FIELD GALAXIES: LUMINOSITY, REDSHIFT, AND ABUNDANCE OF TYPES. PART I. THEORY

JERZY NEYMAN and ELIZABETH L. SCOTT
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

The purpose of the present paper is to use the general theory of the spatial distribution of galaxies published earlier [1], [2], [3] in order to deduce formulas representing the observable distributions, joint, marginal and conditional, of the two important characteristics of field galaxies, namely their apparent magnitude and redshift. These formulas, compared with their empirical counterpart represented by the distributions constructed from the data in the catalogues of galaxies, provide means of testing some of the hypotheses underlying the theory and of estimating certain distributions as they exist in space. Specifically, the problems treated include (i) estimates of the luminosity functions of galaxies of specified types, (ii) the luminosity-redshift relation, (iii) the selection probability, and (iv) the relative abundance of various types of galaxies as they exist in space. Some of the formulas deduced here have been published in [4] without proof.

The paper is divided into two parts, theoretical and empirical. In the present theoretical part I an attempt is made to use only assumptions that are of qualitative character and to deduce results that may be of broader validity. Thus, for example, while discussing the luminosity function of field galaxies, our assumptions regarding it are limited only to the condition that certain integrals are convergent, but no specific parametric form of this function is postulated. The only quantitative hypothesis adopted is that the observations refer to relatively nearby galaxies for which the dimming due to redshift may be allowed for.

In the second part of the paper, to be published elsewhere, the theoretical result of part I will be applied to obtain specific information regarding field galaxies. Using the data of the well-known memoir [5] by Humason, Mayall and Sandage, an attempt will be made to obtain actual estimates of the luminosity functions of the several types of galaxies, the abundances of these types in space, and so forth. Here, then, it will be necessary for us to particularize certain

Work on this paper began at the University of California, Berkeley, with the partial support of the National Science Foundation, Grant G-14648, and was completed while the authors were guests of the Statistics Department, Stanford University, with the partial support of the Office of Naval Research.

assumptions and to adopt parametric interpolation formulas with the hope that, with an appropriate adjustment of parameters, these formulas will yield reasonable approximations to the distributions existing in space.

2. Basic assumptions and notation

The basic assumptions underlying the present paper fall under two separate headings. First there are the assumptions regarding the happenings in space, that is, regarding the stochastic model of the universe. These particular assumptions are exactly those used in our earlier publications quoted above. The second category of the basic assumptions needed in the present paper refers to the connection between the happenings in space and what is to be found in the existing or, possibly, in some future catalogues of galaxies. These particular assumptions have to be explained in some detail. First we do so in operational terms. Next we formulate them mathematically.

(i) Our first assumption is that the astronomer compiling a catalogue of galaxies is able to distinguish field galaxies from those belonging to clusters. Naturally, mistakes in this respect are unavoidable. However, the theory that follows presupposes that such mistakes are rare and ignores them. In fact, the theory presented in this paper applies to field galaxies only. The theoretical counterpart of a "field galaxy" is a "cluster" composed of only one member.

(ii) Whatever may be the equipment available to the astronomer, it is clear that this equipment cannot be sufficient to measure the apparent magnitude and the redshift of all the galaxies existing in space. For these measurements to be feasible, the objects must be sufficiently bright. Also, it is clear that the degree of brightness, as measured by the apparent magnitude, does not determine by itself whether a given object is suitable for observation. For example, the redshift of a compact elliptical galaxy of 13 mag is likely to be measured easily using a modest telescope. On the other hand, this need not be true of a galaxy with the same apparent magnitude if it is irregular and very diffuse.

The basic assumption used in the present paper is that the galaxies are classified into several types and that this classification is sufficiently fine so that, for any two galaxies G_1 and G_2 *belonging to the same type*, the relative ease of measuring the apparent magnitude and the redshift depends upon their apparent magnitude and on nothing else. In other words, it is assumed that, if G_1 and G_2 belong to the same type and have the same apparent magnitude, then the observations of magnitude and redshift for these two galaxies are equally difficult.

(iii) Our third assumption is that the catalogue of field galaxies has been compiled with reference to a specified region R in the sky, free from galactic obscuration, and with reference to the ease of measurements of the two characteristics, the apparent magnitude and the redshift. More specifically, we assume that, after selecting a suitable region R, the astronomer preparing the catalogue goes over a set of uniform survey plates taken over this region, marks all those objects which he feels are reasonably accessible to his instruments, classifies them according to the adopted system of types, and measures the apparent magnitude and redshift for all the objects marked.

It must be obvious that the existing catalogues need not conform exactly to the hypotheses made. In fact, in many cases, the existing catalogues represent combinations of results of several observational programs, each conducted by a different astronomer who may have different ideas not only on the relative ease of making measurements but also on the relative desirability of studying galaxies of different types. For example, cases are on record where particular galaxies were included in certain observational programs for the reason that the appearance and/or location of these objects made them especially interesting.

The theory given below is applicable only to catalogues in which the frequency of occurrences of the kind just described is negligible.

The mathematical counterpart of assumptions (ii) and (iii) is as follows. The basic mathematical assumption is that the inclusion in the catalogue of any field galaxy located in the region R is a random event with a fixed probability, not necessarily the same for all galaxies, and that for any two galaxies these events are independent.

Let s be the number of different types of galaxies considered and let m denote the photographic apparent magnitude of a galaxy. For the tth type, $t = 1, 2, \cdots, s$, and for any m, we postulate the existence of the probability $\Phi_t(m)$ that a field galaxy located in R, belonging to the tth type and having its apparent magnitude equal to m, will be included in the catalogue. This probability will be described as the selection probability.

