THE STATISTICAL STUDY OF THE GALACTIC STAR SYSTEM

R. J. TRUMPLER

UNIVERSITY OF CALIFORNIA, BERKELEY

IN THE FIRST PART of this paper the general problem of describing statistically the constitution of our stellar system will be outlined. In the second a particular problem, the distance distribution of stars in a limited area of the sky, will be taken up to indicate the kind of mathematical statistical problems met with in stellar statistics.

I

Although our galaxy contains in addition to ordinary stars many peculiar objects such as variable stars, binary and multiple stars, star clusters, nebulae, and highly rarefied interstellar matter, we shall limit ourselves here to the distribution of ordinary stars. The statistical description of the system then involves the following eight variables:

Three polar coördinates with the sun as origin define the position in space

Galactic longitude	l
Galactic latitude	b
Distance	r

Two parameters are needed to describe the physical constitution of ordinary stars. We choose for these parameters

Spectral class	S
Absolute magnitude	M

If the stars are classified according to the pattern of absorption lines shown in their spectrum, the spectral classes for the great majority of stars form a continuous sequence. The spectral class S may thus be considered as a variable. Among the ordinary stars we include only those whose pattern of spectral lines fits into this sequence. The absolute magnitude is a logarithmic measure of the amount of light radiated by the star.

The velocity of a star relative to the sun is given by its three components in a rectangular coördinate system

$$\dot{x},\ \dot{y},\ \dot{z}$$

The description of the stellar system as far as ordinary stars are concerned is then supplied by the distribution law

$$F(l, b, r, S, M, \dot{x}, \dot{y}, \dot{z})dl\, db\, dr\, dS\, dM\, d\dot{x}\, d\dot{y}\, d\dot{z}\,,$$

which gives the number of stars for any combination of infinitesimal intervals of the eight variables.

There are three main difficulties with which the astronomer is confronted in trying to establish this complete distribution law.

1. Only for relatively few stars in the immediate vicinity of the sun can the distance r be measured directly by the parallax method, and only for these stars is it possible to obtain the values of all the variables. For the more distant stars the three variables l, b, S alone are directly observable. For these stars, however, certain functions of the descriptive variables are accessible to observation. These are

Apparent magnitude m, a logarithmic measure of the apparent brightness, which is a function of the absolute magnitude M and the distance r.

Apparent color index c, a function of S, M, r.

Radial velocity v_r, the velocity component in the line of sight.

Proper-motion components μ_1, μ_2, giving the star's angular motion, which are functions of the distance r and the velocity components perpendicular to the line of sight.

2. The second difficulty arises from instrumental limitations. An observational quantity can be measured only for stars bright enough to make possible their observation with available telescopes. Apparent magnitude m, position coördinates l, b, and color index c can be determined for stars as faint as the 20th magnitude, spectral classes to the 14th magnitude, and radial velocities to the 12th magnitude. The limits for the material actually available for statistical studies, however, are much lower.

The limit of apparent magnitude, which is a function of r and M, introduces a very complicated selection in the data.

3. The errors of observation are a third cause of trouble. We may for instance determine the proper motions of the stars in a given area by comparing two photographs taken several years apart. All proper-motion results will then have approximately the same probable error. For the larger motions this error may not amount to more than a few per cent, but for the smaller motions the error may be larger than the motions themselves. The observed distribution of proper motions can be statistically corrected for observational errors; but for the small proper motions this correction becomes very uncertain. The use of proper motions is thus limited by their accuracy.

So far only partial solutions of our general problems have been attempted, and these partial solutions may be divided into three groups.

1. The study of the stars in the vicinity of the sun. In a small homogeneous volume surrounding the sun the distribution of the last five variables $F(S, M, \dot{x}, \dot{y}, \dot{z})$ is to be determined. This problem is usually divided into two parts.

a) The absolute magnitude–spectral class distribution $F(S, M)$ giving the number of stars per unit volume for different spectral-class and absolute-magnitude intervals.

b) The velocity law $\Phi_{SM}(\dot{x}, \dot{y}, \dot{z})$ and its variation with S and M. The velocity distribution is generally represented as a normal frequency function, the so-called ellipsoidal distribution, and the nine parameters of this function vary with S and M.

The principal difficulties in this study are to find among the millions of stars observable the few which are in the vicinity of the sun, to allow for incomplete and selective discovery, and to remove the effects of observational errors in the parallax measures. A fair approximation of these two distribution laws has been obtained.

