PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC
103: 1012–1019, September 1991

ON THE DERIVATION OF SELECTION FUNCTIONS FROM REDSHIFT SURVEY DATA

MICHAEL A. STRAUSS[1]

Department of Astronomy, MS 105-24, California Institute of Technology, Pasadena, California 91125

AMOS YAHIL

Astronomy Program, State University of New York, ESS Building, Stony Brook, New York 11794-2100

AND

MARC DAVIS

Department of Astronomy, University of California, Berkeley, California 94720

Received 1991 March 15, revised 1991 June 24

ABSTRACT

We describe a previously unrecognized effect in the derivation of luminosity functions and selection functions from existing redshift survey data, due to binning of quoted magnitudes and diameters. We correct for this effect in the Center for Astrophysics (CfA) and Southern Sky (SSRS) Redshift Surveys. This correction makes subtle but systematic changes in the derived density fields of the CfA survey, especially within 2000 km s^{-1} of the Local Group. The effect on the density field of the SSRS survey is negligible.

Key words: galaxies–redshifts

1. Introduction

The luminosity function of galaxies $\Phi(L)$ is defined as the average number density of galaxies per unit volume of space per unit interval of luminosity L. The luminosity function is thus the starting point for calculating the luminosity density of the Universe, for understanding the distribution of galaxies of different luminosities, and for quantifying selection effects in all surveys of galaxies (e.g., Bingelli, Sandage & Tammann 1988). It is the latter application which will be most relevant for us in the present paper. In order to determine the true density distribution of galaxies from a flux-limited redshift survey, one must compensate for the fact that, at larger distances, only galaxies above an ever larger luminosity enter the sample. Thus, the apparent galaxy density drops with distance as the fraction of the luminosity function lying above this luminosity. This fraction is called the *selection function* $\phi(r)$; thus the density field can be obtained by weighting the galaxy distribution by the inverse of the selection function. Mathematically, the selection function is related to the luminosity function as

$$\phi(r) = \frac{\int_{L_m}^{\infty} \Phi(L)dL}{\int_{L_s}^{\infty} \Phi(L)dL} \quad . \tag{1}$$

Here, $L_m = \max\,(L_s,\, 4\pi r^2 \nu f_{\min})$ is the luminosity a galaxy would have if it were at distance r with observed flux equal to the flux limit of the sample, $f_{\min}$, and $L_s = 4\pi r_s^2\, \nu f_{\min}$ is a minimum luminosity cutoff. The effective frequency of the passband in which the flux is measured ν. The denominator of equation (1) is the mean number density of galaxies with luminosities greater than L_s:

$$n_1 = \int_{L_s}^{\infty} \Phi(L)dL \quad . \tag{2}$$

Thus, the mean number density of galaxies as a function of redshift is simply $n_1\phi(r)$.

The traditional techniques for deriving luminosity functions from redshift survey data use variants of the $1/V_{\max}$ method (e.g., Felten 1977). In a sample of galaxies flux limited to a flux $f_{\min}$, each galaxy at distance r and flux f would still remain in the sample if it were as far away as

$$r_{\max} = r\,(f/f_{\min})^{1/2} \quad , \tag{3}$$

because fluxes scale as the inverse square of distance. Thus, there exists an effective volume $V_{\max}$ within which galaxies of a given luminosity will appear in the sample. The luminosity function $\Phi(L)$ is then calculated as

$$\Phi(L)dL = \sum_i 1/V_{\max} \quad , \tag{4}$$

where the sum is over all the galaxies in the sample with luminosities between L and $L + dL$. Unfortunately, this

[1]Norris Fellow.

method assumes a homogeneous distribution of galaxies, and thus is biased by galaxy inhomogeneities in the volume that is probed by the survey. Fortunately, there exist techniques which derive luminosity functions free of this assumption (Efstathiou, Ellis & Peterson 1988; Bingelli et al. 1988). Yahil et al. 1991 (hereafter Paper 2) discuss a maximum likelihood technique which allows calculation of the galaxian luminosity function $\Phi(L)$ free from the assumption that the volume sampled is homogeneous and apply it to a full-sky flux-limited redshift survey of galaxies detected by the *Infrared Astronomical Satellite (IRAS)*. One need only make the assumption, which is testable a posteriori, that the luminosity function of galaxies is universal, i.e., that it is not a function of position on the sky, distance, or clustering environment. We present this method in Section 2.1 below.

In applying this technique to the Center for Astrophysics redshift survey (Huchra et al. 1983, hereafter CfA), and the Southern Sky Redshift Survey (da Costa et al. 1991, hereafter SSRS), the maximum likelihood function must take into account explicitly the fact that magnitudes and diameters are listed only with finite accuracy. In this paper we demonstrate the importance of this effect in these two samples. In Section 2.1 we present the "standard" maximum likelihood technique, as shown, e.g., in Paper 2. Section 2.2 shows how this technique is modified for the magnitude-limited CfA sample, and Section 2.3 presents results for the diameter-limited SSRS sample. We summarize our results in Section 3.

