Statistics, the Sun, and the Stars

Charles A. Whitney *Harvard University*

The visible stars appear to be scattered at random in space, so it is natural that astronomers should have turned to statistical methods and probabilistic arguments. But this is a relatively recent development. Babylonian and Greek astronomy was based on cycles, and the concepts of probability were rarely used in modern astronomy before the twentieth century.

An early use of the idea of randomness and probability in astronomy was John Michell's mathematical demonstration, in the eighteenth century, that most stars that appear to be very close together in the sky actually are close together in space. William Herschel, a contemporary of Michell, discovered an abundance of very close pairs of stars—*double stars* they were called—but he had no way of telling whether the stars really were paired or merely appeared to be. Herschel hoped that the individual stars in a pair were quite far apart and that they accidentally lay along the same line of sight; from observing such pairs he hoped to detect the motion of the earth about the sun and, ultimately, to

determine the distances of the stars from earth. But Michell computed the likelihood that such apparent pairing could arise in the numbers found by Herschel if the stars were scattered at random through space; the probability was so minute that Michell believed the stars to be physically coupled. He was later proven correct by Herschel himself. Herschel found many pairs in which the members were rotating about a common center of gravity, and he was thus led to the first demonstration of the application of Newton's concept of universal gravitation outside the solar system.

Herschel also attempted to map our Milky Way galaxy by studying how the number of stars seen in a particular direction increased as he went to fainter and fainter limits. For lack of better information, he assumed that all stars would appear the same at a standard distance, and he found that the sun appeared to be at the center of the Milky Way. Astronomers now recognize that this is an illusion—the space between the stars is littered with dust and gas that dim the light of distant stars and at the same time alter its color. Modern data and statistical methods have led to a model in which the Sun is displaced from the center and is embedded in one of our galaxy's spiral arms.

On a grander scale, recent studies have shown that the galaxies are distributed throughout space in a highly irregular pattern interspersed with vast empty regions resembling bubbles. Understanding this pattern is a major task of current cosmology.

Generally speaking, the statistical arguments now used by astronomers in their attempts to unravel the causal connections woven into the sky are of two classes: first, statistical analyses of data, and, second, physical theories based on statistical or probabilistic concepts. An example of each will be presented.

ANALYSES OF DISTURBANCES IN THE SOLAR ATMOSPHERE

The solar atmosphere seethes with activity, and some disturbances, the gentler ones, are especially amenable to statistical analysis. The solar weather, so to speak, is not altogether chaotic, and astronomers are anxious to glean what they can about the regularities of the pattern because such regularities invariably assist the construction of theoretical explanations.

The vapors of the solar atmosphere are intensely hot, so hot, in fact, that they radiate visible light. Yet, their heat is far from uniform; a telescope reveals a welter of evanescent detail that surges and disappears from place to place within brief minutes. Disturbances are strewn in an irregular pattern—the hot and cool areas cover hundreds of miles and their outlines are roughly hexagonal. This pattern is called *granulation* and the hexagon-shaped elements, called *granules,* are evidently bubbles of hot gas welling up from the interior, carrying heat from the center and disturbing the delicate outer layers. Even the most powerful telescope cannot penetrate beneath the solar atmosphere, so astronomers rely on mathematical analysis to assess the solar interior. But this analysis requires an observational check. The detailed nature of the atmospheric fluctuations provides such a check, and it also permits astronomers to probe the interior of the Sun, similar to the way a seismologist may probe the interior of the earth.

BRIGHTNESS

Two types of measurements can be made to infer the structure: the brightness and the velocity of the gas. The earliest studies showed that the pattern of brightness changed drastically every five minutes or so. Quantitative data have been obtained since from series of photographs exposed briefly every 10 or 20 seconds. The exposures are typically one-hundredth of a second, and the best series are taken immediately after sunrise, when the ground is cool and the air is steady. These series often cover several hours, and astronomers at the polar regions during the seasons of the midnight sun have been able to obtain some sequences lasting many dozens of hours.

Comparisons of individual photographs (Figure 1 is an example) separated by longer and longer intervals reveal changes in brightness associated with time lapses. The cross-correlation of the patterns on different films is determined in the following way: a line is specified on the sun's surface, and the intensity of the light at points along the line is determined on each photograph. (Those who are not interested in the measure of similarity or who are familiar with the correlation coefficient may want to skip past the following material to the heading "A Possible Model.")

Table 1 represents five points along a line on the sun's surface. The first row of numbers gives initial brightnesses. These are measured in a standardized manner so that brightness at a point is measured as a departure from the mean in units of the original variability. This standardizing gives these numbers the properties that they average to 0 and their squares average to 1. (In this table, these averages do not work out exactly because of rounding off.) The second row gives standardized brightnesses a little later in time, but at the same positions. The third row gives standardized brightnesses a few minutes later.