2. The study of the space distribution of stars irrespective of their velocities:

$$F(l, b, r, S, M).$$

Most solutions of this problem are based on the assumption that the distribution of absolute magnitudes is independent of the position in space:

$$F(l, b, r, S)\varphi_S(M),$$

where $\varphi_S(M)$ is known from the study of stars in the vicinity of the sun.

Since l, b, S are directly observable, the procedure is to select a number of sample areas of given l, b, and within each area to determine the distance distribution $F(S, r)$ for stars of different spectral-class intervals. However, for the fainter stars for which no spectral classes are available only the distance distribution $F(r)$ of all stars can be determined.

Unfortunately the distance r is not directly observable, and its distribution law must be derived from star counts according to apparent magnitude m, the latter being a function of the distance r and absolute magnitude M.

From the distribution of distances in a given area the star density as function of the distance in the direction of the area is calculated. If many sample areas with different l and b are investigated, the surfaces of equal star density can be constructed which describe the structure of the stellar system.

3. The theory of galactic rotation is concerned with the state of motion in different parts of the star system. The study of this problem involves not only statistical but also dynamical considerations.

II

The main problem in the study of the space distribution of stars is to find the distribution of distances

$$F(r)$$

for a given area of the sky from star counts made according to apparent magnitude m which give the distribution

$$A(m).$$

We shall at first make the assumption that interstellar space is perfectly transparent so that the apparent brightness of a star is inversely proportional to the square of its distance from the observer. Translated into magnitudes this leads to the relation

$$m = M + 5 \log r ,$$

where r is measured in units of 10 parsecs.

In order to make a solution of the problem possible it is further necessary to assume that the distribution of absolute magnitudes

$$\varphi(M)$$

is independent of r and is known from the study of the stars in the vicinity of the sun.

The determination of $F(r)$ from the two known distribution laws $A(m)$ and $\varphi(M)$ leads to an integral equation. I want to show that this fundamental equation of stellar statistics is merely an application of the following general statistical problem: Given the distribution law $\Phi(x, y)$ of two variables x, y; given also the relation between a quantity t and the two variables: $t = t(x, y)$; find the distribution law $\varphi_1(t)$ of t.

The distribution $\Phi(x, y)$ is transformed to two new variables of which one is t. We may choose as new variables x and t, or t and y, or t and u, where u is an arbitrary function of x, y. The distribution $\varphi_1(t)$ is then obtained by integrating the transformed distribution Φ over the variable other than t.

If t, y are chosen as the new variables we have

$$\varphi_1(t) = \int_{-\infty}^{+\infty} \Phi[x = x(t, y), y] \frac{\partial x(t, y)}{\partial t} dy ,$$

where $x(t, y)$ is found by solving $t = t(x, y)$ for x.

Of particular interest is the case where x and y are not correlated, so that

$$\Phi(x, y) = \varphi_2(x) \varphi_3(y) ,$$

and

$$\varphi_1(t) = \int_{-\infty}^{+\infty} \varphi_2[x = x(t, y)] \varphi_3(y) \frac{\partial x(t, y)}{\partial t} dy .$$

Here we have an equation relating the three distribution laws φ_1, φ_2, φ_3, and when any two of these are known, the third can be determined. If the distribution of one of the independent variables $\varphi_2(x)$ or $\varphi_3(y)$ is the unknown, this is an integral equation.

Evidently our problem of finding $F(r)$ from $A(m)$ and $\varphi(M)$ is of this type, since r and M are uncorrelated. It is convenient to introduce instead of r an auxiliary variable

$$y = 5 \log r$$

which is called the distance modulus. The relation between m, M, and y then is very simple

$$m = M + y\,,$$

and the distribution $F(y)$ of distance moduli is found from the integral equation

$$A(m) = \int_{-\infty}^{+\infty} F(y)\varphi[M = m - y]dy\,,$$

which is generally solved by numerical integration in successive approximations. $F(y)$ may then be transformed to the original variable r or translated into star densities at distance r.

The discovery that the passage of star light through interstellar space is obstructed by highly rarefied dark matter has made this solution obsolete. Since the distribution of the absorbing matter within our stellar system is rather irregular, the increase of absorption with distance in the direction of the area must be introduced as an additional unknown.

