STATISTICS OF IMAGES OF GALAXIES WITH PARTICULAR REFERENCE TO CLUSTERING

JERZY NEYMAN AND ELIZABETH L. SCOTT
STATISTICAL LABORATORY

AND

C. D. SHANE
LICK OBSERVATORY
UNIVERSITY OF CALIFORNIA

CONTENTS

GENERAL INTRODUCTION

The paper summarizes the results obtained by the authors in a cooperative study extending over several years and outlines a program of future work. Theoretical results which are not yet published include (i) extension of the model of simple clustering of galaxies, originally developed for a static universe, to cover the possibility that the universe is expanding (sections 6 and 8), (ii) a model of multiple clustering (section 11), and (iii) formulas relating to the distributions of certain characteristics of images of clusters on the photographic plate (section 12). New empirical results, given in section 9, indicate that, probably, the model of simple clustering of galaxies in a stationary universe does not correspond to reality.

While the title of this paper is "Statistics of Images of Galaxies" and while this

Prepared with the partial support of the Office of Naval Research.

title corresponds exactly to the contents of the paper, it must be realized that the ultimate aim of the study is to bring out plausible conclusions about the distribution of galaxies in space. The distribution of images of galaxies on a photographic plate is no more than the adopted means leading to a more distant goal. In these circumstances it is natural that the study fall under three separate headings.

First there is the observational aspect of the problem, consisting in the accumulation of facts. This includes not only the taking of the photographs of the sky, the identifying and counting of images of galaxies but also the study of the various physical factors, other than the distribution of galaxies in space, which influence the appearance of the distribution of images of galaxies on the photographic plate. This observational aspect of the study is wholly within the domain of activities of an astronomer.

The second aspect of the problem is that of a stochastic model of the spatial distribution of galaxies and of the theory of that model. To construct a stochastic model means to build a plausible system of mutually consistent hypotheses, or of postulates, regarding the distribution of galaxies in space and here a close cooperation between astronomers and statisticians is most essential. The theory of the model is the responsibility of the statistician and includes not only the problem of mutual consistency of the postulates but also the deduction of their various consequences, such as the formulas connecting the properties of the distribution of images of galaxies on a photographic plate with what was postulated to be happening in space.

Finally, there is the third aspect of the general problem. It consists in establishing a bridge between the general theoretical formulas deduced to characterize the distribution of images of galaxies on a photographic plate and the actual counts of these images. Here again, in order to attain success, a close cooperation between astronomers and statisticians is necessary. The questions falling under this heading may be exemplified by the following.

In order to deduce the relation between the postulated distribution of galaxies in space and the corresponding distribution of images, the statistician cannot avoid dealing with a probability $\theta(\xi)$ defined as follows. Consider a galaxy that emitted a light signal that reaches the photographic plate at the time when the photograph of the sky is taken. Denote by τ the moment at which this light signal was emitted and by ξ the distance which at that same moment τ separated the galaxy from the observer. Then $\theta(\xi)$ represents the probability that the photograph will contain a noticeable image of the galaxy concerned.

In his work the statistician need not know the exact form of the function $\theta(\xi)$ representing the probability just defined. However, in this case, the statistician's results will contain the symbol $\theta(\xi)$ and, so long as nothing specific is substituted for this symbol, his formulas will not be usable for numerical calculations. In other words, these formulas will not be ready for a comparison with the observational material. Thus, in order to be able to verify whether or not a given stochastic model is consistent with the body of empirical data, it is necessary to substitute for $\theta(\xi)$ some specifically defined function and here the advice of an astronomer, perhaps the advice of an expert group of astronomers and physicists, is most essential. Unfortunately, as things stand now, even the most expert group is likely to give only a very tentative answer regarding $\theta(\xi)$. The point is that, as is well known, a change in the distance ξ influences not only the intensity of light received from a galaxy but also the distribution of the total energy over the spectrum. The larger the value of ξ, the greater proportion of the total energy of light appears concentrated in the red section of the spectrum (red-shift). Since the photographic plate does not respond equally to light of different wave lengths, the phenomenon of reddening of distant galaxies must have an effect on the chances of their being recorded on a photographic plate and, therefore, must be reflected in the formula selected to represent $\theta(\xi)$. However, as will be seen below, this is not all there is to the problem of selecting a realistic function to represent $\theta(\xi)$.

Out of the above three aspects of the general problem considered, the first differs from the other two by the fact that it is entirely beyond the sphere of competence of a statistician. For this reason, the present paper is split into two parts, the observational and the statistical.

PART I. OBSERVATIONAL STUDIES OF THE DISTRIBUTION OF IMAGES OF GALAXIES

1. Introductory remarks

Although a number of surveys of galaxies have been made and have led to important conclusions, the totality of this material leaves much to be desired if the capabilities of existing instruments and the resources of modern statistical methods are to be developed adequately. We shall mention here the principal counts of images of galaxies, either completed or in progress, that are currently useful for distribution studies. Those made for special or limited purposes are omitted. Following this discussion we shall outline the requirements for counts that would seem to meet most of the present needs and at the same time represent an attainable goal.

2. Existing surveys of galaxies

The surveys of current importance may be listed under the names of the investigators as follows:

(1) Harlow Shapley and Adelaide Ames [1] have published a catalogue of 1249 external galaxies brighter than magnitude 13.4. In addition to the equatorial and galactic coordinates of each galaxy the catalogue contains its photographic magnitude, its maximum and minimum diameters and its classification. The list is probably complete to about magnitude 12.6. The systematic errors in magnitude are satisfyingly small, though the individual magnitudes are subject to the almost inevitable large errors arising from photographic methods as applied to objects of this kind. For studies of the distribution of the brighter galaxies this catalogue is by far the most valuable source of data.

(2) Harlow Shapley and some of his associates [2] have published counts of galaxies on plates taken with the Bruce telescope at Bloemfontein and with the Metcalf telescope at Harvard. The limiting magnitude is variable because of the nonuniform conditions under which the plates were obtained, but in general it is near magnitude 18. As one result of these investigations it was demonstrated that clustering is a general characteristic among galaxies. The counts on the Bruce plates supply practically our only information concerning the distribution of the fainter galaxies in high southern declinations. The results of these surveys are not very

well adapted to statistical study because the plates were taken under nonuniform conditions, they are not systematically spaced over the sky, and the images at the greater distances from the optical axis are in general poor.

(3) E. P. Hubble [3] published in 1934 the first carefully planned general survey for statistical purposes. It consisted of counts of galaxies on 1283 plates north of $-30°$ declination taken with the 60-inch and the 100-inch reflectors on Mount Wilson. 765 of the plates were spaced systematically over the sky. The remaining plates were taken for other purposes but they were used in the discussion. The exposure times were not all the same, and this circumstance combined with the two different telescope apertures and other varying conditions required the application of rather large corrections to reduce the counts to the adopted standard. This standard was the limiting magnitude at the central region of a 100″ plate on an Eastman 40 emulsion for a one hour exposure under good conditions of seeing and transparency. The limiting magnitude under these conditions was determined by Hubble to be 20.0. Subsequently [4] these data, together with the counts on additional plates of long exposure with 100-inch telescope, and counts by Mayall [5] with the Crossley reflector of 36-inch aperture, were used by Hubble to determine the numbers of galaxies per square degree to different limiting magnitudes ranging from 18.5 to 21.0. The circumstance that the surveys give only the numbers of galaxies in small disconnected areas of the sky permits their being used only for studies of the largest features of distribution. The actual counts on the plates were made with great care so that it would be difficult to improve on their accuracy. A large amount of careful labor also went into the determination of the limiting magnitudes. But it is now recognized that the stellar magnitude standards available at the time as well as the photographic methods employed for comparing stellar and nebular images introduced substantial errors in the limiting magnitudes. Moreover the large corrections applied to compensate for effects of coma, seeing, etc., materially reduce the value of the observations for some statistical purposes.

(4) Nicholas U. Mayall [5] investigated the distribution of galaxies on 489 plates taken with the Crossley reflector of the Lick Observatory. Of these plates 305 were obtained by earlier observers and 184 by Mayall to fill in gaps and to provide proper standardization. The methods of counting and the general treatment of the material are similar to those used by Hubble. The results of the investigation were in conformity with the general conclusions of Hubble and they provided an additional datum for his relation between numbers of galaxies and apparent magnitudes.

(5) Anders Reiz [6] measured photographically 4666 galaxies within 40° of the north galactic pole. The data are almost entirely based on plates taken with the Bruce telescope at Heidelberg. The magnitudes estimated by direct comparison with stars reach to approximately 16.5 on the faint side but the catalogue is not complete beyond magnitude 14.5. The magnitudes suffer from the rather large errors inherent in the adopted method of measurement. Unfortunately there are some small areas totaling about 4 per cent of the region studied that are not covered by the catalogue. These omissions are not important for the type of analysis carried out by the author but for some other statistical purposes they constitute a serious defect.

(6) C. D. Shane and C. A. Wirtanen [7] are currently conducting a survey to the 18th magnitude based on astrometric plates taken with the Carnegie 20-inch astro-

graph of the Lick Observatory. The survey covers the sky from the north celestial pole to 23° south declination. The plates are systematically spaced and overlap at least one degree on each side. Galaxies are counted in areas 10′ square to the plate limit of about magnitude 18.4. The plate overlaps permit the determination of corrections for the most important types of systematic error. The results for an area comprising one-ninth of the total region have been published. Counts on fewer than one half of the plates still remain to be made.

To supplement this 18th magnitude survey Shane and Wirtanen have started a 15th magnitude survey with a 5-inch Ross lens used on a photographic telescope loaned to the Lick Observatory by the Mount Wilson and Palomar Observatories. As with the 18th magnitude survey the plates are taken with adequate overlaps and on as nearly a uniform basis as possible.

(7) Fritz Zwicky [8] has under way a project to determine from images obtained by schraffier methods with the 18-inch Mount Palomar Schmidt telescope the magnitudes of all galaxies to the 15th. When completed the resulting some 15,000 objects should provide an exceptionally complete catalogue of individual galaxies. It is also understood that a catalogue of about 2,000 galaxies is in preparation by K. Lundmark of the Lund Observatory.

The surveys of galaxies described here provide only a very incomplete answer to the statistical requirements. The means are now at hand, however, to obtain observations rather ideally suited to our needs.

3. Suggested programs of surveys

A very adequate answer to the observational problem would be provided by a program along the following lines.

(1) A catalogue of galaxies complete to about the 14th magnitude in which are listed the types as well as the photoelectric colors in two wave lengths would be a fundamental requirement. A difficulty lies in defining "magnitude" especially for elliptical galaxies since according to current observations it has not yet been possible to determine a limiting brightness for these objects with increasing size of the area measured. William B. Baum of the Mount Wilson and Palomar Observatories is now studying this difficulty. If it can be overcome a catalogue of the approximately 4,000 objects to magnitude 14 could be prepared without prohibitive labor.

(2) A continuous survey over the sky to about the 18th magnitude should provide the material for studying the small scale irregularities in distribution. The counts should be available by squares 10′ or 20′ on a side, the plates should overlap and should be taken on as uniform a system as possible. It is doubtful if the amount of work necessary to make continuous counts over the whole sky to a magnitude fainter than 18.5 would be justified. The limiting magnitudes for different types of galaxies should be carefully determined by comparison with the magnitudes of a list of galaxies measured photoelectrically.

(3) There are good reasons for extending a continuous survey over a limited area to magnitudes fainter than 18.5 in order to increase our understanding of the phenomenon of clustering. It would be well worthwhile to count galaxies in 10′ squares to magnitude 19.5 over an area of perhaps 1500 square degrees around the north galactic pole. The plates already taken with the 48-inch Schmidt telescope at Mount Palomar might be used in this connection. Zwicky [8] has counted galaxies on these

plates but to the best of our knowledge the counts are for special purposes and do not yield a continuous survey.

(4) For extended studies of distribution in depth and possible large-scale inhomogeneities in distribution the fainter galaxies should be studied by sample surveys following the general procedures of Hubble. Perhaps two such surveys using the same sample areas to magnitudes 20.5 and 22.5, or fainter if possible, would be best. The current work at Mount Palomar on magnitude scales to faint limits should form the basis for accurately determining the limiting magnitudes.

(5) The color system of the counts should have its wave length range in the region 6000–7000A. This longer wave length compared to the blue region used in existing surveys permits better penetration of obscuring matter in our own galaxy and in view of the observed reddening of distant galaxies it increases greatly the observable volume of space. But far more important it improves the accuracy of the magnitude corrections for red-shift and reduces their size as follows: First, for blue magnitudes and large red-shifts the correction depends on the very poorly known or even conjectured energy distribution for nearby galaxies in the ultraviolet. For red magnitudes the better known blue, green and yellow portions of the spectrum are used, thus yielding more reliable values of the corrections. Second, when blue magnitudes are used the effect of red-shift is calculated along the steep portion of the spectral energy curve with resulting large corrections. The red magnitudes on the other hand use the curve along the flatter portion near its maximum and thus require much smaller corrections. It was a part of Hubble's [9] plan for future work to use red magnitudes as suggested here. This procedure would enormously enhance the value of the surveys.

The suggested program may be summarized as follows:

(1) A two color catalogue to the 14th magnitude.

(2) Complete counts in the red to magnitude 18.5.

(3) A continuous partial survey in the red to magnitude 19.5.

(4) Sample surveys in the red to two fainter magnitudes, of which one should be near the limit of the largest modern telescopes.

If the surveys are to be used to the best advantage in statistical investigations, a knowledge of the errors is necessary. It is important, therefore, that with each survey there be a special study of the errors determined by such methods as the counting of duplicate plates and counts of the same plates by all of the observers participating in the program.

The amount of labor entailed in this program is not prohibitive with reasonable cooperation between a few observatories, and for a statistical investigation of the structure of the visible universe the results would be invaluable.

PART II. PROBABILISTIC STUDIES OF THE DISTRIBUTION OF IMAGES OF GALAXIES

4. General remarks

The probabilistic studies of the spatial distribution of galaxies conducted thus far by the authors refer to two different stochastic models. Most of the work done concerns the original model ([10] through [17]) in which the over-all distribution of galaxies is visualized as an agglomeration of clusters, the centers of which are distributed in space uniformly or quasi-uniformly. Thus, this particular model may be labeled the model of simple clustering. The same subject was recently treated [18] by McVittie from the point of view of relativity theory. In addition, however, some further work was done on a more general model which may be labeled the model of fluctuations [33], [34]. Finally, some recent results refer to yet another model, the model of multiple clustering.

