

STATISTICS IN ASTRONOMY*

P.C. VAIDYA

1. INTRODUCTION

I am indeed very happy to be invited to deliver the first lecture in memory of Prof. Dr. Gulabbhai R. Desai. I had the good fortune of working with him in the field of mathematics education in Gujarat since 1957. He was an honorary professor of statistics and actuarial mathematics at a renowned college of commerce at Ahmedabad. Though, professionally, he was working in the general insurance sector, basically he was a teacher and a good research worker in mathematics.

When I received this invitation to deliver the memorial lecture, I was a little worried. My area of interest is mathematical physics and astronomy. Dr. Desai's area was statistics and actuarial mathematics. So you would realize why I was a bit worried. But then Gulabbhai and I have worked together for so many years and so there must be some hidden link between our specialities. Thinking on these lines I realized that any science based on experiments and observations must use statistics explicitly or implicitly. I, therefore, decided to unravel the history of astronomy and try to pinpoint the role of statistics in the development of astronomy.

And this is what I am trying to do in what follows.

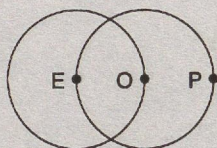
2. KEPLER AND PLANETARY MOTIONS

Now-a-days we use Kepler's laws of planetary motion to discuss the motion of various planets. And we derive these Kepler's laws from Newton's law of gravitation. But actually Kepler derived his planetary laws many many years before Newton's law of gravitation. Then how did Kepler derive his laws of planetary motion? The answer is: he used thousands of observations of positions of planets in the sky recorded by Tycho Brahe, used statistics to reduce their observational error and found the laws which would explain these observations. It is this story of observations, calculations and conclusions which I am going to describe now.

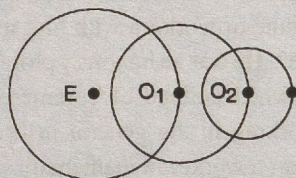
Let us first state Kepler's conclusion expressed in his first law of planetary motion: Every planet moves in an ellipse round the Sun, the Sun being not at the centre but at one focus of the ellipse. But mathematics of planetary motion was known in some form long before Kepler, to the Greeks and the Hindus of ancient

*Prof. G.R. Desai Memorial Lecture delivered on April 14, 1998 at Delhi.

times, and using this mathematics the positions of planets were calculated in advance and recorded in Almanacs (Panchangas). In those days the only closed curve known was a circle. So they began with the planet P moving in a circle with Earth E as centre. But on this assumption the calculated position of the planet did not at all agree with the observed position.



So they modified that assumption. They took a point O to move in a circle round the Earth E and the planet P to move in a second circle round O . Will the calculated position of the planet now agree with the observed position? If it does not, proceed further with your assumption.



A point O_1 moves in a circle round the earth E . Another point O_2 moves in a second circle round O_1 and the planet P moves in a third circle round O_2 . Now calculate the position of the planet in the sky and see if it agrees with the observed position. If it does, well and good, otherwise proceed further in this assumption with an additional circle: and go on adding circles in your assumption till the calculated position agrees with the observed position of the planet. This is the method of epicycles. Before the days of Kepler planetary positions were calculated in advance by this method of epicycles.

The first change in this method was introduced by Copernicus in 1540. He postulated that the centre of our universe is not the Earth but the Sun. So in the epicycle method we should not begin with Earth at the centre of the first circle. We must begin with the Sun at this centre. Mathematically this led to some simplification in calculations because now a smaller number of circles was needed for each planet. But Copernicus had to face great opposition from the Pope because to say that the Earth is not at the centre of the universe was against the religious belief of the time. So, many people including the great astronomer Tycho Brahe did not accept Copernicus' assumption. Tycho Brahe was a very good observer and he decided to observe the positions of the various planets with a view to disprove Copernicus. He devoted the rest of his life in observing the positions of planets and recording them. It appears that at the end of his efforts he was perhaps led to believe that Copernicus was after all right, because before his death he handed over the record of his thousands and thousands of planets' positions to Kepler who was known to be a supporter of Copernicus.

Kepler's first finding was that the observed position (as recorded by Tycho Brahe) of the planet differed from the position as calculated by the epicycle method by about 8 minutes of arc. Now Kepler knew that Tycho was a very talented and accurate observer. So he decided that the difference of 8 minutes of arc implied an

error in the epicycle method, so he further decided to improve the epicycle method. The first thing Kepler did was to assume the planet moving in a single circle, but did not keep the Sun at the centre of the circle, but a little away from the centre. And this off-centric assumption led to very difficult calculations. Taking the Sun at a certain distance away from the centre required laborious calculations lasting for a week. At the end of the week you find that the calculated position does not agree with Tycho's observation. So change the off-centre distance of the Sun and repeat the labour for another week and so on. In this connection Kepler writes in his book *Mysterium Cosmographicum*: "I have carried out such 70 attempts lasting for about 5 years . . ." It is only out of such "trial and error" labour that Kepler ultimately came to his law of elliptic orbit. But then what role did statistics play here? Thereby hangs an important tale of history of astronomy.

