STATISTICAL STUDIES RELATING TO THE DISTRIBUTION OF THE ELEMENTS OF SPECTROSCOPIC BINARIES

ELIZABETH L. SCOTT
UNIVERSITY OF CALIFORNIA

1. Introduction

A study of the relation between the theoretical distribution of the orbital elements of binary stars, on the one hand, and of the distribution of their catalogue values, on the other hand, is of interest in the problem of stellar evolution. One example is presented by the hypothesis of Struve [1] that the evolutionary process of certain binaries involves the emission of streams of gas which encircle the two components of the binary system.

With the spectroscope, the radial velocity (the component of velocity in the line of sight) of a star can be measured. If the star is a member of a binary system then its radial velocity will vary periodically with time in a type of periodic curve which is perfectly determined by Kepler's laws. If, in addition to the two stars, there is present a stream of gas then the measurements of the radial velocity of the bright component will be affected by the absorption of the star's radiation by the gas streaming between and encircling the two components. As a result, the numerical graph of radial velocities will not conform to Kepler's laws and the catalogue elements, representing "best fitting" compromise values obtained by forcing a Keplerian orbit onto the non-Keplerian graph, will be affected by systematic errors. Thus, the Struve effect would contribute to the differences between the distribution of the true elements of binaries and that of their catalogue elements. Struve has found some binaries for which there is strong evidence of a ring of gas.

Theoretical considerations suggest that the true value of ω, the longitude of periastron, which describes the orientation of the major axis of the orbital ellipse, must be distributed uniformly between 0° and 360°. However, it was shown [2], [3] that, at least for some categories of stars, the distribution of the catalogue value of ω is not uniform. It is obvious that the true distribution of ω could be distorted by many causes. For example, stars with some values of ω are easier to identify as binaries than others. This effect will be referred to as selective identifiability. Furthermore, the ordinary procedure of computing orbital elements from observations affected by errors is likely to favor some values of ω at the expense of others. Now, proof that the combined effect of these factors is not sufficient to explain the

This paper was prepared with the partial support of the Office of Naval Research.

nonuniformity of the observed distribution of ω would provide evidence for the necessity of some additional hypothesis such as that of the Struve effect.

The purpose of the present paper is to summarize some results relating to the effects that selective identifiability of binaries and also errors in determining their orbits will have upon the distribution of the catalogue values of the elements. Complete presentation of some of these results is deferred to a forthcoming issue [4] of the Publications in Statistics of the University of California.

2. Selective identifiability of spectroscopic binaries

In this section we shall consider how the probability of classifying a binary star as having variable radial velocity depends upon the various elements of the orbit of the bright component. In so doing, it will be necessary to construct a precise but, unfortunately, oversimplified model of the processes which lead to the conclusion that a star has variable radial velocity.

We consider that n observations,

$$x_1, x_2, \ldots, x_n,$$

of the radial velocity of a star will be taken at times

$$t_1, t_2, \ldots, t_n,$$

respectively, and that the time of observation is chosen at random and thus is uniformly and independently distributed throughout the period of revolution. Let ξ_j denote the star's true radial velocity at time t_j. Then the difference, $x_j - \xi_j$, is the error of the j-th observation. We assume that the errors of observation are completely independent of each other and of the true radial velocities and that they follow a normal distribution with expectation zero and known standard error σ.

Consider a star which has orbital elements

K = semiamplitude of variation in radial velocity,
e = eccentricity,
ω = longitude of periastron,
ξ_0 = radial velocity of center of mass of the system
(ξ_0 is used instead of the usual notation γ).

Then, by considering the laws of motion, it was found [4] that the probability density of ξ, the true radial velocity, is given by, say,

$$(1) \quad p(\xi \mid K, e, \omega, \xi_0) = \frac{(1-e^2)^{3/2}}{2\pi K}\left\{\frac{1}{[1+e\sin(\phi+\omega)]^2} + \frac{1}{[1+e\sin(\theta-\omega)]^2}\right\},$$

where ξ varies between the limits

$$K(e\cos\omega - 1) + \xi_0 \leqq \xi \leqq K(e\cos\omega + 1) + \xi_0$$

and where, for the sake of brevity,

$$(2) \qquad \phi = \arc\sin\left[\frac{\xi - \xi_0}{K} - e\cos\omega\right].$$

Figures 1 to 4 illustrate the general ∪-shaped distribution corresponding to equation (1). Curves of this kind were obtained graphically by Schlesinger [5] for $\omega = 0°, 90°$, etc.

Combining equation (1) with the assumptions above, we find that the absolute probability density of each x, the observed radial velocity, is given by

$$(3)\ p(x \mid K,\ e,\ \omega,\ \xi_0,\ \sigma)$$
$$= \frac{(1-e^2)^{3/2}}{(2\pi)^{3/2}\sigma} \int_{-\pi/2}^{\pi/2} e^{-\lambda^2/2\sigma^2} \left\{ \frac{1}{[1+e\sin(\phi+\omega)]^2} + \frac{1}{[1+e\sin(\phi-\omega)]^2} \right\} d\phi$$

where, for brevity,

$$\lambda = x - \xi_0 - K(\sin\phi + e\cos\omega).$$

The lower part of figure 5 illustrates typical distributions corresponding to equation (3). This figure indicates how the probability density of the true radial velocity ξ combines with the probability density of the error of observation to produce the probability density of x, the observed radial velocity. Note that the probability density of the error of observation is shifted so that it may be more easily compared with that of x.

It will be noticed that, for moderate values of K, the distribution of x differs but slightly from the normal distribution. Thus, it is clear intuitively that, no matter what method of identifying binaries is used, the probability of detecting that such a star has variable radial velocity must be very small. If σ were not known but estimated from the observations x, this probability would be still smaller. It is also clear intuitively that when the eccentricity is small or when ω differs from zero so that the distribution of ξ is not so concentrated at one point, then the probability of detecting that the star has variable radial velocity will increase somewhat. This is because the proportion of time during which the radial velocity of the bright component differs appreciably from that of the center of mass of the binary system will increase so that there is more chance of observing such a divergent velocity.

Although no standard method has been adopted to decide whether a star has constant or variable radial velocity, we postulate that the procedures ordinarily used conform to the "best" (in a certain sense) criterion for identifying a variable. This corresponds to the experience in many fields of application of mathematical statistics that the procedures developed on intuitive bases by persons in the field have later been proved by statisticians to be the best possible.

We want the function of the observable velocities, $x_1, x_2, \ldots, x_n$, which is "best" for testing the hypothesis that the star has constant radial velocity, that is, for testing the hypothesis that $K = 0$. Using the Neyman-Pearson theory of testing hypotheses [6], the test which has the following properties was found [4]:

(a) The probability that a constant velocity star is deemed to have variable radial velocity is a small number α, fixed in advance, no matter what ξ_0, σ and n may be.

(b) Out of all the tests satisfying (a), the probability that a variable velocity star is recognized as having variable velocity is as large as possible in the neighborhood of $K = 0$ in the sense that the first nonzero derivative of this probability, taken with respect to K and evaluated at $K = 0$, is maximum for all e, ω, ξ_0, σ and n.