At this moment it may be useful to mention an important difference between the theoretical approach of the present studies and that of the great majority of others. In the latter, as exemplified by the important work of Hubble, it is frequently recognized that the numbers of images of galaxies in squares on a photographic plate, and also the numbers of galaxies located in unit volumes in space are subject to some sort of random variation. However, no clear-cut statement is given to explain the postulated properties of this variation. Instead, scattered throughout the study, there are assumptions regarding the methods of eliminating the undesirable effects of the random variation on the various averages that are being computed.

There are several disadvantages in this approach. The main one is the lack of clarity: after a study of a paper it is difficult, if not impossible, to visualize the details of the spatial distribution of galaxies implied by the hypotheses adopted. As a consequence of this lack of clarity, it is difficult to avoid inconsistencies in the results.

For the above reasons, the present authors adopted the policy of collecting together all the basic assumptions and of formulating them at the outset of the study into several postulates with as much clarity and precision as practicable. By this device, it may be found easier to test each of the postulates separately. Because of the multiplicity of circumstances that have to be taken under consideration, it is convenient to classify the postulates into three groups: those relating to the distribution of galaxies in space, those establishing the connection between what happens in space and what may be seen on a photographic plate and those relating to what is briefly called "errors in counting." Actually, this last term includes all factors which cause differences between independently made counts on two different plates taken of the same area of the sky and intended to be to the same limiting magnitude.

5. Postulates underlying the stochastic model of simple clustering of galaxies

The first stochastic model considered by the authors [10] is no more than the result of an effort to put into exact terms the ideas about the spatial distribution of galaxies that for some time have been more or less generally adopted in the astronomical literature. These ideas may be exemplified by the following quotation from a paper by Hubble [4] published in 1936. This quotation was selected because, as is well known, originally Hubble rather favored the assumption that the galaxies are uniformly distributed in space.

"While the large-scale distribution (of galaxies) appears to be essentially uniform, the small-scale distribution is very appreciably influenced by the well-known tendency toward clustering. The phenomena might be roughly represented by an originally uniform distribution from which nebulae have tended to gather about various points until now they are found in all stages from random scattering, through groups of various sizes, up to occasional great clusters . . . .

"This representation is purely formal and has no genetic implications. For purposes of speculation, the reverse development seems preferable, namely, that nebulae were formed in great clusters whose gradual evaporation has populated the general field."

In accordance with the above purely qualitative suggestions, the postulates of the theory of simple clustering of galaxies are as follows, assuming ordinary Euclidean space and Newtonian mechanics.

(i) Galaxies occur only in clusters.

(ii) To every cluster there corresponds a random variable ν which represents the number of galaxies belonging to this cluster. The variables ν corresponding to different clusters are mutually independent and have the same distribution characterized by a probability generating function $G_\nu(t)$. The general model does not involve any limitation on the distribution of ν except, of course, that it is capable of assuming only positive integer values $\nu = 1, 2, \cdots$. The variables ν are completely independent from all other variables considered in the model.

(iii) The distribution of positions of galaxies within a cluster is random and is subject to a probabilistic law which is the same for all clusters. More precisely, given a moment T in time and given the coordinates $u_1(T)$, $u_2(T)$, $u_3(T)$ of the position occupied at that time by a cluster center, there exists a function $f^*(\eta, T)$ which represents the conditional probability density of the coordinates $X_1(T)$, $X_2(T)$, $X_3(T)$ of the galaxy. For a fixed T this probability density is postulated to be a continuous function of the distance η between any given point x_1, x_2, x_3 and the center of the cluster. This function f^* may depend on T. Also it is postulated that, given T and the coordinates of all the clusters at that time, a triplet of random coordinates of one galaxy is stochastically independent from that of any other galaxy, and is independent of any other variable considered in the model.

(iv) At each moment T, to every region R in the space (really: to every Borel set) there corresponds a random variable $\gamma(R, T)$ which represents the number of cluster centers which, at time T, are situated in R. The distribution of $\gamma(R, T)$ depends only on the volume (really: measure) of R. If $R_1, R_2, \cdots, R_s, \cdots$ is a sequence of disjoint regions, then the variables $\gamma(R_i, T)$, corresponding to the same moment T, are completely independent. Let R be a region of finite volume and R^* a measurable subset of R. It is postulated that, given that R contains exactly $n > 1$ cluster centers $C_1, C_2, \cdots, C_n$, the probability that R^* will contain some specified $m < n$ of these clusters, say $C_{a_1}, C_{a_2}, \cdots, C_{a_m}$, depends only on m, n and the measures of R and R^* but not on the particular numbers $a_1, a_2, \cdots, a_m$.

The above four postulates determine the general structure of the stochastic model of the spatial distribution of galaxies. So long as the probability generating functions $G_\nu(t)$ of ν, and that of $\gamma(R, T)$ remain unspecified, as well as the probability density $f^*(\eta, T)$, we shall speak of the general model of simple clustering. Here the word "simple" indicates the exclusion of clustering of clusters. It is easy to see how the model of simple clustering may be generalized allowing for clusters of second, third, etc. orders (see section 11). Whenever any one of the three functions postulated is ascribed a specific form, perhaps with some adjustable parameters, then the resulting model will be called a specific model [11].

One of the early interesting results is that, so long as $G_\nu(t)$ remains unspecified, no generality of the model is lost by the assumption that the variables $\gamma(R, T)$ are Poisson distributed,

(1) $$G_\gamma(t|T,\ V(R)) = e^{-\lambda(T)V(R)(1-t)}$$

where $\lambda(T)$ stands for the average number of the cluster centers per unit volume at time T and $V(R)$ for the measure of the region R to which the variable γ refers.

All the earlier papers [10] through [14] contemplated a static universe. Thus, the symbol T referring to a particular moment in time did not appear in the formulas.

6. Generalization to a possibly expanding universe

The originally static model was recently generalized [15], [16], [17] to allow for the possibility of expansion of the universe. This generalization required an assumption regarding the motions of cluster centers and also another assumption regarding the motions of galaxies within a cluster. These assumptions were formulated in accordance with the "cosmological principle" [19] which excludes the possibility of any sort of "privileged" position in the universe. In application to cluster centers, this principle asserts that at every moment T and for every pair C_1 and C_2 of such centers, the velocity of C_2 with respect to C_1 is the same continuous function of the distance between C_1 and C_2. As is well known, for this to be true it is both necessary and sufficient that, with $u_i(T)$ representing the coordinate of a cluster center with respect to an orthogonal system of axes centered at one of them, $i = 1, 2, 3$,

(2) $$u_i(T) = u_i(0)H_1(T)\ , \qquad \frac{du_i}{dT} = u_i(T)h_1(T)$$

with

(3) $$H_1(T) = \exp\left\{\int_0^T h_1(t)dt\right\}.$$

Thus, a function $h_1(T)$ characterizes the process of recession of clusters. Similarly, if the cosmological principle is applied to the motions of galaxies within a cluster, then the coordinates $x_i(T)$ of a galaxy belonging to a cluster centered at $u_i(T)$, $i = 1, 2, 3$, must satisfy the condition

(4) $$x_i(T) - u_i(T) = [x_i(0) - u_i(0)]H_2(T)$$

or

(5) $$x_i(T) = x_i(0)H_2(T) + u_i(0)[H_1(T) - H_2(T)]$$

where

(6) $$H_2(T) = \exp\left\{\int_0^T h_2(t)dt\right\}$$

and the function $h_2(t)$ characterizes the phenomenon of the expansion of clusters. It will be noticed that, if $h_1(t) = h_2(t) \equiv 0$, then the universe is static. Thus, all formulas which relate to the generalized model, admitting the possibility of expansion of the universe in accordance with the assumptions just formulated, reduce to corresponding formulas of the static theory merely by the substitution of zeros for the two functions $h_1(t)$ and $h_2(t)$.

Formulas (1) through (6) imply that the motions of cluster centers and those of galaxies within clusters are deterministic. For any given moment T the randomness in the positions of cluster centers and of galaxies is reducible to the randomness at one standard point, for example, at $T = 0$. Since the origin from which time is measured is arbitrary, we may agree to place it at "the present," that is to say, at the moment of taking a photograph of the sky.

The development of the theory leading to the distribution of images on the photographic plate required the consideration of a particular system of coordinates Y_1, Y_2, Y_3, of a galaxy, labeled the "apparent" coordinates. These are defined as follows. Denote by X_1, X_2, X_3 the (random) coordinates of a galaxy referring to "the present," that is to say, to $T = 0$ and let $-\tau$ be the moment in the past at which the galaxy emitted the light signal that reaches the photographic plate at $T = 0$. Then the "apparent" coordinates Y_1, Y_2, Y_3 are the coördinates of the position occupied by the same galaxy at time $-\tau$. If u_1, u_2, u_3 are the "present" coordinates of a cluster center then it is found that the conditional probability density of the apparent coordinates Y_1, Y_2, Y_3 of a galaxy belonging to the corresponding cluster is given by, say,

$$f(y_1, y_2, y_3; u_1, u_2, u_3) = f^*(\eta)\,|J|\,, \tag{7}$$

where

$$\eta^2 = \sum_{i=1}^{3} \left(\frac{y_i - H_1 u_i}{H_2}\right)^2 \tag{8}$$

$$J = \frac{1}{H_2^3}\left\{1 + \frac{H_2'}{cH_2}\xi + \frac{H_1'H_2 - H_1H_2'}{H_2}\,\frac{u_1y_1 + u_2y_2 + u_3y_3}{\xi}\right\}. \tag{9}$$

Here the functions H_1 and H_2 and their derivatives H_1' and H_2' are evaluated at the point $-\xi/c$ where c is the velocity of light and

$$\xi = \{y_1^2 + y_2^2 + y_3^2\}^{\frac{1}{2}} \tag{10}$$

represents the "apparent" distance between the galaxy and the observer.

In the absence of theoretical considerations ascribing specific forms to the functions $h_1(t)$ and $h_2(t)$ one would probably consider the simplest hypothesis $h_1(t) \equiv h_2(t)$. If the Doppler-like shifts in the spectral lines of distant galaxies found by Hubble and Humason [20], [21] to be proportional to distance are interpreted as indicating velocities, then one might put $h_1(t) \equiv h_2(t) \equiv h =$ constant. Under these assumptions formula (8) reduces to

$$\eta^2 = \sum_{i=1}^{3} (y_i e^{g\xi} - u_i)^2 \tag{11}$$

and formula (9) to

$$J = e^{3g\xi}\{1 + g\xi\} \tag{12}$$

where, for the sake of brevity, we have put $g = h/c$.

7. Postulate regarding the "visibility" of a galaxy on a photographic plate

As mentioned in the Introduction, in order to connect the postulated properties of the spatial distribution of galaxies with what might be observed on a photographic plate it is necessary to introduce a special postulate. Such a postulate may or may not take into account the possible presence of galactic absorbing clouds tending to dim larger or smaller areas of the sky. Thus far, for the sake of simplicity, the authors have been working on the assumption that the available empirical data refer to regions in the sky somewhat affected by uniform galactic absorption but not affected by irregular absorbing clouds. In its simplest form, the postulate adopted regarding the "visibility" of a galaxy on a photographic plate is as follows.

Given an observational setup, with the limiting magnitude m_l for identifying and counting galaxies, and given that the apparent distance of a galaxy is ξ, there exists a probability $\theta(\xi, m_l)$ that a properly oriented photograph will contain an image of the galaxy concerned. Given the distances $\xi_1, \xi_2, \cdots, \xi_s$ of any s galaxies, their "visibilities" on plates having the same limiting magnitude m_l, are mutually independent.

For a given type of galaxies (spirals, elliptical, etc.) and for an "idealized" photographic plate, with strictly constant limiting magnitude, say m_l, over its whole area, the structure of the function $\theta(\xi, m_l)$ may be described as follows. Consider the absolute magnitude M of a galaxy concerned and treat it as a random variable with a probability density $p_M(x)$. Assume that the absolute magnitude M of one galaxy is independent from that of any other and from its position within a cluster. Let $m(\xi, M)$ be the photographic apparent magnitude of a galaxy at distance ξ whose absolute magnitude is M. Including the effect of the familiar phenomenon of reddening of distant galaxies, the relation between M, ξ and $m(\xi, M)$ may be written as

$$m(\xi, M) = M - 5 + 5\log_{10}\xi + \chi(\xi) \tag{13}$$

where $\chi(\xi)$ represents the effects of reddening and depends on several factors such as the spectral response of the plates and the type of galaxy considered. The exact nature of the function $\chi(\xi)$ is uncertain. However, it may be hoped that the series of studies initiated by Stebbins and Whitford will eventually lead to its precise determination. In any case, the assumptions made imply that $\theta(\xi, m_l)$ coincides with the probability that the absolute magnitude of a galaxy is less (brighter) than the limit

$$m_l + 5 - 5\log_{10}\xi - \chi(\xi)\,. \tag{14}$$

Thus

$$\theta(\xi, m_l) = \int_{-\infty}^{m_l+5-5\log_{10}\xi-\chi(\xi)} p_M(x)dx\,. \tag{15}$$

This applies to a single type of galaxies to which the density $p_M(x)$ and the reddening effect $\chi(\xi)$ refer. The corresponding quantity referring to a galaxy chosen at random would be represented by a weighted mean of the same quantities referring to all the several types of galaxies.

We should mention that the postulate that the absolute magnitudes of galaxies are independent of their positions within a cluster may be an open question. For example, Zwicky [22] has interpreted his counts of nebulae in the Coma cluster as showing that "The fainter nebulae are relatively more frequent as the distance from the center of the cluster increases." On the other hand, Mayall [23] has observed several bright nebulae which are more than 1°.5 distant from the center of the cluster and which have red-shifts within the range of those near the center.