In the present part of the paper the assumptions regarding the selection probabilities $\Phi_t(m)$ are limited to the postulate that, as $m \to \infty$, the function $\Phi_t(m)$ tends to zero sufficiently fast for certain integrals to converge. This is the counterpart of the fact that, whatever the astronomer's equipment, there is a limit of faintness beyond which galaxies are difficult to observe and, therefore, are not likely to be included in the catalogue.

One of the purposes of the present part of the paper is to provide means whereby the selection probabilities $\Phi_t(m)$ can be estimated from the data. In part II the relevant formulas are applied to obtain the estimates of $\Phi_t(m)$.

Let $\Lambda_1, \Lambda_2, \cdots, \Lambda_s$ denote the space densities of the s types of galaxies. In other words, Λ_t stands for the expected, or average, number of field galaxies of type t per unit volume in space. Also let $\Lambda = \sum_t \Lambda_t$. Then the quotients

$$\lambda_t = \frac{\Lambda_t}{\Lambda} \tag{1}$$

are positive numbers, adding up to unity, and represent the "space relative abundances" of the different types or, simply, the "space abundances" of these types. The space abundances are another subject of the present study, and must be distinguished clearly from the "catalogue abundances."

Denote by N the total number of field galaxies in the catalogue and by N_t the number of those that belong to type t. Then the fraction

$$n_t = \frac{N_t}{N} \tag{2}$$

is called the catalogue abundance of galaxies of type t. Clearly, the catalogue abundance n_t of a given type is related to the space abundance λ_t. However, this relation is not immediate and is affected by both the luminosity functions and by the selection probabilities for the different types. For example, if some two types t' and t'' are equally abundant, so that $\lambda_{t'} = \lambda_{t''}$, but the galaxies of type t' are generally fainter than those of type t'', it is plausible to suppose that $n_{t'}$ will be smaller than $n_{t''}$, and so forth.

In the sequel we shall consider probability densities of various random variables. The generic notation for the probability density of some variables $X, Y, \cdots, Z$ is $p_{X,Y,\cdots,Z}(x, y, \cdots, z)$, where $x, y, \cdots, z$ stand for particular values of the random variables considered. On several occasions it will be necessary for us to distinguish between the "space distribution" of certain random variables and their "catalogue distribution." This is the same kind of distinction as that discussed above between the space abundances λ_t and the catalogue abundances n_t of the various types of galaxies. In order to distinguish between the space probability density of some variables and their catalogue density, the latter will be marked with an asterisk. Thus $p_X(x)$ stands for the space density of X while $p^*_X(x)$ represents the catalogue density. In other words, $p^*_X(x)$ stands for the conditional probability density of X, given that the galaxy characterized by X has been selected for observation and for inclusion in the catalogue.

With each type of galaxy, we shall associate an appropriate luminosity function. In accordance with our earlier theory, the absolute magnitude of a galaxy will be considered as a random variable, independent of all other random variables in the system. It will be denoted by $\mathcal{M}$ and its particular values by M. The space luminosity function corresponding to the tth type of galaxies will be denoted by $p_{\mathcal{M}}(M|)$. This, then, is the conditional probability density of the absolute magnitude $\mathcal{M}$, given that the galaxy belongs to category t. In the present part of the paper no assumption will be made regarding the luminosity functions other than that certain integrals converge. The results obtained here will be used in part II in order to obtain estimates of $p_{\mathcal{M}}(M|)$ and then a certain parametric form of the function will be adopted.

The apparent magnitude μ of a galaxy at a fixed distance ξ and with redshift z, is given by the familiar formula

$$\begin{aligned} \mu &= \mathcal{M} - 5 + 5\log_{10}\xi + K(z) \\ &= \mathcal{M} - 5 + a\log\xi + K(z), \end{aligned} \tag{3}$$

where the last term represents the dimming due to redshift. For not very distant galaxies this term is close to zero and may be presumed not to affect the chances of including a galaxy in the catalogue. Also, the dimming effect $K(z)$ is approximately known so that the apparent magnitudes given in the catalogues may be corrected for redshift. For purposes of the theory we shall assume that the corrections are perfectly successful and already have been introduced. With this assumption, the term $K(z)$ in (3) disappears.

The coefficient $a = 5 \log_{10} e$ in the second line of formula (3) is introduced for convenience in dealing with natural logarithms rather than with logarithms to the base 10.

Because of the particular connection between μ and $\mathcal{M}$ the space conditional probability density $p_\mu(m|\xi, t)$ of the corrected apparent magnitude of galaxies of type t, given their distance ξ, is simply

$$p_\mu(m|\xi, t) = p_{\mathcal{M}}(m + t - a \log \xi|t). \tag{4}$$

The three term argument $m + 5 - a \log \xi$ on the right is bound to appear frequently in this paper. In order to achieve a typographical simplification, rather than consider the apparent magnitude (corrected for redshift) as usually defined, from now on we shall be concerned with the "modified" apparent magnitude defined as the usual apparent magnitude plus 5. The modified apparent magnitude considered as a random variable will again be denoted by μ and its particular values by m. Also, from now on, the symbol m appearing as the argument of $\Phi_t(m)$ will be understood to mean the modified apparent magnitude.

The distance of a galaxy contemplated as a random variable will be denoted by Ξ and its particular values by ξ. The last random variable to be considered is the redshift

$$Z = c\frac{\Delta\lambda}{\lambda}. \tag{5}$$

The particular values of Z will be denoted by z.