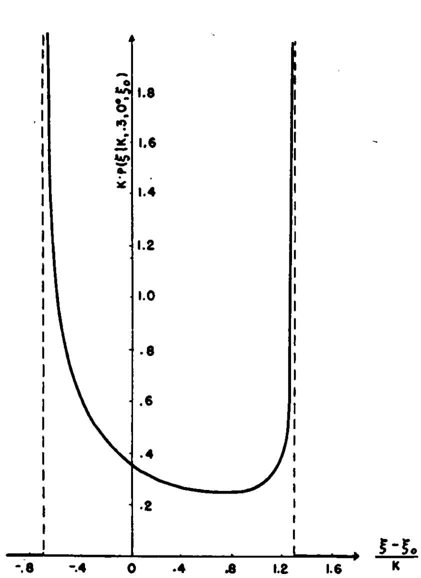

FIGURE 1

$K \cdot p(\xi | K, e = .3, \omega = 0°, \xi_0)$ as a function of $(\xi - \xi_0)/K$

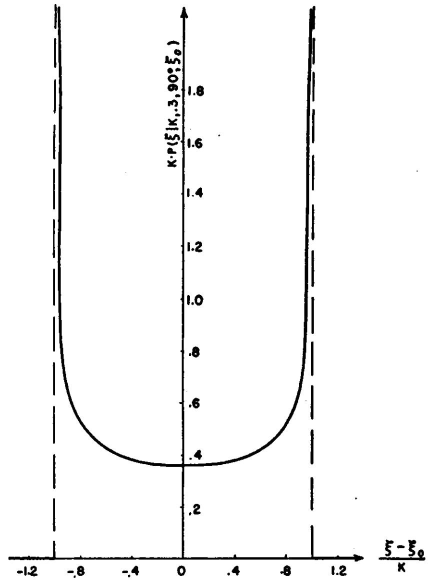

FIGURE 2

$K \cdot p(\xi | K, e = .3, \omega = 90°, \xi_0)$ as a function of $(\xi - \xi_0)/K$

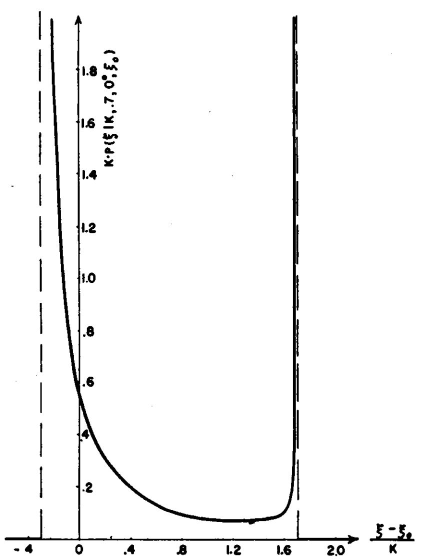

FIGURE 3

$K \cdot p(\xi | K, e = .7, \omega = 0°, \xi_0)$ as a function of $(\xi - \xi_0)/K$

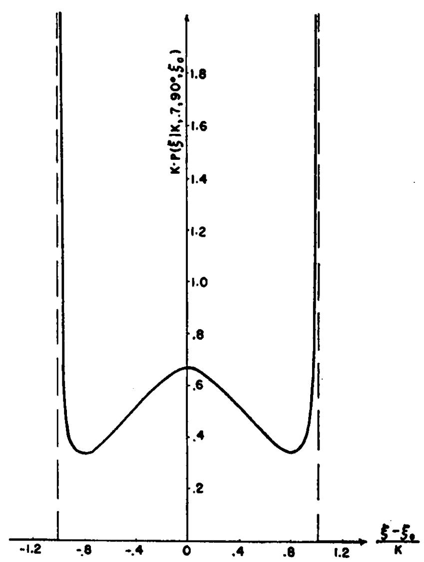

FIGURE 4

$K \cdot p(\xi | K, e = .7, \omega = 90°, \xi_0)$ as a function of $(\xi - \xi_0)/K$

The criterion corresponding to the best test, in this sense, is: Declare that the star has variable radial velocity whenever

$$\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\bar{x})^2 \geq \chi^2_{\alpha,n-1} \tag{4}$$

where $\bar{x}$ is the arithmetic mean of the observed radial velocities, $x_1, x_2, \ldots, x_n$ and

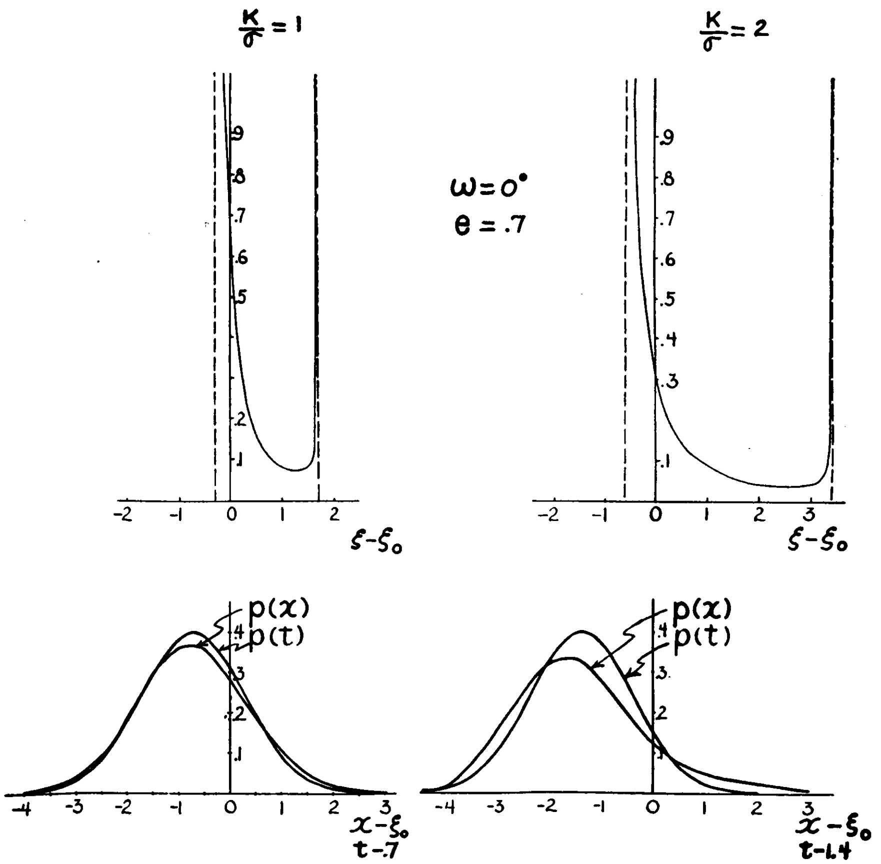

FIGURE 5

Distribution of true radial velocity ξ, of error $t = x - \xi$, and of the observable radial velocity x.

$\chi^2_{\alpha,n-1}$ is the value of the classical χ^2 corresponding to the level of significance α and to $n - 1$ degrees of freedom. This is the test suggested by Trumpler [7] on intuitive grounds. Previously, the author proved [8] that Trumpler's test has the described optimum property for the special case when the orbit is a circle. In this case, however, the computations are greatly simplified.