In an effort to apply the general model of simple clustering [11], the authors used the suggestion of Hubble that the function $p_M(x)$ may be approximated by a normal probability density with a mean M_0 and a variance σ_M^2. An easy transformation then gives

$$\theta(\xi, m_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{L(\xi)} e^{-t^2/2} dt , \tag{16}$$

where

$$L(\xi) = (m_1 - M_0 + 5 - 5 \log_{10}\xi - 5.19 \times 10^{-9}\, \xi)/\sigma_M \tag{17}$$

and ξ is measured in parsecs. It will be seen that $\theta(\xi, m_1)$ depends not so much on m_1 as on the difference $m_1 - M_0$. Since the values of m_1 and M_0 are uncertain, the difference $m_1 - M_0$ was taken to be an adjustable parameter, as well as σ_M. A comparison of the consequences of the model with the data of the Lick Observatory gave good agreement for $1 \leqq \sigma_M \leqq 2$ combined with all the values[1] considered of $m_1 - M_0$, namely,

$$\begin{cases} 31.5 \leqq m_1 - M_0 \leqq 34.5 \\ 1 \leqq \sigma_M \leqq 2 . \end{cases} \tag{18}$$

8. Basic formulas of the theory of simple clustering of galaxies

We reproduce here two formulas which are basic in the theory of simple clustering of galaxies. First, consider s arbitrary disjoint regions in space, say $R_1, R_2, \cdots, R_s$, and denote by N_i the number of galaxies whose present positions lie in R_i, $i = 1, 2, \cdots, s$. The first of the two basic formulas gives the joint probability generating function of the random variables $N_1, N_2, \cdots, N_s$, namely,

$$\begin{aligned} & G_{N_1, N_2, \cdots, N_s}(t_1, t_2, \cdots, t_s) \\ & = \exp\left\{ -\lambda \int \left(1 - G_\nu \left[1 - \sum_{i=1}^{s} P_i(u)(1 - t_i)\right]\right) du \right\} \end{aligned} \tag{19}$$

where the single letter u stands for the triplet of coordinates of a cluster center u_1, u_2, u_3 and the single integral sign for the triple integral extending from $-\infty$ to $+\infty$ for each of the three coordinates u_1, u_2 and u_3. The symbol $P_i(u)$ represents the conditional probability that the present position of a galaxy will be in R_i given that it belongs to a cluster centered at $u = \{u_1, u_2, u_3\}$. Thus

$$P_i(u) = \iiint_{R_i} f^*(\eta) dx_1 dx_2 dx_3 \tag{20}$$

with

$$\eta^2 = \sum_{j=1}^{3} (x_i - u_i)^2 . \tag{21}$$

[1] The numerical values given in this paper are in terms of the distance scale employed up to 1952. If this scale is revised, as suggested by W. Baade, by multiplying the distance to all objects outside the Milky Way system by a factor k, with a corresponding correction in brightness, then the values of the parameters in this paper are automatically revised. Thus, in equation (17), the parameter $m_1 - M_0 + 5$ is increased by $5 \log_{10}k$ because M_0 is decreased (made brighter) by $5 \log_{10}k$. At the same time, σ_M is unchanged while the revised value of the coefficient of ξ is $5.19 \times 10^{-9}k^{-1}$. Instead of writing, for example, $m_1 - M_0 = 31.5$, we would write a new value equal to $31.5 + 5 \log_{10} k$.

Formula (19) will be used below in the summary of results concerned with the possible phenomenon of superclustering of galaxies and also in the comparison of the theory of simple clustering of galaxies with the theory of fluctuations.

The second fundamental formula of the theory of simple clustering refers to the distribution of images of galaxies on photographic plates. Denote by ω_1 and ω_2 two arbitrary regions, overlapping or not. Let $m_1 \leqq m_2$ be the limiting magnitudes of the photographs of these regions. Finally, denote by n_1 and n_2 the numbers of images of galaxies in ω_1 and ω_2 respectively. Then the joint probability generating function of n_1 and n_2 is given by

$$(22)\quad G_{n_1,n_2}(t_1, t_2) = \exp\left\{-\lambda\int\{1-G_\nu[1-p_1(u)(1-t_1)-p_2(u)(1-t_2)-p_3(u)(1-t_1t_2)]\}du\right\}$$

where $p_1(u)$, $p_2(u)$ and $p_3(u)$ represent, respectively, the conditional probabilities that a galaxy from a cluster centered at $u = \{u_1, u_2, u_3\}$ will be visible in ω_1 but not in ω_2, that it will be visible in ω_2 but not in ω_1 and that it will be visible in both ω_1 and ω_2. Thus, with obvious notation

$$(23)\quad p_1(u) = \int_{\omega_1-\omega_1\omega_2} \theta(\xi, m_1) f dy_1 dy_2 dy_3 + \int_{\omega_1\omega_2} \{\theta(\xi, m_1) - \theta(\xi, m_2)\} f dy_1 dy_2 dy_3$$

$$(24)\quad p_2(u) = \int_{\omega_2-\omega_1\omega_2} \theta(\xi, m_2) f dy_1 dy_2 dy_3$$

$$(25)\quad p_3(u) = \int_{\omega_1\omega_2} \theta(\xi, m_2) f dy_1 dy_2 dy_3$$

where f stands for the conditional probability density (7) of the apparent coordinates of a galaxy belonging to a cluster with its present position at $u = \{u_1, u_2, u_3\}$.

Even with the most simplifying assumptions regarding all the intervening functions, formula (22) is too complicated for direct evaluation of the probabilities of particular combinations of values of n_1 and n_2. However, the expressions for the moments of the joint distribution of n_1 and n_2 are easily derived from (22), naturally, under the assumption that the moments of ν are finite. In particular, we have

$$(26)\qquad \bar{n}_1 = E(n_1) = \lambda\bar{\nu}R_{100}\,, \qquad \sigma_{n_1}^2 = \lambda\bar{\nu}R_{100} + \lambda\bar{\nu}(C-1)R_{200}$$

where, in general,

$$(27)\qquad R_{kmn} = \int (p_1 + p_3)^k (p_2 + p_3)^m p_3^n du\,,$$

where $\bar{\nu}$ stands for the expectation of ν and $C = E(\nu^2)/\bar{\nu}$.

The covariance $\sigma_{n_1n_2}$ of the variables n_1 and n_2 is

$$(28)\qquad \sigma_{n_1n_2} = \lambda\bar{\nu}R_{001} + \lambda\bar{\nu}(C-1)R_{110}\,.$$

Two special cases of these formulas are of particular interest. One is when ω_1 and ω_2 are two disjoint equal and similarly oriented squares on the same photographic plate or on two such plates with the same limiting magnitude m. Then $\omega_1\omega_2$ is empty and $p_3(u) = R_{001} = 0$. In this case the quotient

$$Q = \frac{\sigma_{n_1 n_2}}{\sigma^2_{n_1} - \bar{n}_1} = \frac{R_{100}}{R_{200}}, \tag{29}$$

labeled the quasi correlation between n_1 and n_2, is independent of λ and of the moments of the variable ν. Also, using the counts of images of galaxies in equal squares ω_1, ω_2 separated by a fixed angular distance, say β, it is possible to obtain estimates $Q^*(\beta)$ for several values of β. These estimates can then be compared with the corresponding quotients of integrals R_{110} and R_{200} computed on tentative assumption regarding the function θ. A comparison of this kind gives, then, an indication whether or not the hypotheses regarding f and θ are at all tenable.

The other interesting particular case of the formulas for moments of n_1 and n_2 corresponds to the assumption that ω_1 and ω_2 coincide but $m_1 > m_2$. Here $\omega_1 - \omega_1\omega_2$ and $\omega_2 - \omega_1\omega_2$ are empty and thus $p_2(u) \equiv 0$. Also $R_{010} = R_{001}$. Here again λ and the moments of ν can be eliminated to obtain the quotient

$$Q_d = \frac{R_{110}}{\sqrt{R_{200}R_{020}}} = \frac{\sigma_{n_1 n_2} - \bar{n}_2}{\sqrt{(\sigma^2_{n_1} - \bar{n}_1)(\sigma^2_{n_2} - \bar{n}_2)}}, \tag{30}$$

to be termed the quasi correlation in depth. The quasi correlation in depth can be estimated empirically from counts of images of galaxies in the same solid angles ω but taken to varying limiting magnitudes. A comparison with the theoretical values of Q_d may then eliminate unsuccessful guesses regarding the function f. In particular, it is hoped that the comparisons of the empirical and theoretical values of (29) and (30) computed for varying sizes of ω will serve to choose between the two alternative hypotheses regarding the expansion of the universe. Some results of such empirical studies are already available and are reproduced in the next section.

9. Some numerical results

Numerical investigations involve many important details which are difficult to summarize in a paper like the present. For this reason only the briefest account is given here and the reader is referred to the original publications [11], [13] and [14].

The main numerical results concern the empirical quasi correlations (29) which were computed using Shane and Wirtanen's counts [7] first in $1° \times 1°$ and then in $10' \times 10'$ squares on the plates of the Lick Observatory. The first set of these quasi correlations, for the larger squares, were then compared with the theoretical quasi correlations computed on the specific assumptions that the universe is static and that the probability density

$$f^*(\eta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^3 e^{-\eta^2/2\sigma^2}, \tag{31}$$

where σ was treated as an adjustable parameter. The process of adjusting the value of σ is illustrated in figure 1. Here filled circles represent the empirical quasi correlations for successive values of $\beta = k°$, $k = 1, 2, 3, 4, 5$. The five curves represent the theoretical values computed as quotients of the integrals R_{110}/R_{200} using trial values of σ indicated in figure 1. Since the two integrals depend on $m_1 - M_0$ and σ_M, it was necessary to substitute numerical values for these parameters and the curves shown in figure 1 correspond to $m_1 - M_0 = 32.5$ and $\sigma_M = 0.85$ mag, which is Hubble's original estimate. The trial values of σ characterizing the particular curves are measured in units of 10^5 parsecs.

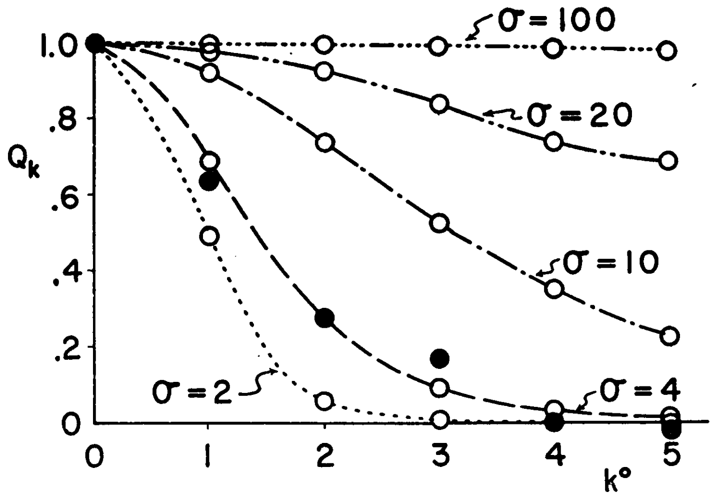

FIGURE 1

Process of fitting the value of σ.

It will be seen that $\sigma = 4 \times 10^5$ parsecs gives a very good fit and appears to be reasonably well determined. The difficulty is that the best fitting value of σ depends very much on the assumptions made regarding the values of $m_1 - M_0$ and of σ_M. The situation is illustrated in figure 2. We find that the empirical quasi correlations are fitted about equally well by assuming any one of the following combinations of the three parameters.

TABLE I

THE VALUES OF σ, FOR VARIOUS COMBINATIONS OF VALUES OF PARAMETERS $m_1 - M_0$ AND σ_M, WHICH PROVIDE EXCELLENT AGREEMENT WITH EMPIRICAL QUASI CORRELATIONS

$m_1 - M_0$	σ_M		
	0.85	1.25	2.00
31.5	2.8	3.7	7.5
32.5	4.2	5.4	10.0
33.5	6.4	7.5	14.0

As mentioned, all the above corresponds to counts in $1° \times 1°$ squares on the Lick Observatory plates.

If the theoretical model with the indicated specialization corresponds to reality, then a system of values of the three parameters $m_1 - M_0$, σ_M and σ must exist with which the theoretical quasi correlations would fit the empirical ones computed not only for $1° \times 1°$ squares but for squares of any other given size, for example for $10' \times 10'$ for which the counts are available. A conscientious search for such values of the three parameters was conducted in the Statistical Laboratory and figure 3

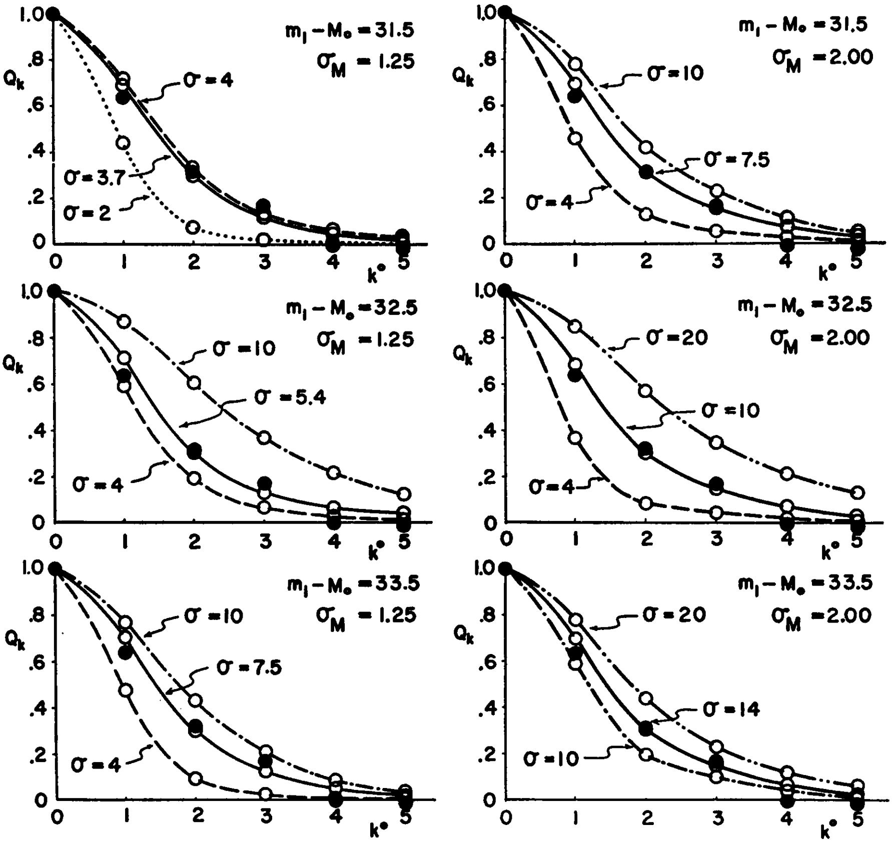

FIGURE 2

Best fitting theoretical quasi correlations.

exhibits a sample of the results obtained. It will be seen that the adopted value of $\sigma_M = 1.25$ is the same in each of the four panels of the figure. On the other hand, the panels differ in the trial values of $m_1 - M_0$. In each case the quantity measured on the axis of abscissae is the angular distance between the centers of squares. The quantity measured on the axis of ordinates is the quasi correlation. The filled circles represent the empirical quasi correlations computed for $1° \times 1°$ squares, and the crosses give the empirical quasi correlations for $10' \times 10'$ squares. In each panel there are also two pairs of curves representing the theoretical quasi correlations. One pair corresponds to the numerical value of σ that fits satisfactorily the empirical quasi correlations for $1° \times 1°$ squares. The other pair corresponds to the value of σ which fits approximately the data for the $10' \times 10'$ squares. The upper curve within

