Mon. Not. R. astr. Soc. (1971) **155**, 95–118.

A METHOD OF ALLOWING FOR KNOWN OBSERVATIONAL SELECTION IN SMALL SAMPLES APPLIED TO 3CR QUASARS

D. Lynden-Bell

(Received 1971 February 12)

SUMMARY

A useful method is developed which minimizes numerical fluctuations when deriving luminosity functions and density evolution from data subject to observational selection. The distribution of ratios of radio power to optical power for quasars is derived from the 3CR quasars; this is a steeply rising function in agreement with the observation that most quasars are radio quiet. The density evolution of the quasi-stellar sources is derived direct from the data and a possible interpretation is an exponential decay with cosmic time. The density evolution is then used to determine the optical luminosity function of quasars. There appears to be a sharp upper limit to the optical luminosity; however, at the faint end the luminosity function is still increasing exponentially, so that most quasar light comes from the dimmest quasars.

1. INTRODUCTION

Dim objects can only be seen nearby and bright objects are rare, so, in a sample complete to a given apparent magnitude, both classes may contribute significantly. Even when the observed distribution of intrinsic luminosities has been found for such a sample, there is still a problem in deriving the luminosity function. This function describes how many objects of each luminosity there are in a given volume rather than in the observed sample which has been biased by the cut off in apparent magnitude. The problem becomes worse when the number density of the whole class of objects is varying along the line of sight in an unknown manner.

On the assumption that the luminosity function is of the same form at all points along the line of sight but with a normalization that varies as the number density, we show how both the luminosity function and the number density may be deduced directly from the observations of the sample.

In the Appendix we show that the method adopted is the best possible.

The method is not restricted to luminosity functions and where other observational selections arise, as for objects selected for high proper motion, it may well prove useful. We shall demonstrate how to use the method on quasars where there is selection according to both ratio and optical apparent brightness, but we begin with a somewhat simpler problem in which this double selection reduces to a single selection.

Following Schmidt (1) we define:

(i) f_0 the optical flux f_ν at the wavelength to which 2500 Å would be red-shifted due to the recession of the object, the observer being situated at the Earth. The

units of f_0 are Watts $(m)^{-2}$ $(Hz)^{-1}$. In Schmidt's notation (1) f_0 was called $f(2500)$.

(ii) f_R the radio flux that would be observed under similar circumstances at the red-shifted 500 MHz frequency (instead of 2500 Å wavelength). Schmidt called this $f(500)$.

(iii) The total fluxes emitted by the object at 2500 Å and 500 MHz are called F_0 and F_R respectively. The units of these are Watts $(Hz)^{-1}$.

(iv) The log radio to optical power ratio, $\log_{10}(f_R/f_0)$, we denote by P. On the assumptions that the fluxes are emitted isotropically from the source and that any corrections for absorption have been applied so that the fluxes f_0 are already freed from absorption, then the diminution of flux with distance is the same in the radio and optical regions, so

$$P = \log(f_R/f_0) = \log(F_R/F_0).$$

Notice that P is similar in nature to a photometric colour like R–U on the *UBVRI*-photometric system. Radio astronomers may think of P as being similar to a spectral index determined all the way from the radio to the optical. In fact for a fictitious source with a power law spectrum $f_\nu \propto \nu^{-\alpha_1}$ all the way from the radio to the optical P would be $6{\cdot}38\alpha_1$.

(v) $S(178)$ is the actually observed 3CR flux at 178 MHz.

(vi) z is the redshift of the source.

Schmidt (1) has deduced each of these quantities save P from the observations. His results, and P which may be readily deduced, are repeated in Table I together with some quantities described later. F_0 and F_R depend on the cosmological model and we shall illustrate our methods by considering only the model with Hubble's constant $H_0 = 100\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, zero cosmical constant $\Lambda = 0$ and unit acceleration parameter, $q_0 = 1$.

Let us now consider the problem of determining the distribution of P among the quasars. The data are observed fluxes of 40 3CR quasi-stellar sources. Unfortunately, these do not form an unbiased sample of quasars in any given volume but have been selected to be both bright enough in the radio region to be included in 3CR and bright enough to have their optical identifications confirmed with redshifts. The data are insufficient to remove this bias by considering only that small volume within which all objects as dim as the intrinsically dimmest quasars can be seen. Within such a volume there is only one bright quasar 3C 273 and the number of quasars left in that sample could be counted on one hand. We can not remove the bias caused by selection of apparently bright objects without further assumption. However, Schmidt (2) has earlier found that the distribution of the numbers of quasars with P is independent of their absolute optical flux F_0. This he deduced from his finding that the redshift distribution of a sample of solely optically selected quasi-stellar objects at 18th magnitude was the same as the redshift distribution of the 18th magnitude 3CR quasi-stellar sources. (We use the terms quasi-stellar object and quasar interchangeably to cover objects with similar optical properties to the quasi-stellar sources. Quasars may or may not have been detected at radio wavelengths: quasi-stellar sources must have been.) Throughout our discussion we shall assume that the redshifts of quasars are cosmological so that they may be used as some possibly non-linear measure of distance. Schmidt's finding on the P distribution, together with our assumption

on the normalization of the luminosity function (called density evolution when applied to quasars), leads to a luminosity function, Φ, of the form:

$$\Phi = D(z)\,\phi_1(P)\,\phi_2(\log F_0).$$

Here ϕ_1 and ϕ_2 are distribution functions of their arguments and $D(z)$ is the comoving number-density of quasars as a function of redshift. We now define comoving densities and volumes. For each source it is convenient to define V, the comoving volume of that part of the Universe nearer than the source at the current epoch. In Milne's nomenclature this is the volume determined from a world map current at the Earth now, not a volume determined from a world picture in which distant objects are represented as they were the light travel time ago. A useful unit of volume for these investigations in the cubic giga parsec = $(\mathrm{Gpc})^3 = (10^9\,\mathrm{pc})^3 = 10^{27}\,(\mathrm{pc})^3$. We define the comoving cubic giga parsec $C(\mathrm{Gpc})^3$ as the volume of that fraction of the Universe which, when expanded to the present cosmic epoch, fills one cubic giga parsec of the current world map. Notice that a volume of one $C(\mathrm{Gpc})^3$ expands with the Universe. The comoving volumes V nearer than each quasar are calculated from the redshifts following the method described in reference (1) and are recorded in Table I. The comoving number-density of quasars is the number per unit comoving volume; thus, if quasars were born with the Universe and lived for ever their comoving density would be constant in spite of the expansion of the Universe. For such a Universe $D(z)$ would be independent of z which is far from being the case in this Universe. To avoid confusion of volumes V with photometric visual magnitudes we always refer to the latter as the V-photometric band.

Returning to our expression for Φ above we see that for objects situated at a distance corresponding to redshift z the comoving number density of quasars with log power ratios in the range P to $P+dP$ and with log optical flux between $\log F_0$ and $\log F_0 + d\log F_0$ is

$$\Phi\,dP\,d\log F_0 = D(z)\,\Phi_1(P)\,\Phi_2(\log F_0)\,dP\,d\log F_0.$$

This product form of luminosity function implies that the distribution of P is uncorrelated with apparent optical magnitude since that is a function of $\log F_0$ and z, each of which are uncorrelated with P in the above expression. This independence of the P distribution to a selection by optical apparent magnitude means that the only bias in our observed sample arises through the radio selection that the source must be bright enough to be included in 3CR. We quantify this selection by asking of each source by what factor $\beta \geqslant 1$ could all its radio flux have been reduced before its $S(178)$ flux would be reduced below the 9 flux unit limit of 3CR? The factor β is clearly $S(178)/9$. An object that had all its radio flux reduced in this way would have radio flux $f_{R\,\min} = f_R/\beta$ and a log power ratio $P_m = P - \log\beta$. Thus the selection on the P value of each object due to the 3CR limit of 9 f.u. is $P \geqslant P_m$. The values of P_m are calculated for each object and included in Table I. They differ considerably object to object.

We can now pose our basic problem in observational selection. Given a number of observed values of P each associated with a quasar and each with a known limit P_m such that that quasar would not have been included in the sample observed unless $P \geqslant P_m$, can we deduce the true distribution of P among quasars? Since optical selection is irrelevant under our assumptions and since 3CR claims completeness to 9 f.u. there is no further bias in our sample other than $P \geqslant P_m$. In

TABLE I

$z = \Delta\lambda/\lambda$ = redshift, $S(178)$ is the 3CR flux, α the spectral index, and the other quantities a explained in the text. For a change in the Hubble constant from 100 km s^{-1} Mpc^{-1} all volumes V_0, V_R change by a factor $(H/100)^{-3}$ and F_0 changes by a factor $(H/100)^{-2}$. However, the ma conclusions of this paper are invariant to changes in the Hubble constant since they depend on ratio