In 1963 Prof. Gingerich, Professor of Astronomy and History of Astronomy at Harvard University, while reading this book *Mysterium Cosmographicum*, got an idea, "Why not test Kepler's calculations on modern computers?" Gingerich took a week to prepare the computer program for Kepler's calculations. But once the program was ready the computer took 8 seconds to carry out the calculations. If the modern computer took as long as 8 seconds for a single calculation, one can imagine the patience and perseverance of Kepler who carried out, single-handed, such 70 calculations! But here again Gingerich had another surprise from the computer. In order to cover the discrepancy of 8 minutes through trial and error, Kepler had to make 70 trials whereas the computer needed only 9 trials! May be Kepler made mistakes in these laborious calculations, so that the mistake of one attempt got cancelled in a subsequent attempt and so the final correct result required many more attempts!

But let us not get prejudiced about this great calculator because the story is yet incomplete. After about 8 years (*i.e.*, in 1971 or so) Gingerich decided to study the whole case afresh. Now instead of following the calculations as printed in the book he decided to study the original calculations carried out by Kepler in his note-books. But the problem was: "where to find these notes and notebooks?"

After the death of Kepler his notes and other papers came in the possession of his son. After the death of the son these manuscripts were with an astronomer, who prepared a list of those mss. and published it in 1674. After this the notes passed many hands and ultimately they reached the National Museum in Vienna. Finally the Russian Empress Katherine 2nd purchased all these papers in 1773 and now they are safely stored in the Leningrad Science Academy. Gingerich obtained microfilms of these papers and was simply wonderstruck on studying these papers. On studying these notes Gingerich found that Kepler's "trial and error" calculations were as precise and correct as by modern computers, but he had some other problem which was not mentioned in the book.

Why then did Kepler need those 70 calculations? Because he found that in order to compare his calculated position with the observed position given by Tycho, he needed to refine Tycho's observations statistically. One calculates the position of a planet on a particular day. Now Tycho had recorded day-to-day positions of a planet and given thousands of positions. So one should take the observed positions in a time-interval surrounding the day of calculated position and use statistical methods to arrive at the mean observed position on that day. Thus Kepler played a double role in these calculations: use the fine trial and error method to calculate the position of the planet and at the same time use statistical methods to refine the recorded observed position of the planet. When Gingerich tried this double method on a modern computer, he found that now the computer did not take 8 seconds but took full one week to fulfill the double role!

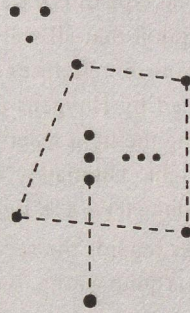
So the study of the history of astronomy revealed that even in deriving the simple Kepler's laws of planetary motions, statistics played a decisive role.

3. HERSCHEL AND THE MILKY WAY

William Herschel is known to many as the discoverer of the planet Uranus. But that discovery was only a chance-discovery. It was, so to say, a bye-product of his main work about estimating the size of our galaxy—the Milky Way. The very word "estimating" used here would suggest that this work used statistics and that is what we are describing now. Herschel estimated the size of our galaxy using a new unit of distance which we have called a Vyadhmeter (व्याध्मिटर). Let us begin with this new unit.

Names of the days of the week are related to planetary astronomy. Similarly names of Hindu months are related to stellar astronomy. A look at the night sky shows some widely scattered stars. But some stars appear quite near each other forming groups or bunches of stars. Such bunches of stars are called constellations (Nakshatra, नक्षत्र). Looking at these bunches of stars one can imagine certain well-known shapes. Ancient man has named these constellations on the basis of their imagined shape. In the Indian month of Kartika, the Krittika Nakshatra (constellation Pleiades) is seen overhead at about midnight. In the month of Mrigshirsh, the Mrig Nakshatra (constellation of Orion) will be overhead, in the month of Magh, the Magha (मघा) Nakshatra will be overhead and so on. What a fine system of recognising stars! If you know the Hindu month on your day of observation, then looking at the midnight sky, you would recognise the bunch of stars over your head. Well, if you know one person in a village, by and by you will come to know many more persons. In the same way if we could recognise one constellation in the sky then it becomes quite easy to recognise several other constellations with its help. It is for this reason that in India astronomy is a subject popular not only among the learned, it is also popular among laymen.