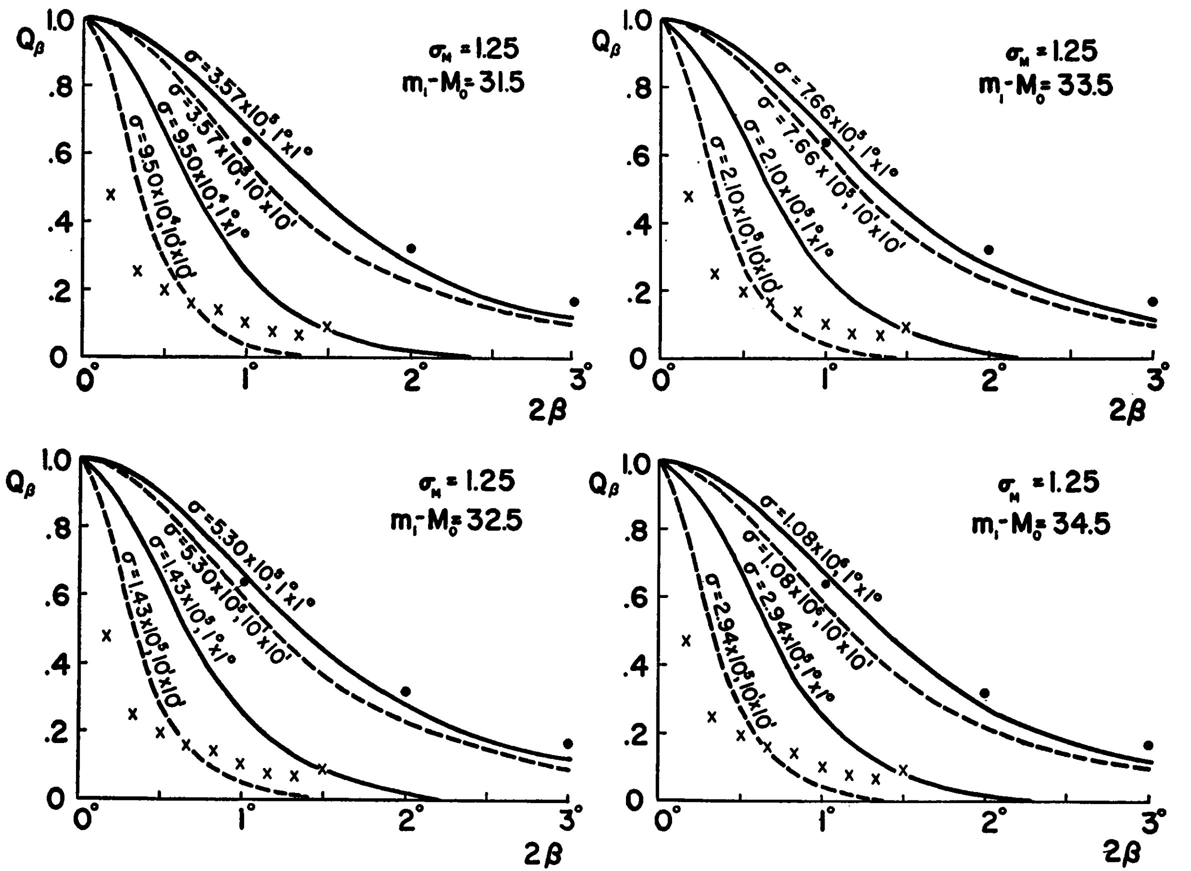

FIGURE 3

Empirical quasi correlations between counts of images of galaxies $1° \times 1°$ and $10' \times 10'$ squares compared with theoretical counterparts corresponding to $\sigma_M = 1.25$ and to varying trial values of the parameters $m_1 - M_0$ and σ.

each pair corresponds to the formula for $1° \times 1°$ squares and the lower curve corresponds to the formula for $10' \times 10'$ squares.

It will be seen that none of the pairs of trial values of $m_1 - M_0$ and σ gives anything like a satisfactory fit to both series of empirical quasi correlations, and that, invariably, the value of σ that fits the data for $10' \times 10'$ squares is substantially smaller than that fitting the quasi correlations for $1° \times 1°$ squares.

This lack of consistency indicates that something is wrong with the model on which the curves in figure 3 were computed. Qualitatively one might perhaps say that the analysis of the $10' \times 10'$ data indicates the presence on the Lick Observatory plates of an excessive (compared to the prediction of the model) number of clusters of galaxies with small angular dimensions (small σ). The same general conclusion was reached on different grounds, as a result of the analysis published in [13]. The modification of the original model which might remove the inconsistency noted is currently the subject of numerical studies. The particular hypotheses within the model that attract the authors' special attention are (i) the status of the universe, stationary versus expanding; (ii) the simple versus the multiple clustering of galaxies (see section 11) and (iii) the hypothetical model of "errors" in counting images of galaxies (section 16).

A separate paper [14] was given to the analysis of a statistic K, suggested by Zwicky [24] for the study of the expansion of the universe and for the verification of the hypothesis of intergalactic absorbing clouds. It was found that, qualitatively, the behavior of K is very much the same whether the universe is expanding or not. Also, at least over the 3 mag. range of limiting apparent magnitudes covered by numerical study, the behavior of K is consistent with the assumption that no intergalactic absorbing clouds are present.

10. Expected number of images of galaxies per square degree

The classical attempts by Hubble and others to study whether the Doppler-like shifts in the spectra of galaxies indicate actual velocities of recession were all based on counts of images of galaxies on plates taken to increasing limiting magnitudes. This circumstance justifies a closer examination of the first formula (26) giving the expected number of images of galaxies within a region ω on an idealized plate with a fixed limiting magnitude. The properties of this formula do not correspond exactly to some of the widespread beliefs. In the following analysis we shall assume that ω is a square with angular dimensions $2\alpha \times 2\alpha$, that the expansion of clusters, if any, is going on at the same rate as the recession of cluster centers, so that $h_1(t) = h_2(t)$, and that counts are made of only one type of galaxies, characterized by a definite function $\chi(\xi)$ representing the effect of reddening and by a single function $\theta(\xi, m_1)$ exhibited in (15). The generalization of our conclusions to the case of several different types of galaxies will be immediate.

Because of the assumption $h_1(t) \equiv h_2(t) \equiv h(t)$, say, we have $H_1(t) \equiv H_2(t) \equiv H(t)$, say, and formulas (8) and (9) reduce to

$$\eta^2 = \sum_{i=1}^{3} \left[y_i H^{-1}\left(-\frac{\xi}{c}\right) - u_i \right]^2, \tag{32}$$

$$J = H^{-3}\left(-\frac{\xi}{c}\right)\left\{1 + \frac{\xi}{c} h\left(-\frac{\xi}{c}\right)\right\}. \tag{33}$$

Changing the notation for the expected number of galaxies brighter than the limit m_1 from $\bar{n}_1$ to $\bar{n}(m_1)$ and using formula (26) we obtain

$$\bar{n}(m_1) = \lambda\bar{\nu} \int_{-\infty}^{+\infty} \left\{ \iiint_{\omega} f^*(\eta) J \theta(\xi, m_1) dy_1 dy_2 dy_3 \right\} du_1 du_2 du_3 \tag{34}$$

or, changing the order of integration and noticing that the integral of $f^*(\eta)$ with respect to u_1, u_2, u_3 from $-\infty$ to $+\infty$ must be equal to unity,

$$\bar{n}(m_1) = \lambda\bar{\nu} \iiint_{\omega} J \theta(\xi, m_1) dy_1 dy_2 dy_3 \, . \tag{35}$$

Since the integrand depends only on ξ while ω represents a solid angle $2\alpha \times 2\alpha$ with its vertex at the observer, a change to polar coordinates gives

$$\bar{n}(m_1) = 4\lambda\bar{\nu}\, \alpha \sin \alpha \int_0^{\infty} \xi^2 J \theta(\xi, m_1) d\xi \tag{36}$$

where J is given by (33) and $\theta(\xi, m_1)$ by (15). Here $\lambda\bar{\nu}$ is equal to the average number of galaxies per unit volume in space and we note that $\bar{n}(m_1)$ does not depend on whether the galaxies are uniformly distributed in space or are clustered.

As is well known, formula (36) has the remarkable property that, if $\chi(\xi) \equiv h(t) \equiv 0$, that is to say, in the absence of reddening and in a stationary universe, then the substitution

$$m + 5 - 5 \log_{10}\xi = t \tag{37}$$

leads to

$$\log_{10}\bar{n}(m) = \frac{3}{5} m + \log_{10}(\lambda\bar{\nu}) + \log_{10} \left\{ C \int_{-\infty}^{+\infty} e^{-Dt} \int_{-\infty}^{t} p_M(x) dx \, dt \right\} , \tag{38}$$

where C and D are known positive constants. This equation indicates that, if counts of images of galaxies are made on plates with increasing limiting magnitudes m_1, $m_2, \cdots, m_s$, then the empirical values of $\log_{10}\bar{n}(m_i)$ plotted against the corresponding m_i must align themselves about a straight line with its slope equal to 0.6 and with its intercept depending on $\lambda\bar{\nu}$ and on the luminosity function $p_M(x)$. It is occasionally believed that this alignment of empirical points along a line with 0.6 slope can be treated as an indication not only of the stationarity of the universe, but also of the absence of clustering. However, the above analysis [10] shows that, in the absence of expansion and of red shift, the phenomenon would be observed irrespective of whether the galaxies are clustered or not.

Formula (15) combined with (36) indicates that, in the presence of reddening which causes the function $\chi(\xi)$ to be positive, but in the absence of expansion, the empirical points $(m_i, \log_{10}\bar{n}(m_i))$ would fall somewhat below the line (38), probably aligning themselves about a curve. Also, formula (33) combined with (15) and (36) indicates that, if reddening were absent and yet expansion present, so that $\chi(\xi) = 0$ but $h(t) > 0$, then the empirical points $(m_i, \log_{10}\bar{n}(m_i))$ would fall above the line (38), also probably about a curve. In recognition of the phenomenon of reddening, this suggested the following method of studying the problem of expansion.

Let m_i be the limiting magnitude of a plate corresponding to the light of a galaxy not affected by reddening. Owing to reddening, a galaxy with its apparent magnitude exactly equal to m_i will be too faint to be identified on the plate. Thus, one may think of determining for the same plate another limiting magnitude, say $m^*(m_i)$, to be labeled the limiting magnitude corrected for the effect of reddening, such that the galaxies affected by reddening but brighter than $m^*(m_i)$ would be

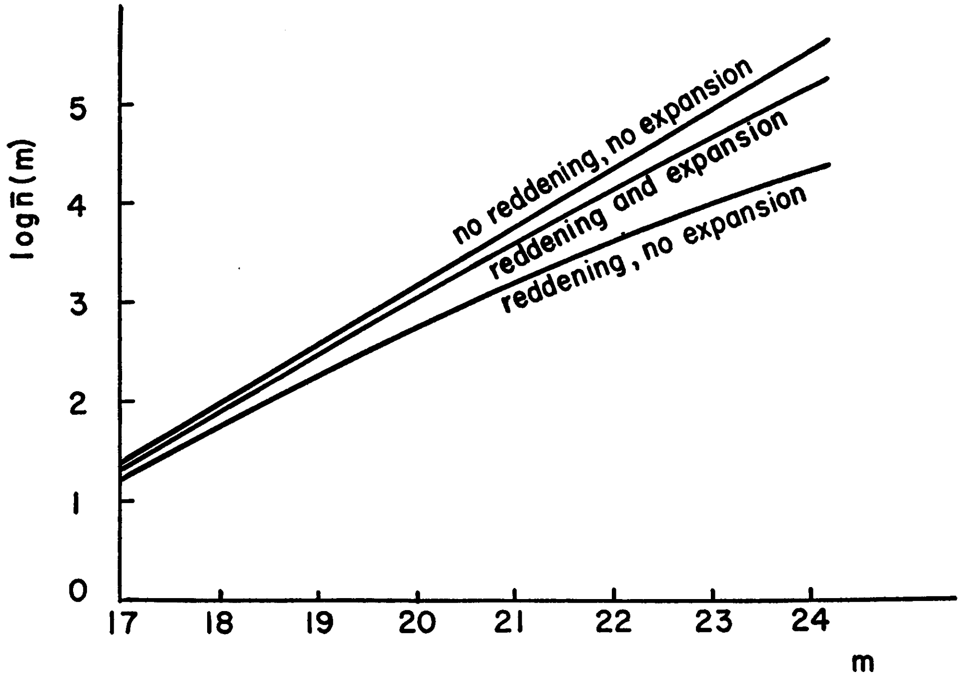

FIGURE 4

Effect on various assumptions regarding reddening of galaxies and the state of the universe on the expected number of images of galaxies per square degree on a plate taken to a limiting magnitude m_i.

visible on this plate. Unfortunately the quantity $m^*(m_i)$ so defined would also depend on the intensity of reddening, that is to say, on the distance of the galaxy. However, one might attempt another definition of $m^*(m_i)$ by means of the equation

$$\int_0^\infty \xi^2 \int_{-\infty}^{m+5-5\log_{10}\xi-\chi(\xi)} p_M(x)dx\, d\xi = \int_0^\infty \xi^2 \int_0^{m^*+5-5\log_{10}\xi} p_M(x)dx\, d\xi \,. \tag{39}$$

With reference to (36), (15) and (33) it will be seen that this equation requires that, in the absence of expansion, the corrected limiting magnitude $m^*(m_i)$ ascribe to $\bar{n}[m^*(m_i)]$, evaluated on the assumption of no reddening, the same value $\bar{n}(m_i)$ which corresponds to the limiting magnitude m_i with the effect of reddening. If efforts to evaluate $m^*(m_i)$ with the above properties were successful, then one might think of plotting the empirical $\log_{10}\bar{n}(m_i)$ against $m^*(m_i)$. If the plotted points significantly deviate from the line (38) with slope 0.6, then this may be considered as indicating the expansion of the universe.

Unfortunately, as is visible from (39), the corrected limiting magnitude $m^*(m_i)$ must depend not only upon the effect of reddening $\chi(\xi)$ but also upon the luminosity function $p_M(x)$ of the galaxies. Figure 4 illustrates the dependence of $\log_{10}\bar{n}(m_i)$ on m_i when the luminosity function $p_M(x)$ conforms with the ideas of Hubble, that is to say, is normal with mean $M_0 = -14.2$ and with standard deviation equal to 0.85. The straight line corresponds to the assumptions of no reddening effect and a static universe. The lower curve corresponds to the assumption that the effect of reddening is represented by

$$\chi(\xi) = 5.15 \times 10^{-9}\xi \tag{40}$$

but there is no expansion. Finally, the upper curve is based on assumption (40) and, in addition, that the universe is expanding, namely [25] that $h(x) = 530$ km/sec/10^6 parsecs. Surprisingly, this upper curve is hardly distinguishable from a straight line. In all cases, the region ω is a square degree and it is assumed that $\lambda\nu = 10^{-17}$ per cubic parsec.