Source	z	$S(178)$	$-\alpha$	$-\log f_0$	P	P_m	$\log F_0$	V	V_0	V_R	$N(V_m)$	
273	0·158	67	0·26	27·63	3·36	2·49	23·80	0·29	48·5	3·80	4·3	3·
323·1	0·264	9·0	0·66	29·17	3·85	3·85	22·70	1·04	7·84	1·04	0·24	1·
249·1	0·311	11·5	0·89	28·75	3·46	3·35	23·27	1·54	23·3	2·02	1·05	1·
277·1	0·320	12·0	0·94	29·62	4·33	4·21	22·42	1·63	4·25	2·28	1·34	1·
48	0·367	47	0·44	29·12	4·58	3·86	23·04	2·23	15·7	1·33	0·74	1·
351	0·371	11·0	0·63	28·61	3·39	3·30	23·56	2·33	35·7	2·92	2·4	1·
215	0·411	10·0	1·01	29·84	4·45	4·40	22·42	2·86	4·25	3·24	3·0	1·
47	0·425	20	0·72	29·74	4·74	4·39	22·55	3·06	5·72	7·25	10·7	1
147	0·545	58	0·47	29·57	5·09	4·28	22·93	5·12	12·7	29·6	65	1
334	0·555	10·0	0·73	29·12	3·80	3·75	23·40	5·31	28·5	5·9	11·3	1·
275·1	0·557	16·0	0·91	30·29	5·13	4·88	22·20	5·32	—	—	—	—
345	0·594	10·0	0·28	29·02	3·81	3·76	23·56	6·09	35·7	6·90	16·6	2·
263	0·643	13·0	0·75	29·14	3·93	3·77	23·51	7·08	33·2	10·1	39	1·
207	0·683	10·0	0·72	29·91	4·59	4·54	22·79	7·80	9·57	8·75	28	1·
380	0·691	57	0·72	29·24	4·67	3·87	23·47	8·08	31·4	26·6	490	1
254	0·734	19·0	0·90	29·80	4·71	4·39	22·96	9·05	13·4	17·1	75	1
138	0·759	18·5	0·37	29·91	4·92	4·61	22·88	9·60	11·5	19·2	53	1
175	0·768	16·0	0·92	28·99	3·82	3·57	23·81	9·80	48·0	16·0	109	1·
286	0·848	21	0·05	29·57	4·68	4·31	23·32	11·6	25·0	27·6	300	1
454·3	0·860	15·0	0·20	29·12	4·06	3·84	23·78	11·9	47·4	20·0	195	1·
196	0·871	59	0·73	29·74	5·18	4·36	23·17	12·2	19·6	48·5	175	1
309·1	0·905	17·0	0·36	29·39	4·34	4·06	23·55	13·0	35·2	23·2	256	1·
336	0·927	13·5	0·89	29·67	4·43	4·25	23·29	13·5	23·8	18·4	153	1·
288·1	0·961	9·5	0·94	29·95	4·55	4·53	23·05	14·4	15·9	15·0	95	1·
245	1·029	9·5	0·52	29·61	4·27	4·25	23·44	16·0	30·0	16·5	113	1·
2	1·037	15	0·82	30·51	5·33	5·11	22·55	16·1	—	—	—	—
287	1·055	14·0	0·33	29·81	4·66	4·47	23·27	16·6	23·2	24·8	258	1
186	1·063	13·5	1·21	29·66	4·38	4·20	23·42	16·9	29·2	22·4	230	0·
208	1·109	16·0	0·84	29·63	4·47	4·22	23·49	18·0	32·2	27·4	360	1·
204	1·112	9·5	1·07	29·96	4·54	4·52	23·16	18·1	19·3	18·6	155	1·
181	1·382	13·0	0·96	30·06	4·79	4·63	23·24	24·5	—	—	—	—
268·4	1·400	9·0	0·73	30·16	4·75	4·75	23·16	24·9	—	—	—	—
298	1·439	44	0·99	29·41	4·67	3·98	23·94	26·0	56·6	62·6	1750	1
270·1	1·519	12·0	0·71	30·15	4·86	4·74	23·24	27·9	—	—	—	—
205	1·534	12·5	0·91	29·71	4·43	4·29	23·69	28·5	42·4	35·0	640	1·
280·1	1·659	11·0	1·11	30·49	5·14	5·05	22·87	31·3	—	—	—	—
454	1·757	9·0	0·76	30·02	4·59	4·59	23·50	33·5	—	—	—	—
432	1·805	11·5	1·06	29·79	4·46	4·35	23·75	34·8	45·5	39·2	820	1·
191	1·952	10·5	0·83	29·96	4·59	4·52	23·65	38·2	40·3	42·1	870	1
9	2·012	15·0	1·04	29·98	4·77	4·55	23·66	39·7	40·8	52·0	920	1

order to avoid a notational chaos later we abstract the mathematical problem and solve it with a different nomenclature for the variables P and P_m.

We wish to know the distribution $\phi(L)$ of an observable L (corresponding to P) in a population. We have a number of individual samples of the population each taken under a separate restriction $L \geqslant M$ but with no further bias. How do we deduce $\phi(L)$ from the known pairs of values (L, M)? The true distribution in L is of course independent of our observational selection represented here by M (see Fig. 1). Let $\psi(L_1)$ be the true fraction of sources with $L \geqslant L_1$ so that

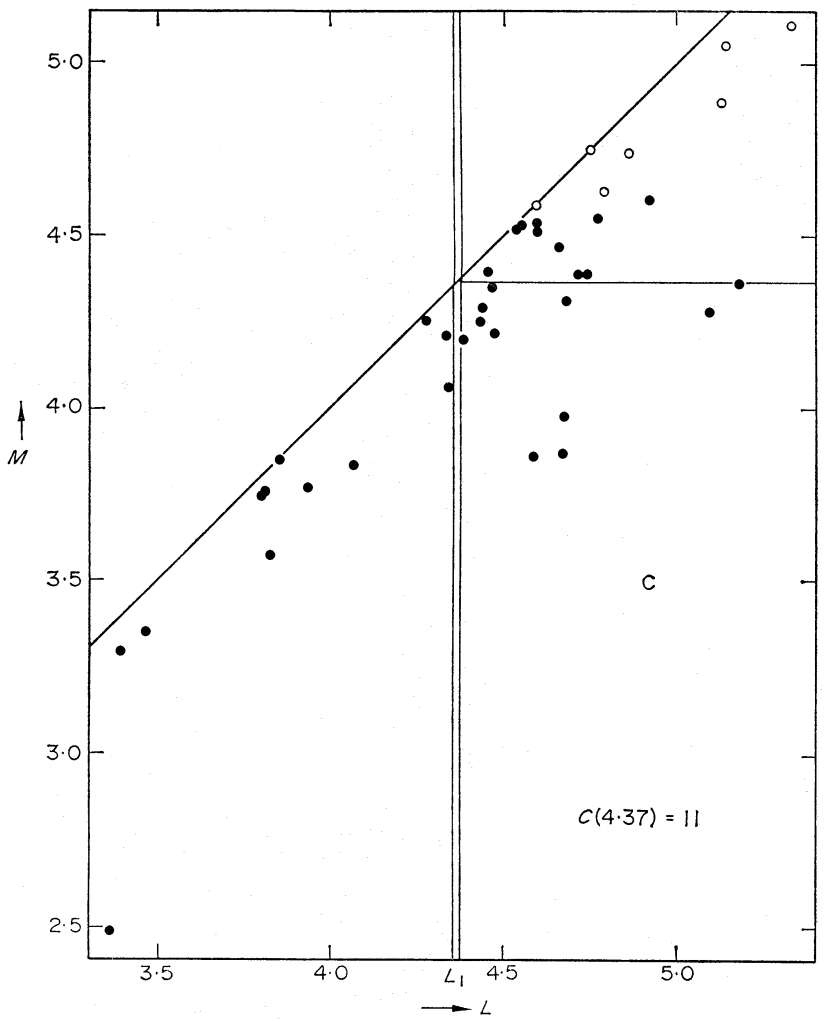


FIG. 1. *Plot of the* 40 3*CR quasars in the L, M plane.* $L = P = \log (F_R/F_0)$ $M =$ *limiting value of L beyond which an object of the same optical flux is rejected from* 3*CR as being too radio weak. The number of points in the box is denoted by C, the number* (o) *in the infinitesimal column is dX.*

$-\psi' = -d\psi(L)/dL$ is the desired true distribution function ϕ. ψ is merely the corresponding cumulative distribution. Let $X(L_1)$ be the observed number of points with $L \geqslant L_1$; then $-dX$ is the number of points in the infinitesimal column between L_1 and $L_1 + dL_1$. The ratio of this number to the number of points $C(L_1)$ in the rectangular area C_1 of Fig. 1, is the ratio of the increment $(-d\psi)$ to the cumulative distribution function, to $\psi(L_1)$ itself. Note that this is true for any slice between M and $M+dM$, and, because ψ is independent of M, we may add these slices to obtain the stated result, which we may write

$$\frac{dX}{C} = \frac{d\psi}{\psi}. \tag{1}$$

Integrating we have

$$\psi(L_1) = A \exp \int \frac{dX}{C} = A \exp \int_{L_0}^{L_1} \frac{1}{C(L)} \frac{dF(L)}{dL}\, dL. \tag{2}$$

A is the constant of integration.

In practice one chooses L_0 to be a value of L well within the observed range of points. With L_0 chosen, A determines the normalization of ψ, so if desired we may choose it to make $\psi(-\infty) = 1$. However, when parts of the distribution have not been observed it is best to use a non-normalized distribution and a convenient one is got by choosing an L_0 and then using $A = 1$. In our application ψ is the cumulative distribution function of the logarithm of radio to optical power ratio.

In practice our observations give us discrete points so X will jump at discrete values L_i at which quasars are actually observed. Thus $-dX/dL$ will be a sum of Dirac delta functions

$$-dX/dL = \sum_i \delta(L-L_i). \tag{3}$$

However, when expression (3) is substituted into formula (2) we find that $C(L_i)$ has to be evaluated. That is the number of points in the box C has to be evaluated again and again when one point (with $L = L_i$) is on the side of the box. When the number of points in C is not large, error could be generated in the choice of whether to include, exclude, or half include the point on the edge. To discover the correct weight to give such an edge point we discuss the transition from the continuous to the discrete distribution by assuming that $-dX/dL$ becomes large in the neighbourhood of each of the points L_i with unit integral across each neighbourhood. To evaluate formula (2) we wish to know the behaviour of $C(L)$ when L is close to L_i. Clearly $C(L)$ will include only the fraction

$$x = X(L)-X(L_i^+)$$

of the point at L_i where L_i^+ is taken just to the right of where the ith point contributes to X. If we define $C^-(L_i)$ to be the value of C at L_i with the contribution of the point i itself omitted, then in the neighbourhood of $L = L_i$ we have

$$C(L) = C^-(L_i)+x.$$

In formula (2) we require the contribution from the neighbourhood of $L = L_i$ to the integral

$$\int_{L_i^-}^{L_i^+} \frac{1}{C}\frac{dX}{dL}\,dL = \int_0^1 \frac{dx}{C^-+x} = \log\left(\frac{C^-}{C^-+1}\right) \tag{4}$$

where we remember that the number of points contributing to X increases by 1 as we cross L_i. The result (4) may be substituted into formula (2) to yield the expression

$$\psi(L) = A\prod_i{}' \left(\frac{C^-(L_i)}{C^-(L_i)+1}\right). \tag{5}$$

Where $\prod_i{}'$ is the product over all i with $L_0<L_i<L$ whenever $L_0<L$, and $\prod_i$ is the inverse of the product over those i with $L<L_i<L_0$ when $L_0>L$. For $L = L_0$ $\prod'(\ )$ is replaced by unity. There is still lack of definiteness in formula (5) when $L = L_i$, but this is physically correct for $\psi(L)$ is a histogram and at the steps of the histogram the function takes all values between the bottom and the top of the step. To draw the histogram one merely joins the points corresponding to the $\psi(L_i^-)$ to those corresponding to $\psi(L_i^+)$. Thus formula (5) is the solution to our problem.