If the star actually is a binary with specified elements, K, e and ω, then the probability of identifying it by criterion (4) was found [4] to be represented approximately by, say,

$$\beta\left(\frac{K}{\sigma}, e, \omega, n\right)=\frac{1}{\Gamma(\gamma)}\left(\frac{\delta}{\delta+1}\right)^{\gamma}\sum_{r=0}^{\infty}\frac{\Gamma(\gamma+r)}{(\delta+1)^{r}r!}P_{n-1+2r}, \tag{5}$$

where P_{n-1+2r} is the probability that the classical χ^2 with $n-1+2r$ degrees of freedom exceeds the value $\chi^2_{\alpha,n-1}$ and

$$\gamma=\frac{n(n-1)\mu_2^2(\xi)}{(n-1)\mu_4(\xi)-(n-3)\mu_2^2(\xi)},$$

$$\delta=\frac{2\sigma^2\gamma}{(n-1)\mu_2(\xi)},$$

in which $\mu_2(\xi)$ and $\mu_4(\xi)$ are the second and fourth, respectively, central moments of the true radial velocity ξ, namely,

$$\mu_2(\xi) = \tfrac{1}{2}K^2(1-e^2)(1-E_2),$$

$$\mu_4(\xi) = \tfrac{3}{8}K^4[4(1-e^2)^{3/2}-(1-e^2)^2(3-4E_2+E_4)]$$

with

$$E_r=\frac{e^r\cos r\omega}{[1+(1-e^2)^{1/2}]^r}\qquad\text{for}\qquad r=2,4.$$

Sampling experiments indicate that the error introduced by the approximation in (5) is negligible, being $\leqq .01$.

Figure 6 and table I illustrate formula (5) and show how the probability of identifying a binary depends very sharply on all of the parameters involved, that is, on K/σ, e, ω and n. Since the probability depends upon all these parameters at once, formula (5) cannot be illustrated very conveniently by graphs; in fact, the dependence on n is shown for two sets of e and ω only: for the circle, $e = 0$, the case in which the binary is most easily identified, and for one of the most difficult cases, $e = 0.9$ and $\omega = 0°$.

Our intuitive considerations are now verified: the probability of identifying a binary increases as ω departs from 0° or 180°, as the eccentricity decreases and, also, as K/σ increases. The dependence on eccentricity is especially strong so that as e tends to one, the probability of detecting variable velocity decreases to α, the preassigned level of significance.

Comparing the curves in the upper left section of figure 6 with the observed distribution of the catalogue values of ω [2], [3], it will be noticed that the observed nonuniformity cannot be the effect of selective identifiability alone.

Figure 6 and table I may be used to read out the proportion of binaries of particular categories which one may expect to identify with a given observing program. Also, they may be used to design an observing program so that the probability of identifying a given type of binary has a specified value. The curves in the lower right section of figure 6 illustrate a fact which seems interesting and, perhaps, unexpected. Whenever the number of plates is increased by any factor, say m, and, at the same time, there is a decrease in precision by the same amount (so that σ is increased by the factor $\sqrt{m}$) then the probability of identifying a binary is

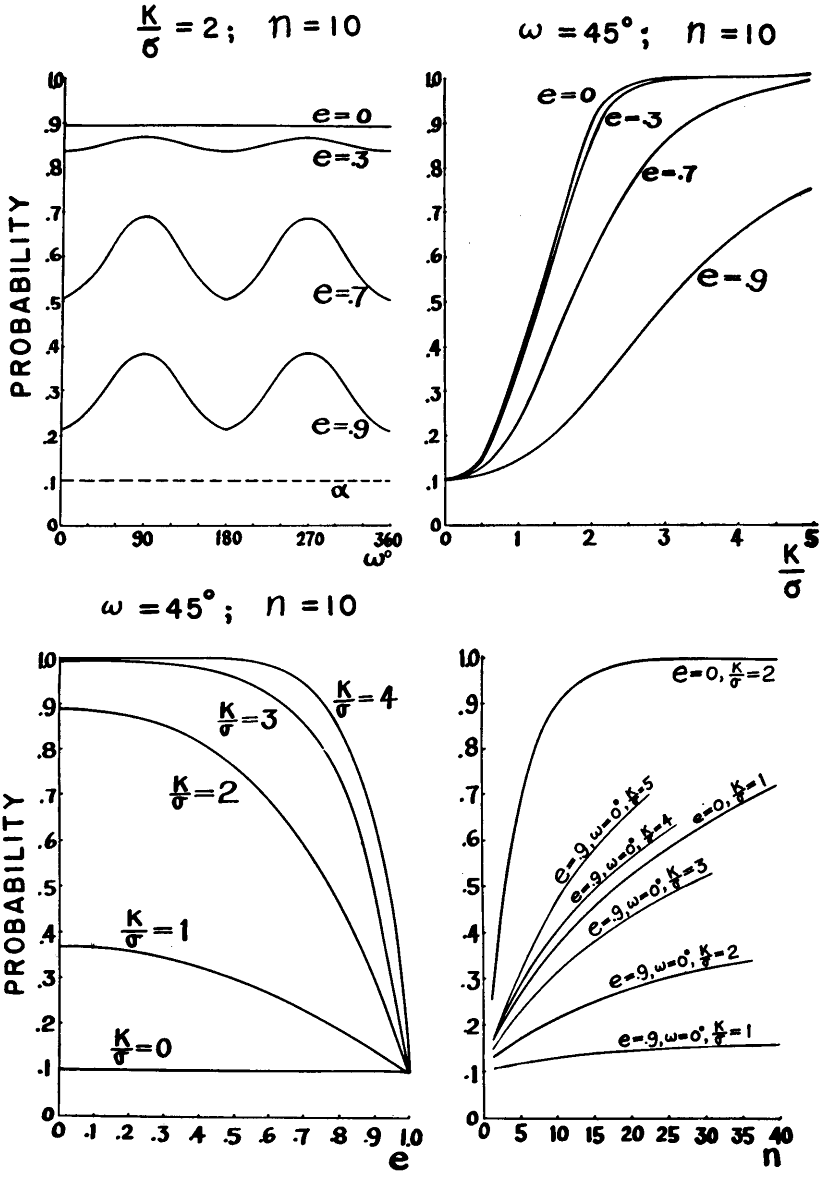

FIGURE 6

Probability of identifying a spectroscopic binary as a function of K/σ, e, ω and n

