beta = 1$. Let

$$\phi = \alpha' V_{12} \beta - \frac{1}{2} \lambda (\alpha' V_{11} \alpha - 1) - \frac{1}{2} \mu (\beta' V_{22} \beta - 1).$$

where $\lambda$ and $\mu$ are Lagrangian multipliers. Then taking the partial derivatives with respect to $\alpha$ and $\beta$ and equating to null vectors we have

$$\frac{\partial \phi}{\partial \alpha} = O, \; \frac{\partial \phi}{\partial \beta} = O \Rightarrow$$

$$V_{12} \beta - \lambda V_{11} \alpha = O \quad \text{(a)}$$

$$-\mu V_{22} \beta + V_{21} \alpha = O. \quad \text{(b)}$$

From the conditions $\alpha' V_{11} \alpha = 1$ and $\beta' V_{22} \beta = 1$ we have $\lambda = \mu = \alpha' V_{12} \beta$. Putting ($a$) and ($b$) together we have

$$\begin{bmatrix} -\lambda V_{11} & V_{12} \\ V_{21} & -\lambda V_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = O.$$

For a non-null solution for this equation the coefficient matrix must be singular or its determinant must be zero. That is

$$\begin{vmatrix} -\lambda V_{11} & V_{12} \\ V_{21} & -\lambda V_{22} \end{vmatrix} = 0.$$

Since the maximum correlation is available for the maximum value of $\lambda = \alpha' V_{12}\beta$ the maximum canonical correlation between $u$ and $w$ is available from the largest $\lambda$ satisfying the above determinantal equation. Let $\lambda_1$ be the largest root. Take this $\lambda_1$ and solve equations ($a$) and ($b$) to obtain the corresponding $\alpha_1$ and $\beta_1$. Normalize through the relations $1 = \alpha' V_{11}\alpha$ and $1 = \beta' V_{22}\beta$. Use those normalized $\alpha_1$ and $\beta_1$ to write $\lambda_1 = \alpha_1' V_{12}\beta_1$. This is the largest canonical correlation and the corresponding predictor $w = \beta_1' X_{(2)}$ is the best predictor for predicting $u = \alpha_1' X_{(1)}$ or vice-versa. Now start with the second largest root of the determinantal equation, say $\lambda_2$ and proceed.

# Exercises

**4.3.1.** Evaluate the principal components in $X' = (x_1, x_2, x_3)$ with the covariance matrix

$$V = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 4 \end{bmatrix}.$$

**4.3.2.** Check whether the following $V$ can represent a covariance matrix. If so evaluate the principal components in the corresponding vector.

$$V = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}.$$

**4.3.3.** Check whether the following $V$ can represent a covariance matrix. If so compute the canonical correlation between $\{x_1, x_3\}$ and $\{x_2, x_4\}$ where $X' = (x_1, x_2, x_3, x_4)$ and

$$V = \begin{bmatrix} 1 & -1 & 1 & 1 \\ -1 & 4 & 0 & -1 \\ 1 & 0 & 2 & 1 \\ 1 & -1 & 1 & 3 \end{bmatrix}.$$

**4.3.4.** Check whether the following $V$ can represent a covariance matrix. If so compute the canonical correlation between $\{x_1, x_4\}$ and $\{x_2, x_3\}$ where $X' = (x_1, x_2, x_3, x_4)$ and

$$V = \begin{bmatrix} 3 & 1 & -1 & 1 \\ 1 & 2 & 0 & -1 \\ -1 & 0 & 2 & 1 \\ 1 & -1 & 1 & 3 \end{bmatrix}$$

**4.3.5.** For $X$,an $n \times 1$ vector and $A = A'$ and $n \times n$ matrix show that

$$\max_{X'X=1}[X'AX] = \lambda_n \text{ and } \min_{X'X=1}[X'AX] = \lambda_1$$

where $\lambda_n$ and $\lambda_1$ are the largest and smallest eigenvalues of $A$.