It is interesting to note that, while significant deviations of the empirical points $[m^*(m_i), \log_{10}\bar{n}(m_i)]$ from line (38) do indicate either that the universe is expanding, or that the effect of reddening was not entirely accounted for by the function $\chi(\xi)$ used in the computations, or that the function $m^*(m_i)$ was not properly adjusted to the density $p_M(x)$, the reverse conclusion need not be true. In fact, situations are possible in which the empirical points $[m_i, \log_{10}\bar{n}(m_i)]$ align themselves around a straight line with the slope 0.6 even though the apparent photographic magnitudes are affected by the reddening of galaxies and even though the universe is expanding. The authors were not able to produce an example in which this alignment is exact. On the other hand, it is easy to indicate one in which the lack of exact alignment is so insignificant as to defy empirical detection.

In constructing this example we shall make no assumption regarding the function $h(t)$ and no assumption regarding $\chi(\xi)$. After leaving this much arbitrary, we shall make a specific assumption regarding the luminosity function $p_M(x)$. Namely we shall assume that it corresponds approximately to the ideas of Zwicky [24] and that, for values of t not exceeding a certain limit μ, the distribution of absolute magnitude is exponential,

$$\int_{-\infty}^{t} p_M(x)dx = e^{a(t-\mu)} . \tag{41}$$

For values of $t > \mu$ the above integral will have to be assumed equal to unity. Here a is a positive number. Upon substituting (41) into (36), we obtain

$$\bar{n}(m_i) = 4\lambda\bar{\nu}\alpha \sin\alpha \left\{ e^{a(m_i+5-\mu)} \int_{\xi(m_i)}^{\infty} \xi^2 J\, e^{-a[5\log_{10}\xi+\chi(\xi)]}\, d\xi + \int_0^{\xi(m_i)} \xi^2 J\, d\xi \right\} \tag{42}$$

where $\xi(m_i)$ is determined by the equation

$$m_i + 5 - \mu = 5\log_{10}\xi(m_i) + \chi[\xi(m_i)]. \tag{43}$$

Since $\chi(\xi) \geqq 0$, it follows that

$$\xi(m_i) \leqq 10^{(m_i+6-\mu)/5} \tag{44}$$

and, provided μ is a rather large number, $\xi(m_i)$ is very small. In these circumstances the second integral in (42) will be negligible and it will be seen that, for all values of the limiting apparent magnitude m_i that may come under consideration, the $\log_{10}\bar{n}(m_i)$ will depend linearly on m_i, the coefficient of m_i being equal to $a \log_{10}e$. Thus, with an appropriate value of a and with μ of order of 42 or so, the empirical points $[m_i, \log_{10}\bar{n}(m_i)]$ corresponding to $m_i < 22$ will lie close to a straight line with its slope equal to 0.6, quite irrespective of whether the effect of reddening was eliminated or not and whether the universe is expanding, contracting or static.

The above conclusions apply to the particular luminosity function given by (41) and may perhaps be used in order to verify whether or not this function, with arbitrary $a > 0$, could be used to represent $p_M(x)$.

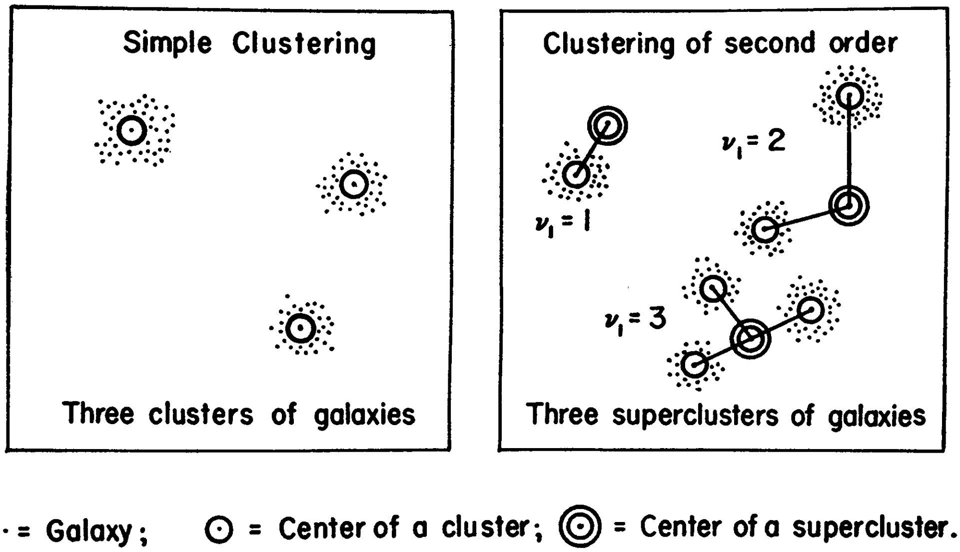

FIGURE 5

Concepts of simple clustering and of clustering of second order.

11. Generalization of the original model: Model of multiple clustering

Because of the observations, originally made by Shapley [26] and more fully confirmed by Shane and Wirtanen [7], that the clusters of galaxies tend to congregate in small groups, it appeared interesting to deduce formulas allowing for the possibility of "superclustering" or "multiple" clustering. Actually formulas of this kind were deduced [27] and the relevant theory is being prepared for publication. This theory is concerned with clustering of an arbitrary sth order. For the sake of brevity we give below only the basic formula referring to clustering of the second order. This formula corresponds to formula (22) of the theory of simple clustering.

In order to visualize the model of clustering of the second order, it is convenient to return to that of simple clustering as described above and to begin by making a few substitutions of terms. Here the inspection of figure 5 may be helpful.

(i) Instead of the old term "cluster center" use "center of supercluster."

(ii) Instead of speaking of a variable ν = number of galaxies per cluster, consider the variable ν_1 = number of clusters per supercluster.

(iii) Instead of speaking of galaxies within a cluster, speak of cluster centers within a supercluster.

(iv) Instead of the function $f^*(\eta)$ representing the conditional probability density of the present coordinates of a galaxy given the present coordinates of the center of the cluster to which this galaxy belongs, use the function f_1^* to represent the conditional probability density of the coordinates of a cluster center, given the coordinates of the center of the supercluster to which the cluster belongs.

The coordinates of the center of a supercluster will be denoted by $u = \{u_1, u_2, u_3\}$, and those of a cluster center by $v = \{v_1, v_2, v_3\}$. Accordingly, it will be convenient to complete the symbol for f_1^* by writing $f_1^*(v, u)$. The hypotheses originally adopted for cluster centers and galaxies will now be transferred without change to centers of superclusters and cluster centers respectively.

In addition, we attach to each cluster a random variable ν_2 representing the number of galaxies belonging to this cluster. Also we introduce another probability density, say $f_2^*(x, v)$. This will be the conditional probability density of the present coordinates of a galaxy, given the coordinates v of the center of cluster to which this galaxy belongs. The original postulates regarding the distribution of galaxies within a cluster are transferred to the present model and, in particular, it is assumed that the arguments of f_1^* and f_2^* are distances between points v and u and between x and v, respectively. Naturally, the distributions of ν_1 and ν_2 need not coincide. Also f_1^* and f_2^* need not be the same functions. The symbol $f_2(y, v)$ will be used to denote the conditional probability density of the apparent coordinates of a galaxy given that it belongs to a cluster with the present coordinates of its center equal to v.

With these postulates and notation, the probability generating function of the variables n_1 and n_2 representing the numbers of images of galaxies in regions ω_1 and ω_2 on two photographic plates is given by

$$(45)\quad \log G_{n_1, n_2}(t_1, t_2) = -\lambda \int \left\{ 1 - G_{\nu_1}\left[\int f_1^*(v, u) G_{\nu_2}\{1 - p_1(1-t_1) - p_2(1-t_2) - p_3(1-t_1t_2)\} dv\right] \right\} du$$

where $p_1(v)$, $p_2(v)$ and $p_3(v)$ are obtained from formulas (23), (24) and (25) by substituting in them v instead of u and f_2 instead of f. Both integrals in (45) are triple integrals extending from $-\infty$ to $+\infty$ for each variable of integration.

Although the idea of multiple clustering is interesting, the model of second order clustering is less attractive than the original model of simple clustering because, even with the most simplifying specializations, the new model is likely to depend on at least six adjustable parameters. It will be noticed that, according to the hypotheses adopted regarding the functions $h_1(t)$ and $h_2(t)$, formulas (45) will apply either to the static or the expanding universe. However, this formula is deduced on the assumption that, if the universe is expanding at all, then the rate of expansion of the superclusters coincides with that of the recession of centers of superclusters.

In the deduction of formula (45) an essential role is played by formula (19) which is used to characterize the spatial distribution of cluster centers.

Thus far, only the general theoretical model of multiple clustering was studied and very little was done to adjust this model to the observable phenomena. This adjustment, and also the relevant numerical results, must be relegated to a future publication. However, certain qualitative considerations, which may be considered as evidence in favor of the hypothesis of multiple clustering, deserve recording in the present paper. These considerations indicate the possibility that the assumption of multiple clustering of galaxies will account for certain numerical discrepancies which caused the authors a considerable amount of grief.

The discrepancies in question concern the function $f^*(\eta)$ which, in the theory of simple clustering, represents the density of galaxies per unit volume as a function of the distance from the cluster center. This function can be estimated by two independent methods, directly [7] through the analysis of identified clusters (see section 11), and indirectly through the analysis of quasi correlations (see section 8) on the basis of the theory of simple clustering [11]. The two estimates disagree and

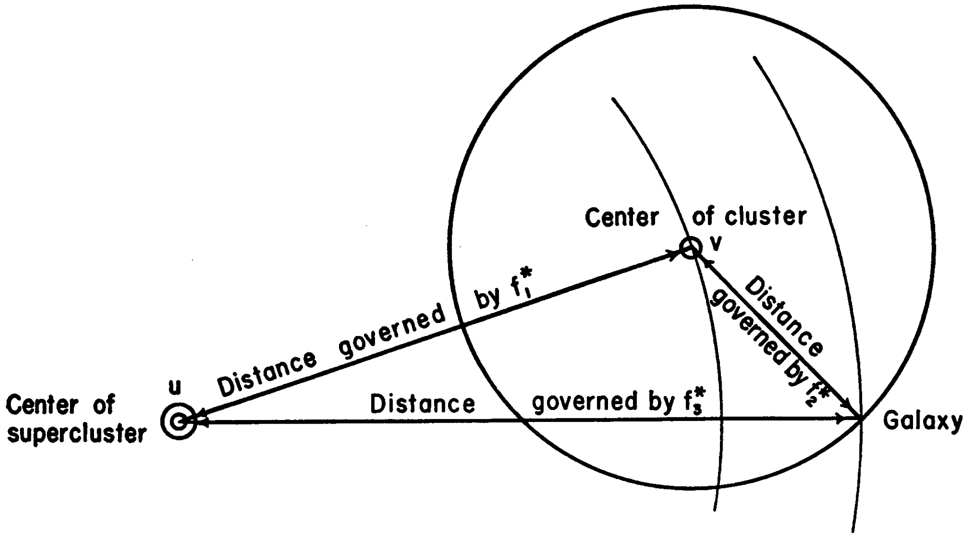

FIGURE 6
Flatness of the function f_3^*.

the function obtained by the method of quasi correlations appears much flatter than that obtained by the direct method. The construction of a synthetic plate [13] provided further evidence that the function $f^*(\eta)$ obtained by the method of quasi correlations may be too flat.

Originally it was thought that the discrepancy might be due to an unlucky choice of specialization of $f^*(\eta)$ used in paper [11]. However, subsequent numerical work indicated that, even if one adopts an alternative specialization (a weighted average of several normal densities rather than a single normal density), the indicated discrepancy continued to persist. Now it is easy to see that, qualitatively at least, the phenomenon of multiple clustering tends to account for this discrepancy.

In order to see this, assume that the actual distribution of images of galaxies on a photographic plate reflects multiple clustering and conforms with formula (45). Assume further that this fact is ignored and that formula (29), based on the theory of simple clustering, is used to estimate the density $f^*(\eta)$. Starting with formula (45) it is easy to deduce that the left hand side of equation (29) is a weighted mean of two quasi correlations Q_c and Q_s, one depending on the actual clustering of galaxies and the other on the clustering of clusters. More specifically, we may write

$$Q = w_c Q_c + w_s Q_s . \tag{46}$$

Here Q_c stands for the quasi correlation that would have been observed if (i) $\nu_1 \equiv 1$, so that there is no multiple clustering and (ii) if the internal structure of the clusters were governed by the density $f_2^*(x, v)$, so that

$$Q_c = \frac{\int p_1(v)p_2(v)dv}{\int p_1^2(v)dv} \tag{47}$$

with

$$p_i(v) = \int_{\omega_i} \theta(\xi, m) f_2(y, v) dy \ . \tag{48}$$

In order to explain the meaning of Q_s, we must introduce two new symbols. The first of these is $f_3^*(x, u)$ which represents the conditional probability density of the present coordinates X_1, X_2, X_3 of a galaxy, given that it belongs to a supercluster centered at $u = \{u_1, u_2, u_3\}$. We have

$$f_3^*(x, u) = \int f_1^*(v, u) f_2^*(x, v) dv \ . \tag{49}$$

It will be noticed that, most probably, the density $f_1^*(v, u)$, governing the structure of a supercluster, is very substantially flatter than the density $f_2^*(x, v)$ governing the structure of a cluster. Formula (49) implies that $f_3^*(x, u)$ is even flatter than $f_1^*(v, u)$. Also, this fact can be seen intuitively in figure 6. Using $f_3^*(x, u)$ and the formulas of section 6, it is easy to obtain the corresponding probability density $f_3(y, u)$ of the apparent coordinates Y_1, Y_2, Y_3 of a galaxy, given that this galaxy belongs to a supercluster centered at u. Now, denote by $\pi_i(u)$ the conditional probability

$$\pi_i(u) = \int_{\omega_i} \theta(\xi, m) f_3(y, u) dy \tag{50}$$

that a galaxy belonging to a supercluster centered at u will have an image visible within the square ω_i on the photographic plate. Then,

$$Q_s = \frac{\int \pi_1(u)\pi_2(u)du}{\int \pi_1^2(u)du} \ . \tag{51}$$

In other words, Q_s represents the quasi-correlation that would have been observed if there were no multiple clustering and if the structure of the clusters were governed by the very flat density $f_3^*(x, u)$.