Note that the second row is almost the same as the first, but the third row differs a great deal from the first. One way to measure similarity between two sets of five numbers is to multiply the two standardized numbers of each pair, add the five products, and divide by five to get the average. The underlying idea is that this average will be near 1 if there is high similarity, near 0 if there

Table 1 The idea of correlation

			Points		
Time	*1*	*2*	*3*	*4*	*5*
1	1.5	0.5	0	− 0.5	− 1.5
2	1.4	0.7	0	− 0.7	− 1.4
3	1.0	− 1.4	0.4	− 1.0	1.0

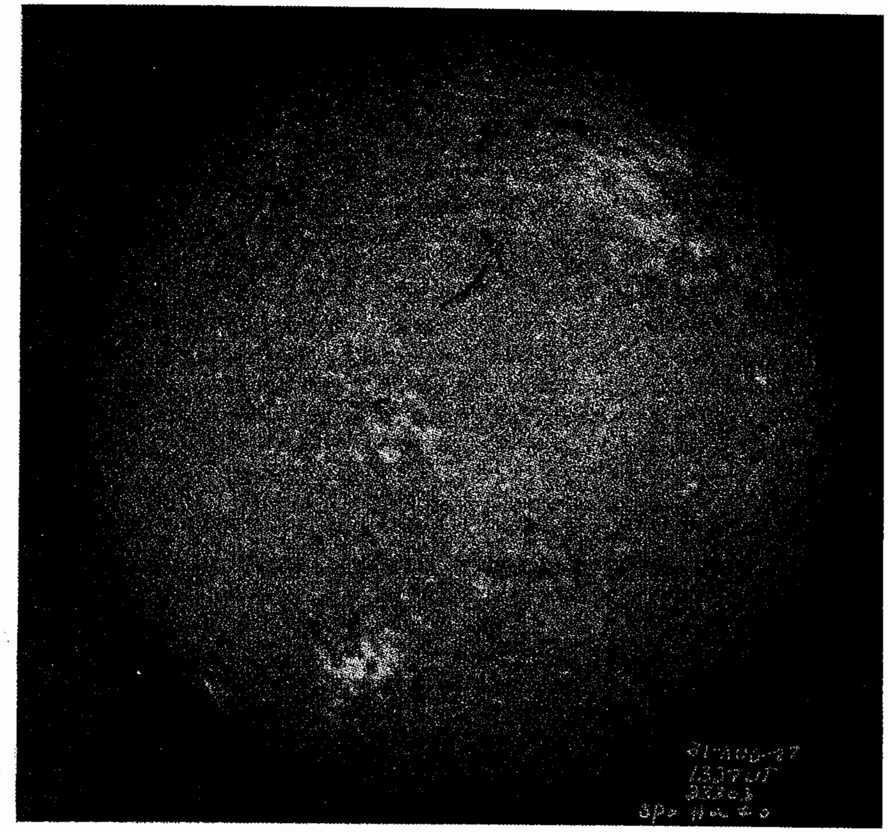

Figure 1 *Photograph of the sun's surface showing fine details that have been subjected to statistical analyses of various sorts. The light and dark areas reveal regions of different magnetic intensity. Source: National Solar Observatory.*

is little connection, and near −1 if there is high dissimilarity. The average product is called the *correlation coefficient.*

Let's first look at an extreme case of similarity; when we compare the first set of five numbers with itself, we should get perfect similarity. The correlation coefficient is

$$\tfrac{1}{5}[1.5 \times 1.5 + 0.5 \times 0.5 + 0 \times 0 + (-0.5) \times (-0.5) + (-1.5) \times (-1.5)] = 1,$$

as we said it would be. The similarity of a set of such standardized numbers with itself when measured this way always gives a value of 1, just as in this example. The correlation coefficient can range from + 1 to − 1. If we can't predict the values at one time from those at another any better than we would by guessing, the correlation coefficient gives the value 0.

Correlating the first set of brightnesses with the second, we get

$$\tfrac{1}{5}[1.5(1.4) + 0.5(0.7) + 0(0) + (-0.5)(-0.7) + (-1.5)(-1.4)] = .98.$$

This pair is very highly correlated, although, of course, slightly less than the original numbers with themselves.

Correlating the first set with the third gives us

$$\tfrac{1}{5}[1.5(1.0) + 0.5(-1.4) + 0(0.4) + (-0.5)(-1.0) + (-1.5)(1.0)] = -.04,$$

a slightly negative value, but not far from 0. The example is primarily for illustrative purposes, as the correlation ordinarily stays positive.

If every standardized brightness were replaced by its negative, we would find a coefficient of -1 between the original and the new values. Let's try it with the first set of numbers:

$$\tfrac{1}{5}[1.5(-1.5) + 0.5(-0.5) + 0(0) + (-0.5)(+0.5) + (-1.5)(1.5)] = -1.$$

Although the correlation coefficient ranges between -1 and 1, other ways of standardizing are possible. But the -1, 1 interval is conventional and convenient.

A Possible Model. This correlation will be positive if bright points on one film correspond with bright points on the other film—as will be the case if the films were separated by a very short interval of time. The correlation will be negative if brighter points on one correspond to darker points on the other, and if we cannot forecast one set from another better than guessing will do, the correlation will be 0.

As the time interval increases, the correlation has been found to decrease steadily toward 0 without actually going negative. The value of the correlation is reduced to about 1/2 in an interval of five minutes and to somewhat less than 1/4 in ten minutes.