We designate by $a(r)$ the absorption in magnitudes between a star at distance r and the observer; the absorption coefficient at distance r is then the first derivative of $a(r)$. Since the absorption varies with the wave-length of light, both the absorption law $a(r)$ and the distance distribution of stars $F(r)$ can be determined if we measure the apparent magnitudes of the stars in two different wave-length intervals. The visual magnitude m_v measures the brightness of a star in yellow light (for faint stars it is generally obtained from photographs taken through a yellow filter); the photographic magnitude m_p measures the brightness in the blue-violet.

The relations between apparent magnitude and distance are now

$$m_v = M_v + 5\log r + a(r)\,,$$
$$m_p = M_p + 5\log r + (1 + \chi)a(r)\,.$$

$a(r)$ represents the absorption in visual light; $1 + \chi$ is the ratio between photographic and visual absorption. χ appears to be constant and its value is fairly well known.

Star counts according to visual magnitude and according to photographic magnitude give the distribution laws

$$A(m_v) \qquad \text{and} \qquad A(m_p)\,.$$

Assuming that M_v and M_p are not correlated with r, we know their distributions

$$\varphi(M_v) \qquad \text{and} \qquad \varphi(M_p)$$

from the study of stars in the vicinity of the sun. The problem is to find

$$F(r) \qquad \text{and} \qquad a(r)\,.$$

The distance modulus is now defined as

$$y = 5 \log r + a(r) ,$$

and the absorption when represented as a function of y is designated as $\alpha(y)$. The relations between the variables then become

$$m_v = M_v + y ,$$
$$m_p = M_p + y + \chi\alpha(y) = M_p + f(y) .$$

The first of these is the same as in the former case when absorption was neglected, and the distribution $F(y)$ is determined from the star counts $A(m_v)$ by the same integral equation as before; the only difference is that the distance modulus y has now another meaning and that its relation to the distance r is unknown.

As we take up the photographic magnitudes m_p, to which the second relation applies, we know the distributions of all three variables: $A(m_p)$ from the star counts, $\varphi(M_p)$ from the study of the nearest stars, and $F(y)$ from the solution of the visual star counts; what is unknown, however, is $f(y)$, that is, the functional relation of the variables.

The general statistical problem before us is the reverse of that discussed at first: Given the distribution $\Phi(x, y)$ of two variables, given also the distribution $\varphi_1(t)$ of a function t of these two variables, what is the functional relation of t with x, y?

This problem can be solved under the following restrictions:

a) When the functional relation is of the form

$$f(t) = g(x, y) ,$$

where g is known;

b) When the functional relation is of the form

$$f(y) = g\,(t, x) ,$$

where g is known, and when x and y are not correlated so that $\Phi(x, y) = \varphi_2(x)\,\varphi_3(y)$.

The latter case, where $\varphi_1(t)$, $\varphi_2(x)$, $\varphi_3(y)$ are given whereas $f(y)$ is unknown, corresponds to our astronomical problem. The solution is obtained by introducing an auxiliary variable η defined by

$$\eta = f(y) .$$

The relation between η, t, x is then known:

$$\eta = g(t, x) ,$$

and the distribution law of η, $\psi_3(\eta)$ is found by solving the integral equation

$$\varphi_1(t) = \int_{-\infty}^{+\infty} \psi_3(\eta)\varphi_2[x = x(t, \eta)] \frac{\partial x(t, \eta)}{\partial t} d\eta ,$$

in which $\varphi_1(t)$ and $\varphi_2(x)$ are known.

Since η is a function of y only, the frequency for corresponding intervals of η and y must be the same:

$$\psi_3(\eta) \frac{d\eta}{dy} = \varphi_3(y) .$$

If ψ_3 and φ_3 are given in analytical form, this is a differential equation for the determination of $\eta = f(y)$. One pair of corresponding values of η and y must be known to obtain the integration constant.

When ψ_3 and φ_3 are given numerically as tables, as is generally the case in astronomy, each distribution is integrated numerically

$$F(\eta) = \int_{\eta_0}^{\eta} \psi_3(\eta) d\eta$$

$$G(y) = \int_{y_0}^{y} \varphi_3(y) dy ,$$

where η_0, y_0 are given by the initial condition. Corresponding values of η and y are then found by interpolating pairs of arguments for which $F = G$.

This method has been successfully applied to determine the distance distribution $F(r)$ and the absorption law $a(r)$ from star counts according to visual and photographic magnitudes.

February, 1946