2. The Derivation of the Selection Function

2.1 *The "Standard" Maximum Likelihood Technique*

The maximum likelihood technique treats the selection function $\phi(r)$ as the basic quantity and the luminosity function $\Phi(L)$ as a derived quantity. The results of this paper parameterize the selection function in the same way as in Paper 2 and in Strauss et al. (1992, Paper 4):

$$\phi(r) = \min\left[1, \left(\frac{r}{r_s}\right)^{-2\alpha}\left(\frac{r^2 + r_*^2}{r_s^2 + r_*^2}\right)^{-\beta}\right] . \quad (5)$$

However, the maximum likelihood technique, and the discretization effects described in the following sections, hold for any parameterization and even for nonparametric variants of the method (Efstathiou et al. 1988).

We wish to derive the selection function of a sample independent of the density field. Thus, we will take as given the measured distance r of each galaxy and ask for the selection function consistent with its observed luminosity L. Define $f(L|r)dL$ as the probability that a galaxy at a *given* distance r has a luminosity between L and $L + dL$. Equivalently, $f(L|r)dL$ is the luminosity *distribution* of galaxies in the sample at distance r. It is given by the value of the luminosity function at L multiplied by dL, divided by the integral of the luminosity function over all luminosities the galaxy could have, and remain in the sample:

$$f(L|r)dL \equiv \frac{dF(L|r)}{dL}\,dL = \begin{cases} \dfrac{\Phi(L)dL}{\int_{L_m}^{\infty}\Phi(L)dL} & \text{if } L \geq L_m \\ 0 & \text{otherwise} \end{cases} . \quad (6)$$

Notice that this expression is independent of n_1, the mean density of the sample. The luminosity function may be expressed in terms of the derivative of the selection function (eq. (1)), so the luminosity distribution may be written

$$f(L|r)dL = -\left[\frac{\partial\phi}{\partial r}\frac{\partial r}{\partial L}\right]\Bigg|_{r=r_{\max}} \frac{1}{\phi(r)}\,dL , \quad (7)$$

where $r_{\max}$ is defined above in equation (3). The product of the function in equation (7) over all galaxies in the sample is proportional to the probability that the assumed selection function is correct, again given the distances to each galaxy. The maximum likelihood technique of Paper 2 consists of maximizing this product with respect to the parameters of the selection function, α, β, and r_*. The implicit assumption is made that the range dL is the same for all galaxies, and thus is an irrelevant multiplicative constant in the maximization of the probability.

2.2 *The CfA Sample*

The CfA survey contains complete redshift information for all galaxies in the Zwicky et al. 1961–68 catalog with $m_Z \leq 14.5$ and in two regions of the sky: $\delta > 0°$, $b > 40°$, and $\delta > -2.5°$, $b < -30°$. The magnitudes of the vast majority of the galaxies in the catalog are listed to a single decimal place. Let us assume that the measured magnitudes are exact and are simply subject to a round-off error to the nearest 0.1 magnitude. The calculated luminosity of a galaxy with distance r and listed magnitude m is then proportional to:

$$L = 4\pi\, r^2\, \nu 10^{0.4(m_{\rm lim} - m)} , \quad (8)$$

where $m_{\rm lim} = 14.5$ is the quoted magnitude limit of the sample. But, in fact, the true magnitude of the source lies in the range $[m - 0.05,\ m + 0.05]$, and thus the true luminosity lies in the range $[L_1, L_2] = [10^{-0.02}L,\ 10^{+0.02}L]$. Thus, the interval dL, which was assumed to be constant in Paper 2, is not the same for each galaxy and must be explicitly reintroduced into the likelihood function. Thus we modify equation (7) to read:

$$f'(L|r)dL = -\frac{\phi(r_{\max,1}) - \phi(r_{\max,2})}{(r_{\max,1} - r_{\max,2})\phi(r)}\frac{\partial r}{\partial L}\Bigg|_{r=r_{\max}} dL , \quad (9)$$

where $r_{\max,1,2}$ correspond to L_1 and L_2, respectively (eq. (3)). In essence, we have replaced the derivative of the selection function in equation (7) by its slope determined over a finite interval. When this is done, the effective magnitude limit becomes $m_{\rm lim} = 14.55$, with a corresponding effect on the calculation of r_{max}.

We calculate the selection function of the CfA sample, using the parameterization of equation (5), and maximizing the likelihood, given either by equation (7) or equation (9). Redshifts are placed in the Local Group barycenter frame using the correction of Yahil, Tammann & Sandage 1977, and objects in the Virgo, Ursa Major, and Coma clusters are collapsed to a common redshift (cf. Paper 2, Paper 4). Galaxies between $r_s = 500$ km s^{-1} and $R_{max} = 8000$ km s^{-1} are used. In each case, n_1, the average density of galaxies with luminosities greater than L_s, is calculated using equation (13) of Paper 2. In the upper panel of Figure 1 we plot the quantity $n_1\phi$, which is the mean number density of galaxies in the sample as a function of redshift, for two cases: the solid line is calculated ignoring the discretization of the magnitudes (eq. (7)) while the dashed line takes the discretization explicitly into account (eq. (9)). In the latter case, the small number of galaxies in the CfA with magnitudes listed to two decimal places are rounded off to the nearest decimal place. The derived selection function parameters are listed in Table 1. The two selection functions are clearly

Table 1
Selection Function Parameters

Survey	α^a	β	r_*	n_1^b
			km s^{-1}	
CfA[c]	0.350	8.206	11050	0.122
CfA[d]	0.345	8.914	12420	0.092
SSRS[c]	0.379	3.765	8480	0.065
SSRS[d]	0.368	3.764	8500	0.060

[a] Best-fit parameters of the selection function to the form given in Eq. 5.
[b] Density in units of (100 km s^{-1})$^{-3}$.
[c] Without correction for discreteness effect.
[d] With correction for discreteness effect.