TABLE I

PROBABILITY OF DETECTING VARIABILITY OF RADIAL VELOCITY FOR VARIOUS VALUES OF n, e, ω AND $\frac{K}{\sigma}$

$\omega°$	e	$n=5$							$n=10$							$n=15$						
		$\frac{K}{\sigma}=0$	0.5	1	2	3	4	5	$\frac{K}{\sigma}=0$	0.5	1	2	3	4	5	$\frac{K}{\sigma}=0$	0.5	1	2	3	4	5
0	.0	.10	.14	.27	.68	.91	.98	1.00	.10	.15	.37	.89	1.00	1.00	1.00	.10	.18	.45	.97	1.00	1.00	1.00
	.3	.10		.25	.62	.86	.95	.98	.10		.34	.84	.98	1.00	1.00	.10		.41	.93	1.00	1.00	1.00
	.7	.10		.17	.36	.53	.66	.74	.10		.21	.50	.74	.86	.92	.10		.24	.61	.85	.94	.98
	.9	.10	.10	.12	.17	.23	.28	.30	.10	.11	.13	.21	.31	.40	.46	.10	.11	.14	.25	.38	.49	.57
45	.0	.10	.14	.27	.68	.91	.98	1.00	.10	.15	.37	.89	1.00	1.00	1.00	.10	.18	.45	.97	1.00	1.00	1.00
	.3	.10		.25	.63	.88	.96	.99	.10		.34	.85	.99	1.00	1.00	.10		.42	.94	1.00	1.00	1.00
	.7	.10		.19	.42	.64	.78	.86	.10		.23	.60	.85	.95	.98	.10		.27	.71	.94	.99	1.00
	.9	.10	.11	.13	.22	.34	.45	.54	.10	.11	.15	.30	.49	.64	.74	.10	.11	.16	.35	.59	.76	.86
90	.0	.10	.14	.27	.68	.91	.98	1.00	.10	.15	.37	.89	1.00	1.00	1.00	.10	.18	.45	.97	1.00	1.00	1.00
	.3	.10		.26	.65	.90	.97	.99	.10		.35	.87	.99	1.00	1.00	.10		.43	.95	1.00	1.00	1.00
	.7	.10		.20	.48	.75	.89	.96	.10		.25	.69	.94	.99	1.00	.10		.30	.81	.98	1.00	1.00
	.9	.10	.11	.14	.28	.46	.63	.75	.10	.11	.16	.38	.66	.85	.94	.10	.12	.18	.47	.78	.94	.98

usually *decreased*. Thus, for example, it is less efficient to obtain 10 plates each of which has a standard error of $2\sigma_0$ than to take only 5 plates with standard error of σ_0 each. In other words, two plates taken independently and, as far as possible, simultaneously on each of five randomly selected nights give a better chance of identifying a binary than ten plates taken singly on ten different nights. In interpreting this result, one must emphasize the condition of independence of the plates taken in quick succession on the same night. If the two plates are taken by the same instrument then this condition cannot be satisfied exactly because of the so called "night error." However, if the night error is not large, the condition may be satisfied approximately.

3. Effects of the determination of the elements

The present section is concerned with the next step in establishing the connection between the true elements and the catalogue elements of spectroscopic binaries. For this purpose, it is necessary to extend the model we have been considering so as to include the processes used by the astronomer in the determination of the orbital elements.

We consider that all those stars which are classified, after n observations of radial velocity, as having variable radial velocity are observed further in an effort to determine the period and the other elements. The method of determining the period consists, strictly, of: make a guess at the period, obtain a few observations in an attempt to verify this guess, then guess again, and so forth. It is possible that no satisfactory estimate of the period can be found even when the number of observations available is increased to a total of N (perhaps, $N = 40$).

Even though a binary is identified as having variable radial velocity and a period determined, the *catalogue* elements determined for it need not be the *true* elements, K, e, ω and ξ_0. The calculation of the elements is influenced by the errors in observing the radial velocity and by any error made in determining the period. The methods used contain an essential step which consists in drawing a "smooth" curve through the observations of radial velocity plotted against the phase of the observation. Then, either by identifying this curve as representing one of the possible Keplerian orbits or by certain geometric computations, the estimates of the elements are deduced. When the observations are numerous and relatively accurate, corrections are computed which mitigate the effects of any errors in judgment, but the effects of the errors in the observations, of course, persist.

We postulate that a total of N observations are available for estimating the elements K, e, ω and ξ_0 and that the combined effects of the weather, of the efforts to find the period and, once the period has been determined, of the efforts to fill the gaps in the velocity curve will tend to produce a uniform distribution of the times of observation with respect to the period.

There are, then, three steps which must be accomplished before the catalogue elements are obtained.

(i) The first step is that the binary be identified as having variable radial velocity. In the last section we studied the probability $\beta(K/\sigma, e, \omega, n)$ that this step will be accomplished.

(ii) The next step is the determination of a period. Let $\psi(P, K/\sigma, e, \omega, N, \text{var.})$ denote the probability that an estimate of the period is obtained with N or fewer observations, given that the binary has been identified as having variable velocity. In addition, let $\Phi(Q|P, K/\sigma, e, \omega, N, \text{var.})$ denote the probability density of the estimate, say Q, of the period when the true elements are P, K, e and ω, the standard error of observation is σ, and given that the binary has been identified as variable and a period determined with N or less observations.

(iii) When the first two steps are completed and N observations are available, then estimates of the elements K, e, ω and ξ_0 are determined. Let L, f, v and η_0 stand for the catalogue elements corresponding to K, e, ω and ξ_0, respectively. Also, let $\Lambda(L, f, v, \eta_0|P, K/\sigma, e, \omega, \xi_0, N, \text{var.}, Q)$ denote the probability density of the estimates L, f, v and η_0 relative to all the conditions enumerated at the end of (ii) plus the condition that the true velocity of the center of mass is ξ_0 and that the estimated period is Q.

It is obvious that all of the probabilities, β, ψ, Φ and Λ, enter into the relation between the observable distribution of the catalogue elements Q, L, f, v and η_0 and the unknown distribution of the true elements P, K, e, ω and ξ_0. Unfortunately, at the present time, there is no information about ψ, Φ, and Λ. The remainder of the present section is given to a study of the probability density

$$(6) \qquad \Lambda_1 (L, f, v, \eta_0|P, K/\sigma, e, \omega, \xi_0, N, \text{var.}) = \Lambda (L, f, v, \eta_0|P, K/\sigma, e, \omega, \xi_0, N, \text{var.}, P)\,.$$

Thus, Λ_1 is defined as the probability density of L, f, v and η_0 relative to all the conditions previously specified with the additional hypothesis that there is no error in estimating the period. As a justification for the special interest in Λ_1, the argument may be advanced that, whenever an astronomer announces the period of a spectroscopic binary, it may be expected that, owing to the usual checks already described, this estimated period is approximately equal to the true period.

Since the usual procedure for estimating K, e, ω and ξ_0 is not strictly analytic, no analytic study of its properties seems possible and the probability density Λ_1 was studied by means of a sampling experiment.