However, in some applications one needs not only $\psi(L)$ but also the M distribution. For instance we might apply the method to star counts down to a given limiting apparent magnitude in order to determine the density distribution. We could then put the volume of the cone defined by the angular field of the survey and the sphere centre the observer and through the star in question equal to L. $\phi_1(L)$ would then be the density distribution of the stars as a function of volume along the cone about the line of sight. In such an application M has a physical meaning, for the limiting volume out to which a star would have been included in the sample down to a definite apparent magnitude is a linear function of its absolute magnitude. Thus in this application M values are directly related to absolute magnitudes and their true distribution would give us a knowledge of the luminosity function. For such an application we need the true distribution of the points with M corrected for the 'missing' points with $L<M$. If $Y(M)$ is the cumulative distribution of the observed points and $\Psi(M)$ is the required true cumulative distribution which has been so corrected, then

$$\frac{d\Psi}{dM} = \frac{dY}{dM}\Big/\psi(M)$$

since the required correction to each infinitesimal row between M and $M+dM$ is $1/\psi(M)$ where this ψ is normalized to unity for M small. However this formula is not always the most convenient and the symmetry of our problem allows us to write a formula corresponding to (2) as follows:

$$\Psi(M_1) = B \exp \int \frac{dY}{C} = B \exp \int_{M_0}^{M_1} \frac{\frac{dY}{dM}}{C(M)}\, dM.$$

In the discrete case this leads to

$$\Psi(M) = B \prod_i{}' \left(\frac{C^-(M_i)}{C^-(M_i)+1}\right) \tag{6}$$

with a definition of $\prod_i{}'$ as that given earlier but with M for L etc. A generalization of the whole method for the case of a known gradual observational selection rather than a sharp cut off is given in Section 4.

2. THE DISTRIBUTION OF THE RADIO TO OPTICAL POWER RATIO OF THE QUASI-STELLAR SOURCES

The material used here is that given by Schmidt (**2**) for the 40 known 3CR quasars. These are plotted in the LM diagram as Fig. 1. The function $C(L)$ giving the numbers of points in the box C as a function of the position of its edge is plotted as Fig. 2. Note that C is zero between 3·46 and 3·58. It is reasonably clear from the distribution of the points in Figs 1 and 2 that this zero is merely a numerical fluctuation. Now zeros in C^- within the observed range of points cause a breakdown in the method as outlined above for a clear physical reason. The information about the cumulative distribution so far is transmitted to the next step of lower L by the ratio of the number of points in the infinitesimal column at L_1 to the number of points in C^-. A zero in C^- allows us to determine the distribution ψ within each of the two sets of data on either side but gives

zero for the relative normalization. For this reason we can only rigorously apply the method for $L > 3{\cdot}58$. This leads to the heavy histogram of Fig. 3. To see whether the data for $L < 3{\cdot}58$ might imply anything different we note that the zero of C^- between $3{\cdot}46$ and $3{\cdot}58$ can be interpreted as a numerical fluctuation. It is then sense to replace C^- by a smoother integer function without this zero. We did this by moving the values of the L_i at which C^- increased or decreased in such a manner that the smoother C^- always increased towards its peak around $L = 4{\cdot}5$ and decreased thereafter. The smoother C^- histogram used is shown dotted and displaced in Fig. 2. The distribution, ψ, derived from the moved L_i

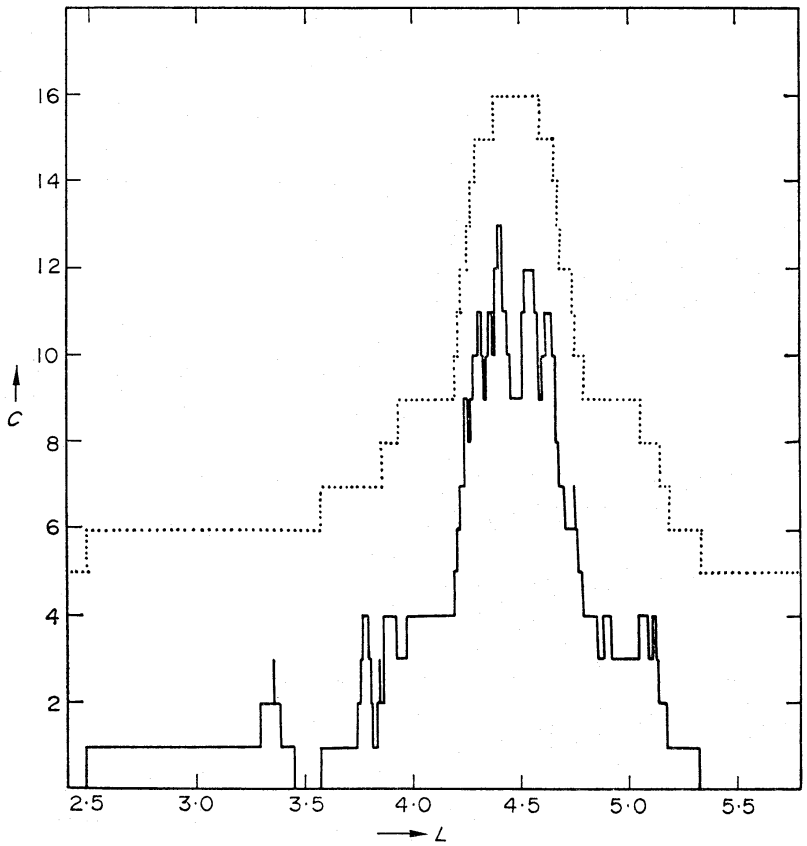


FIG. 2. *The integer function C giving the number of points in the box C of Fig. 1 as a function of the position of its edge. The dotted function, displaced for clarity, is C smoothed to remove numerical fluctuation.*

and the resulting smoother $C(L)$ is shown dotted in Fig. 3. Apart from the increased range to lower L there are no remarkable differences between the ψ histogram derived with the smoothed C^- and that which was derived rigorously with the correct values. Thus where our smoothing can be checked it has not changed the result.

Fig. 3 shows the histograms determined by formula (5) applied to our data to give the distribution of the radio to optical power ratio. The dotted histogram is determined from the smoothed version of C. In both cases we chose $L_0 = 4{\cdot}3$ and $A = 1$. The straight line which gives a rather crude fit to the histogram is $\psi(P) \propto 10^{-(p/0{\cdot}58)}$. Since that is a law of exponential type the non-cumulative distribution function $\phi = -d\psi/dL$ will be of the same form. We have derived it in the range $3{\cdot}5 < P < 5{\cdot}0$. Schmidt (1) has derived $\psi \propto 10^{-p}$ for the much

larger range $2 < P < 4 \cdot 6$ from a comparison of the distribution of optical magnitudes of 3CR quasars and a sample of optically selected quasars. As his method covers a larger range of radio to optical power ratio it is probably more accurate over the large range than our derivation of Fig. 3 which rests on 3CR quasars alone.

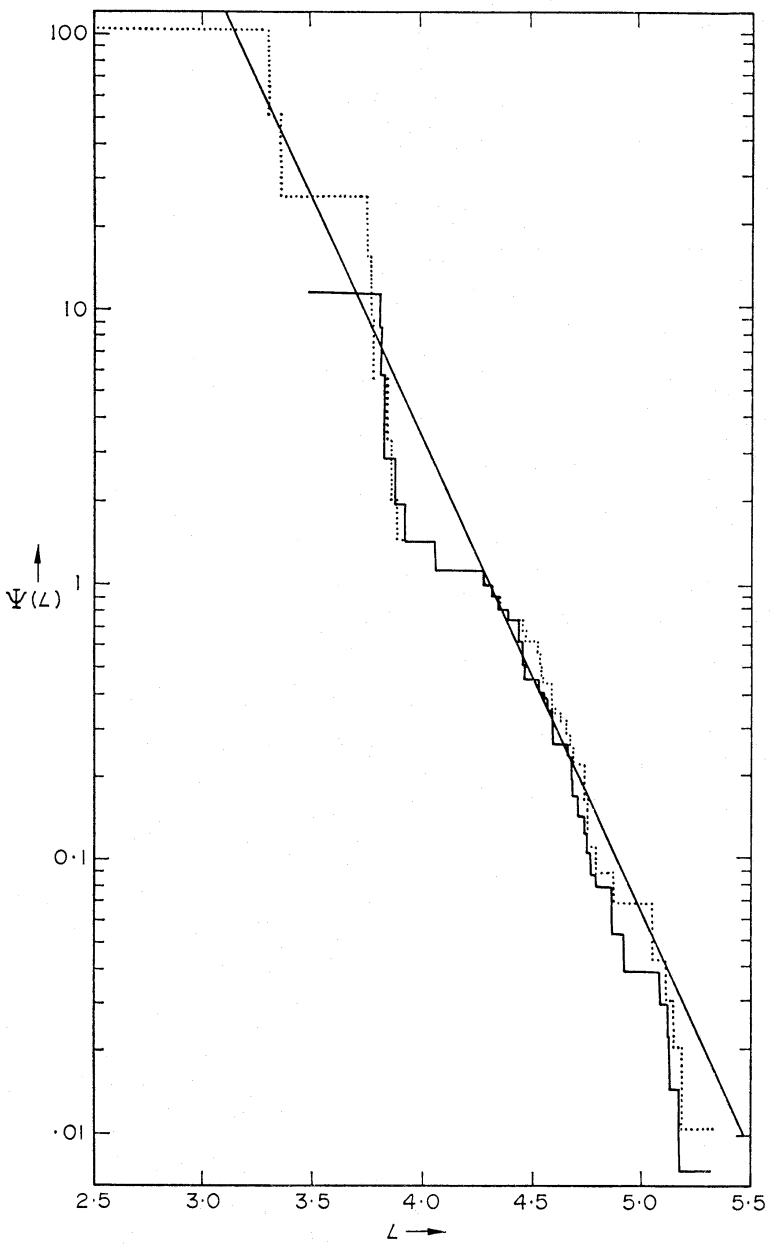


FIG. 3. *The histogram of the cumulative radio to optical power ratio distribution. $\psi(L)$ is the number of quasars whose P is greater than L. The normalization of ψ is arbitrary but the method by which the histogram was derived corrects for observational selection.*