The weights w_c and $w_s = 1 - w_c$ in (46) are proportional to

$$E(\nu_1)[E(\nu_2^2) - E(\nu_2)] \int p_1^2(u)du \quad \text{and} \quad [E(\nu_1^2) - E(\nu_1)]E^2(\nu_2) \int \pi_1^2(u)du \ , \tag{52}$$

respectively.

It must now be obvious that, if an effort is made to use the formulas of simple clustering and to adjust a single probability density $f^*(\eta)$ so that the quasi correlations defined by it fit the weighted mean (46) of the quasi correlations, one defined by $f_2^*(x, v)$ and the other by a much flatter density $f_3^*(x, u)$, then the adjusted function $f^*(\eta)$ will be intermediate between $f_2^*(x, v)$ and $f_3^*(x, u)$ and probably quite substantially flatter than $f_2^*(x, v)$. Therefore, the observed difference in the flatness of the two independent estimates of $f^*(\eta)$ is indirect evidence in favor of the hypothesis that superclustering is a real phenomenon. In addition, it is possible that the same phenomenon of superclustering will account for the discrepancy exhibited in figure 3 between the theoretical and the empirical quasi correlations computed for the $10' \times 10'$ squares.

12. Probability distributions of certain characteristics of images of clusters

Up to the present time, a few empirical distributions of the characteristics of clusters have been published, mainly by Shapley and by Zwicky. However, as the technique of identifying images of clusters is developed and larger telescopes are made available, the publication of the relevant empirical data is likely to increase. In fact, something of this sort appears foreshadowed in the recent paper [24] by Zwicky. One would hope that such data would be very powerful in studying the phenomena of clustering. Therefore, a study was undertaken [16], [28] to develop formulas, based on the model of simple clustering, giving the distributions of several important characteristics of images of clusters. Only the briefest account of one of the results obtained can be given here.

Consider a cluster of galaxies with its center at the present distance u. In accordance with the above theory, this cluster has a random number ν of galaxies. However, not all of these galaxies need have visible images which can be identified as galaxies on a photographic plate taken at a given limiting magnitude m_1. Thus, it is necessary to consider a new random variable, say ν^*, representing, for the same cluster, the number of galaxies that have images visible and countable on the plate ("visible" galaxies, for short). We shall say that a cluster itself is "visible" only if the corresponding value of $\nu^* \geqq 1$. Naturally, the distribution of ν^* must depend on that of ν, as well as on the structure of clusters as determined by the density $f^*(\eta)$ and on whether the universe is expanding or not.

Denote by $\theta^+(u, m_1)$ the conditional probability, given the present distance u of a cluster center (rather than the distance ξ to the galaxy itself), that a galaxy belonging to this cluster will be "visible" on the photographic plate with the limiting magnitude m_1. We have

$$\theta^+(u, m_1) = \iiint_{-\infty}^{+\infty} \theta(\xi, m_1) f(y_1, y_2, y_3; u, 0, 0) dy_1 dy_2 dy_3 \,. \tag{53}$$

This integral may be somewhat simplified by a change to polar coordinates.

The probability generating function of ν^*, conditioned by $\nu^* \geqq 1$, is then given by

$$G_{\nu^*}(t|\nu^* \geqq 1) = 1 - \frac{\int_0^\infty u^2\{1 - G_\nu[1 - \theta^+(u, m_1)(1 - t)]\} du}{\int_0^\infty u^2\{1 - G_\nu[1 - \theta^+(u, m_1)]\} du} \,. \tag{54}$$

Now consider a cluster with ν^* visible galaxies. Denote by Φ the angular distance between the direction from the observer towards the center of the cluster and the direction from the observer to the apparent position of one of the visible galaxies. The ν^* distances Φ may be ordered in decreasing order so that $\Phi_1 \geqq \Phi_2 \geqq \Phi_3 \geqq \cdots \geqq \Phi_{\nu^*}$. Now denote by ρ the value of Φ_1, the largest of the angles, and describe it as the apparent angular radius of the cluster. For any positive number x denote by $F_\rho(x)$ the conditional distribution function of ρ, given $\nu^* \geqq 1$, that is to say, the conditional probability, given that the cluster contains at least one visible galaxy, that the apparent angular radius ρ will not exceed the number x. It is found that

$$(55) \qquad F_\rho(x) = P\{\rho \leqq x|\nu^* \geqq 1\}$$
$$= 1 - \frac{\int_0^\infty u^2 \left(1 - G_\nu\{1 - \theta^+(u, m_1)[1 - F_\Phi^*(x|u)]\}\right) du}{\int_0^\infty u^2\{1 - G_\nu[1 - \theta^+(u, m_1)]\}du},$$

where

$$(56) \qquad F_\Phi^*(x|u) = \frac{1}{\theta^+(u, m_1)} \iiint_{\omega(x)} \theta(\xi, m_1)f(y_1, y_2, y_3; u, 0, 0)dy_1dy_2dy_3,$$

is the conditional distribution function of the angular distance Φ of a galaxy from its cluster center, given that this galaxy is visible and given that the cluster center is at distance u from the observer. It is just this observable function $F_\Phi^*(x|u)$ that provides the "direct" estimate (mentioned in section 10) of the function $f_\Phi^*(\eta)$ for any cluster whose distance u is known. The symbol $\omega(x)$ appearing in (56) represents the circular cone defined by $0 \leqq \Phi \leqq x$ and $0 \leqq \xi < \infty$.

As is to be expected, the distribution of the apparent radius of a visible cluster depends upon that of the number of galaxies per cluster and on the internal structure of clusters as characterized by the density $f^*(\eta)$. Also, the distributions (55) and (56) depend on whether or not the universe is expanding.

Formulas analogous to (55) and (56) but applying specifically to static universe were reported by one of the present authors at the Cleveland meeting of the American Astronomical Society in December, 1951. A formula for the distribution of ρ (with different notation) was recently published by Zwicky [24]. However, Zwicky's formula was deduced without taking into account the randomness of the number of galaxies per cluster and the effect of dimming with distance.

13. Problem of interlocking of clusters

In all the studies reported above, the central problem was to deduce the various properties of the distribution of images of galaxies or of images of clusters on photographic plates from what was assumed to be happening in space. The theory so developed may be used for comparison with the results of observation in order to verify certain of the hypotheses made or in order to estimate some parameters involved in a specific model. In addition to these studies, a certain amount of attention was given to problems of an entirely different kind, concerned, so to speak, with the external appearance of the spatial distribution of galaxies.

Although the model studied was built up using the concept of a cluster as a basis, the actual role of this concept is not predetermined, and depends upon the actual phenomena. Consider two extreme possibilities. One extreme is that galaxies are grouped in fairly tight systems separated from each other by great distances. The other extreme possibility is that such groupings of galaxies as exist are loose and interlocking, and that the situation would be more adequately described by the term continuous field of galaxies with numerous local concentrations. These two possibilities are illustrated in figure 7.

In the first case, the "conceptual cluster" as postulated in the model would have an identifiable phenomenal counterpart and would be an important element in any theory of the distribution of galaxies. In the second case, the role of the concept of a cluster would dwindle to that of an easily replaceable element in an interpolatory structure invented to approximate the actual distribution of galaxies. The question as to whether or not one should use the theory of clustering to represent the distribution of galaxies would then be the same kind of question as to whether one should use a polynomial (rather than, say, a trigonometric polynomial) to approximate a function of which the real nature is unknown.

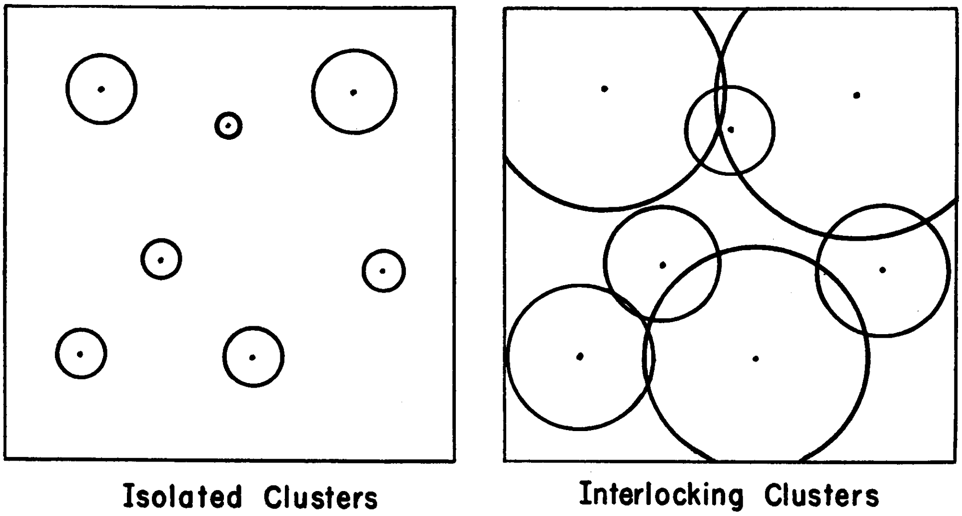

FIGURE 7

Concept of interlocking of clusters.

It follows that, in order to elucidate the actual role of clustering and also in order to gain a clearer picture of the spatial distribution of galaxies, it is interesting (i) to fit the theory of clustering to the observational data just as accurately as possible and (ii) to obtain from the fitted model numerical characteristics describing the degree to which the individual clusters interlock or overlap.

Among the many difficulties involved in this problem there is the conceptual one of formulating a mathematical problem so that its solution would provide an answer to the astronomical question about the degree of interlocking of clusters. In papers [12] and [17] the following problems were treated.

Consider a particular cluster to be termed the selected cluster Γ with its center at a point C. Denote by ν_1 the (random) number of galaxies belonging to Γ. We shall number these galaxies in the decreasing order of their distances from the center C. Thus the most distant galaxy will be called the first, and the least distant the ν_1-st. Let η_i denote the (random) distance of the ith galaxy from C. Thus $\eta_1 \geqq \eta_2 \geqq \cdots \geqq \eta_{\nu_1} \geqq 0$. For any positive number x denote by $S(x)$ the sphere of radius x centered at C.

If a galaxy not belonging to the cluster Γ happens to be within the sphere $S(\eta_k)$ then we shall say that this galaxy is imbedded in Γ to the depth k.

In order to characterize the degree of interlocking of clusters, the following random variables were studied.

(i) The variable μ_k = number of galaxies belonging to the cluster Γ_1, which is nearest to Γ itself, that are imbedded in Γ to the depth k.

(ii) The variable τ_k = number of clusters having at least one galaxy imbedded in Γ to the depth k.

(iii) The variable ϵ_k = number of clusters with all of their galaxies imbedded in Γ to the depth k.

For each of these three variables it was possible to deduce the probability generating functions based on the theory of simple clustering.

It will be seen that the three variables μ_k, τ_k and ϵ_k do characterize the degree of interlocking of clusters. Thus, for example, if $\tau_1 = 0$ then the selected cluster is entirely isolated. Now, if we should find that the estimates obtained from a reasonably fitting specialized model imply that $P\{\tau_1 = 0\} = .99$, then, confronted with the choice between the two extreme possibilities envisaged in figure 7, we would probably have no hesitation in choosing the one on the left, corresponding to tight clustering. On the other hand, if, for example, $P\{\tau_1 = 0\} = 10^{-6}$ we would choose the alternative on the right.

As was mentioned before, the present estimates of the constants involved in the particularized model [11] are extremely shaky, mainly because of the uncertainty regarding the luminosity function of the galaxies. Furthermore, it is not impossible that several of the assumptions underlying these estimates are not true. The assumptions of a static universe and of a quasi-uniform distribution of cluster centers are particularly doubtful. Thus at the present time nothing like a reliable indication is available as to which of the two alternatives exhibited in figure 7 is closer to reality. In particular, on the original estimates obtained in [11] the probability $P\{\tau_1 = 0\} = .0017$. However, the results obtained in [13] and also those illustrated in figure 4 indicate that the original estimate of what may be called the "size" of a cluster is too large and that about one-half of its value would be more realistic. If we adopt this value, then we obtain $P\{\tau_1 = 0\} = .204$, and it is seen that, after allowing for the possibility of multiple clustering, we may well arrive at an estimate of the probability $P\{\tau_1 = 0\}$ which is more favorable to the concept of an isolated cluster than the ones already calculated.

While the three variables μ_k, τ_k and ϵ_k do characterize the degree of interlocking of clusters, it is obvious that this characterization is incomplete. Consider for example a large cluster Γ, of some 1000 galaxies, in which there are imbedded a dozen small clusters, $\Gamma_1, \Gamma_2, \cdots, \Gamma_{12}$, of 5 to 10 galaxies each. Further, assume that no other cluster "invades" the above cluster Γ. In these circumstances the small clusters $\Gamma_1, \Gamma_2, \cdots, \Gamma_{12}$ would probably be considered as local irregularities in the structure of Γ and there would be no objection to treating Γ as an isolated cluster. Yet, for this particular cluster $\tau_1 = 12$ and $\epsilon_1 = 12$. Thus, it appears that for a more satisfactory characterization of the degree of interlocking of clusters some other random variables need be studied. Accordingly, a forthcoming paper [29] gives the probability generating function of two such additional variables, of which one, ψ_k, is defined as the number of clusters that contain at least one galaxy within $S(\eta_k)$ and at least one galaxy outside of $S(\eta_1)$.