To be completely realistic one should postulate that for a given ξ the redshift Z has a conditional probability distribution. The problem was first studied by Hubble [6] who found that the conditional dispersion of Z given ξ is of the order of 200 km/sec. More recently, using a different method and a considerably larger amount of data, the problem was reexamined [7]. It was found that the data are consistent with the assumption that $\sigma_Z = 0$ and that, with probability of error not exceeding 5 in 100, one can assert that $\sigma_Z \leqq 150$ km/sec. In these circumstances, while recognizing the desirability of a theory treating Z as a random variable with a nondegenerate conditional distribution given ξ, in this paper we postulate that Z is simply proportional to Ξ, so that, for all ξ we have $z = H\xi$, where H is the Hubble constant. As a result, the study of the distribution of Z becomes equivalent to that of the distribution of Ξ and, for example, the problem of the magnitude-redshift relation is equivalent to the problem of the magnitude-distance relation, subject only to a change in scale.

3. The fundamental formula of the theory of clustering

The present study is essentially based on the fundamental formula of the theory of simple clustering as published in [1] and then generalized in [2]. The formula gives the joint probability generating function of the numbers N_1,

$N_2, \cdots, N_r$ of galaxies located in a certain number r of disjoint regions in space, say $R_1, R_2, \cdots, R_r$, and, in a sense, "successful." The general theory leaves the definition of "success" free. In our case, there will be several types of "success," which will be discussed in some detail. One kind of "success" may be that the galaxy in question be of type *SO and* that it have its apparent magnitude between limits 13 to 14, *and* that it be included in the catalogue. Another kind of success may be that the galaxy be of type *EO and* that it be included in the catalogue, and so forth. We shall denote by s the total number of different "successes" of the above kind and by N_{ij} the number of galaxies located in R_i and having success of the kind j. The bold face letters **N** and **v** will stand for the vectors $\mathbf{N} = (N_{11}, N_{12}, \cdots, N_{rs})$ and $\mathbf{v} = (v_{11}, v_{12}, \cdots, v_{rs})$. The fundamental formula gives then the probability generating function

$$G_{\mathbf{N}}(\mathbf{v}) = \exp\left(-\lambda \int \left\{1 - G_\nu\left[1 - \sum_{i=1}^{r} \sum_{j=1}^{s} p_{ij}(\mathbf{u})(1 - v_{ij})\right]\right\} d\mathbf{u}\right) \tag{6}$$

of the random variables N_{ij}, where λ stands for total density of cluster centers per unit volume, $\mathbf{u} = (u_1, u_2, u_3)$ denotes the coordinates of a cluster center, the integral on the right is a triple integral over $-\infty < u_1, u_2, u_3 < +\infty$, where $p_{ij}(\mathbf{u})$ stands for the probability that a galaxy from a cluster centered at $\mathbf{u}$ will be found in the region R_i and that it will have "success" of kind j, and, finally, where $G_\nu(\tau)$ is the probability generating function of the number ν of galaxies per cluster.

For the present needs this general formula has to be specialized in accordance with the various limitations of the study. Namely, we take into account that we are concerned with field galaxies only, that is, with galaxies belonging to clusters of just one member each. Thus, for each "cluster" considered, $\nu = 1$ with probability unity. It follows that $G_\nu(\tau) \equiv \tau$. As a result, formula (6) reduces to

$$G_{\mathbf{N}}(\mathbf{v}) = \exp\left[-\Lambda \sum_{i=1}^{r} \sum_{j=1}^{s} (1 - v_{ij}) \int p_{ij}(\mathbf{u})\, d\mathbf{u}\right]. \tag{7}$$

Inspection of formula (7) indicates that the numbers N_{ij} are mutually independent random variables, each following a Poisson law with expectation $E(N_{ij}) = \Lambda \int p_{ij}(\mathbf{u})\, d\mathbf{u}$.

In order to evaluate this integral we must deal with the conditional probability density $f(\mathbf{x} - \mathbf{u})$ of the coordinates $\mathbf{X} = (X_1, X_2, X_3)$ of the galaxy, given that it belongs to the cluster centered at $\mathbf{u}$. Also, it is necessary to specify the meaning of success and the probability, say $\theta_{ij}(\mathbf{x})$ of the "jth success" in region R_i, given that the galaxy is located at $\mathbf{x} = (x_1, x_2, x_3)$. Whatever this probability may be,

$$p_{ij}(\mathbf{u}) = \int_{R_i} f(\mathbf{x} - \mathbf{u})\theta_{ij}(\mathbf{x})\, d\mathbf{x}, \tag{8}$$

where the integral is again in three dimensions, extending over the region R_i. Substituting (8) into $E(N_{ij})$ and changing the order of integration, we have

$$E(N_{ij}) = \Lambda \int_{R_i} \theta_{ij}(\mathbf{x})\, d\mathbf{x} \int f(\mathbf{x} - \mathbf{u})\, d\mathbf{u}. \tag{9}$$

Because f is a probability density, its integral over the whole space must be equal to unity. Therefore (9) reduces to

$$E(N_{ij}) = \Lambda \int_{R_i} \theta_{ij}(\mathbf{x})\, d\mathbf{x}. \tag{10}$$

4. Space and catalogue abundances of different types of galaxies

Our first and simplest specialization of the general concept of "success" of galaxies leads to the joint probability distribution of the catalogue numbers N_t of galaxies belonging to the specified type t.