For a large number of cases, $N = 40$ synthetic observations were produced and then, assuming the correct period known, the other elements were estimated from a plot of the synthetic observations *without* knowledge of the true elements used to produce these observations. Figure 7 is intended to illustrate how the sampling experiment was set up. Each of the two halves of the figure corresponds to a separate combination of values of the true elements. The curves in the upper part of the figure show the true radial velocity as a function of phase (labeled from 0° to 360°, for convenience) and are computed from known formulae and tables. In order to produce the first "synthetic observation," we begin by selecting at random a number between 0 and 360 to represent the phase of the first observation. (This may be accomplished by reading from a table of random numbers, for example, from [9].) Next, the error of this observation is determined by choosing at random a normal deviate of appropriate σ. (This may be accomplished with the aid of a table of random normal deviates, for example, [10].) Now this error of observation

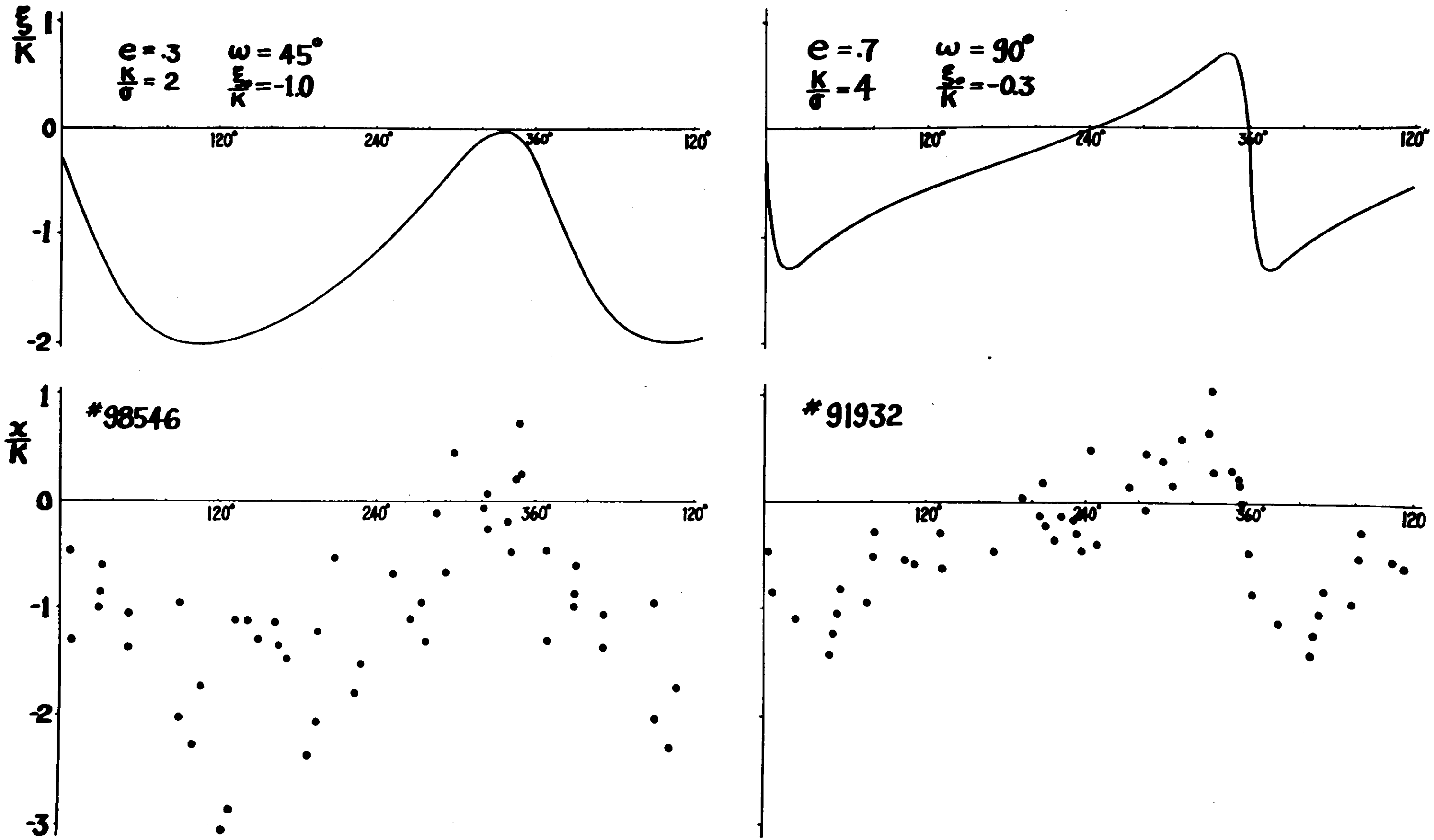

FIGURE 7

Production of synthetic observations of radial velocity is added to the true radial velocity corresponding to the phase of the first observation and the result is the first of the synthetic observations of radial velocity. This process is repeated $N = 40$ times to obtain the 40 observations which are plotted. The lower part of figure 7 gives the plots obtained in the two cases shown in the upper part of the figure. The only identifying material on the plot is a new random number used for coding and for shuffling the plots.

The plot is used to determine the estimates of the elements by comparing it in turn with each of a dense set of true velocity curves until the best fitting curve is found. The author tried to make all decisions in the same way that they are usually made, as described in the literature. In particular, whenever a circular orbit is possible, it was adopted as the estimate.

The reader will realize that this sampling experiment is laborious and very time consuming. It is still in progress. At the time of writing, all of the results obtained referred to single line binaries (the spectrum of the brighter component only is discernable on the spectrogram) and most of the results referred to $K/\sigma = 2$. In addition, some sampling experiments were made with $K/\sigma = 0$, with $K/\sigma = 4$ and $K/\sigma = 6$. For $K/\sigma = 2$ the same number 30 experiments were made for each of the following 15 combinations of e and ω

e	ω°
0	—
.1	0, 45, 90
.3	0, 30, 45, 60, 90
.7	0, 45, 90
.9	0, 45, 90.

The value of ξ_0 was chosen at random separately for each of the 450 sampling experiments performed. The reader will notice that for reasons of symmetry the values of ω considered range from 0° to 90° only.

The results of the study reportable now refer to the distribution of the estimate, say f_1, of the eccentricity when the true period is known. Since the methods of obtaining the estimate f_1 do not depend, in any way, on the value of ξ_0 nor of P, we need consider as parameters only K/σ, e, ω and N. Table II shows the results referring to the case $K/\sigma = 2$ and $N = 40$. The particular columns of this table refer to various combinations of the true eccentricity e and the true longitude of periastron ω. The rows of the main body of the table, that is, all rows except the last, give the observed frequency of the different values of f_1 obtained. There are entries in the last row in only the last six columns, referring to large values of the eccentricity. These entries represent the frequencies of those cases where, even when the true period was assumed known, it was impossible to make any estimate of the orbital elements. Figure 8 shows one such case. Although table II lists these frequencies, they are not taken into account in the further computations for the reason that we are now discussing step (iii) in the determination of the catalogue elements. Step (iii) is concerned with the determination of the elements K, e, ω and ξ_0 for those stars which already (i) have been identified as binaries and (ii) for which a period has been found. The particular sampling experiments referred to in the last line of table II are precisely those for which the variability of radial velocity would not even be detected with 40 observations, much less with 10 observations [this was checked by computing the criterion (4) for identifying a binary]. In fact, the last line is in qualitative agreement with the probability $\beta(K/\sigma,\ e,\ \omega,\ n)$ of

TABLE II

FREQUENCY DISTRIBUTION OF THE ESTIMATED ECCENTRICITY f_1 FOR VARIOUS COMBINATIONS OF THE TRUE ECCENTRICITY e AND THE TRUE LONGITUDE OF PERIASTRON ω WHEN $K/\sigma = 2$ AND $N = 40$ OBSERVATIONS