14. Model of fluctuating density of galaxies

The numerical data reproduced in the preceding section indicate that the concept of a cluster of galaxies, at least the concept of an isolated cluster, may have no identifiable phenomenal counterpart. In this connection, it is natural to consider a theory of the spatial distribution of galaxies independent of the concept of a cluster but leaving room for the existence of local concentrations. The development of such a theory was initiated by Chandrasekhar, Münch and Limber [30], [31], [32]. Unfortunately [33], certain of the formulas advanced for the purpose of estimating the unknown parameters of the theory contain some misunderstanding. The formulas in question refer to the numbers $n(\omega)$ of images of galaxies in squares ω on the photographic plate and ascribe to $n(\omega)$ impossible properties. Thus, for example, the formulas for the expectations $E[n(\omega)]$ and $E[n^2(\omega)]$ of $n(\omega)$ and of $n^2(\omega)$, respectively, require that, as $\omega \rightarrow 0$, these two quantities tend to zero but that the quotient

$$C(\omega) = E[n^2(\omega)]/E^2[n(\omega)] \tag{57}$$

tend to a finite limit, which is impossible. In fact, since $n(\omega)$ is necessarily a nonnegative integer number, the first power of $n(\omega)$ cannot be greater than the second, and we have

$$n(\omega) \leqq n^2(\omega)\,, \tag{58}$$

$$E[n(\omega)] \leqq E[n^2(\omega)] \quad \text{or} \quad 1 \leqq \frac{E[n^2(\omega)]}{E[n(\omega)]}\,, \tag{59}$$

and the quotient (57) can never be less than the reciprocal of $E[n(\omega)]$, that is,

$$C(\omega) = \frac{1}{E[n(\omega)]}\,\frac{E[n^2(\omega)]}{E[n(\omega)]} \geqq \frac{1}{E[n(\omega)]}\,. \tag{60}$$

Thus, if $E[n(\omega)]$ tends to zero as $\omega \rightarrow 0$, then at the same time the quotient $C(\omega)$ must tend to infinity. This point is quite important, because the formulas of the theory discussed are supposed to hold good *for small squares* ω and the function of the unknown parameters which is equated to $C(\omega)$ for purposes of estimation, is independent of ω. If the estimation of these parameters is attempted on a number of sets of data, with different sizes of ω, say $\omega_1 > \omega_2 > \omega_3 > \cdots$, then the same function of the parameters will be equated to corresponding quantities $C(\omega_1)$, $C(\omega_2)$, $\cdots$ which will tend to infinity. Consequently, the estimates obtained will reflect the fortuitous circumstance that in the particular sets of data the counts of images of galaxies were made in $3^\circ \times 3^\circ$ squares or in $1^\circ \times 1^\circ$ squares, or in $10' \times 10'$ squares, etc. Similar remarks apply to some other formulas involved in the theory discussed.

Because of the above circumstances, an independent attempt was made [34] to build up the theory of fluctuations. The basic assumptions adopted are as follows.

(i) *Existence postulate.* To every region S in space (really, to every Borel set S) there corresponds a random variable $N(S)$ representing the number of galaxies contained in S. Thus $N(S)$ is necessarily a nonnegative integer.

(ii) *Postulate of stationarity.* If two systems of regions $S_{11}, S_{12}, \cdots, S_{1m}$ and S_{21},

$S_{22}, \cdots, S_{2m}$ are congruent, then the joint distribution of $N(S_{11}), N(S_{12}), \cdots, N(S_{1m})$ is identical with that of $N(S_{21}), N(S_{22}), \cdots, N(S_{2m})$, irrespecitve of the relative locations of the two systems of regions.

(iii) *Postulate of complete additivity.* If the regions $S_1, S_2, \cdots, S_m, \cdots$ are all disjoint, then the variable $N(\sum S_i)$, corresponding to their union, is equal to the sum of the numbers $N(S_i)$, so that $N(\sum S_i) = \sum N(S_i)$.

(iv) *Postulate of finiteness of moments.* If a region S has a finite volume (measure) V, then the variable $N(S)$ has a finite expectation and a finite variance $\sigma^2_{N(S)}$.

(v) *Postulate of positiveness and monotonicity of correlation.* Let L be an arbitrary line in space and P_1 and P_2 two points on L separated by a distance $\xi > 0$. Let $C(P_i, V_i)$ be a cube of volume V_i centered at P_i, for $i = 1, 2$. It is postulated that the correlation $R(\xi, V_1, V_2)$ between $N[C(P_1, V_1)]$ and $N[C(P_2, V_2)]$ is never negative. Also it is postulated that for fixed V_1 and V_2 and for fixed orientations of the two cubes, $R(\xi, V_1, V_2)$ is a nonincreasing function of ξ.

The above postulates imply that

$$E[N(S)] = \bar{\rho} V(S) , \tag{61}$$

where $\bar{\rho}$ is a constant and $V(S)$ represents the volume of S, that

$$\lim_{V \to 0} \frac{\sigma_N^2[C(P, V)]}{V} = \bar{\rho}\beta^2 , \tag{62}$$

where β^2 is a constant at least equal to unity, and also that

$$\lim_{\substack{V_1 \to 0 \\ V_2 \to 0}} \frac{R(\xi, V_1, V_2)}{\sqrt{V_1 V_2}} = \Gamma(\xi) \geqq 0 , \tag{63}$$

where $\Gamma(\xi)$ is a finite nonincreasing function defined for all $\xi > 0$. In turn, for every region S_1 and S_2 the variances of $N(S_1)$ and $N(S_2)$ and their correlation are expressable in terms of ρ, β and $\Gamma(\xi)$. Also, it is easy to deduce formulas for the moments of the first two orders of the random variables $n(\omega)$.

The theory of fluctuations as sketched above contains the theories of simple and of multiple clustering as particular cases. In fact, if the actual distribution of galaxies conforms with the postulates of the theory of simple clustering then, using formula (19), it is easy to find that

$$\bar{\rho} = \lambda \bar{\nu} , \qquad \beta = 1 \tag{64}$$

and

$$\Gamma(\xi) = 2\pi \frac{\overline{\nu^2} - \bar{\nu}}{\nu} \int_0^\infty x^2 f^*(x) \int_{-1}^{+1} f^*(\sqrt{\xi^2 - 2\xi xy + x^2}) dy\, dx . \tag{65}$$

Similarly, if the actual distribution of galaxies conforms with the postulates of the theory of multiple clustering of the second order, then

$$\bar{\rho} = \lambda \bar{\nu}_1 \bar{\nu}_2 , \qquad \beta = 1 \tag{66}$$

and

$$\Gamma(\xi) = 2\pi\bar{\nu}_2 \frac{\overline{\nu_1^2} - \bar{\nu}_2}{\nu_1} \int_0^\infty x^2 f^*(x) \int_{-1}^{+1} f_3^*(\sqrt{\xi^2 - 2\xi xy + x^2})dy\, dx \tag{67}$$

$$+ 2\pi \frac{\overline{\nu_2^2} - \overline{\nu_2}}{\overline{\nu_2}} \int_0^\infty x^2 f_2^*(x) \int_{-1}^{+1} f_2^*(\sqrt{\xi^2 - 2\xi xy + x^2})dy\, dx$$

with

$$f_3^*(x) = 2\pi \int_0^\infty t^2 f_2^*(t) \int_{-1}^{+1} f_1^*(\sqrt{x^2 - 2xyt + t^2})dy\, dt\,. \tag{68}$$

A slight change in notation has been introduced in formulas (65), (67) and (68). Previously the symbol $f_i^*(a, b)$ was used to denote the probability density f_i^* depending on the distance between two points $a = \{a_1, a_2, a_3\}$ and $b = \{b_1, b_2, b_3\}$, for $i = 1, 2$. In the three formulas mentioned, instead of $f_i^*(a, b)$ we wrote $f_2^*(\sqrt{\sum(a_i - b_i)^2})$ and it is hoped that this change will not cause confusion.

The authors are indebted to Professor John W. Tukey for the suggestion that the equality $\beta = 1$ will hold good for all spatial distributions of dimensionless particles which satisfy the following condition. Suppose that a region S is known to contain $n \geqq 2$ particles. In order that $\beta = 1$ it is necessary and sufficient that the conditional probability that the positions of any two particles in S coincide be equal to zero. The proof of this proposition is immediate.

15. Test for the expansion of clusters of galaxies

In connection with the discussion of section 6 regarding the possibility that the universe is expanding, the question arose whether or not it is possible to test the expansion of clusters. A test was actually developed and published [35]. It is based on the following assumptions. (i) The peculiar velocities of galaxies within a cluster have random components following a symmetrical normal law. It was shown that this assumption, combined with an appropriate distribution of positions of galaxies, is consistent with the assumption that this latter distribution of positions is constant in time, so that the cluster neither expands nor contracts. (ii) It was postulated that in addition to the above *random* components, the peculiar velocities of galaxies within a cluster may have a *nonrandom* component, directed away from the center of the cluster and, for each galaxy, proportional to the distance between that galaxy and the center of the cluster. The coefficient of proportionality was denoted by h. The hypothesis tested was that $h = 0$, the alternatives being that $h > 0$.

For each cluster of galaxies the observable (approximately) random variables may be (a) the angular distances, say ϕ, of particular galaxies from the estimated direction towards the center of the cluster, (b) the apparent magnitudes of the galaxies and (c) their radial velocities. Using these variables and the above assumptions (i) and (ii) a locally best test of the hypothesis $h = 0$ was deduced. The power of this test is being investigated.

16. Concluding section: Some important outstanding problems

In an effort to formulate a brief summary of the studies outlined above, the authors arrive at the following description. All the studies already published, and those still in progress, have for their general purpose the invention of a chance machinery of the spatial distribution of galaxies which, combined with the adopted mechanisms of transmission of light and of other relevant phenomena, could account for the observable characteristics of the distribution of images of galaxies on the photographic plates. While some progress in this general direction has been achieved, there are still many outstanding problems which, probably, will require the efforts of more than one generation.

(i) Probably the most difficult and at the same time the most important outstanding problem was mentioned in the Introduction and then again in section 7. It concerns the dependence of the apparent photographic magnitude $m(M, \xi)$ of a galaxy on its absolute magnitude M and on its distance ξ. The expression for $m(M, \xi)$ is needed on the alternative assumptions that the universe is static and that it is expanding. Obviously, this problem is one in astrophysics rather than in statistics.

(ii) The next important problem we wish to mention may be labeled the problem of stochastization of the theory of relativity. This problem was already attacked by McVittie [18] and it is understood that further work in this direction is contemplated.

(iii) The third outstanding important problem is purely statistical and consists in developing a convincing stochastic model of the distribution of the light absorbing clouds that are present within the Milky Way galaxy. Interesting contributions to this subject are due to Ambartzumian [36] and to Chandrasekhar and Münch [37]. Also, it may be mentioned that at the Lick Observatory there is work in progress on counts of stars made on the same plates on which the images of galaxies have already been counted. The comparison of the two systems of counts is likely to lead to interesting conclusions.

(iv) The last outstanding problem we wish to mention is somewhat less inspiring than the preceding ones, but is just as important. In fact, a satisfactory solution of this last problem is a prerequisite to the success of any statistical treatment of the distribution of images of galaxies on photographic plates.

The problem consists in building a convincing stochastic model of the machinery of "errors of counting" images of galaxies. We include the words "errors of counting" in quotation marks because this term is meant to cover not only the errors of counting in the strict sense of the words, but also the effects of the night-to-night variation of the transparency of the atmosphere, the variation in the quality of emulsion from one plate to the next, the variation in the emulsion from one part of the plate to another part and, finally, all the complexities involved in the performance of the telescope. The details of the latter, named the off-center effect and the east-west effect, are described in [7].

Whenever a statistician is faced with the situation of studying some variables ξ and η that are not directly observable but for which there are given the values X and Y of measurements of ξ and η, the customary procedure is to assume that $X = \xi + u$ and $Y = \eta + v$, where u and v are "errors of measurement." It is customary to postulate that these errors of measurement are independent of the quantities measured and independent of each other. Finally, there usually is the ubiquitous assumption of normality.

This kind of error model (error model I) was tried [11] by the authors in relation to counts of images of galaxies. The same model was used by Limber [32]. The undisputed advantage of this model is its simplicity. However, if one is interested not so much in obtaining from calculations just a number, but a number connected in an intelligible way with the phenomena studied, then the mere simplicity of the formulas is not enough and the adequacy of the model becomes important.

Model I was tested and had to be abandoned because its consequences are in conflict with the observations. For example, denote by ξ and η the numbers of images of galaxies in two adjoining $10' \times 10'$ squares on the photographic plates and by x_i, y_i the estimates of these quantities obtained on the ith count performed on the same plate, for $i = 1, 2$. If model I corresponded to reality then $x_1 - x_2$ and $y_1 - y_2$ would be independent. In actual fact, these differences are dependent and the effect of this dependence is felt quite strongly when one considers simultaneously [14] counts made in $10' \times 10'$ squares and in $1° \times 1°$ squares. Yet, in order to make a comprehensive test of any hypothesis regarding the spatial distribution of galaxies, it is essential to compare its consequences with the observations relating to varying sizes of the squares in which the counts are made and to varying limiting magnitudes. Thus, for example, with reference to figure 3 it is essential to know whether the discrepancies illustrated indicate the inadequacy of the model of simple clustering in a stationary universe or to the inadequacy of the error model applied.

Also the second error model considered had to be abandoned immediately. Here it was assumed that to every image of a galaxy, present on the photographic plate, there corresponds a constant probability of this image being counted. Also, it was assumed that the counting of one galaxy is independent of the counting of any other.

The next model, labeled model II, is a modification of the preceding one. Here it is postulated that if a square degree on a photographic plate contains N images of galaxies then the number X of images counted in that square is a binomial variable with exponent equal to N and with a probability δ that is treated as a random variable. In fact, it is postulated that to each of 36 one degree squares on a photographic plate and to each "session" of counting there corresponds a random variable δ, independent of all others and following the same Beta distribution.

In the early period, when the authors dealt only with counts in $1° \times 1°$ squares, corrected as indicated in [7], no contradictions were found between model II and the observations. However, when more data were accumulated and, particularly, when the authors became concerned with counts in $10' \times 10'$ squares, this model also broke down because it appeared that the Beta distribution applicable to counts made during one period of some six months will not fit the counts made during another such period. Still, in the absence of anything better, model II continues to be used, with misgivings.