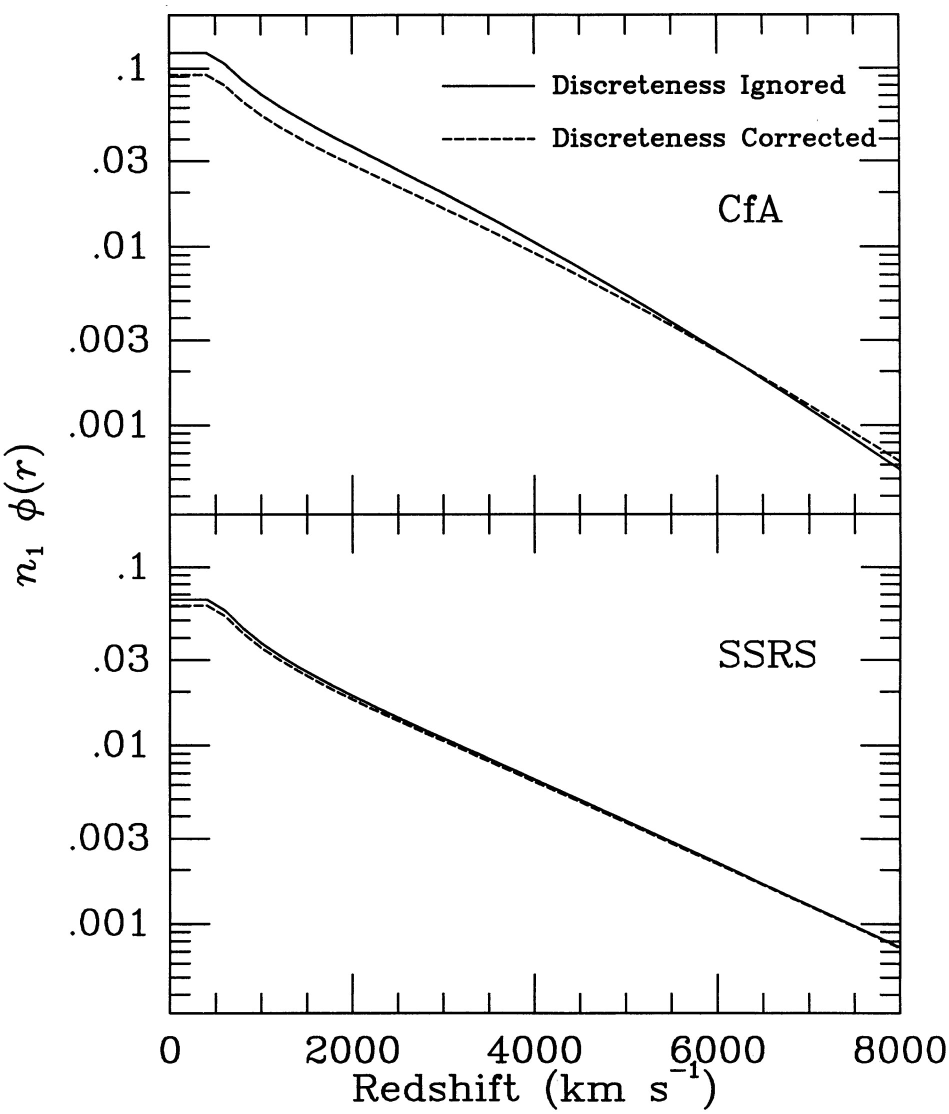

FIG. 1–The quantity $n_1\phi$, which is the mean number density of galaxies in a sample in a homogeneous universe, is given for the CfA (upper panel) and SSRS (lower panel) redshift surveys. In each case, the solid curve is derived ignoring the effects of discretization, while the solid curve corrects for it.

different; which one is right? Paper 2 presents a powerful test of the goodness of fit of the selection function. As $f(L|r)$ is the probability distribution of luminosities at any given distance, the *cumulative* probability $F(L|r) = \int_L^\infty f(L'|r)\,dL'$ must be homogeneously distributed, independent of redshift, in the interval [0, 1]. Any significant deviations from this imply either that we have used the wrong selection function or that the sample is incomplete. Figure 2 shows the distribution of F as a function of r when discretization is ignored. In this case, equation (7) can be integrated immediately to yield

$$F(L|r) = \frac{\phi(r_{\max})}{\phi(r)} \; . \tag{10}$$

The dramatic stripes seen in the figure are caused by the discreteness of the listed magnitudes, which causes discreteness in the quantity $(r/r_{\max})^{1/2}$. The uppermost stripe consists of galaxies with Zwicky magnitude $m_Z = 14.5$ (for which $F \equiv 1$ by definition), the second stripe includes galaxies with $m_Z = 14.4$, and so on. The distribution of F is anything but uniform. Equation (10) cannot be generalized to take into account the discreteness explicitly, so we proceed as follows: we replace each magnitude m by a random magnitude in the interval $[m - 0.05, m + 0.05]$, with distribution given by equation (6) (using, of course, the selection function parameters given in the second line of Table 1). Figure 3 shows the resulting scatter plot of F, as calculated by equation (10). The stripes have disappeared, as they must. The resulting distribution of F in different distance ranges is shown in Figure 4. The curves are normalized to unity throughout, and the error bars are given by Poisson statistics. No significant deviations from a homogeneous distribution are seen, implying both that our fit to the selection function is adequate and that the sample is complete to the quoted limiting magnitude.