True e	0	.1			.3					.7			.9		
True $\omega°$		0	45	90	0	30	45	60	90	0	45	90	0	45	90
Estimate f_1 0	15	14	10	10	7	4	6	2	5	2	1	0	0	0	0
.1	14	10	11	14	9	5	2	11	7	0	0	0	0	0	0
.2	1	3	7	2	5	9	8	10	9	0	5	2	0	0	2
.3		1	0	3	6	6	6	3	2	0	4	8	2	0	0
.4		1	2	1	2	5	4	1	4	1	0	2	0	0	0
.5		1			1	1	1	1	2	5	4	2	0	2	0
.6							2	1	0	6	3	9	1	2	1
.7							1	1	1	6	7	4	1	1	3
.8										0	0	1	2	1	2
.9										1	0	0	2	3	6
Can't										9	6	2	22	21	16

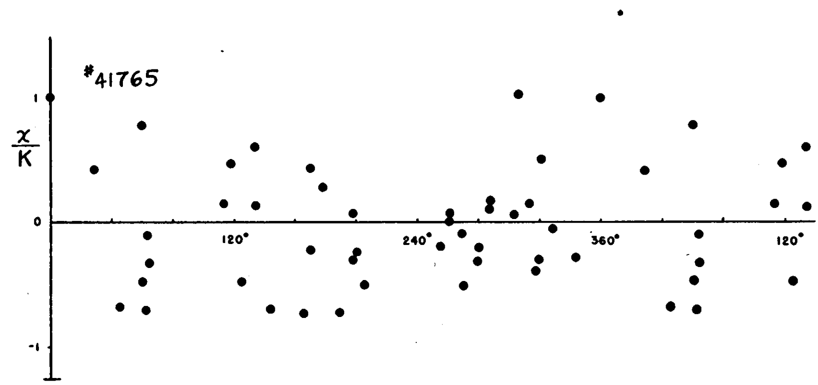

FIGURE 8

A case where it was impossible (entered in table II as "Can't") to make any estimate of the orbital elements.

True elements are: $e = .9$, $\omega = 45°$, $K/\sigma = 2$, $\xi_0/K = -0.14$.

identifying that the binary is variable: the larger the eccentricity and, also, the closer ω is to 0°, the more difficult it is to identify the binary.

The dependence of the frequency distribution of f_1 on the true eccentricity e is strongly reflected in table II. On the other hand, with the possible exception of $e = .3$, there is no noticeable dependence on the value of ω. When a test was applied, no definite evidence of dependence on ω was found. For this reason, further study of the distribution of f_1 was based on the presumption that it is independent of ω. Consequently, further computations were based on the figures obtained by adding together all of the columns of table II which refer to the same value of e. These figures are shown in table III in the columns marked "Observed frequency."

As far as direct information about how the variability of the estimate f_1 depends upon the true eccentricity e is concerned, the figures just described are the real evidence obtained. Inspection of these figures shows that, especially when the value of e is large, the estimate f_1 has surprisingly large dispersion. For example, when $e = .7$, in 3 cases out of 73 the value of f_1 was zero. The empirical distributions of f_1 change smoothly as e passes through successive values; this is in spite of a certain bumpiness in the particular columns. This suggests the possibility of using just one formula, depending on e, to represent approximately the distribution of f_1. This formula is based on the following remarks.

TABLE III

COMPARISON OF FREQUENCY OBSERVED IN SAMPLING EXPERIMENT AND COMPUTED FROM APPROXIMATING FORMULA (7)

e	0		.1		.3		.7		.9	
f_1	Obs. Freq.	Comp. Freq.	Obs. Freq.	Comp. Freq.	Obs. Freq.	Comp. Freq.	Obs. Freq.	Comp. Freq.	Obs. Freq.	Comp. Freq.
0	15	20.2	34	29.1	24	18.4	3	1.3	0	0.0
.1	14	6.2	35	40.0	34	44.5	0	5.3	0	0.3
.2	1	2.2	12	15.3	41	36.2	7	7.2	2	0.9
.3		0.9	4	4.5	23	24.5	12	8.4	2	1.6
.4		0.4	4	1.0	16	14.5	3	9.0	0	2.5
.5			1	0.2	6	7.4	11	9.3	2	3.1
.6					3	3.2	18	9.2	4	4.0
.7					3	1.0	17	8.6	5	4.8
.8						0.2	1	7.6	5	5.4
.9							1	5.7	11	5.7
1.0								1.4		2.7
Can't							17		59	
Total	30		90		150		90		90	

Each of the empirical distributions of f_1 obtained has but one maximum and extends, at least in principle, from zero to unity. This suggests that a reasonable approximation to the observed distribution can be obtained by using the Pearson type I distribution, say,

$$p(f_1) = \frac{f_1^{m_1-1}(1-f_1)^{m_2-1}}{B(m_1, m_2)}. \tag{7}$$

The two parameters, m_1 and m_2, are connected with the expectations of the first two powers of f_1 by the following easy formulae

$$\begin{cases} E(f_1) = \dfrac{m_1}{m_1+m_2}, \\ E(f_1^2) = \dfrac{m_1+1}{m_1+m_2+1} E(f_1). \end{cases} \tag{8}$$

Thus, if we knew the expectations of f_1 and of f_1^2, for any given value of e, the corresponding values of the parameters m_1 and m_2 would be easy to find.

In order to study the dependence of $E(f_1)$ and $E(f_1^2)$ on e, the average value of f_1 and the average value of f_1^2 were computed for each e. It was found that the dependence on e can be represented by the following formulae

$$E\,(f_1 | e,\, \omega,\, K/\sigma = 2,\, N = 40) = 0.0567474 + 0.422982\, e + 0.297820\, e^2\,,$$

$$E\,(f_1^2 | e,\, \omega,\, K/\sigma = 2,\, N = 40) = 0.010885 + 0.611834\, e^2\,.$$

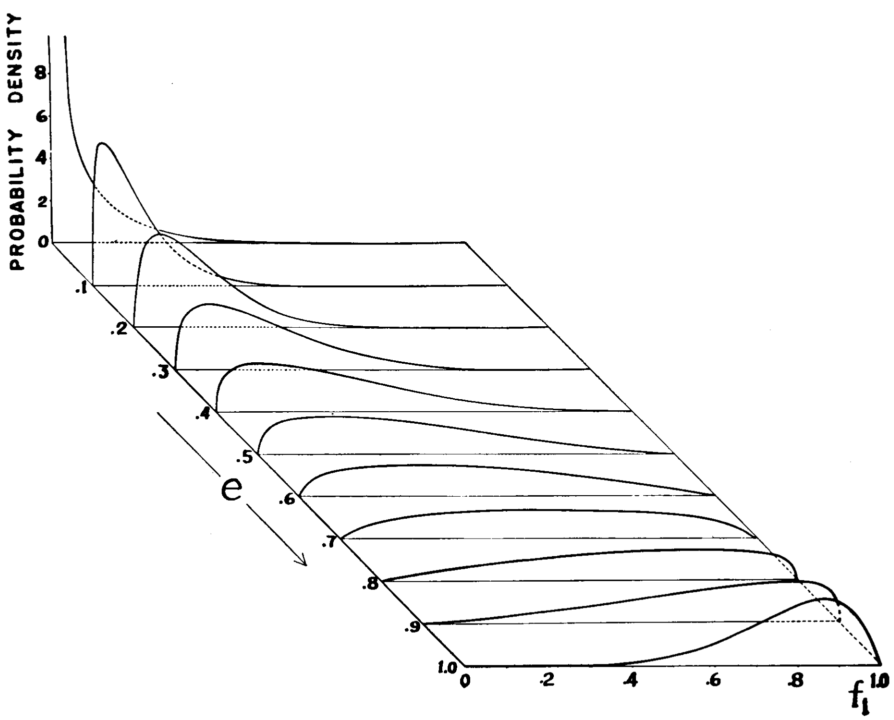

FIGURE 9

Approximation to the probability density of the estimated eccentricity f_1 as a function of the true eccentricity when $K/\sigma = 2$.