In principle, it is possible to count all, or practically all, the images of galaxies present on a photographic plate. This has been done in the past and a good description of the results is to be found in a paper [5] by Mayall. The process is illustrated in figure 8 reproducing sections of two different photographic plates taken from the same region in the sky. The plate below has a fainter limiting magnitude than the plate above. Thus all the objects, stars and galaxies, visible on the upper plate appear also on the one below and their images are much more distinct. By making a reasonably complete census of images of galaxies on the plate with the larger

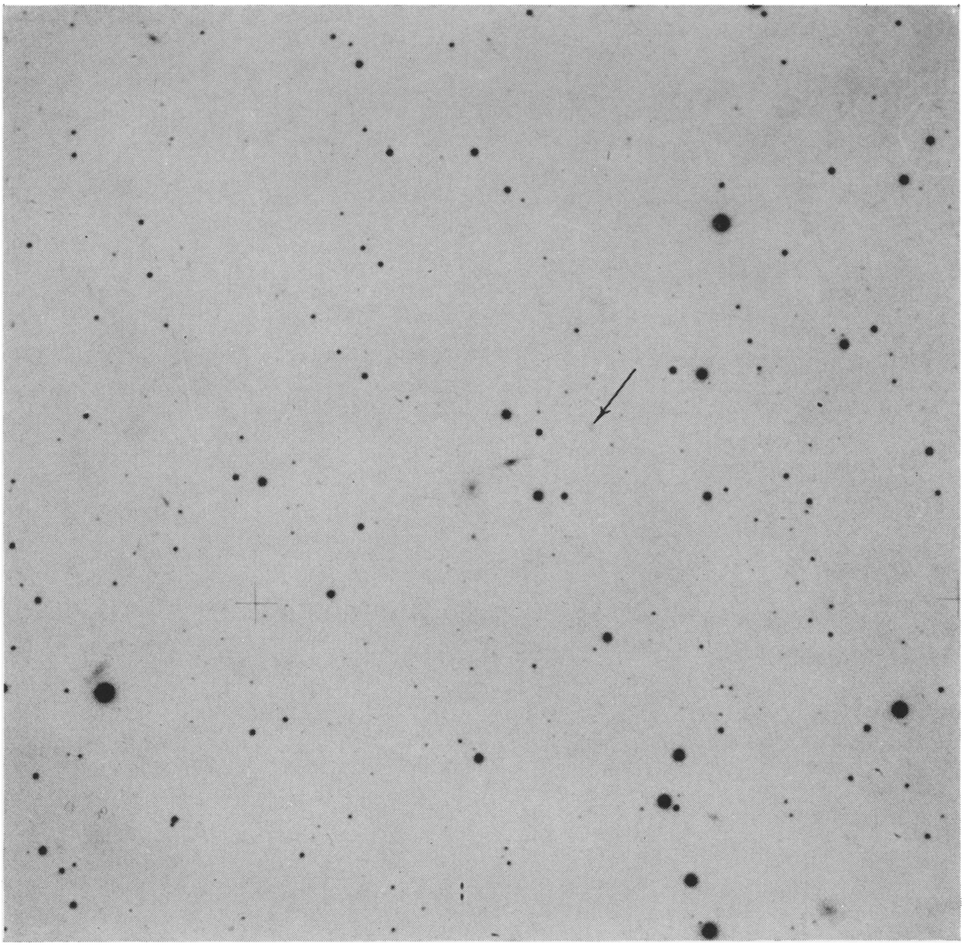

Figure 8a

Section of a plate taken by the 20″ astrographic telescope of the Lick Observatory.

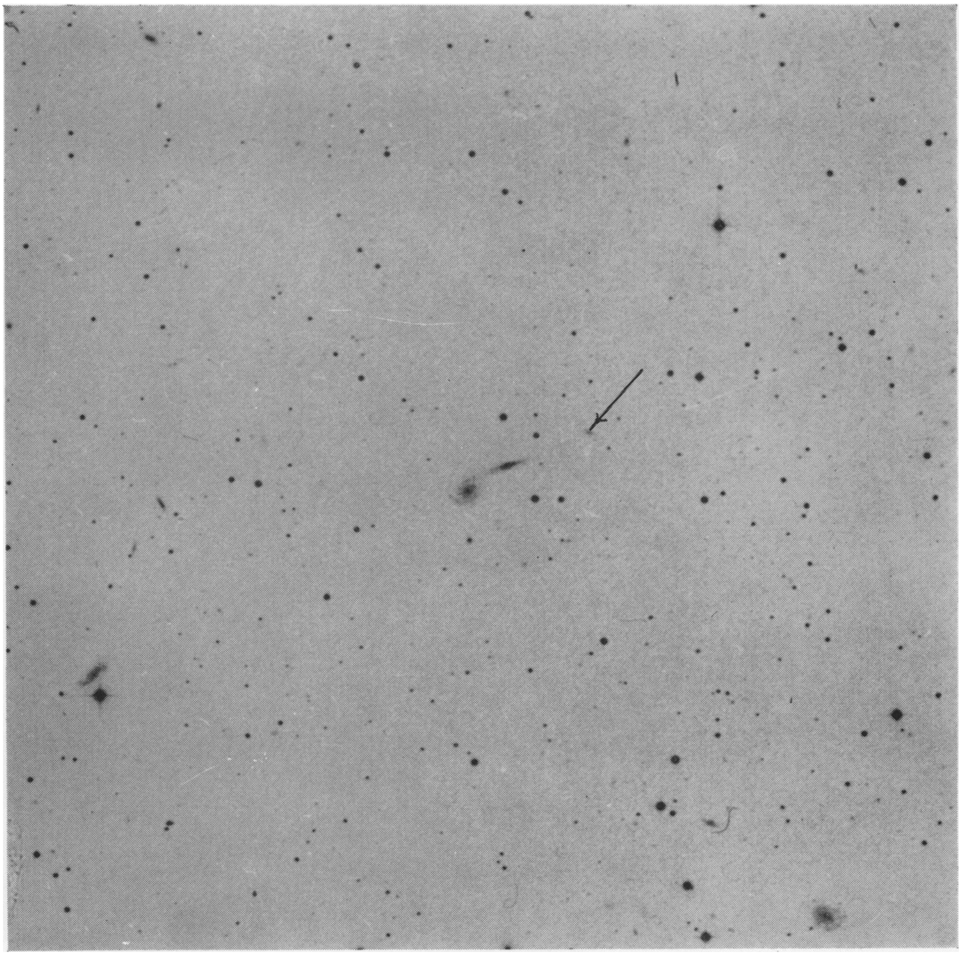

FIGURE 8b

Section of a plate taken by the 48″ Schmidt telescope of the Mount Palomar Observatory.

limiting magnitude and by measuring the coordinates of these images, it is possible to inspect the corresponding places on the plate with the brighter limiting magnitude and to trace the faint images of galaxies that may have been originally overlooked. Also, this method allows one to weed out the images of faint stars or imperfections of the emulsion that may have been counted as galaxies. Table II, based on data taken from Mayall, indicates the order of magnitude of mistakes in counting.

TABLE II

ERRORS IN COUNTING IMAGES OF GALAXIES

Total images of galaxies on 47 plates	1228
Images missed	571
Spurious images counted	81

It will be seen that, of the 1228 images of galaxies actually present on the photographic plates taken to a smaller limiting magnitude, over 46 per cent were missed on the original count. The inspection of figure 8 indicates that it is unrealistic to assume that the probability of being counted is the same for every image of a galaxy. Thus, for example, the two bright spiral galaxies in the central portions of two parts of figure 8 could hardly be missed by any observer. On the other hand (see arrows) the figure exhibits some very faint objects which can be detected on the "weaker" plate after it is compared with the "stronger" one but not so easily without such comparison. The conclusion is that a more realistic model might be built up by assuming that the probability of an image of a galaxy being counted is not constant but depends on this galaxy's apparent magnitude or surface brightness. Also, the results of counts and recounts of the same plates indicate the desirability of including in the model of the concept of "personal" limiting magnitude of the observer which should be considered as random, varying in value from one day to the next. Another necessary assumption within a realistic model must concern the variability of the limiting magnitude of the plate, caused by variation in observing conditions.

The assumptions that might be profitably included in a realistic model of errors in counting could be further multiplied. However, in so doing it is necessary to remember two paramount requirements. One is that, in order to be useful, the model must be sufficiently uncomplicated (this word "uncomplicated" fits the situation better than the customary word "simple") so as to lend itself to numerical computations. The other requirement is that a model be experimentally justifiable. In order to determine whether a given model is justifiable or not, it is unavoidable to compare it with relevant observations. In the present case, the adequacy of an error model might be tested on a reasonable number of duplicate plates taken from the same region of the sky, to the same intended limiting magnitude and counted independently at least twice. With this kind of data it may be possible to evaluate separately the component error due to the plate and the component error due to the observer. The work on this problem is in progress.

The authors are very grateful to several persons in the Statistical Laboratory who have carried out the computations described in this paper. They are especially indebted to Mrs. Jeanne Lovasich and Mrs. Marija Vasilevskis.

REFERENCES

[1] HARLOW SHAPLEY and ADELAIDE AMES, "A survey of the external galaxies brighter than the thirteenth magnitude," *Annals of the Harvard College Observatory*, Vol. 88, No. 2 (1932), pp. 43–75.

[2] HARLOW SHAPLEY and collaborators. The results of these investigations are contained in a large number of publications of the Harvard College Observatory.

[3] E. P. HUBBLE, "The distribution of extragalactic nebulae," *Astrophysical Jour.*, Vol. 79 (1934), pp. 8–76.

[4] E. P. HUBBLE, "Effects of red shifts on the distribution of nebulae," *Astrophysical Jour.*, Vol. 84 (1936), pp. 517–554.

[5] N. U. MAYALL, "A study of the distribution of extragalactic nebulae based on plates taken with the Crossley reflector," *Lick Observatory Bull.* No. 458, Vol. 16 (1934), pp. 177–198.

[6] ANDERS REIZ, "A study of external galaxies with special regard to the distribution problem," *Annals of the Observatory of Lund*, No. 9 (1941), pp. 1–94.

[7] C. D. SHANE and C. A. WIRTANEN, "The distribution of extragalactic nebulae," *Astr. Jour.*, Vol. 59 (1954), pp. 285–304.

[8] F. ZWICKY, see "Annual Report of the Director of the Mount Wilson and Palomar Observatories," pp. 23–24, *Carnegie Institution of Washington Year Book*, No. 51 (1951–1952).

[9] E. P. HUBBLE, "Explorations in space. Cosmological program for the Palomar telescopes," *Proc. Amer. Phil. Soc.*, Vol. 95 (1951), pp. 461–470.

[10] J. NEYMAN and E. L. SCOTT, "A theory of the spatial distribution of galaxies," *Astrophysical Jour.*, Vol. 116 (1952), pp. 144–163.

[11] J. NEYMAN, E. L. SCOTT and C. D. SHANE, "On the spatial distribution of galaxies: a specific model," *Astrophysical Jour.*, Vol. 117 (1953), pp. 92–133.

[12] J. NEYMAN and E. L. SCOTT, "Frequency of separation and of interlocking of clusters of galaxies," *Proc. Nat. Acad. Sci.*, Vol. 39 (1953), pp. 737–743.

[13] E. L. SCOTT, C. D. SHANE and M. D. SWANSON, "Comparison of the synthetic and actual distribution of galaxies on a photographic plate," *Astrophysical Jour.*, Vol. 119 (1954), pp. 91–112.

[14] J. NEYMAN, E. L. SCOTT and C. D. SHANE, "The index of clumpiness of the distribution of images of galaxies," *Astrophysical Jour. Supp.* No. 8, Vol. 1 (1954), pp. 269–294.

[15] J. NEYMAN, "Probabilistic theory of clustering of galaxies with particular reference to the hypothesis of an expanding universe," *Proceedings of the International Congress of Mathematicians*, 1954, Vol. 2, Amsterdam, North-Holland Publishing Co., 1954, pp. 298–299.

[16] E. L. SCOTT, "Distributions of certain characteristics of clusters of galaxies, with particular reference to the hypothesis of an expanding universe," *Proceedings of the International Congress of Mathematicians*, 1954, Vol. 2, Amsterdam, North-Holland Publishing Co., 1954, pp. 303–304.

[17] J. NEYMAN, "Sur la théorie probabiliste des amas de galaxies et le problème de verification de l'hypothese de l'expansion de l'univers," *Annales de l'Inst. Henri Poincaré*, Vol. 14 (1955), pp. 201–244.

[18] G. C. MCVITTIE, "Relativity and the statistical treatment of the distribution of galaxies," *Astr. Jour.*, Vol. 60 (1955), pp. 105–115.

[19] H. BONDI, "*Cosmology*," Cambridge, The University Press, 1952.

[20] E. P. HUBBLE, "The luminosity function of nebulae, I," *Astrophysical Jour.*, Vol. 84 (1936), pp. 158–179.

[21] E. P. HUBBLE, "The luminosity function of nebulae, II," *Astrophysical Jour.*, Vol. 84 (1936), pp. 270–295.

[22] F. ZWICKY, "The Coma cluster of galaxies," *Publications Astr. Soc. of Pacific*, Vol. 63 (1951), pp. 61–71.

[23] C. D. SHANE, "Lick Observatory Report," *Astr. Jour.*, Vol. 58 (1953), pp. 245–253.

[24] F. ZWICKY, "Neue Methoden der kosmologischen Forschung," *Helvetica Physica Acta*, Vol. 26 (1953), pp. 241–254.

[25] E. P. HUBBLE, "The law of red shifts," *Monthly Notices of the Roy. Astr. Soc.*, Vol. 113 (1953) pp. 658–666.

[26] H. Shapley, "Luminosity distribution and average density of matter in twenty-five groups of galaxies," *Proc. Nat. Acad. Sci.*, Vol. 19 (1933), pp. 591–596.

[27] J. Neyman and E. L. Scott, "On the high order clustering of particles," in preparation.

[28] E. L. Scott, "Probability distributions of several characteristics of images of clusters of galaxies on photographic plates," in preparation.

[29] A. Shapiro, "Further contributions to the problem of interlocking of clusters," in preparation.

[30] S. Chandrasekhar and G. Münch, "The theory of the fluctuations in brightness of the Milky Way, V," *Astrophysical Jour.*, Vol. 115 (1952), pp. 103–123.

[31] D. N. Limber, "The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field," *Astrophysical Jour.*, Vol. 117 (1953), pp. 134–144.

[32] D. N. Limber, "Analysis of counts of the extragalactic nebulae in terms of a fluctuating density field, II," *Astrophysical Jour.*, Vol. 119 (1954), pp. 655–681.

[33] J. Neyman and E. L. Scott, "On the inapplicability of the theory of fluctuations to galaxies," *Astr. Jour.*, Vol. 60 (1955), pp. 33–38.

[34] J. Neyman and E. L. Scott, "Spatial distribution of galaxies—analysis of the theory of fluctuations," *Proc. Nat. Acad. Sci.*, Vol. 40 (1954), pp. 873–881.

[35] J. Neyman and E. L. Scott, "On the problem of expansion of clusters of galaxies," *Studies in Mathematics and Mechanics*, presented to Richard von Mises by Friends, Colleagues and Pupils, New York, Academic Press, 1954, pp. 336–345.

[36] V. A. Ambartzumian, "K Teorii Fluktuatsii v Vidimom Raspredelenii Zvezd Na Nebe," Akad. Nauk. Armianskoi S.S.R., *Soobsheniia Biurakanskoi Observatorii*, Vol. 6 (1951), pp. 5–61.

[37] S. Chandrasekhar and G. Münch, "The theory of fluctuations in brightness of the Milky Way," *Astrophysical Jour.*, Vol. 112 (1950), pp. 380–398; Vol. 114 (1951), pp. 110–122; Vol. 115 (1952), pp. 94–123.