A model for this process has been suggested. It assumes that at random times, averaging about five minutes apart, new granules appear and gradually cool. Each is replaced by another at the next random time. If the cooling is slow and the replacement is slow in coming, the correlation is slow in going to 0. If a replacement comes rapidly, the correlation goes to 0 rapidly. Theoretical work not given here shows that this model will create correlations that reduce from 1 to 1/2 in about five minutes on the average. This agreement with the facts lends some support to the model.

VELOCITY

The other type of measurement, velocity, has been more exciting in its consequences; it is also more difficult to obtain but a vast amount of data has been

accumulated, and these data show a behavior markedly different from that c brightness changes associated with granulation. Correlations can be measure for the velocity in much the same way they are for the brightness; they als decline with time, but they dip down and become negative before they ris again to 0. In fact, a detailed plot of the velocity at a particular point on th sun's surface shows an oscillation that is quite striking in its apparent regulai ity. Figure 2 shows one example; it is not perfectly regular, but there is no douł of an actual oscillation. Why are oscillations seen in the velocity, while th brightness pattern is irregular in both space and time?

The answer to this question remains incomplete, but studies of the detaile frequency composition of these oscillations are beginning to clarify th phenomenon—and to show its real complexity. These studies have shown th: the oscillations are amazingly similar to waves of musical notes emitted by violin string. Overtones are present, but they are relatively weak. In fact, ove half the energy is contained in oscillations whose periods are confined betwee two and six minutes. As a terrestrial example, we might say that a slow motio movie of surf breaking on a beach would show about the same degree c periodicity.

Astronomers were astonished by the discovery of this regularity because the had come to think of the sun's atmosphere as the seat of mere chaos; also, th brightness variations had shown no such periodic oscillations, and astronome had assumed that the pattern of the upward and downward motions woul closely mimic the irregularity of the brightness changes.

The periodicity of the motions, and its sharp contrast with the randomne of the brightness, showed at once that astronomers were observing two di ferent parts of the solar atmosphere; it has since been proven that the brightne variations are produced low in the visible atmosphere while the observed m tions take place in the upper layers of the atmosphere. And, what is more, th nature of the motions alters with increasing height in the atmosphere—th

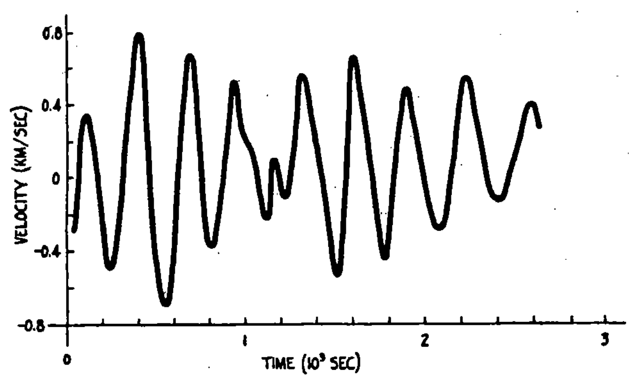

Figure 2 *Measurements of the vertical component of the velocity of gas in the sun atmosphere are plotted here for a single point on the sun. The marked periodici of the motions is typical of the solar photosphere, and it has been the object of ma correlation studies.*

average period shortens by a factor of two, and at the greatest observable height (several thousand miles above the "surface") the velocity fluctuations become quite chaotic, resembling noisy static without any pronounced periodicity.

Why? Astronomers assume that we are witnessing the upward flight of very long "sound waves" in the solar atmosphere—waves that are generated deep in the solar atmosphere, perhaps by the rising granules. Some waves are trapped in that atmosphere, predominantly those with periods of about five minutes; waves of shorter period escape quickly to the upper levels, where they predominate. Waves of very long period die out quickly, and, in fact, they are not easily excited by the granules, so they are very weak at all levels.

Even this brief explanation makes it clear that a study of the frequencies of these oscillations may reveal several features of the solar envelope: the nature of the deeper disturbances that generate these sound waves, the rate at which waves of different periods are dissipated as they propagate, and the extent to which the solar atmosphere is capable of trapping waves of different periods.

PROBLEMS

1. What role have calculations of improbability played in the theory of double stars?
2. What are the two types of statistical arguments used by astronomers?
3. Refer to Table 1. We are interested only in the measurements taken at time 2. How does the zero entry at point 3 compare to the average brightness at all points?
4. Using Table 1, calculate the correlation coefficient for the second set of measurements with the third.
5. From the model described in the text, would you expect the correlation coefficient for times 1 and 3 to be larger than that for times 2 and 3? Did your calculation above meet your expectation? Explain.
6. Refer to Figure 2. How many oscillations fail to dip below 0? Below −0.4?
7. Draw an analogue of Figure 2 for the correlation coefficient of measurements of brightness of the sun's atmosphere made at several points on the sun.
8. The velocities presented in Figure 2 show marked periodicity. Would you expect similar graphs from other measuring points in the solar atmosphere? Explain your answer.
9. a. Describe the differences in the observed brightness and velocity patterns of gas in the sun's atmosphere.
   b. Give a partial explanation of this difference.