By substituting the values of these formulae into (8), a set of m_1 and m_2 was computed. The resulting curves are shown in figure 9, which gives the approximation to the probability density of the estimated eccentricity f_1 when the true eccentricity e is 0, .1, .2, . . . , 1.0. Figure 9, then, may be interpreted as representing approximately the dependence of the distribution of f_1 on the value of e. The goodness of the approximation may be judged from table III which, in addition to the observed frequencies of f_1, gives the frequencies computed using the above approximation. While, in certain cases, the agreement between the observed and expected figures in not very good, this should not be judged too severely because a very large number of frequencies was approximated at one time using only five parameters.

4. Remarks on the distribution of eccentricity

We have let $\beta(K/\sigma, e, \omega, n)$ denote the probability that a binary with specified elements is identified as variable, and then we let $\psi(P, K/\sigma, e, \omega, N, \text{var.})$ denote the probability that an estimate of the period is obtained. In addition, we let $\Phi(Q|P, K/\sigma, e, \omega, N, \text{var.})$ denote the relative probability density of the estimate Q of the period and $\Lambda(L, f, v, \eta_0|P, K/\sigma, e, \omega, \xi_0, N, \text{var.}, Q)$ denote the joint relative probability density of the estimates L, f, v and η_0 of the semiamplitude, the eccentricity, the longitude of periastron and the radial velocity of center of mass, respectively.

Now, if we denote the unknown distribution of the true elements by, say, $g(P, K, e, \omega, \xi_0)$ and the observable distribution of the catalogue elements by, say, $h(Q, L, f, v, \eta_0)$, we have the relation

$$h(Q, L, f, v, \eta_0) = \frac{\int \cdots \int g\beta\psi\Phi\Lambda \, dP\, dK\, de\, d\omega\, d\xi_0}{\int \cdots \int g\beta\psi \, dP\, dK\, de\, d\omega\, d\xi_0}, \tag{9}$$

where the range of integration is over the extreme limits of each variable. Equation (9) is concerned with the joint distribution of the five estimates Q, L, f, v and η_0. If all of the probability functions β, ψ, Φ and Λ were known then, since h is observable, the solution of the integral equation (9) would provide the unknown distribution g, that is, would rectify the catalogue distribution h. The denominator of the right hand side of (9) is a constant, and is the probability that a binary chosen at random will have estimates of its elements appearing in the catalogue. This probability may be used to rectify the total of the number of binaries listed in the catalogue, that is, to obtain an estimate of the true total number of binaries in the vicinity of the sun.

Integrating both sides of (9) with respect to any four of the five estimates, Q, L, f, v and η_0, provides the probability density of the remaining estimate. Thus, for example, the probability density of the estimate of the eccentricity is, say,

$$H(f) = \int_0^\infty \int_0^\infty \int_0^{2\pi} \int_{-\infty}^\infty h(Q, L, f, v, \eta_0)\, dQ\, dL\, dv\, d\eta_0 . \tag{10}$$

The preliminary results on the distribution of true eccentricity which it is possible to report now are based on the assumption that the true period is always found (so that $\psi = 1$ and $Q = P$) and refer to the case when $K/\sigma = 2$. Since there are several hundred stars which have been announced as spectroscopic binaries but for which no period has been found, it should be emphasized that these preliminary results are not to be taken as an estimate of the distribution of true eccentricity since they are still distorted by the unknown effects of failure to find a period.

If we assume also that the distribution of the true elements may be factored into, say,

$$g(P, K, e, \omega, \xi_0) = C_1(e)\, C_2(\omega)\, G_3(P, K, \xi_0), \tag{11}$$

with $G_2(\omega)$ representing the uniform distribution, then we may use the results of sections 2 and 3 to obtain a preliminary estimate of $G_1(e)$. Putting (9) and all the assumptions just made into (10) and integrating, we obtain for the distribution of the estimated eccentricity when $K/\sigma = 2$, say,

$$(12)\quad H\left(f_1 \,\middle|\, \frac{K}{\sigma} = 2,\, n,\, N\right) = \frac{\int_0^1 G_1(e) \int_0^{2\pi} \beta\left(\frac{K}{\sigma} = 2,\, e,\, \omega,\, n\right) \Lambda_1\left(f_1 \,\middle|\, \frac{K}{\sigma} = 2,\, e,\, N\right) de\, d\omega}{\int_0^1 G_1(e) \int_0^{2\pi} \beta\left(\frac{K}{\sigma} = 2,\, e,\, \omega,\, n\right) de\, d\omega}.$$

We have made use of the fact that the relative probability density Λ_1 of f_1 is independent of P, of ω and of ξ_0.

Values of $\beta(K/\sigma = 2, e, \omega, n)$ may be read from figure 7 or table I while $\Lambda_1(f_1 | K/\sigma = 2, e, N = 40)$ is illustrated in figure 9 and may be computed from table III. Strictly speaking, $H(f_1 | K/\sigma = 2, n, N)$ is not an observable distribution because K is an unknown. However, a study of the observed distribution of the catalogue eccentricity of single line binaries of spectral types O, B and A (for which σ is large so that the case $K/\sigma = 2$ is of interest)[1] revealed that the distribution of f for L = 5–15, 15–25, 25–35, . . . km/sec do not differ essentially. But the results of the sampling experiment described in section 3 indicate that $L - K$, the error in estimating K, is almost always nonnegative for small values of e and nonpositive for large values of e. When $K/\sigma = 2$, this effect is small although noticeable. On the other hand, for $K/\sigma \geqq 4$, the error $L - K$ is negligible. For this reason, and because the number of stars listed in the catalogue is not large, we shall combine the figures pertaining to all values of L for early type, single line binaries in order to obtain the observed distribution of f for $K/\sigma = 2$. The solid line in figure 11 shows this observed distribution of f.

Reasonable values for n and N seem to be $n = 10$ and $N = 40$ and we shall adopt these in order to obtain some preliminary numerical results.

Figure 10 shows a number of possible solutions for $G_1(e | K/\sigma = 2)$, the probability density of the true eccentricity under the assumptions we have made. Figure 11 shows the resulting distribution of f_1, the estimated eccentricity assuming the true period known. The computed values of $H(f_1)$ were obtained by substituting the numerical values of the corresponding $G_1(e)$ into (12) and performing the integrations numerically. Each of the trial functions $G_1(e)$ shown in figure 10 produces a satisfactory fit of the observed distribution. Remembering that the observed distribution is poorly determined, especially for medium and large values of the observed eccentricity, we cannot claim that any one of the suggested solutions is "better" than the others. Some other probability densities would fit as well or better; the particular trial densities illustrated were chosen because of their distressing diversity.