In this case we consider just one region R in space, the solid angle subtended by the region in the sky for which the catalogue is being compiled. Thus, in the present case $r = 1$. For each galaxy within this region R we consider s different exclusive kinds of success, the tth of them consisting in (a) the galaxy belonging to type t and (b) the galaxy being included in the catalogue. According to the preceding section, the number N_t of galaxies successful in this particular sense is a Poisson variable, independent of N_i if $i \neq t$, and has expectation computable from a formula analogous to formula (10), namely,

$$E(N_t) = \Lambda \int_R \theta_t(\mathbf{x})\, d\mathbf{x}, \tag{11}$$

where $\theta_t(\mathbf{x})$ stands for the probability that a field galaxy located at $\mathbf{x}$ belongs to type t (the probability of this is λ_t) and that it will be included in the catalogue compiled in accordance with the assumptions enumerated in section 2. It follows that

$$\theta_t(x) = \lambda_t \int_{-\infty}^{+\infty} \Phi_t(m) p_{\mathcal{M}}(m - a \log \xi)\, dm \tag{12}$$

and, because $\Lambda\lambda_t = \Lambda_t$,

$$E(N_t) = \Lambda_t \int_{-\infty}^{+\infty} \Phi_t(m) \int_R p_{\mathcal{M}}(m - a \log \xi)\, d\mathbf{x}\, dm, \tag{13}$$

where ξ is the distance of the point $\mathbf{x}$ from the origin. Taking into account that the region R is a solid angle with its vertex at the origin of coordinates, the integral over R is easily calculated. We begin by introducing the polar coordinates

$$\begin{aligned} x_1 &= \xi \cos \varphi \cos \psi \\ x_2 &= \xi \cos \varphi \sin \psi \\ x_3 &= \xi \sin \varphi \end{aligned} \tag{14}$$

with $0 < \psi \leqq 2\pi$ and $|\varphi| < \pi/2$. Then, denoting by $\bar{R}$ the region of variation of φ and ψ within R, we obtain

$$\int_R p_{\mathcal{M}}(m - a \log \xi)\, d\mathbf{x} = k \int_0^{\infty} \xi^2 p_{\mathcal{M}}(m - a \log \xi)\, d\xi, \tag{15}$$

where

$$k = \iint_{\bar{R}} \cos d\varphi\, d\psi \tag{16}$$

is a numerical constant depending upon the region in the sky covered by the catalogue. Substituting (15) in (13) and replacing ξ by a new variable of integration τ connected with it by the formula

$$m - a \log \xi = \tau, \tag{17}$$

we obtain

$$E(N_t) = \frac{k}{a} \Lambda_t I_t J_t = \frac{k}{a} \Lambda_t x_t \tag{18}$$

(say), with

$$I_t = \int_{-\infty}^{+\infty} \Phi_t(m) e^{3m/a} \, dm \tag{19}$$

and

$$J_t = \int_{-\infty}^{+\infty} e^{-3\tau/a} p_{\mathcal{M}}(\tau|t) \, d\tau. \tag{20}$$

At this point we make our restrictive assumptions regarding $\Phi_t(m)$ and $p_{\mathcal{M}}(M|t)$ to the effect that both integrals (19) and (20) converge.

It will be noticed from (18) to (20) that the distribution of N_t depends both upon the selection probabilities and upon the luminosity functions of the different types of galaxies. It follows that the space abundances of the different types cannot be estimated without the earlier estimation of $\Phi_t(m)$ and of $p_{\mathcal{M}}(M|t)$. If this is done then the estimates of the abundances λ_t are obtained as follows.

Because the catalogue numbers N_t are all independent Poisson variables, their sum N is also a Poisson variable. The conditional distribution of the N_t given N is known to be multinomial with probability generating function

$$\begin{aligned} G_{N_1, N_2, \cdots, N_s}(v_1, v_2, \cdots, v_s) &= \left[\sum_{i=1}^{s} \frac{E(N_i)}{E(N)} v_i\right]^N \\ &= \left[\sum_{i=1}^{s} \frac{\lambda_i x_i}{x.} v_i\right]^N, \end{aligned} \tag{21}$$

where, for the sake of brevity, $x. = \sum_i \lambda_i x_i$.

This being the case, the maximum likelihood estimate $\hat{\lambda}_t$ of the space abundance of type t field galaxies is

$$\hat{\lambda}_t = \frac{N_t/x_t}{\sum_{i=1}^{s} N_i/x_i} = \frac{n_t/x_t}{\sum_{i=1}^{s} n_i/x_i}. \tag{22}$$

The precision of this estimate is, of course, of considerable interest. It can be measured by the asymptotic standard error of $\hat{\lambda}_t$ or, equivalently, by its square, the asymptotic variance. In computing the latter it is necessary to be clear about the sources of variation in (22). First there are the catalogue abundances n_i and we have

$$E(n_i) = \frac{x_i\lambda_i}{x.}$$

(23) $$\text{Var}\,(n_i) = \frac{1}{N}\frac{x_i\lambda_i}{x.}\left(1 - \frac{x_i\lambda_i}{x.}\right)$$

$$\text{Cov}\,(n_i, n_j) = -\frac{1}{N}\frac{x_i\lambda_i}{x.}\frac{x_j\lambda_j}{x.}.$$

In addition to the catalogue abundances n_i, the estimate of the space abundance λ_t depends upon the quantities x_i defined by (18) to (20) and dependent upon the selection probabilities $\Phi_t(m)$ and upon the luminosity functions $p_M(M|t)$ for each of the galaxy types considered. Whatever may be the a priori considerations regarding these functions, it is clear that eventually they will have to be estimated from the data subject to chance variation. Thus the quantities x_i have to be considered as random variables also. The distribution of x_i will depend upon the formulas used to approximate $\Phi_t(m)$ and $p_M(M|t)$ and also on the method employed in estimating them. This will be dealt with in part II of this paper. For the present we notice that for $i \neq j$, the quantities x_i and x_j will have to be estimated from the distributions of magnitude and redshift of galaxies of two different types and thus, given N_i and N_j, the random variables x_i and x_j must be independent. Also, they are not correlated with the n_k for $k = 1, 2, \cdots, s$. Taking this into account and denoting the asymptotic variance of x_i by $\sigma^2_{x_i}$, we can write the general formula for the asymptotic variance of $\hat{\lambda}_t$ as follows, say