2.3 *Diameter-Limited Surveys*

Existing diameter-limited redshift surveys of galaxies (da Costa et al. 1991; Dressler 1991, hereafter SPS) suffer from similar truncations in the listed diameters; diameters in the ESO and UGC catalogs are listed to the nearest 0.1 arc minute. For magnitude or flux-limited surveys,

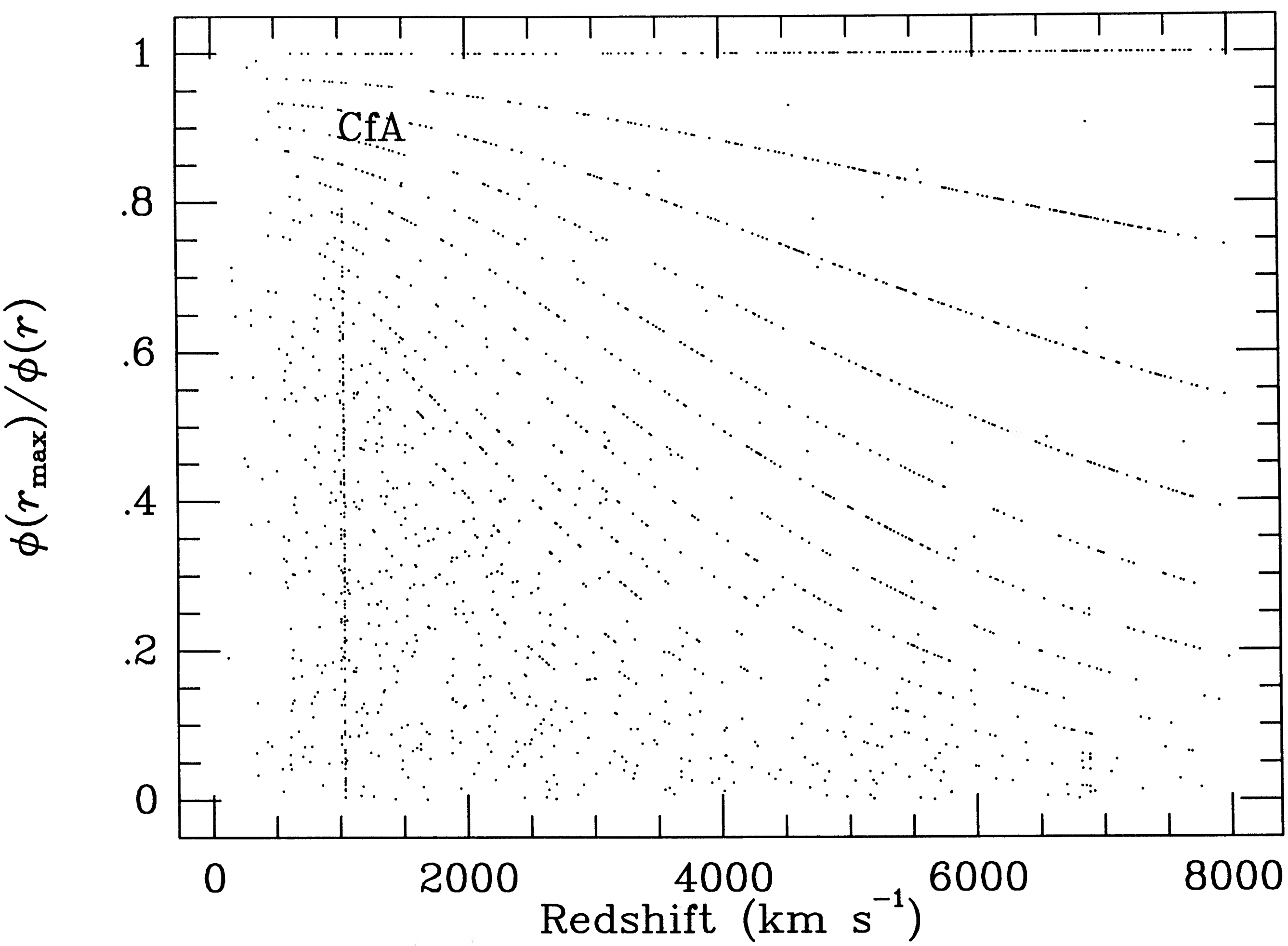

FIG. 2–A scatterplot of F, the cumulative luminosity probability distribution, versus redshift in the CfA sample. The dramatic stripes are caused by the discretization. The few points not falling along the stripes represent galaxies with more precise magnitudes listed. The vertical feature at ~ 1000 km s^{-1} is due to the Virgo cluster, collapsed in redshift space.