We conclude that, even assuming there is no error in the period, we are not able to rectify the catalogue distribution of eccentricity convincingly when $K/\sigma = 2$.

[1] A study of the relation between the standard error σ and the rotational velocity of a star, now in progress in cooperation with H. F. Weaver, indicates that with low dispersion spectrograms a typical value of σ for B stars is 10 km/sec.

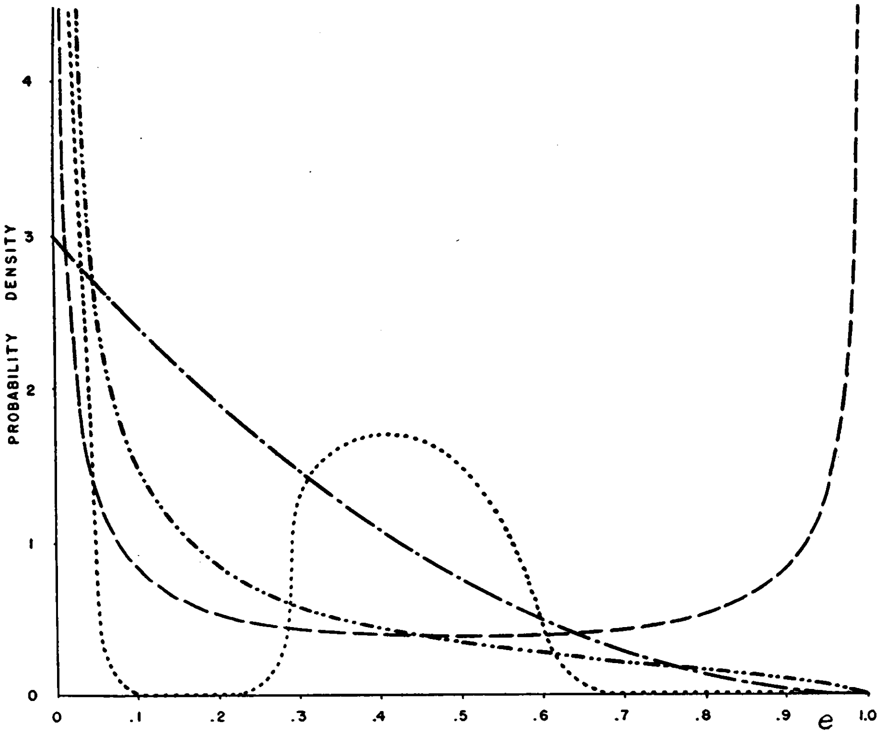

FIGURE 10

Possible solutions for the probability density of the true eccentricity e when $K/\sigma = 2$

$3(1 - e)^2$	$0.286\, e^{-.75}\, (1 - e)^{.5}$
$0.135\, e^{-.75}\, (1 - e)^{-.75}$	repeated guessing

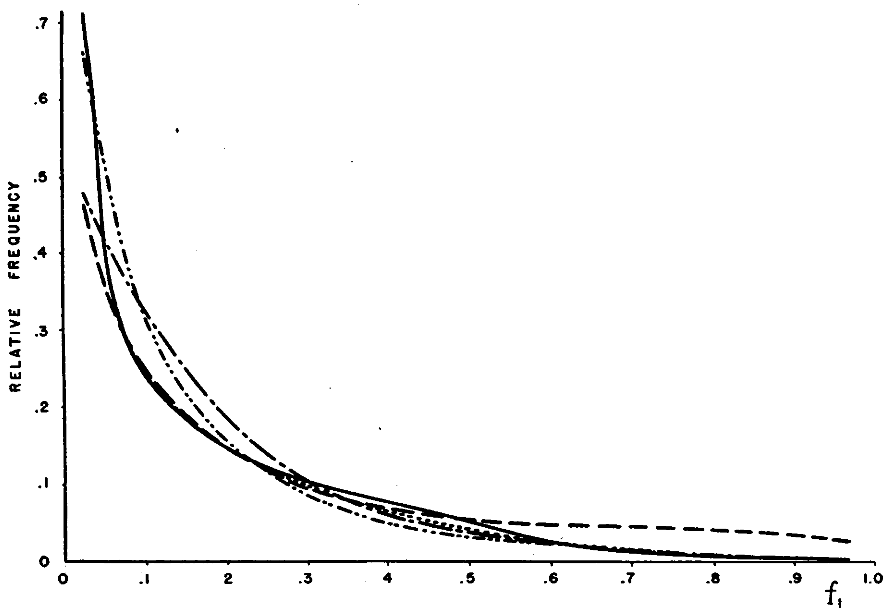

FIGURE 11

Comparison of observed distribution of catalogue eccentricity and computed distributions using the trial probability densities of e shown in figure 10.

observed relative frequency

The tremendous range in the possible solutions for $G_1(e|K/\sigma = 2)$ is due to the large dispersion in the distribution Λ_1 of the estimate f_1. It is anticipated that this range will be substantially decreased when larger values of K/σ are considered. If so, the solutions for $G_1(e|K/\sigma = 4)$ and for $G_1(e|K/\sigma = 6)$ may be of aid in obtaining $G_1(e|K/\sigma = 2)$.

That it is of some importance to determine which of the solutions for $G_1(e|K/\sigma = 2)$ is the correct rectification of the distribution of catalogue eccentricity is shown by the fact that the estimates of the proportion of binaries catalogued for $K/\sigma = 2$, assuming that the true period is always determined, ranges from .43 to .86!

5. Acknowledgments

The author is deeply indebted to Professor J. Neyman for aid and encouragement in the study of this problem. Hearty thanks are due Miss Jeanne Linton and Mrs. Eloise Putz for help in the computations and Mr. C. L. Chiang for drawing the figures.

REFERENCES

[1] O. STRUVE, *Stellar Evolution*, Princeton University Press, Princeton, 1950.

[2] E. L. SCOTT, "Distribution of the longitude of periastron of spectroscopic binaries," *Astrophysical Jour.*, Vol. 109 (1949), pp. 194–207.

[3] ———, "Further note on the distribution of the longitude of periastron," *Astrophysical Jour.*, Vol. 109 (1949), pp. 446–451.

[4] ———, "Contribution to the problem of selective identifiability of spectroscopic binaries," *Univ. of California Publ. in Stat.*, to be published.

[5] FRANK SCHLESINGER, "A criterion for spectroscopic binaries, with an application to ρ Leonis," *Astrophysical Jour.*, Vol. 41 (1915), pp. 162–167.

[6] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London*, Vol. 231-A (1933), pp. 289–337.

[7] R. J. TRUMPLER, "Statistical problems of spectroscopic binaries," abstract in *Annals of Math. Stat.*, Vol. 18 (1947), p. 462.

[8] E. L. SCOTT, "A note on the problem of binary stars," abstract in *Annals of Math. Stat.*, Vol. 18 (1947), p. 461.

[9] L. H. C. TIPPETT, "Random sampling numbers," *Tracts for Computers*, No. 15, Cambridge University Press, Cambridge, 1927 and 1950.

[10] P. C. MAHALANOBIS, "Tables of random samples from a normal population," *Sankhyā*, Vol. 1 (1933), pp. 289–328.