In the month of Mrigshirsh, the constellation Mriga (Orion) is known to village farmers as Deer, because it has the shape of a deer. Four stars at the vertices of a quadrilateral make up the four legs of the deer and the small bunch of stars just outside the quadrilateral represents its head. In the belly of the deer three stars are seen in a line. The hunting instinct of the early man suggested that the three stars in a line represent an arrow used to kill the deer. But then where is the hunter? If we produce the line joining the three stars, we shall note that there is a very bright star on this line. The early man imagined this star as the hunter whose arrow killed the deer. This bright star has been given the name Vyadh (व्यूहो), the hunter. (Its English name is Sirius). The Vyadhmeter mentioned earlier refers to this star.

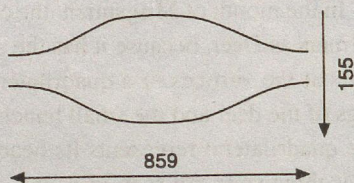


Among all the stars seen in the sky with the naked eye, this star Vyadh (Sirius) is the brightest. Farther the stars are from us, less bright they appear. William Herschel used this simple fact to estimate distances of stars. He assumed that basically all stars are equally bright. If some stars appear brighter than others, this is due to their being at lesser distances from us. The brightness of a star varies inversely as the square of its distance from us. If a star is half as bright as Vyadh, then its distance from us must be four times the Earth-Vyadh distance, *i.e.*, its distance will be 4 Vyadhmeters. If the brightness of a star is one-tenth of the brightness of Vyadh, its distance will be 100 Vyadhmeters. In this way by comparing the observed brightness of a star with that of Vyadh, we can work out its distance in Vyadhmeters. (Of course readers must have realized by now that a Vyadhmeter represents the distance of Sirius from the Sun.)

Herschel divided the visible sky into 3400 parts and counted the number of stars in each part. Fixing his attention on the stars in one such part, he compared the brightness B of each star with that of Sirius. If the brightness of Sirius is taken as 1, the brightness of each star will be less than 1. So each B will be less than 1. He calculated the average of these brightnesses B in one part and used this average brightness to find the distance of this part in Vyadhmeters. Thus he carried out the population census of stars visible in his telescope. According to his estimate there are 7.5 crores to 10 crores of stars. These stars are arranged in a bun-shaped region of space. We call this region our galaxy or the Milky Way (Sanskrit name is Aakash Ganga, DeekaeM e iebiee). As per calculations of Herschel, the diameter of the Milky Way is 850 Vyadhmeters and the central thickness is 155 Vyadhmeters.

There are two types of errors in Herschel's estimates of the size of the Milky Way. One was that he had not found the length of his 'yard-stick'—the Vyadhmeter,

the distance of Vyadh from the Sun. The second type of error enters through his assumption that all stars are equally bright. Of course certain experiments were performed by Huygens and Newon by comparing the light received from Vyadh with sunlight. Ultimately it was found that 1 Vyadhmeter = 118 light years.



As regards the second assumption observations have shown that this assumption is quite wrong. A question naturally arises: if the basic assumption of Herschel is not correct, what faith can we have in his final conclusion? It is a fact that Herschel did not have full information about stars and that whatever information he had was not quite correct. Yet the dimensions that he gave for our galaxy (the Milky Way) were very near the correct values. What could be the reason for this? Of course, Herschel's method will not give the correct value of the distance when applied to a single star but Herschel did not take individual stars. He had divided the sky into 3,400 parts, had considered all stars in each part and had found the average distance of stars in this part. That is why he did not give distances of individual stars, but gave the dimensions of the whole galaxy of stars. Herschel used statistical methods in his calculations and so his average results are largely accepted.

4. CONCLUDING REMARKS

The above two illustrations of the use of statistics in astronomy represent two classes of cases for this use. In the case of Kepler's law, the celestial object observed was a single planet. To study the motion of this planet, thousands of observations of the position-coordinates were recorded and statistics was needed to calculate the exact observed position of the planet on a particular night.

But in the case of Herschel's work, not one but a large number of celestial objects (*viz.* stars of our Milky Way) were objects of observation. At the very outset of observation one has to formulate a statistically consistent plan of observations and so statistics dominates the entire astronomical work of Herschel. As a matter of fact, Hershel's pioneering efforts in planning statistically consistent observations have helped later astronomers to study the structure of the entire universe consisting of millions and millions of galaxies, our Milky Way being one such galaxy.

So these are the two distinct ways in which statistics is used in astronomical studies.