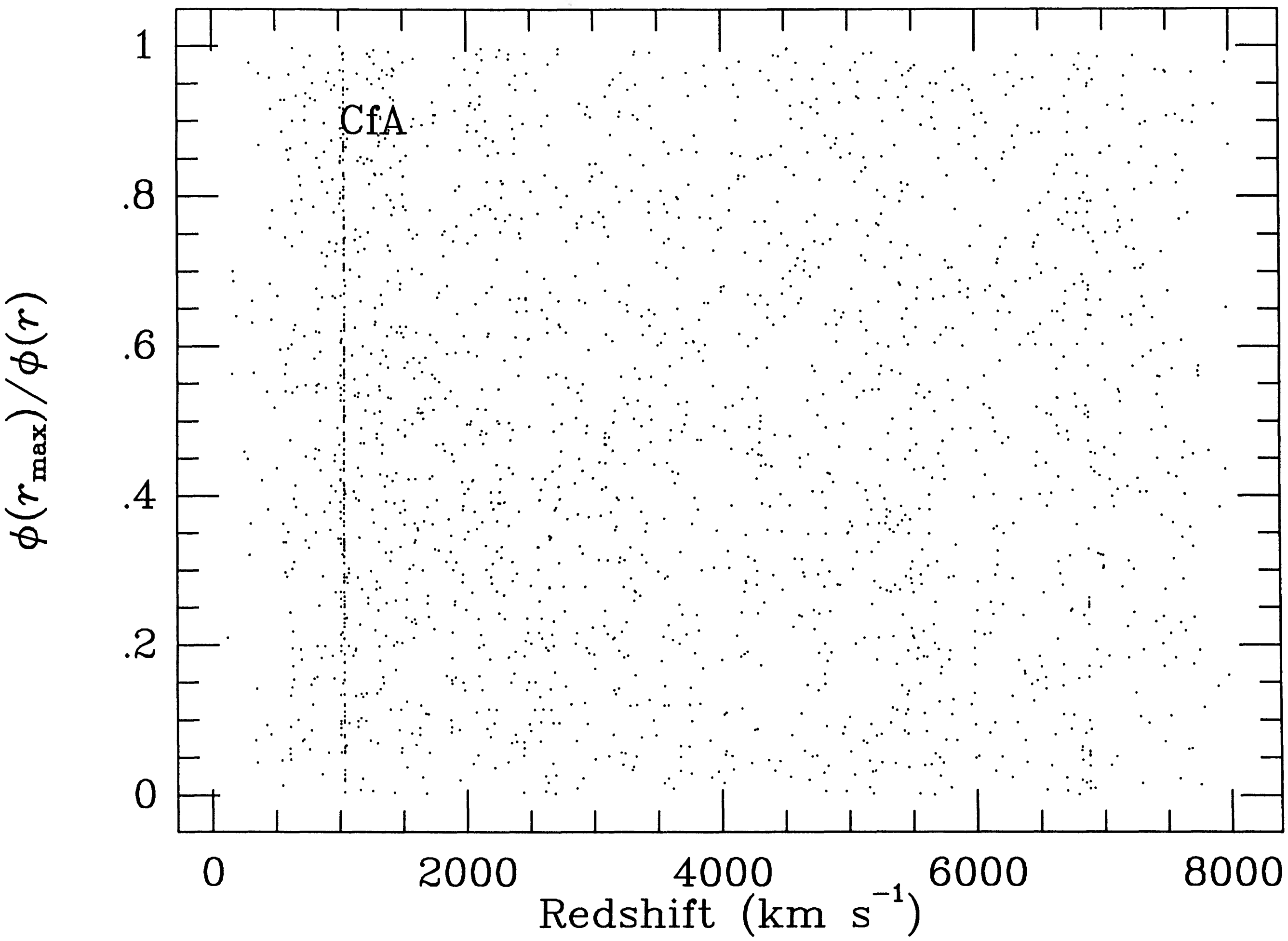

FIG. 3–As in Figure 2, but with the discretization artificially taken out, as described in the text.

the redshift and flux are combined in the intrinsic quantity $L = 4\pi r^2 v f$. In the case of angular diameters, the corresponding quantity is spatial diameter, or $D = \theta r$. If a listed value of θ actually corresponds to a range [$\theta - 0\overset{'}{.}05$, $\theta + 0\overset{'}{.}05$], one can modify the probability density exactly as in equation (9). However, although the discretization affects both major and minor axis diameters, both the SSRS and SPS are designed to be complete to $1\overset{'}{.}26$ in a derived quantity, a "face-on" diameter, as given in da Costa et al. 1988 (see also de Vaucouleurs, de Vaucouleurs & Corwin 1976), which is a complicated function of the major and minor axis diameters. Thus, the effects of discreteness are not as straightforward to correct. For this reason, rather than using the analogue of equation (9) to solve for the selection function, we replace the major-axis diameter of each galaxy with a random diameter in the range [$\theta - 0\overset{'}{.}05$, $\theta + 0\overset{'}{.}05$], with a distribution consistent with the selection function, in analogy with what was done to the CfA magnitudes to create Figure 3. This of course must be done iteratively, so that the input and output selection functions from this procedure are consistent. After this "jittering" of diameters, the sample will be incomplete near $1\overset{'}{.}26$, because there are no galaxies with face-on diameters below $1\overset{'}{.}26$ to be jittered to larger values. Thus, we limit the sample to galaxies with face-on diameters greater than $1\overset{'}{.}3$. The resulting distribution of F for different ranges of distances is shown in Figure 5 for the SSRS survey. As in the case of the CfA, the distribution is consistent with the flat line, implying that the sample is complete and that the resulting selection function is correct. The selection function is plotted in the lower panel of Figure 1, together with the function derived not taking into account the discretization effect. The parameters are listed in Table 1. The selection function is changed only slightly by the correction for discreteness in the major axis diameters, because the correction to face-on diameters (da Costa et al. 1988) already blurs much of the discreteness.