3. THE DENSITY EVOLUTION OF THE QUASI-STELLAR SOURCES

In this section we derive the function giving the numbers of quasi-stellar sources per unit comoving volume as a function of redshift or of time. Again the *C* method allows us to derive a cumulative density distribution as a histogram without the assumption of any special mathematical form. However we shall

assume that there has been no intrinsic evolution in the nature of quasars themselves, that although they were more numerous in the past they were not different. Deductions of the density evolution are universe dependent. Our purpose here is to illustrate useful methods of data reduction on interesting data rather than to explore all logical possibilities so we limit ourselves to the Universe with $\Lambda = 0$ and $q_0 = 1$. This is a closed universe whose radius varies cycloidally with cosmic time (see Sandage 1961 (**3**)).

The material for discussion needs to be free of hidden observational selection. Whereas in Section 2 we could use all 40 3CR quasars because the function sought was by explicit assumption uncorrelated with optical magnitude, now we must recognize and quantify the selection. It is probable that all 3CR quasars brighter than $18{\cdot}4^{m}$ in the V-photometric band are known since 33 were known in 1967 and no more have been found since. Data for these 33 have been given by Schmidt (**1**); translated in $C(\mathrm{Gpc})^3$ using a Hubble constant of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ the comoving volumes nearer than each object in the $q_0 = 1$, $\Lambda = 0$ universe are given in Table I. For each source we also give V_R the volume out to which that source could have been moved before its radio flux at 178 MHz (observed) hit the limit of the 3CR catalogue. Further we define V_0 the volume out to which the object could have been moved before its observed optical flux at 2500 Å emitted wavelength fell to $10^{-30}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$. This is the strict definition of the optical limit adopted rather than the approximate limit of $18{\cdot}4^{m}$ in the V-photometric band. The smaller of V_R and V_0 is called V_m the maximum volume out to which the source would have been included in our list. Reference (**1**) gives details of the computations of V_m.

We now wish to correct the observed distribution with V for the observational selection $V \leqslant V_m$. To do this we plot the 33 3CR quasars in a V, V_m diagram, Fig. 4. This diagram is analogous to the L, M diagram of Section 1, Fig. 1, but

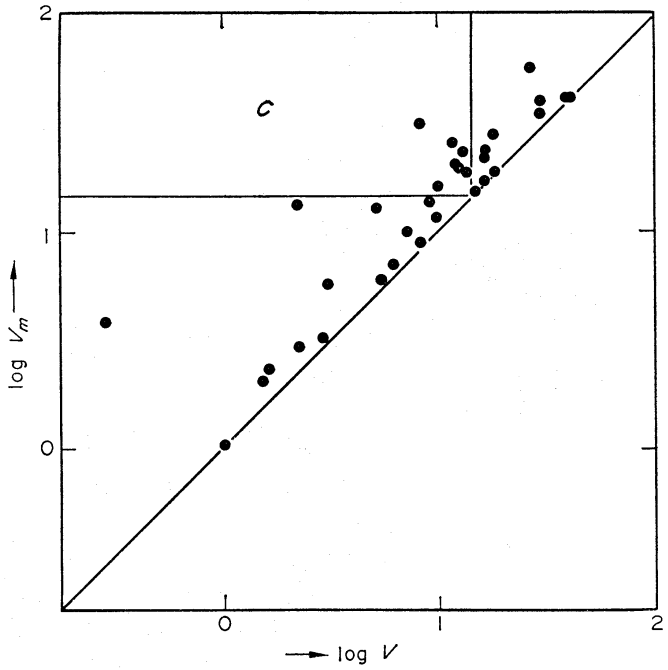


FIG. 4. *Plot of log V_m, the volume out to which a particular quasar would have been bright enough to be included in our list, against log V, the volume nearer than that quasar in the $\Lambda = 0$, $q_0 = 1$ cosmological model.*

whereas the range of L was $L \geqslant M$, the range of V is $0 \leqslant V \leqslant V_m$. Analogously to the method of Section 1 we find the true number of quasars $N(V)$ within the comoving volume V.

$$N(V) = A \prod_i{}' \left(\frac{C^-(V_i)+1}{C^-(V_i)} \right) \tag{7}$$

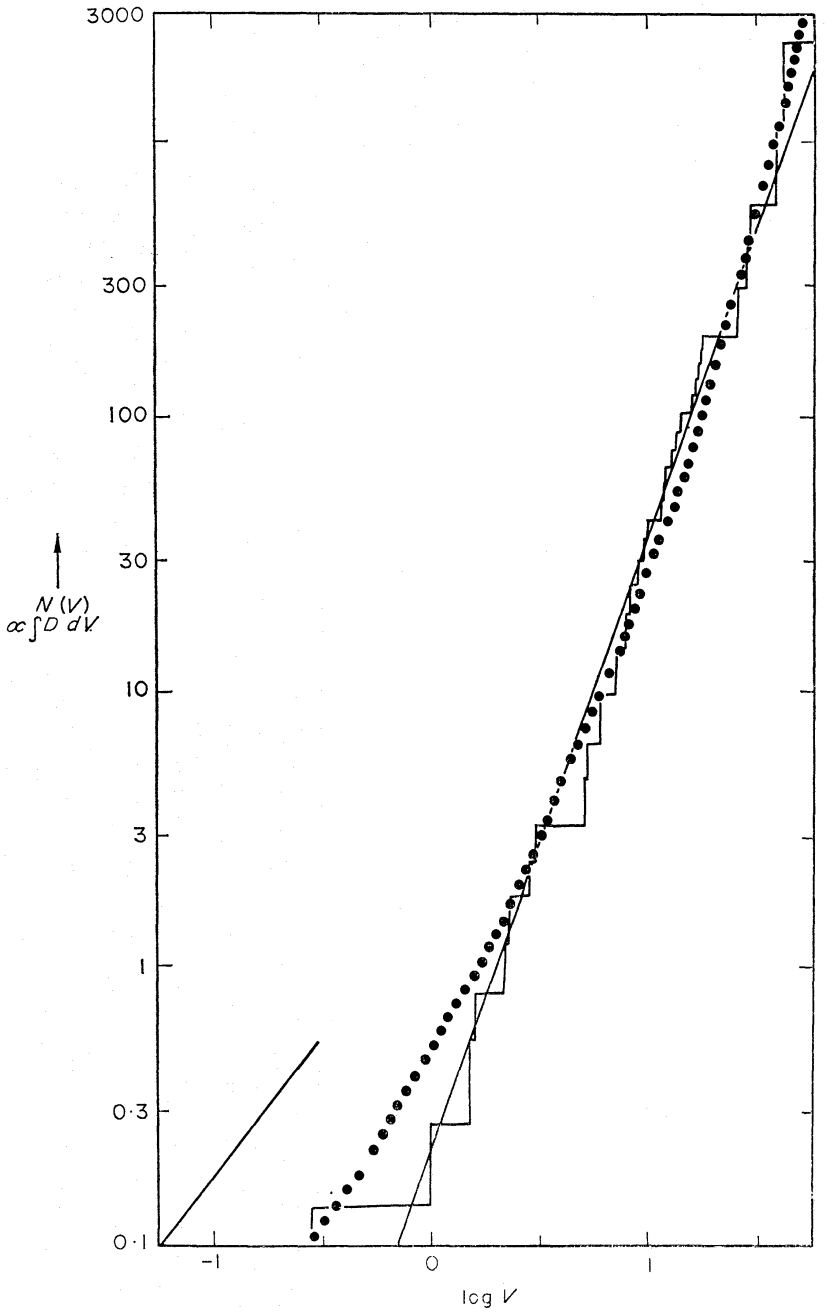


FIG. 5. *The derived histogram showing the number of quasars $N(V)$ within volume V about the Sun. The derivation corrects for observational selection. Normalization of N is arbitrary but relative numbers are correctly represented (on the hypothesis of density evolution). The fitted line is $N \propto V^{2\frac{1}{2}}$. dN/dV is a measure of the density of quasars per comoving volume so for no density evolution the histogram should show the gradient drawn in the lower left corner. The dotted curve is Schmidt's law* (1) $N \propto z^3(1+z)^2$ *converted to an $N(V)$ relationship by Table II.*

where $C^-(V_i)$ is the number of points in the area C_i when the edge is on $V = V_i$ and the ith point is not counted in C^-. $\prod_i'$ is the product over all points i between some chosen V_0 and V, if $V \geqslant V_0$, and is the inverse of the product over all such points if $V < V_0$. Taking log $V_0 = 0{\cdot}8$ $C(\mathrm{Gpc})^3$ and $A = 10$, N is plotted logarithmically as a histogram in V as Fig. 5.