(24) $$\sigma^2_{\hat\lambda_t} = \sum_{i=1}^{s}\left[\left(\frac{\partial\hat\lambda_t}{\partial n_i}\right)^2 \text{Var}\,(n_i) + \left(\frac{\partial\hat\lambda_t}{\partial x_i}\right)^2 \sigma^2_{x_i}\right] + 2\sum_{i=1}^{s-1}\sum_{j=i+1}^{s}\frac{\partial\hat\lambda_t}{\partial n_i}\frac{\partial\hat\lambda_t}{\partial n_j}\text{Cov}\,(n_i n_j),$$

where the derivatives have to be evaluated at $n_i = x_i\lambda_i/x.$. We have

(25) $$\frac{\partial\hat\lambda_t}{\partial n_i} = -\frac{x.\lambda_t}{x_i}, \qquad \frac{\partial\hat\lambda_t}{\partial x_i} = \frac{\lambda_t\lambda_i}{x_i}$$

for $i \neq t$, and

(26) $$\frac{\partial\hat\lambda_t}{\partial n_t} = \frac{x.}{x_t}(1 - \lambda_t), \qquad \frac{\partial\hat\lambda_t}{\partial x_t} = -\frac{1}{x_t}\lambda_t(1 - \lambda_t).$$

It follows that

(27) $$\sigma^2_{\hat\lambda_t} = \frac{x.}{N}\frac{\lambda_t}{x_t}\left[1 - 2\lambda_t + \lambda_t x_t \sum_{i=1}^{s}\frac{\lambda_i}{x_i}\right] + \lambda_t^2\left[(1 - 2\lambda_t)\left(\frac{\sigma_{x_t}}{x_t}\right)^2 + \sum_{i=1}^{s}\left(\frac{\lambda_i\sigma_{x_i}}{x_i}\right)^2\right].$$

It will be seen that, if the x_i are known, so that $\sigma_{x_i} = 0$, and if they are all equal, then $\hat\lambda_t = n_t$ and (27) reduces to the familiar formula $\lambda_t(1 - \lambda_t)/N$.

5. Catalogue distribution of magnitude and distance of field galaxies

Because of the independence of the catalogue abundances of particular types of galaxies, the problem of the catalogue distribution of magnitude and distance can be treated separately for particular types. Thus, all considerations in the present section refer to field galaxies of one particular type and the subscript t referring to the type of galaxies considered will be omitted throughout.

Our first purpose is to deduce the formula giving $p^*_{\Xi,\mu}(\xi, m)$, the joint catalogue probability density of the distance Ξ and of the modified apparent magnitude μ. In other words, our purpose is to find the conditional density of these variables, given that the galaxy to which they refer has been included in the catalogue.

Let N denote the total number of galaxies (of the specified type) in the catalogue. For $i = 1, 2, \cdots, N$ let $\{(\xi_i, \xi_i + \delta_i]\}$ be a sequence of N nonoverlapping but otherwise arbitrary intervals all located on the axis of distances between zero and infinity. Also, let $\{(m_i, m_i + \Delta_i]\}$ be N arbitrary nonoverlapping intervals on the axis of apparent magnitudes, $-\infty < m_i < m_i + \Delta_i < +\infty$, for $i = 1, 2, \cdots, N$.

We begin by using the fundamental formula (7) in order to obtain the conditional probability (given that the catalogue contains exactly N galaxies of the specified type) that the ith of these galaxies will have its distance between ξ_i and $\xi_i + \delta_i$ and its modified apparent magnitude between m_i and $m_i + \Delta_i$. When this probability is obtained, it will be divided by the product of all δ_i and all Δ_i. Then a passage to the limit, when the increments δ_i and Δ_i tend to zero, will yield the desired probability density.

In order to apply formula (7) we divide the solid angle R into $N + 1$ disjoint regions. The ith of the first N regions, say R_i, is composed of that part of R characterized by the distance ξ from the observer between the limits

$$\xi_i < \xi \leqq \xi_i + \delta_i, \qquad i = 1, 2, \cdots, N. \tag{28}$$

The last region R_{N+1} represents the remainder of R, after the removal of $R_1, R_2, \cdots, R_N$. (Obviously R_{N+1} is composed of $N + 1$ disjoint regions but this circumstance is entirely immaterial.)

For each of the first N regions R_i we define the "success" of a galaxy as a compound event, consisting of the galaxy having modified apparent magnitude between the limits m_i and $m_i + \Delta_i$ and of its being included in the catalogue. The "success" of a galaxy in R_{N+1} will be defined as consisting of this galaxy's being included in the catalogue, irrespective of the magnitude it may have.

Denoting, as formerly, by $N_1, N_2, \cdots, N_{N+1}$ the numbers of galaxies successful in each of the $N + 1$ regions, we can use formula (7) to write the joint probability generating function of these variables. This probability generating function can then be used to compute the probability $P\{\cap_{i=1}^{N}(N_i = 1) \cap (N_{N+1} = 0)\}$ that in the region R there will be included in the catalogue exactly N galaxies of which exactly one satisfies the condition

$$\xi_i < \Xi \leqq \xi_i + \delta_i, \qquad m_i < \mu \leqq m_i + \Delta_i. \tag{29}$$