3. Discussion and Conclusions

As Figure 1 shows, the effect of discretization is a subtle one, slightly changing the slope and amplitude of the derived selection function. It is vitally important, however, in deriving the true density field of galaxies from the

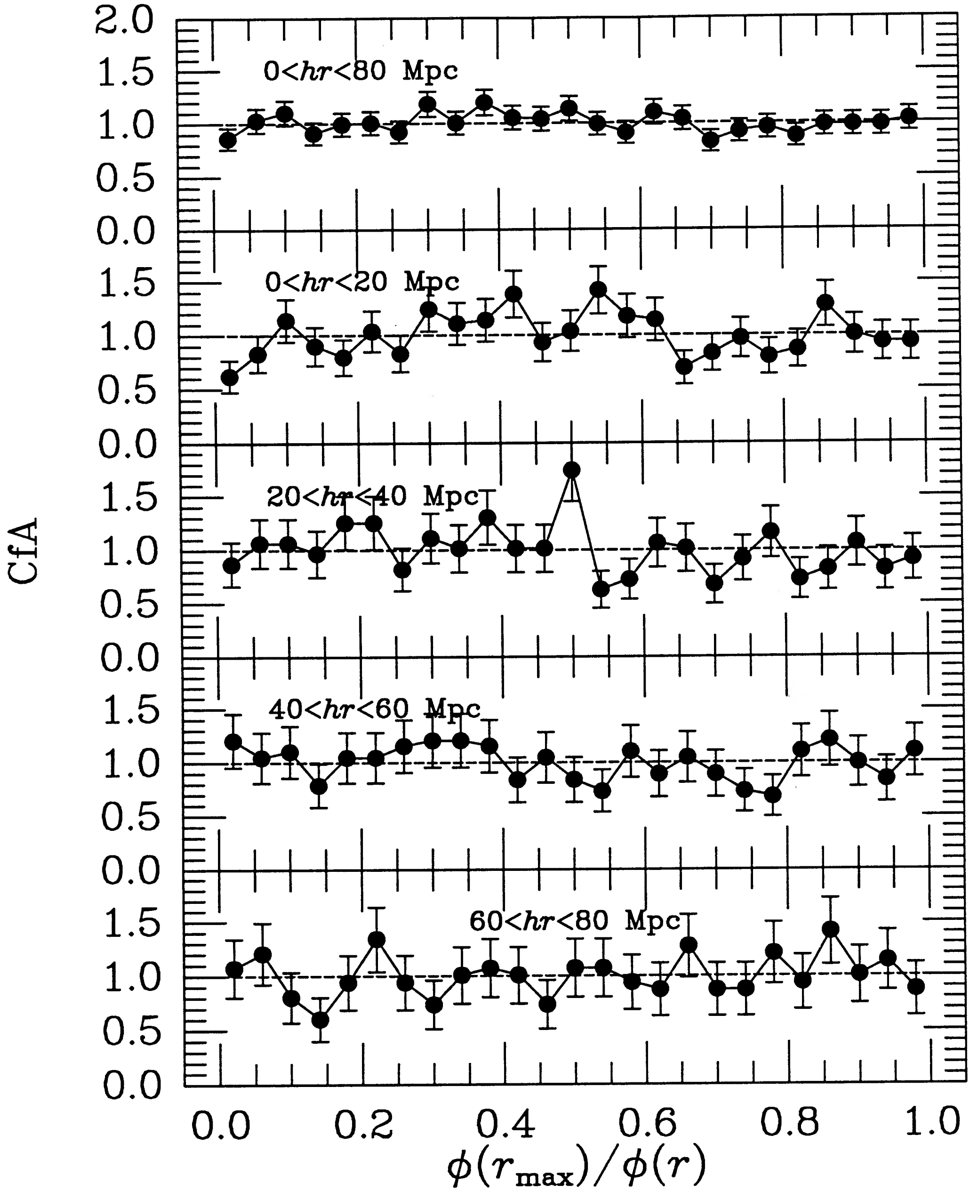

FIG. 4–The distribution of F for galaxies in different distance ranges. The points have been normalized to a mean of unity, and the error bars are Poissonian. Deviations from the unit line indicate either incompleteness or a poor fit to the selection function.

sample. In a uniform universe, the number of galaxies expected in a radial shell is given by the volume of the shell times its mean density, $n_1\phi$. Thus, the fractional density in shells is given by the ratio of the number of galaxies in a redshift bin to $4\pi r^2\ \Delta r\ n_1\phi$, where r is the redshift of the bin and Δr is its width. Figure 6 shows the average density relative to the mean on radial shells. The upper panel is for the CfA, and the lower panel is for the SSRS. The solid line in each case is calculated using the selection function assuming no discretization, while the dashed line uses the correct selection function. There are systematic differences in the derived density field in both cases. The derived overdensity of the Local Supercluster ($r \approx 1000$ km s^{-1}) is enhanced in the CfA when the correct selection function is used, and the region at larger distances comes closer to mean density. These differences are critical in evaluating, for example, the gravitational pull of these structures on the Local Group (Davis & Huchra 1982; Pellegrini & da Costa 1990) or in comparing the density fields traced by optical and IRAS galaxies (Paper 4). As discussed in the latter paper, the comparison of the IRAS catalog to that of the CfA survey is quite good, but there remains a problem with the comparison of the IRAS and SSRS catalogs, probably related to the nature of the diameter errors in the SSRS selection (Santiago & Strauss 1991). The density field of the SSRS sample is almost unaffected by the discreteness correction, for the reason mentioned at the end of the