If there were no density evolution $N(V)$ would be proportional to V. In Fig. 5 the histogram is well fitted by a straight line of slope $2\frac{1}{4}$ so $N(V) \propto V^{2\frac{1}{4}}$. This implies a density evolution such that $D = dN/dV \propto V^{1\frac{1}{4}}$ throughout the observed range of points $0{\cdot}6 \leqslant V \leqslant 50$. Also plotted is Schmidt's relationship $N(V) \propto z^3(1+z)^2$ deduced from the same data on the assumption that $N \propto z^3(1+z)^n$. This particular mathematical law predicts more quasars at high

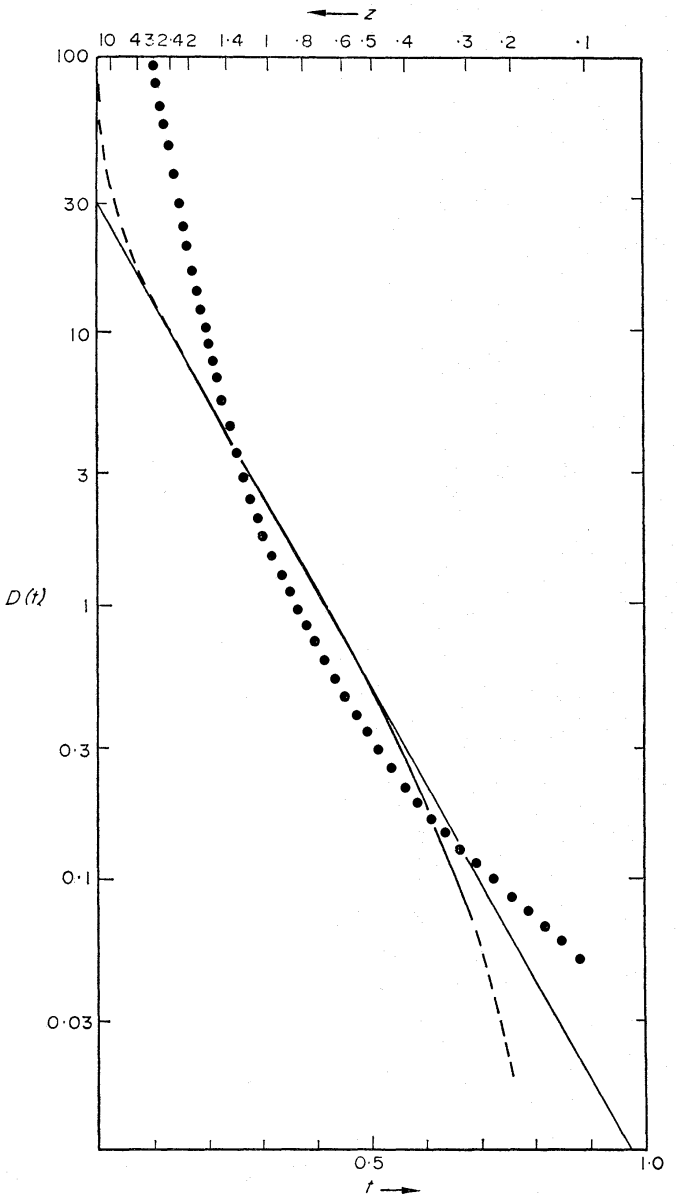


FIG. 6. *The derived evolution as a function of cosmic time, in the* $q_0 = 1$ $\Lambda = 0$ *cosmology, is close to exponential decay with a half life of* $0{\cdot}085$ *of the age of the Universe. The dotted line shows* $D \propto (1+z)^6$ *which has been derived from optically selected quasars* (2).

redshift than $N(V) \propto V^{2\frac{1}{2}}$. It is not yet clear whether extrapolation of this new law would still predict too many quasars with $z = 2{\cdot}5$. That problem is not very clear cut because the data at these redshifts are subject to considerable undetermined observational selection; no 3CR redshift in our sample is greater than 2·011 (3C9).

In Fig. 6 we convert our density law $D \propto V^{1\frac{1}{2}}$ into a density redshift relationship and a density time relationship by re-expressing V as a function firstly of redshift and then of time using Table II which is taken from Sandage's tabulations (3), (4) of observational properties of Friedman universes. At small redshift $D \propto V^{1\frac{1}{2}}$ must break down if the local density of quasars is non-zero. Fig. 6 shows that the $D(t)$ relationship is close to linear when drawn logarithmically. If $D(t)$ is taken to be a dying exponential there is no difficulty at low redshift and the gradien of our logarithmic plot yields a half life of 0·085 of the current age of the Universe.

TABLE II

Properties of the $q_0 = 1$, $\Lambda = 0$ Universe. $V(z)$ the volume in $C(Gpc)^3$ within which objects have redshifts less than z. ($H = 100$ km s^{-1} Mpc^{-1}). $t(z)$ the time seen back to at a redshift of z as a fraction of the present time $R(z)$ the radius of the Universe at the time $t(z)$.

z	$V(z)$	$t(z)$	$R = \frac{1}{1+z}$
0	0	1·00	1·00
0·05	0·0121	0·919	0·953
0·10	0·0851	0·848	0·910
0·15	0·254	0·786	0·870
0·20	0·525	0·731	0·834
0·25	0·910	0·683	0·800
0·30	1·40	0·639	0·770
0·35	2·02	0·600	0·741
0·40	2·71	0·565	0·715
0·45	3·45	0·534	0·690
0·50	4·34	0·504	0·667
0·60	6·20	0·454	0·625
0·70	8·28	0·412	0·589
0·80	10·4	0·376	0·556
0·90	12·9	0·345	0·526
1·00	15·3	0·317	0·500
1·20	20·1	0·275	0·455
1·40	24·9	0·240	0·416
1·60	29·8	0·210	0·385
1·80	34·6	0·186	0·358
2·00	39·4	0·167	0·333
2·20	43·8	0·150	0·313
2·40	47·8	0·138	0·294
2·60	51·0	0·124	0·278
2·80	56·0	0·115	0·264
3·00	59·6	0·106	0·250
3·50	68·2	0·088	0·222
4·00	75·4	0·076	0·200
5·00	88·6	0·058	0·167
10	128	0·0230	0·091
100	222	0·0008	0·0099
1000	250	$2{\cdot}6 \times 10^{-5}$	10^{-3}
∞	266	0	0

The small number of quasars at small redshift makes it hard to extrapolate $D(t)$ to the present epoch with any real accuracy.

We should point out that if quasars form in galaxies, there is a possibility of error in the $D(t)$ graph for $z < 0{\cdot}3$. Rather weak nearby quasars may be placed in the lists of N galaxies, not quasars, if fuzz is visible around the optical image even when the same objects further away would not be resolved and may be called quasars. A systematic observational selection of this sort would have steepened our $D(t)$ curve for $t > 0{\cdot}6$.

4. THE LUMINOSITY FUNCTION OF THE QUASI-STELLAR SOURCES—GENERALIZATION OF THE C METHOD

For each source we have a volume V_m out to which it could have been seen within our sample. For that source we may now calculate from our $N(V)$ line the number $N(V_m)$ which is proportional to the number of quasars in V_m. These numbers are tabulated in Table I. Then the quasar we are considering represents a fraction $\propto 1/N(V_m)$ of the quasars in the volume V_m. To determine the luminosity function of the quasars we must weight each observed object with its true frequency in the class of quasars rather than its observed frequency (which favours bright objects). This weighting is properly performed if instead of giving each quasar unit weight we give each point the weight $1/N(V_m)$.

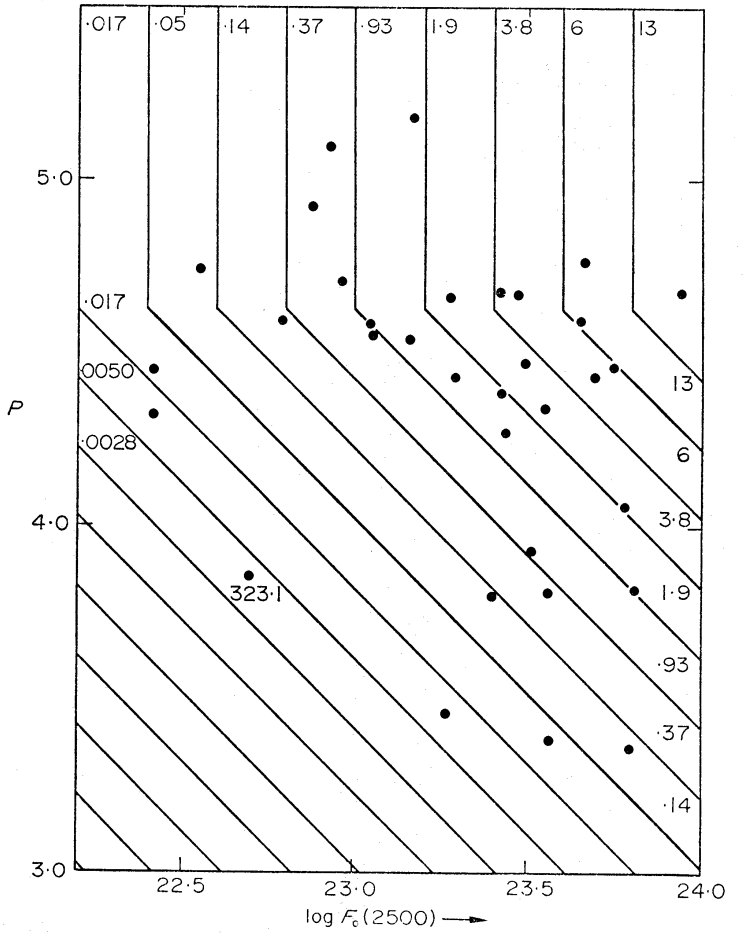


FIG. 7. *Plot of quasars in the P, log F_0 diagram. The lines of constant weight are labelled with values of the inverse weights $\rho(L, M)$. The weights are necessary to obtain a true luminosity function from a sample selected by observed apparent fluxes.*

Fig. 7 plots the 33 points in a diagram with the logarithm of the radio to optical power ratio plotted upwards and log F_ν at 2500 Å emitted wavelength plotted to the right. F_ν is the absolute emitted flux in Watts meter^{-2} Hz^{-1}. From Table I we see that the points have very different weights differing by factors of up to 7600. Most of this weighting is systematic over Fig. 7 the points of low weight (large $N(V_m)$) being at the top right-hand corner. Some weights are very large; 3C323·1 has a weight greater than all other points combined. Clearly a summation of weighted points will give a luminosity function dominated by numerical fluctuations in the points of high weight. This is why Schmidt found that the optical luminosity function was ill determined (1). Here we determine the luminosity function using our assumption that the radio to optical power ratio is independent of optical power. Although the large weights are a handicap if the data are treated by a poor statistical method nevertheless treated properly the weights vary systematically in just the direction that one would like for an accurate determination of the luminosity function. Enormous volumes have been surveyed for the rare bright objects, small volumes for the common dim ones. Thus quasar data are ideally suited for determining a luminosity function over a wide range of magnitudes from sparse data.