This probability is

$$(30)\qquad \Lambda^N \prod_{j=1}^{N} \int p_j(\mathbf{u})\, d\mathbf{u} \exp\left[-\Lambda \sum_{i=1}^{N+1} \int p_i(\mathbf{u})\, d\mathbf{u}\right]$$

where $p_j(\mathbf{u})$ means the probability that a galaxy from a cluster centered at $\mathbf{u}$ will be in R_j and will be "successful" in it. For $j = 1, 2, \cdots, N$ we have

$$(31)\qquad p_j(\mathbf{u}) = \int_{R_j} f(\mathbf{x} - \mathbf{u}) \int_{m_j}^{m_j+\Delta_j} \Phi(m) p_{\mathcal{M}}(m - a \log \xi)\, dm\, d\mathbf{x},$$

with $\xi^2 = x_1^2 + x_2^2 + x_3^2$. Also

$$(32)\qquad p_{N+1}(\mathbf{u}) = \int_{R_{N+1}} f(\mathbf{x} - \mathbf{u}) \int_{-\infty}^{+\infty} \Phi(m) p_{\mathcal{M}}(m - a \log \xi)\, dm\, d\mathbf{x}.$$

Performing calculations exactly similar to those leading to formulas (8) to (10), we obtain

$$(33)\qquad \int p_j(\mathbf{u})\, d\mathbf{u} = k \int_{m_j}^{m_j+\Delta_j} \Phi(m) \int_{\xi_j}^{\xi_j+\delta_j} \xi^2 p_{\mathcal{M}}(m - a \log \xi)\, d\xi\, dm, \qquad j = 1, 2, \cdots, N$$

and

$$(34)\qquad \int p_{N+1}(\mathbf{u})\, d\mathbf{u}$$

$$= k \int_{-\infty}^{+\infty} \Phi(m) \left[\int_0^{\infty} \xi^2 p_{\mathcal{M}}(m - a \log \xi)\, d\xi - \sum_{i=1}^{N} \int_{\xi_i}^{\xi_i+\delta_i} \xi^2 p_{\mathcal{M}}(m - a \log \xi)\, d\xi\right] dm.$$

The unconditional probability that the catalogue compiled for the region R will contain exactly N galaxies of the type considered is

$$(35)\qquad \frac{1}{N!}\left[\Lambda \int p_0(\mathbf{u})\, d\mathbf{u}\right]^N \exp\left[-\Lambda \int p_0(\mathbf{u})\, d\mathbf{u}\right],$$

where

$$(36)\qquad \int p_0(\mathbf{u})\, d\mathbf{u} = k \int_{-\infty}^{+\infty} \Phi(m) \int_0^{\infty} \xi^2 p_{\mathcal{M}}(m - a \log \xi)\, d\xi\, dm.$$

The value of this integral already has been calculated in the process of obtaining (18). Dividing (30) by (35) and numbering the N galaxies in some particular order (there are $N!$ different ways of doing so) we find that the conditional probability, given the total number N of the specified type of galaxies in the catalogue, that the ith of them will satisfy (29) for $i = 1, 2, \cdots, N$, is represented by the formula

$$(37)\qquad \prod_{i=1}^{N} \frac{\int p_i(\mathbf{u})\, d\mathbf{u}}{\int p_0(\mathbf{u})\, d\mathbf{u}} \exp\left\{-\Lambda\left[\sum_{j=1}^{N+1} \int p_j(\mathbf{u})\, d\mathbf{u} - \int p_0(\mathbf{u})\, d\mathbf{u}\right]\right\}.$$

Now divide (37) by the product $\prod_{i=1}^{N} \delta_i \Delta_i$ and pass to the limit as each of the increments δ_i and Δ_i tends to zero. With reference to (33), (34), and (36), it will be seen that the limit of the indicated ratio exists and is equal to the product

$$\prod_{i=1}^{N} C\, \xi_i^2\, \Phi(m_i) p_{\mathcal{M}}(m_i - a \log \xi_i), \tag{38}$$

where C is the reciprocal of the double integral appearing in the right side of (36). According to what was said before, the product (38) represents the joint probability density of the modified apparent magnitudes and of the distances of N galaxies, given that these N galaxies have been included in the catalogue. Because this probability density is represented by a product in which each factor corresponds to a different galaxy, the conclusion is that the couples of random variables (μ_i, Ξ_i) are mutually independent. Thus, dropping the subscript i, the general formula for the catalogue probability density of μ and Ξ of a field galaxy can be written as

$$p^*_{\mu,\Xi}(m, \xi) = C\, \xi^2 \Phi(m) p_{\mathcal{M}}(m - a \log \xi). \tag{39}$$

Formula (39) is basic for the present section and has several interesting implications.

5.1. *Catalogue distribution of the apparent magnitude.* Integrating (39) for ξ from zero to infinity and using (13) to (18), we obtain the catalogue probability density of the corrected modified apparent magnitude μ, namely,

$$p^*_\mu(m) = \frac{\Phi(m) e^{3m/a}}{\int_{-\infty}^{+\infty} \Phi(x) e^{3x/a}\, dx}. \tag{40}$$

This formula is remarkable for the reason that, perhaps contrary to intuitive expectation, it shows that the catalogue distribution of the apparent magnitude corrected for dimming is completely determined by the selection probability $\Phi(m)$ and has no relation to the distribution of absolute magnitude in space. Thus, while it would be futile to attempt to use the catalogue distribution of apparent magnitude to estimate the luminosity function, the empirical catalogue distribution of apparent magnitudes of field galaxies of specified type can be used to estimate the selection probability $\Phi(m)$.

Strictly speaking, this can be done only up to a constant factor

$$\Phi(m) = p^*_\mu(m) e^{-3m/a} \int_{-\infty}^{+\infty} \Phi(x) e^{3x/a}\, dx. \tag{41}$$

However, in cases where a catalogue has been compiled making an effort to have it complete at least for the brightest galaxies, the difficulty may be avoided by postulating that for very bright galaxies $\Phi(m) = 1$, which determines the constant factor. The details of estimating $\Phi(m)$ will be dealt with in part II of the present paper.