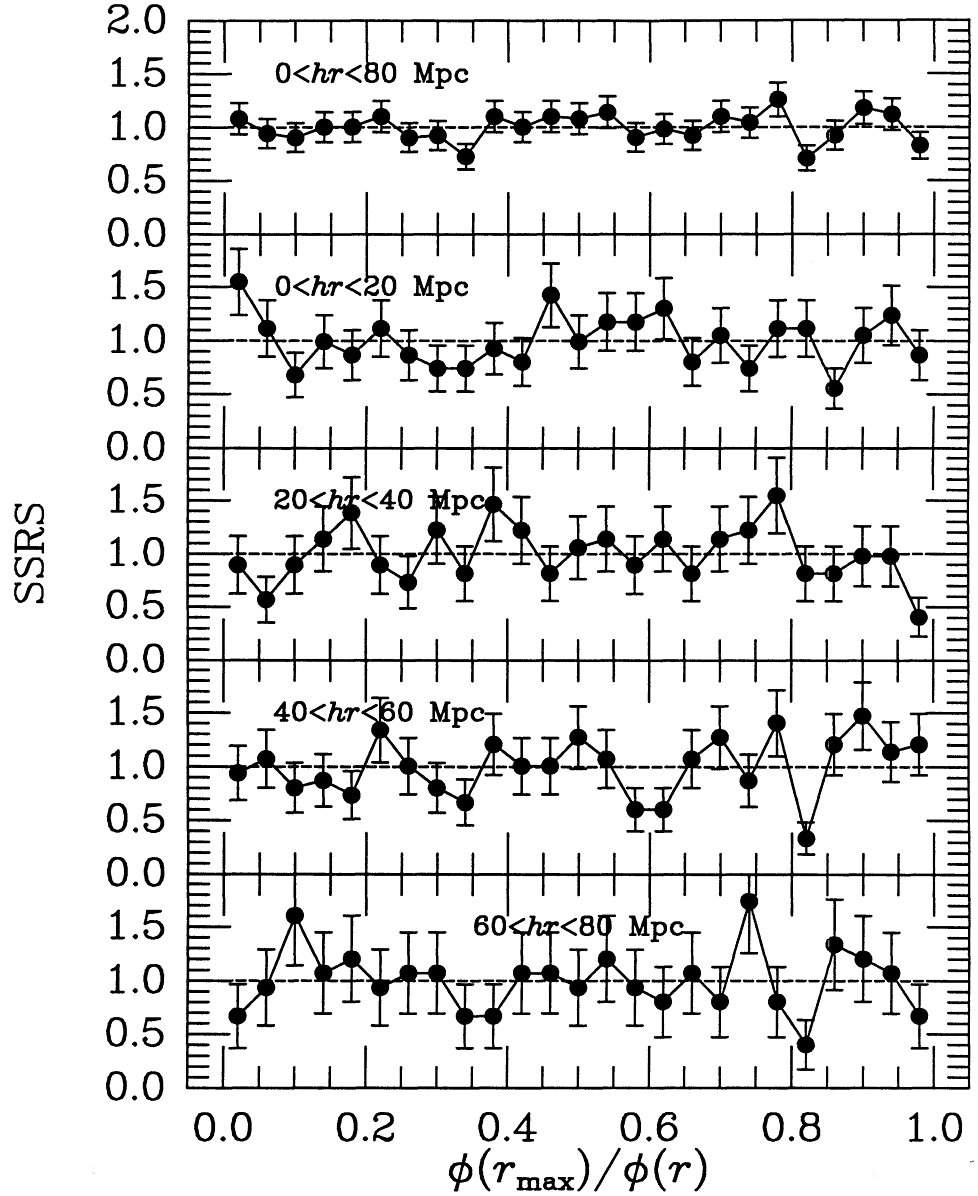

FIG. 5–As in Figure 4, for the SSRS sample. The discreteness effect has been artificially removed, in a similar way to Figure 3.

previous section.

A related issue to the effect of magnitude discretization is the effect of magnitude errors. It might be argued that, in the face of the 0.3-magnitude errors thought to exist in the CfA (Huchra 1976), it is meaningless to worry about the effect of 0.1-magnitude discretization. Efstathiou et al. 1988 show that such errors in fact do not bias the estimate of the density field, if the magnitude error is independent of the magnitude (i.e., constant fractional flux errors). The biases in the density field due to constant magnitude and constant flux errors will be discussed in a paper in preparation (Santiago & Strauss 1991).

We thank John Huchra and Basilio Santiago for useful discussions. This work has been supported under the IRAS extended mission program, grants from the NSF, and NASA grant NAG 51228. M.A.S. acknowledges the support of a Norris postdoctoral fellowship.