We notice that in forming an optical luminosity function we will be adding the points of Fig. 7 in columns. By hypothesis the true distribution of points has the same distribution across each row, but for a normalization. However, this true distribution has been spoiled by the observational selection, so that we see as observed points only a fraction of the true distribution. Our problem is to provide the best estimate of the true distribution.

This problem is similar to the observational selection problem solved in Section 1. Consider the following generalization of that problem. Points in an L, M diagram (like Fig. 1) have been observed subject to a known observational selection such that there is a known probability $\rho(L, M)$ of an object at (L, M) being observed. Given that the *true* distribution of objects is the same (but for normalization) along each row M = constant find the best estimate of that distribution function $\phi(L) = -d\psi/dL$ from the observed points. This problem is almost mathematically equivalent to the problem of determining the optical luminosity function from the points and weights of Fig. 7 so we shall solve the abstract form first. Note that if $\rho(L, M)$ were zero for $L < M$ and unity $L \geqslant M$ the problem would reduce to that of Section 1 which we solved by the C method. We now look for a generalization of that method.

We define $X(L_1)$ and $\psi(L_1)$ as in Section 1 and we look for an expression $C(L_1)$ so that equation (1) shall still hold in the form

$$\frac{d\psi(L_1)}{\psi(L_1)} = \frac{dX(L_1)}{C(L_1)}. \tag{1}$$

Clearly dX is a number of points in an infinitesimal column at L_1; by adding numbers of points (unweighted) in this column we have the least fluctuation in this small number. These points may have very different weights $1/\rho$. Now $d\psi/\psi$ is the probability ratio for weighted points in the column to those to the right of the column and it is by hypothesis the same on each row. Since X has been defined in terms of unweighted numbers, the points contributing to C from the row M must be given the weights $\rho(L_1, M)/\rho(L, M)$. That is the points to the right of the infinitesimal column must have their weights reduced by a

reduction factor $\rho(L_1, M)$ equal to the inverse weight of the point on the same row but in the infinitesimal column. We therefore deduce that the expression for the generalized C is

$$C(L_1) = \int_{-\infty}^{\infty} \int_{L_1}^{\infty} \frac{\rho(L_1, M)}{\rho(L, M)} n(L, M)\, dL\, dM. \tag{8}$$

Here $n(L, M)$ is the observed number density of points at L, M. Notice that the expression reduces to the old one for the problem of Section 1. From equation (1) we deduce that

$$\psi = \exp \int \frac{dX}{C}$$

or following the discussion of Section 1

$$\psi = A \prod_i{}' \left(\frac{C^-(L_i)}{C^-(L_i)+1} \right) \tag{9}$$

where C^- leaves out the point on the edge as before but is otherwise defined by equation (8). An analysis of the best possible estimate of ψ from given data is performed in the Appendix where it is shown that the expression (8) though not in principle the best is in practice close to it. In fact expression (8) is the first step in a rapidly convergent iterative sequence outlined in the Appendix.

In the determination of the optical luminosity function $\log F_0$ plays the role of L and $P = \log (f_R/f_0)$ the role of M. Note that in the application of Section 1 P played the role of L. However such minor confusions apart, application of the method is straightforward once we know the inverse weight function $\rho(L, M)$. We would like to choose $\rho(L, M)$ to be $N(V_m)$ the inverse of the weights of the points in Fig. 7. However V_m is not a function of L and M alone. V_m is the least of V_0 and V_R. The volume V_R out to which the object would have been included in the radio catalogues is not merely dependent on the radio absolute magnitude $\log F_R = P + \log F_0$, but also on the spectral index α which varies from source to source and is not a function of F_0 and P. The C method as developed so far can only cope with a systematic weighting dependent on L and M so our aim is to split the known weights $1/N(V_m)$ into a systematic part that depends on $\log F_0$ and P only, and a further weighting w peculiar to each object. The systematic weighting can then be dealt with correctly by identifying its weights with $1/\rho(L, M)$ while we hope to make the peculiar weights so close to unity that adding weighted points does not significantly increase the numerical fluctuations.

The procedure for determining V_R as given by Schmidt (1) is firstly to determine z_R the redshift out to which the object in question would still have been included in 3CR. This is done by comparing two formulae for the flux f^* at 500 MHz (emitted frequency) which would be observed when the object is placed so far away in the Universe that it is at the limiting flux for inclusion in 3CR.

$$\log f^* = \log F_R - 2 \log z_R - 53{\cdot}03$$

$$\log f^* = -26 + \log (9 \times 1{\cdot}16) + \alpha \log \left(\frac{500}{178} \right) - (1+\alpha) \log (1+z_R).$$

The first formula is the standard conversion from an absolute flux to an apparent one while the second ensures that the 178 MHz flux that would be observed

is $9 \times 1 \cdot 16$ flux units the limit of 3CR. Combining these formulae we have

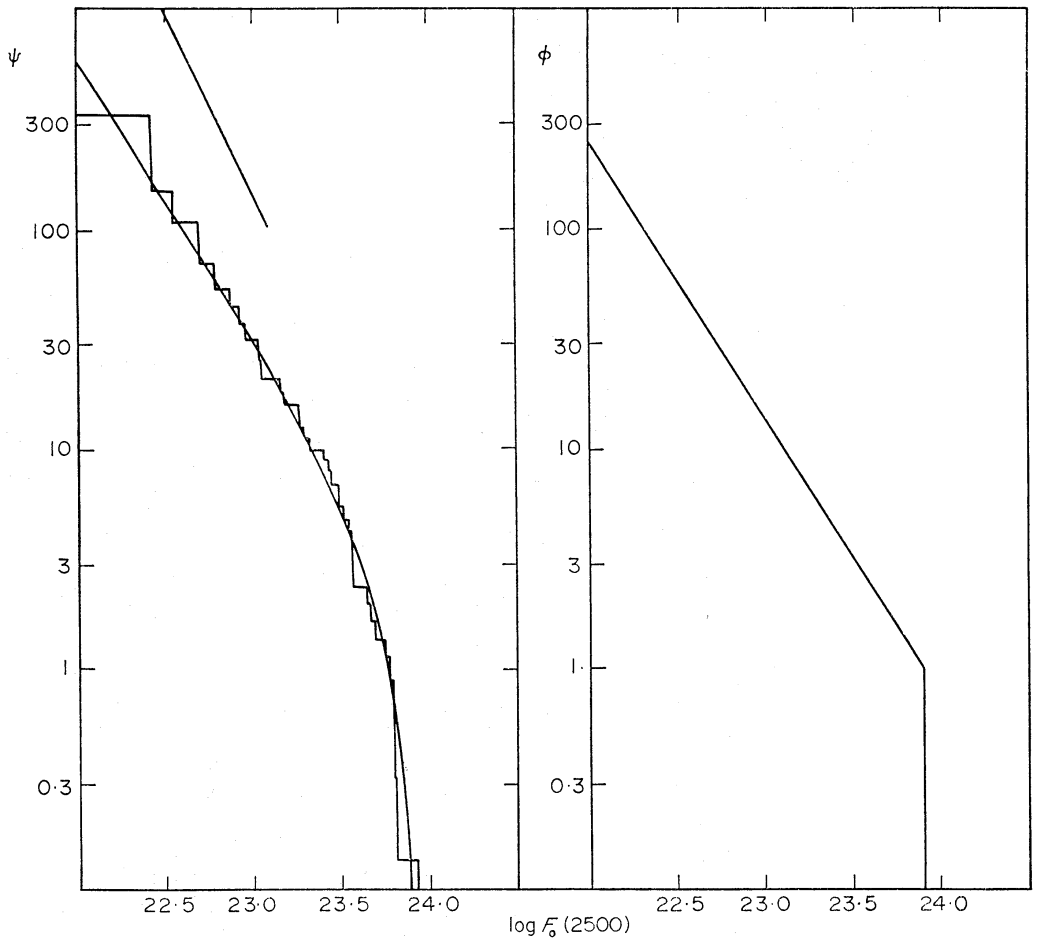
$$\frac{z_R^2}{(1+z_R)^{1+\alpha}} = \frac{F_R \, 10^{-27 \cdot 03}}{9 \times 1 \cdot 16 \left(\frac{500}{178}\right)^{\alpha}}.$$

If α is known this formula determines z_R as a function of F_R.