5.2. *Catalogue distribution of the absolute magnitude of field galaxies.* Formula (39) can be used to calculate the catalogue joint distribution of two variables, the modified apparent magnitude μ and the absolute magnitude $\mathcal{M}$ of field galaxies. This latter distribution determines the marginal catalogue distribution of $\mathcal{M}$. Our first step consists in passing from the system of variables (μ, Ξ) to the system $(\mu, \mathcal{M})$, where

$$\mathcal{M} = \mu - a \log \Xi. \tag{42}$$

Following the familiar steps we obtain

$$p^*_{\mu,\mathcal{M}}(m, M) = \frac{c}{a}\left[\Phi(m)e^{3m/a}\right]\left[p_{\mathcal{M}}(M)e^{-3M/a}\right]. \tag{43}$$

This again is a somewhat surprising result. Because of the indicated factorization, the absolute magnitude and the apparent magnitude of galaxies in a catalogue are mutually independent. Furthermore, the catalogue distribution of the absolute magnitude, given by

$$P^*_{\mathcal{M}}(M) = cp_{\mathcal{M}}(M)e^{-3M/a}, \tag{44}$$

is independent of the selection probability $\Phi(m)$. As a result, so long as the compilation of a catalogue of field galaxies of a specified type depends only upon their apparent magnitude, the distribution of absolute magnitude among the selected galaxies will be the same whether this selection is made for measurement with a 200 inch or with the limited capabilities of a much smaller telescope!

If one acts on the assumption that the redshift Z of a galaxy does not differ very much from the product $H\xi$ of the Hubble constant and the distance (which, of course, is a frequent assumption and is adopted in the present paper), then the absolute magnitude M of a galaxy with modified apparent magnitude m and redshift Z can be estimated by the familiar formula

$$\begin{aligned} M &= m + a \log H - a \log Z \\ &= m + 5 \log_{10} H - 5 \log_{10} Z. \end{aligned} \tag{45}$$

Using estimates of this kind, an empirical counterpart of $p_{\mathcal{M}}(M)$ can be constructed. By multiplying each ordinate by exp $[3M/a]$ and by norming, a direct estimate of the space probability density $p_{\mathcal{M}}(M)$ of the absolute magnitude is obtained, namely

$$p_{\mathcal{M}}(M) = c^{-1}e^{3M/a}p^*(M). \tag{46}$$

A particular case of formula (46) was first found by Malmquist [8] with reference to his study of the space distribution of stars. Malmquist's assumption included the hypothesis that the space luminosity function of stars is normal. This implied that the catalogue luminosity function is also normal with a slightly different mean. The present formula (46) was first published in [4] without proof. It indicates that no a priori assumption regarding the space luminosity function is necessary and that the shape of this curve may be judged directly from the shape of the catalogue distribution of $\mathcal{M}$, multiplied by the factor exp $[3M/a]$. An analysis of this kind will be found in part II of the present paper.

5.3. *Catalogue distribution of the distance of galaxies.* The integral of (39) for m from $-\infty$ to $+\infty$ yields the catalogue distribution of the distance Ξ of a galaxy from the observer,

$$p^*_{\Xi}(\xi) = c\xi^2 \int_{-\infty}^{+\infty} \Phi(m)p_{\mathcal{M}}(m - a \log \xi)\, dm, \tag{47}$$

or, substituting the value of the norming factor c,

$$p_{\Xi}(\xi) = a \frac{\xi^2 \int_{-\infty}^{+\infty} \Phi(m) p_{\mathcal{M}}(m - a \log \xi)\, dm}{\int_{-\infty}^{+\infty} \Phi(m) e^{3m/a}\, dm \int_{-\infty}^{+\infty} e^{-3t/a} p_{\mathcal{M}}(t)\, dt}. \tag{48}$$

Formula (48) points to a fallacy occasionally encountered in the literature. This fallacy consists in treating the product $c\xi^2$ as representing the space probability density of the distance Ξ of field galaxies, valid for all ξ from zero to infinity. This, of course, is nonsense. The probability density integrated over the whole range of its validity must yield unity. However, the integral of $c\xi^2$, taken from zero to infinity, diverges. In other publications, $c\xi^2$ is treated as the catalogue density of distance Ξ, on the assumption that this catalogue is complete up to a certain limiting magnitude m_1. This assumption of completeness is equivalent to that of $\Phi(m)$ being equal to unity for $m < m_1$, and perhaps to zero elsewhere. On this assumption (48) yields

$$p^*_{\Xi}(\xi) = \frac{3\xi^2 \int_{-\infty}^{m_1 - a \log \xi} p_{\mathcal{M}}(t)\, dt}{\int_{-\infty}^{+\infty} e^{-3t/a} p_{\mathcal{M}}(t)\, dt}, \tag{49}$$

a formula which does not reduce to a simple product of ξ^2 by a constant but involves a dependence on the space luminosity function $p_{\mathcal{M}}(M)$ of the galaxies considered. In general, of course, the catalogue distribution of the distance Ξ depends on both the probability of selection and on the space luminosity function. As far as the space distribution of Ξ is concerned, it is degenerate: whatever the preassigned number ξ_1 the probability that Ξ will exceed ξ_1 is equal to unity. The product $c\xi^2$ represents the conditional space probability density of the distance Ξ given that the field galaxy concerned is at a distance between some fixed limits A and B, so that $A \leqq \Xi \leqq B$.