REFERENCES

Bingelli, B., Sandage, A., & Tammann, G. A. 1988, ARA&A, 26, 509

da Costa, L., Pellegrini, P., Davis, M., Meiksin, A., Sargent, W., & Tonry, J. 1991, ApJS, 75, 935 (SSRS)

da Costa, L. N., et al. 1988, ApJ, 327, 544

Davis, M., & Huchra, J. 1982, ApJ, 254, 437

de Vaucouleurs, G., de Vaucouleurs, A., & Corwin, H. G., Jr. 1976, Second Reference Catalogue of Bright Galaxies (Austin, University of Texas Press)

Dressler, A. 1991, ApJS, 75, 241 (SPS)

Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, MNRAS, 232, 431

Felten, J. E. 1977, AJ, 82, 861

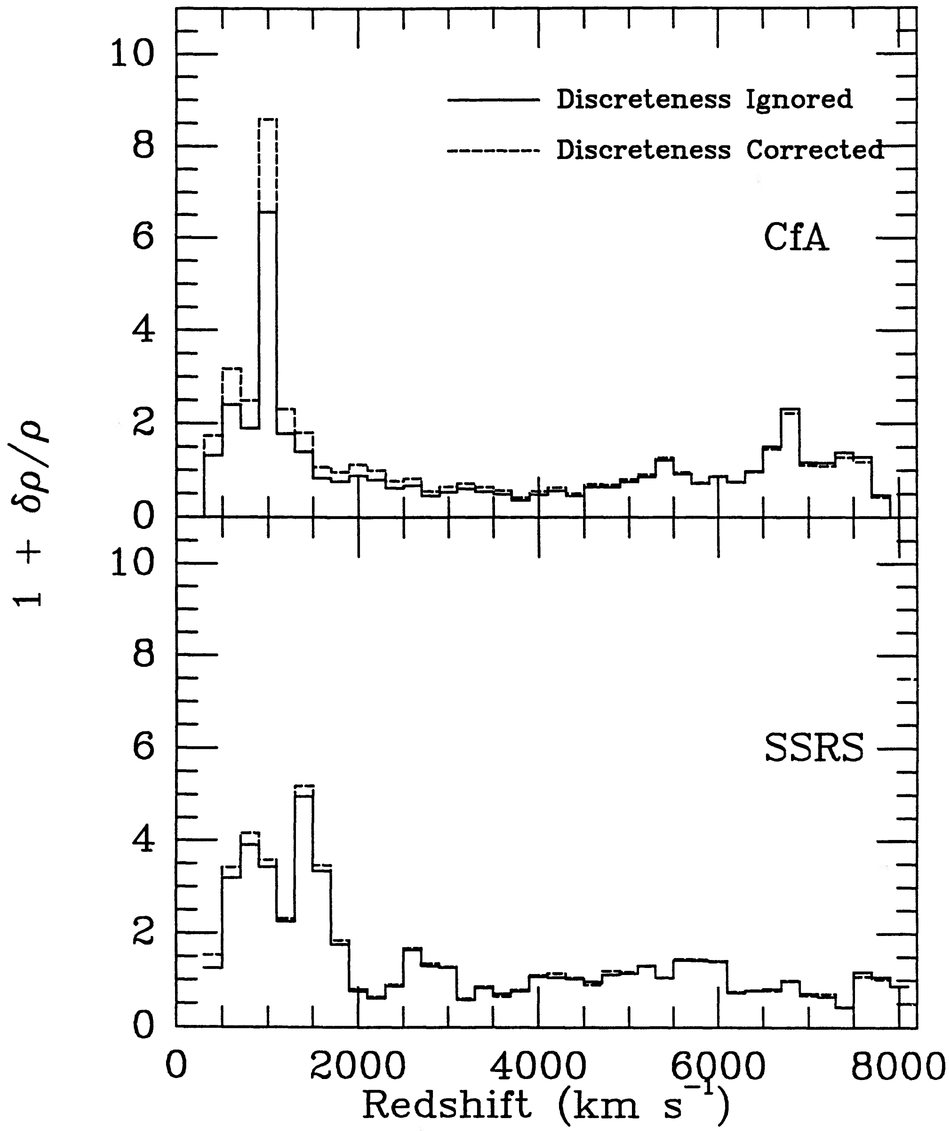

FIG. 6–The fractional density relative to the mean on radial shells is given for the CfA (upper panel) and SSRS (lower panel) redshift surveys. In each case, the solid curve is derived ignoring the effects of discretization, while the solid curve corrects for it.

Huchra, J. 1976, AJ, 81, 952

Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, ApJS, 52, 89 (CfA)

Pellegrini, P. S., & da Costa, L. N. 1990, ApJ, 357, 408

Santiago, B. X., & Strauss, M. A. 1991, in preparation

Strauss, M. A., Davis, M., Yahil, A., & Huchra, J. P. 1992, ApJ, February 10 issue (Paper 4)

Yahil, A., Strauss, M. A., Davis, M., & Huchra, J. P. 1991, ApJ, 372, 380 (Paper 2)

Yahil, A., Tammann, G. A., & Sandage, A. 1977, ApJ, 217, 903

Zwicky, F. et al. 1961–68, Catalog of Galaxies and Clusters of Galaxies (Pasadena, California Institute of Technology)