Secondly V_R is found from z_R by consulting Table II which gives the $V(z)$ function for the $q_0 = 1$ Universe. Finally, the weights $N(V_m)$ are found by taking V_m the least of V_0 and V_R and consulting the line in Fig. 5. Notice that if we set $\alpha = -1$ for every source then z_R^2 would be a linear function of F_R just as the equivalent optical quantity z_0^2 is a linear function of F_0. Since the spectral indices are not far different from $\alpha = -1$ we define a new quantity N^* the value that $N(V_m)$ would take if we arbitrarily set $\alpha = -1$ in the above derivation. N^* is a systematic weighting that only depends on position in the log F_0, P diagram of Fig. 7. The lines drawn across that diagram are isobars—lines of constant N^*. The ratios $w = N^*/N(V_m)$ are given in Table I. Notice that all the w are close to unity showing that nearly all the weighting is included in the systematic part N^* and that the further weighting is not very important. To apply the C method we choose $\rho = N^*$ and everywhere that we count points in the method we replace the number of points by the sum of their w's. Thus n is replaced by wn in formula (8) and X is defined as the sum of the weights w of all points to the right of some vertical line. Calculation of the function C is somewhat tedious since instead of counting the number of points in an area we have to add all their systematic weight ratios $\rho(L_1, M)/\rho(L, M)$ with weights w. Furthermore the systematic weight ratios change with the position L_1 of the infinitesimal column of Fig. 1. For repeated use it is clearly advisable to write a simple computer program to determine C from formula (8) generalized to include w, and then for generating

$$\psi = \prod_i{}' \left(\frac{C^-(L_i)}{C^-(L_i)+w_i}\right) \tag{10}$$

but for one application to 33 points the computation took one morning using a slide rule. Fig. 8 shows the resulting histogram for the cumulative optical luminosity function derived from the C method. The quasar luminosity function drops sharply at a flux of log $F_0 = 24$ and quasars brighter than that must be very rare. However in the range $23 \cdot 5 > \log F_0 > 22 \cdot 5$ the luminosity function is increasing exponentially with very large numbers at small optical absolute luminosity. To demonstrate this more clearly we have fitted a simple analytic function to the histogram for the cumulative function and have differentiated the function to obtain the non-cumulative function $\phi = -d\psi/dL$ which is plotted as Fig. 9. The perfectly sharp upper cut-off to the luminosity function that results must be an artefact since some exceedingly rare very bright non-3CR quasars are known above this limit; their existence does not contradict a rapid fall in the luminosity function beyond log $F_0 = 24$. The upper cut off in the optical luminosity function is not due to lack of a single quasar. No less than eight zags of the histogram of Fig. 8 fall below the straight line whereas a cut off from lack of data should disappear after one zag of the integral histogram. However, it is possible that this upper cut off is a property of the $q_0 = +1$ Universe since this end of the histogram must be quite cosmology sensitive. The luminosity function of Fig. 9 is



(FIGS. 8 and 9)

FIG. 8. *The cumulative optical luminosity function ψ giving the total number of quasars with log optical output at 2500 Å greater than log F_0. Notice the cut off at log $F_0 \simeq 24$. The empirical curve fitted to the data is $\psi = \{exp\ -[2{\cdot}74\ (log\ F_0 - 23{\cdot}92)] - 1\}$ for log $F_0 < 23{\cdot}92$, and zero for log $F_0 > 23{\cdot}92$. Normalization of ψ is arbitrary.*

FIG. 9. *The non-cumulative optical distribution function ϕ giving the numbers of quasars per unit range in log F_0, derived from the analytical fit to Fig. 8. The normalization of ϕ is arbitrary.*

very similar to that derived by Schmidt (**2**) from a sample of quasars selected by optical properties alone.

Collecting our results together the luminosity function giving the number of quasars per comoving cubic giga parsec per decade of ratio to optical power ratio per decade of optical flux is assumed to be of the form

$$KD(t)\,\phi_1(P)\,\phi_2(\log F_0) \tag{12}$$

where ϕ_1, ϕ_2 and $D(t)$ are the functions found from Section 2 and Figs 9 and 6 respectively. We take

$$\phi_1 = 10^{-(P-4{\cdot}55)/0{\cdot}58} \tag{13}$$

where P is the logarithm of the radio to optical power ratio and

$$\phi_2(x) = \begin{cases} 10^{-(x-23{\cdot}90)/0{\cdot}80} & x \leqslant 23{\cdot}90 \\ 0 & x > 23{\cdot}90 \end{cases} \tag{14}$$

where $x = \log F_0$.

K is a normalization constant and $D(t)$ is defined by the solid line of Fig. 6 which is normalized so that $D(0{\cdot}42) = 1$. It is possible to check the assumption of the form (12) by testing the goodness of fit of our final answer to the observed data. The agreement can be seen to be quite good from the Table 3(A) given in the accompanying note which is mainly concerned with the extrapolation of the luminosity function to the luminosities of N galaxies. The distribution of luminosities which is observed will be considerably different. For a survey covering an area of A steradians the observed distribution of luminosities expressed as a function of intrinsic flux will be

$$\chi(P, \log F_0) = \frac{A}{4\pi} \int_0^{V_m} KD(t(V))\, \phi_1(P)\, \phi_2(\log F_0)\, dV$$

where V_m is the maximum volume out to which an object of given $\log F_0$ and P will be included in our list. Since $D(t(V)) \propto dN(V)/dV$ we find that

$$\chi(L, \log F) = \frac{KA}{4\pi} \frac{N(V_m)}{\left(\frac{dN}{dV}\middle|_{t\,=\,0{\cdot}42}\right)} \phi_1(P)\, \phi_2(\log F_0). \qquad (15)$$

In the region where optical selection is more important than radio selection, V_m is a function of the absolute optical flux F_0. Elsewhere we replace $N(V_m)$ by $N^*(V_m)$ which is a function of the radio flux $P + \log F_0$ and the resulting luminosity function predicts the sums of weights w rather than the numbers of objects. By a tedious fitting of expression (15) to the observed total number of points (or rather total weight) in Fig. 7 we find the normalization constant K to be $0{\cdot}24\ C(\mathrm{Gpc})^{-3}$ from the 3CR quasars for which $A/4\pi$ is approximate $0{\cdot}43$. Our final answer for the luminosity function is $0{\cdot}24 D(t)\, \phi_1(P)\, \phi_2(\log F_0)$ quasars per $C(\mathrm{Gpc})^3$ per unit increase of P per unit increase in $\log_{10} F_0$, where ϕ_1 and ϕ_2 are given by equations (13) and (14) and $D(t)$ is given by Fig. 6 or approximately $D(t) = 31e^{-\lambda t}$ where t is measured in units of the current age of the Universe and λ is $\log_e 2/0{\cdot}085 = 8{\cdot}1$. The ranges over which this luminosity function have been derived are

$$\left.\begin{array}{l} 3{\cdot}5 < P < 5{\cdot}0 \\ 22{\cdot}5 < \log F_0 < 23{\cdot}9 \end{array}\right\} \text{ and } \left\{\begin{array}{l} 0{\cdot}2 < t < 0{\cdot}7 \\ 2 > z > 0{\cdot}3. \end{array}\right.$$

As we stated previously there is evidence that for a wider range in P towards smaller values, ϕ_1 increases considerably less steeply as P decreases (1). The following facts may be deduced from our luminosity function. Defining a radio bright quasar as one for which $\log F_0 > 22{\cdot}3$ and $P > 3{\cdot}3$ then there were about 2.10^5 such quasars in the Universe at $t = 0{\cdot}317$ which is the time at which quasars at $z = 1$ are now seen, but by the time $t = 0{\cdot}64$ ($z = 0{\cdot}3$) there were only about 13 000. The total number of radio bright quasars there have ever been can be calculated provided we assume a lifetime for each. If $Q(t)$ is the total number of quasars at epoch t and the lifetime of each is T in units in which the age of the Universe is one, then the number of once-radio-bright quasars dead or alive is $T^{-1} \int Q(t)\, dt$. For our exponential density law this gives $3.10^5\ T^{-1}$; so for lifetimes T of 10^{-3} of the age of the Universe the number of dead or alive once-radio-bright quasars is of the order of 3.10^8. Since there are at least 30 times as many radio quiet quasars as radio bright ones the total number of dead quasars will be comparable to the number of non-dwarf galaxies in the Universe

(3) which in this model is about 2.10^{10}. This assumes that the radio-quiet quasars also have lives of the order of 10^{-3}. Shorter lives will give still greater numbers of dead quasars. Quasars probably make a substantial contribution to the radio background noise. Indeed extrapolation of our luminosity function over all optical fluxes and to radio fluxes any lower than log $F_R \simeq 25{\cdot}3$ would give a brighter background than is observed. The lowest radio flux in our 3CR quasar sample is log $F_R = 26{\cdot}55$ so there is room for a change of slope of the function before the background is exceeded. Indeed Schmidt's $\phi(P) \propto 10^{-p}$ in place of our $10^{-p/0{\cdot}58}$ was derived over the wider range of P and is in the right direction to avoid conflict with the background. The above results assume a $\Lambda = 0$, $q_0 = 1$ Universe. Some of the more speculative implications of our luminosity function are considered in the accompanying note.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr Maarten Schmidt for stimulating the development of the C method and its use. But for the difficulties of international cooperation during postal strikes this paper would have been written jointly. I thank Dr Toomre for discussions of the C method, Dr Sandage for photometric data and Dr J. Matthews for referring me to the method of maximum likelihood. The referee has contributed greatly to the readability of the paper by demanding notation reform, and the removal of a cluttering mathematical symbolism.

California Institute of Technology
Hale Observatories
Carnegie Institution of Washington.

On leave from:
Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, Sussex
and The University of Sussex.

Received in original form 1970 *November* 11.

REFERENCES

(1) Schmidt, M., 1968. *Astrophys. J.*, **151**, 393.
(2) Schmidt, M., 1970. *Astrophys. J.*, **162**, 371.
(3) Sandage, A. R., 1961. *Astrophys. J.*, **133**, 355.
(4) Sandage, A. R., 1961. *Astrophys. J.*, **134**, 916.

APPENDIX

ON THE BEST COMBINATION OF THE OBSERVATIONS

Points in an L, M diagram are distributed according to some unknown non-normalized probability distribution which is of the form $\phi(L)\eta(M)$. However, there is an observational selection so that only a fraction $\rho(L, M)$ of the points at L, M are observed. It is assumed that $\rho(L, M)$ is known so the non-normalized probability distribution for observed points is of the form $\rho(L, M)\,\phi(L)\,\eta(M)$. Given an observed number density of points $n(L, M)$ our problem is to recover the best possible estimates of the unknown distribution functions $\phi(L)$ and $\eta(M)$.

The probability of observing $n(L, M)\ \Delta L\Delta M$ points in the box with edges L and $L+\Delta L$, M and $M+\Delta M$ will be

$$\left(\frac{\rho(L, M)\phi(L)\eta(M)\Delta L\Delta M}{\int\int \rho(L', M')\phi(L')\eta(M')\, dL'\, dM'}\right) n(L, M)\Delta L\Delta M.$$

The probability of observing all the points as they are, is therefore a giant product, Π, of such factors corresponding to each area $\Delta L\Delta M$. Evidently

$$\text{Log } \Pi = \Sigma\{\log [\rho(L, M)\, \phi(L)\, \eta(M)\, \Delta L\Delta M] - \log \int\int \rho\phi\eta dL'dM'\}\, n(L, M)\, \Delta L\Delta M$$

where the sum extends over all areas $\Delta L\Delta M$. To determine the most probable distributions $\phi(L)$, $\eta(M)$ from the given observations $n(L, M)$ we use the method of maximum likelihood. This states that if all distributions ϕ, η are equally likely *a priori*, then the most probable pair of distributions is that pair for which one would calculate that the observed $n(L, M)$ are the most likely. Our problem is therefore reduced to maximizing Π or equivalently log Π over all possible pairs of distributions $\phi(L)$, $\eta(M)$.