5.4. *The magnitude-distance relation.* The term magnitude-distance relation is customarily used to describe either the catalogue regression function of the distance Ξ on the apparent magnitude μ or, vice versa, the catalogue regression of the apparent magnitude μ on the distance Ξ. Operationally, the first of these interpretations means the average distance for a given apparent magnitude and the second the average apparent magnitude for a given distance. Both can be obtained from formula (39).

Dividing (39) by (40) we obtain the conditional catalogue distribution of Ξ given $\mu = m$,

$$p^*_{\Xi}(\xi | m) = c\xi^2 p_{\mathcal{M}}(m - a \log \xi), \tag{50}$$

where c is a norming constant. The kth moment of this distribution is

$$E(\Xi^k | m) = c \int_0^{\infty} \xi^{2+k} p_{\mathcal{M}}(m - a \log \xi)\, d\xi, \tag{51}$$

and simple calculations yield

(52) $$E(\Xi^k|m) = e^{km/a}\frac{Q_k}{Q_0} = 10^{km/5}\frac{Q_k}{Q_0}$$

where, generally,

(53) $$Q_k = \int_{-\infty}^{+\infty} e^{-(3+k)t/a} p_{\mathcal{M}}(t)\, dt.$$

Formula (52) is very interesting because it indicates that, irrespective of the process of selection on apparent magnitudes and irrespective of what might be the luminosity function $p_{\mathcal{M}}(M)$ of the galaxies considered, the average of any kth power of the distance Ξ of galaxies in the catalogue, with the same magnitude m, is proportional to the kth power of $10^{m/5}$. The only thing that depends upon the luminosity function is the coefficient of proportionality Q_k/Q_0.

The conditional variance $\sigma_{\Xi}^2(m)$ of Ξ given $\mu = m$ is of interest because its size is relevant to the question whether the apparent magnitude of a galaxy might be taken as a distance indicator. Simple calculations yield

(54) $$\sigma_{\Xi}^2(m) = 10^{2m/5}\frac{Q_2Q_0 - Q_1^2}{Q_0^2}.$$

Thus, the standard error of distance Ξ for a galaxy with its catalogue apparent magnitude m is proportional to $10^{m/5}$ and the ratio

(55) $$\frac{\sigma_{\Xi}(m)}{E(\Xi|m)} = \frac{(Q_2Q_0 - Q_1^2)^{1/2}}{Q_1}$$

is independent of m. It follows that the relative error in estimating the distance Ξ of a galaxy by its average value for a given $\mu = m$ is independent of the apparent brightness of the galaxy and is determined by the space luminosity function $p_{\mathcal{M}}(M)$ as indicated in (55) and (54). This result suggests investigating the random variable, say X, representing the proportional deviation of the value of Ξ from its conditional expectation given m, so that

(56) $$\Xi = (1 + X)\, E(\Xi|m) = (1 + X)e^{m/a}\frac{Q_1}{Q_0}.$$

Using (51) and performing easy calculations, we find

(57) $$p_X^*(x|m) = cp_{\mathcal{M}}\{a[\log Q_0 - \log Q_1 - \log(1 + x)]\},$$

a distribution which is independent of m. This result may be summed up as follows: whatever may be the modified apparent magnitude of a field galaxy, the conditional catalogue distribution of

(58) $$X = \frac{\Xi - E(\Xi|m)}{E(\Xi|m)}$$

representing the deviation of the galaxy's distance from its conditional expectation $E(\Xi|m)$, measured in terms of this expectation, is always the same and is determined by the luminosity function of the type of galaxies considered.

In the second part of this paper, after obtaining an estimate of $p_{\mathcal{M}}$, we shall calculate $p_X^*(x|m)$ and the quotient (55). It will be seen that the apparent magnitude of a field galaxy is a very poor distance indicator indeed.

Dividing (39) by (47), we obtain the catalogue conditional probability density of the apparent magnitude, given that the distance Ξ of a field galaxy is equal to ξ (or equivalently, that its redshift Z has the value $H\xi$),

$$p_\mu^*(m) = c\Phi(m)p_{\mathcal{M}}(m - a \log \xi), \tag{59}$$

where c is a norming factor depending upon ξ. It is obvious that the probability density (59) depends both upon the selection probability $\Phi(m)$ and upon the space luminosity function $p_{\mathcal{M}}(M)$. In part II of this paper, after estimating $\Phi(m)$ and $p_{\mathcal{M}}(M)$, we shall return to the study of (59).

REFERENCES

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

[2] J. NEYMAN, "Sur la théorie probabiliste des amas de galaxies et la vérification de l'hypothèse de l'expansion de l'univers," *Ann. Inst. H. Poincaré*, Vol. 14 (1955), pp. 201–244.

[3] J. NEYMAN and E. L. SCOTT, "Statistical approach to problems of cosmology," *J. Roy. Statist. Soc. Ser. B*, Vol. 20 (1958), pp. 1–43.

[4] ———, "Large scale organization of the distribution of galaxies," *Handbuch der Physik*, Vol. 53 (1959), pp. 416–444.

[5] M. HUMASON, N. U. MAYALL and A. R. SANDAGE, "Redshifts and magnitudes of extragalactic nebulae," *Astr. J.*, Vol. 61 (1956), pp. 97–162.

[6] E. HUBBLE, "The luminosity function of nebulae. II. The luminosity function as indicated by residuals in velocity-magnitude relations," *Astrophys. J.*, Vol. 84 (1936), pp. 270–295.

[7] J. NEYMAN and E. L. SCOTT, "Estimation of the dispersion of the redshift of field galaxies," *Astr. J.*, Vol. 66 (1961), pp. 148–155.

[8] K. G. MALMQUIST, "On some relations in stellar statistics," *Lund Medd.* I, No. 100 (1922) *Ark. Mat. Astro. Fys.*, Vol. 16 (1922), pp. 1–52.