**4.3.6.** Let $A$ and $B$ be $n \times n$ symmetric matrices where $|B| \neq 0$. Let $X' = (x_1, ..., x_n)$. Then show that

$$\max_{X'BX=c} [X'AX] = \lambda_n c$$

where $\lambda_n$ is the largest root of the determinantal equation $|A - \lambda B| = 0$ and

$$\min_{X'BX=c} [X'AX] = \lambda_1 c$$

where $\lambda_1$ is the smallest root of the determinantal equation above.

**4.3.7.** Let $X' = (x_1, ..., x_n), b' = (b_1, ..., b_n)$ and $A = A' > 0$ be $n \times n$, where $A$ and $b$ are constants. Then show that

$$\min_{X'b=c}[X'AX] = \frac{c^2}{b'A^{-1}b}.$$

**4.3.8.** Let $X$ be $p \times 1$, $Y$ be $q \times 1$, $A$ be $p \times q$, $p \leq q$. Let $B = B' > 0$ be $p \times p$ and $C = C' > 0$ be $q \times q$. Then show that

$$\min_{X'BX=1, X'CX=1}[X'AX] = |\lambda_1|$$

and

$$\max_{X'BX=1, X'CX=1}[X'AX] = |\lambda_p|$$

where $|\lambda_p|$ is the largest and $|\lambda_1|$ is the smallest root of the determinantal equations

$$|C^{-1}A'B^{-1}A - \lambda^2 I| = 0$$
$$\text{or}|C^{-\frac{1}{2}}A'B^{-1}AC^{-\frac{1}{2}} - \lambda^2 I| = 0.$$

**4.3.9.** Let $X, Y, C = C' > 0$ be as in Exercise 4.3.8. Then for fixed $X$ show that

$$\max_{Y'CY=1}[X'AY] = \sqrt{X'AC^{-1}A'X}.$$

**4.3.10.** Let $X$ be $p \times 1$, $Y$ be $q \times 1$, $B = B' > 0$ be $q \times q$. Let $A$ be $p \times q$. Then for fixed $Y$ show that

$$\max_{X'BX=1}[X'AY] = \sqrt{Y'A'B^{-1}AY}.$$

## References

[1] Mathai, A.M. (1993): The residual effect of growth-decay mechanism and the distributions of covariance structures, *The Canadian Journal of Statistics*, **21(3)**, 277-283.

[2] Mathai, A.M. (2005): A pathway to matrix-variate gamma and normal densities. *Linear Algebra and Its Applications*, **396**, 317-328.

[3] Mathai, A.M. and Davis, T.A. (1974): Constructing the sunflower head. *Journal of Mathematical Biosciences*, **20**, 117-133.

[4] Haubold, H.J. and Mathai, A.M. (1994): Solar nuclear energy generation and the chlorine solar neutrino experiment. *American Institute of Physics, Conference Proceedings*, **320**, pp.102-116.

[5] Haubold, H.J. and Mathai, A.M. (1995): A heuristic remark on the periodic variation in the number of solar neutrinos detected on Earth. *Astrophysics and Space Science*, **228**, 113-134.

[6] Mathai, A.M. and Haubold, H.J. (1988): *Modern Problems in Nuclear and Neutrino Astrophysics*, Akademie-Verlag, Berlin.

[7] Mathai, A.M. and Haubold, H.J. (2007): Pathway model, superstatistics, Tsallis statistics and a generalized measure of entropy, *Physica A*, **375**, 110-122.

[8] Mathai, A.M. and Haubold, H.J. (2007a): On generalized entropy measures and pathways, *Physica A*, **385**, 493-500.

[9] Mathai, A.M. and Haubold, H.J. (2007b): On entropic, distributional, and differential pathways. *Bull. Astr. Soc. India*, **35**, 669-680.

[10] Tsallis, C. (1988): Possible generalizations of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, **52**, 479-487.