Since $n(L, M)$, ΔL, ΔM are all fixed in this maximization, we may maximize log Π^* instead of log Π where

$$\log \Pi^* = \log \Pi - \Sigma \log (\Delta L\Delta M)\, n(L, M)\, \Delta L\Delta M,$$

log Π^* has the advantage that it may be replaced by an integral

$$\log \Pi^* = \int\int \{\log [\rho(L, M)\, \phi(L)\, \eta(M)] - \log \int\int \rho(L', M')\, \phi(L')\, \eta(M')\, dL'\, dM'\}\, n(L, M)\, dL\, dM$$

$$\delta \log \Pi^* = \int\int \left\{\frac{\delta\phi}{\phi} + \frac{\delta\eta}{\eta} - \frac{\int\int \rho(L', M')[\phi(L')\, \delta\eta(M') + \delta\phi(L')\, \eta(M')]\, dL'\, dM'}{\int\int \rho\phi\eta\, dL'\, dM'}\right\} \times n(L, M)\, dL\, dM$$

hence

$$\delta \log \Pi^* = \int \left\{\frac{\int n(L, M)\, dM}{\phi(L)} - \frac{\int \rho(L, M)\, \eta(M)\, dM \int\int n dL''\, dM''}{\int\int \rho\phi\eta\, dL'\, dM'}\right\} \delta\phi(L)\, dL + \text{a similar term in } \delta\eta(M).$$

Since ϕ and η are non-normalized distributions they are not subject to constraints so $\delta\phi(L)$ and $\delta\eta(M)$ may be chosen independently in each range of L and M respectively. Hence the ϕ and η that maximize Π and log Π^* will be those that make the coefficients of $\delta\phi(L)$ and $\delta\eta(M)$ vanish,
i.e.

$$\frac{\int n(L, M)\, dM}{\phi(L)} = \int \rho(L, M)\, \eta(M)\, dM\, \frac{J^*}{K^*} \qquad \text{(A1)}$$

and

$$\frac{\int n(L, M)\, dL}{\eta(M)} = \int \rho(L, M)\, \phi(L)\, dL\, \frac{J^*}{K^*}. \tag{A2}$$

Where J^* and K^* are the normalization constants

$$J^* = \int\int n(L, M)\, dL\, dM$$

$$K^* = \int\int \rho(L, M)\, \phi(L)\, \eta(M)\, dL\, dM.$$

Substituting the value of $\eta J^*/K^*$ obtained from A2 in A1, we obtain a strange integral equation for ϕ

$$\frac{\int n(L, M)\, dM}{\phi(L)} = \int \frac{\rho(L, M) \int n(L', M)\, dL'}{\int \rho(L', M)\, \phi(L')\, dL'}\, dM. \tag{A3}$$

For an initial acquaintance with this strange equation we specialize to the problem considered in Section 1 in which there are complete observational selection against all points with $M > L$ and complete observations with $M \leqslant L$. Then ρ is always zero or unity and there are only observations with $M \leqslant L$, so equation A3 reduces to

$$\frac{dX}{d\psi} = \frac{\int_{-\infty}^{L} n(L, M)\, dM}{\phi(L)} = \int_{-\infty}^{L} \frac{\int_{M}^{\infty} n(L', M)\, dL'}{\int_{M}^{\infty} \phi(L')\, dL'}\, dM \tag{A4}$$

where following the notation of Section 1

$$X(L) = \int_{L}^{\infty} \int_{-\infty}^{\infty} n(L', M)\, dM\, dL'$$

and

$$\psi(L) = \int_{L}^{\infty} \phi(L')\, dL'.$$

Note that we shall shortly need X defined for the more general problem so we have left the upper limit of the M integration at ∞ although in the current problem $n(L', M)$ is zero for $M > L'$. On multiplying equation (A4) by $\phi(L) = d\psi/dL$ and integrating from L to infinity we obtain

$$X = -\int_{L}^{\infty} \frac{d\psi}{dL_1} \int_{-\infty}^{L_1} \frac{\int_{M}^{\infty} n(L', M)\, dL'}{\psi(M)}\, dM\, dL_1$$

on integrating by parts

$$X = \psi(L) \int_{-\infty}^{L_1} \frac{\int_{M}^{\infty} n(L', M)\, dL'}{\int_{M}^{\infty} \phi(L')\, dL'}\, dm + \int_{L}^{\infty} \psi(L_1) \frac{\int_{L_1}^{\infty} n(L', L_1)\, dL'}{\psi(L_1)}\, dL_1. \tag{A5}$$

Now

$$X-\int_L^\infty \int_M^\infty n(L', M)\, dL'\, dM$$

is just what we called $C(L)$ in Section 1 and it is just this term that appears when the last term of equation (A5) is transferred to the L.H.S. If we further use equation (A4) to simplify the penultimate term of (A5) we have

$$C(L) = \psi(L) \frac{dX}{d\psi} \tag{A6}$$

which is equation (1) of Section 1.

We have now established that the simple C method gives the best possible result, but what of the generalization of the C method used in Section 4? In order to cast equation (A3) into something resembling the equations of the C method, it is natural to perform the same formal sequence of operations that led us from (A4) to (A5).

Performing that sequence on (A3) we obtain the analogue to (A5).

$$X = \psi \frac{dX}{d\psi} + \int_L^\infty \psi(L_1) \int \frac{\dfrac{\partial \rho(L_1, M)}{\partial L_1} \displaystyle\int n(L', M)\, dL'}{\displaystyle\int \rho(L', M)\, \phi(L')\, dL'} dM\, dL_1.$$

This suggests that the best generalization of C to the weighted problem is

$$C^* = X - \int_L^\infty \left(\psi \int \frac{\dfrac{\partial \rho}{\partial L_1} n dL'}{\displaystyle\int \rho \phi dL'} dM \right) dL_1 = \psi \frac{dX}{d\psi}. \tag{A7}$$

This formula has the severe drawback that C^* depends on the unknown solution ψ and its differential ϕ. Since $d\rho/dL_1$ does not appear elsewhere in our considerations, it is instructive to perform an integration by parts on L_1; we then find

$$C^* = X - \int_L^\infty \int \left(\frac{\phi(L_1)\, \rho(L_1, M)}{\displaystyle\int \rho(L', M)(L')\, dL'} \int n(L', M)\, dL' \right) dM\, dL_1 + \psi(L) \int \frac{\rho(L, M) \displaystyle\int n dL'}{\displaystyle\int \rho\phi\, dL'} dM. \tag{A8}$$

Consider the interpretation of the middle term:

$$\frac{\phi(L_1)\, \rho(L_1, M)\, dL_1}{\displaystyle\int \rho(L', M)\, \phi(L')\, dL'}$$

is the probability, given an observed point at M, that it will lie with an L value between L_1 and $L_1 + dL_1$ on the assumption that the distribution we seek is $\phi(L)$. $\int n(L', M)\, dL'\, dM$ is the number of points observed between M and $M + dM$. Hence the second term as a whole is the expected number of points to the

right of a line drawn at some value L given the distribution of points with M. The actual number of those points is of course X so it is natural to call the number expected on the basis of the probability distribution ϕ, $E_\phi(X)$. Equation (A8) now reads

$$C^* = \Delta_\phi(X) + \int_L^\infty \phi(L_1)\, dL_1 \int \frac{\rho(L, M) \int n(L', M)\, dL'}{\int \rho(L', M)\, \phi(L')\, dL'}\, dM \qquad \text{(A9)}$$

where $\Delta_\phi X$ is the change of X from its expected value on the basis of the probability distribution ϕ

$$\Delta_\phi X = X - E_\phi(X).$$

In practice we may expect $\Delta_\phi(X)$ to be small since we are trying to choose a ϕ which gives the best representation of these same points. Thus the main contribution to C^* will come from the last term of equation (A9). Writing that term in an unnecessarily complicated but enlightening way we have

$$\int \int_L^\infty \left\{ \frac{\rho(L, M)}{\rho(L_1 M)} \left[\frac{\rho(L_1 M)\, \phi(L_1)}{\int \rho(L', M)\, \phi(L')\, dL'} \right] \int n(L', M)\, dL' \right\} dL_1\, dM. \qquad \text{(A10)}$$

The term in square brackets is the probability distribution for points at a given M. If we weight that probability distribution by $\rho(L, M)/\rho(L_1 M)$ we are performing exactly the same weighting operation* that we did to the actual numbers of points themselves in Section 4. Thus if C is defined in general by

$$C(L) = \int \int_L^\infty \frac{\rho(L, M)}{\rho(L_1, M)}\, n(L_1, M)\, dL_1\, dM$$

then we may write

$$C^* = \Delta_\phi(X) - \Delta_\phi(C) + C \qquad \text{(A11)}$$

where $\Delta_\phi(C)$ is C minus the value of C expected on the basis of the probability distribution ϕ and the distribution with M of the actual points.

Equation (A11) immediately suggests a rapidly convergent method of calculating the best possible ϕ from a given set of points. Since we expect the Δ's to be small we neglect them in the first approximation. We can therefore replace C^* by C and solve equation (A7) in the form

$$\psi = \exp \int \frac{dX}{C}.$$

This is the approximation used in Section 4. However, with the cumulative distribution known to first approximation, we can smooth it and differentiate it to obtain our first estimate of ϕ. This estimate may then be used to calculate the small terms $\Delta_\phi(X)$ and $\Delta_\phi(C)$. On this basis we obtain a second approximation to C^* which can be used in equation (A7) to yield a second approximation to ψ etc.

* The dummy variables L and L_1 are transposed here because we have got $C(L)$ whereas $C(L_1)$ is defined by equation (8).