Astr 511: Galaxies as galaxies

University of Washington
Mario Jurić & Željko Ivezić

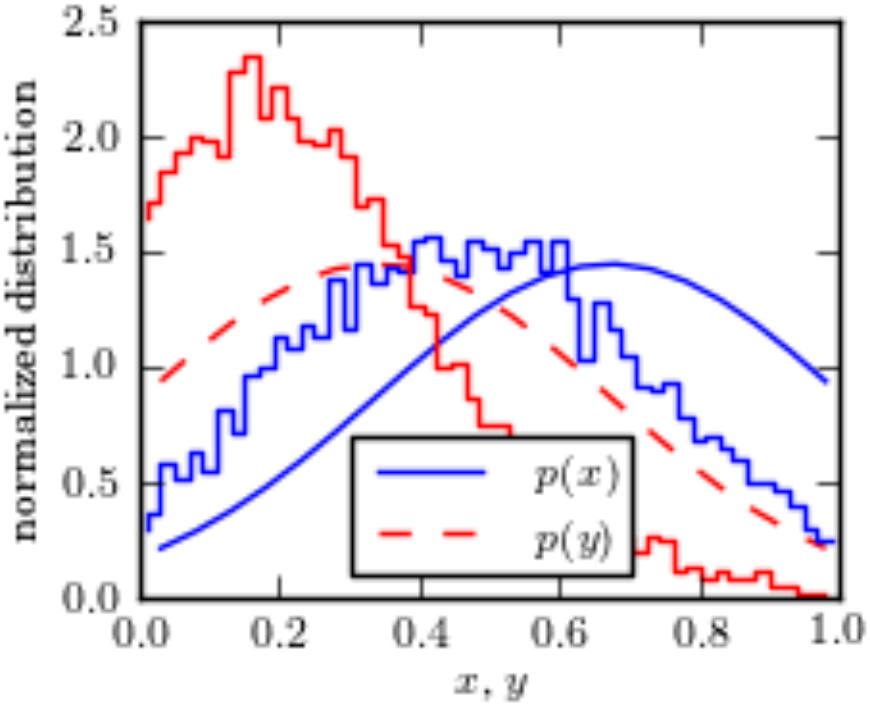
Lecture 5:

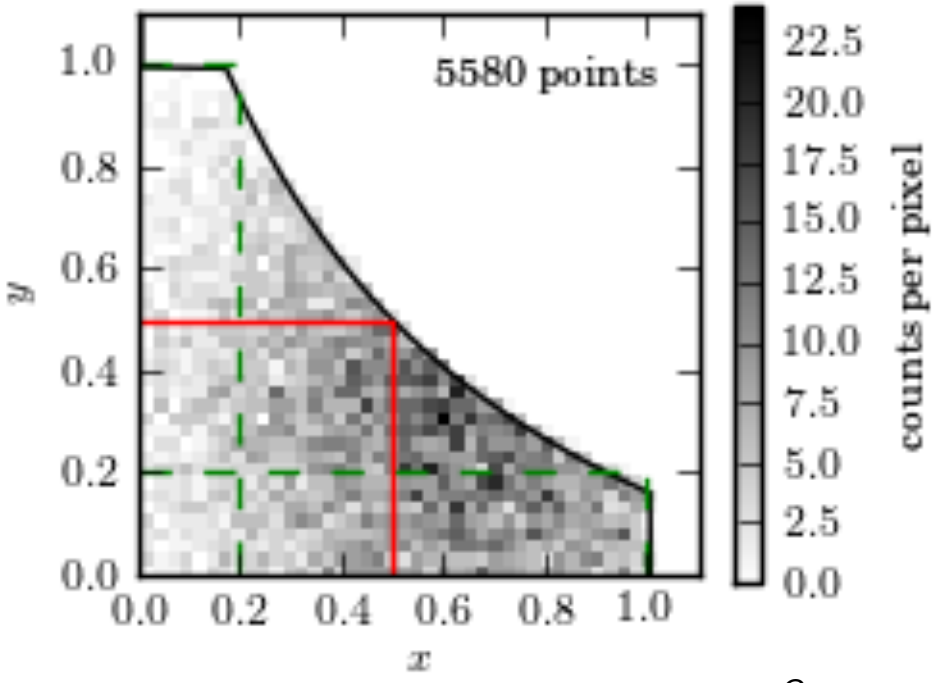
Luminosity and mass functions of galaxies: II

Outline

- Luminosity function: basic concepts
- Stellar mass function in the Milky Way
- **Methods for estimating LF from data**

- Toy example: the x-y "measurements" are truncated (similar to a flux-limited sample in astronomy).
- Histograms of measured quantities are biased for a truncated sample.
- Unbiased distributions can be obtained using various methods (V/V_{max} method, C^- method, maximum likelihood method)





The C^- method for estimating LF

- Lynden-Bell (1971, MNRAS 155, 95); a non-parametric method that works for separable LFs, $\Psi(L,z) = \Phi(L)n(z)$

- practically all non-parametric methods can be reduced to the C^- method (Petrosian 1992)

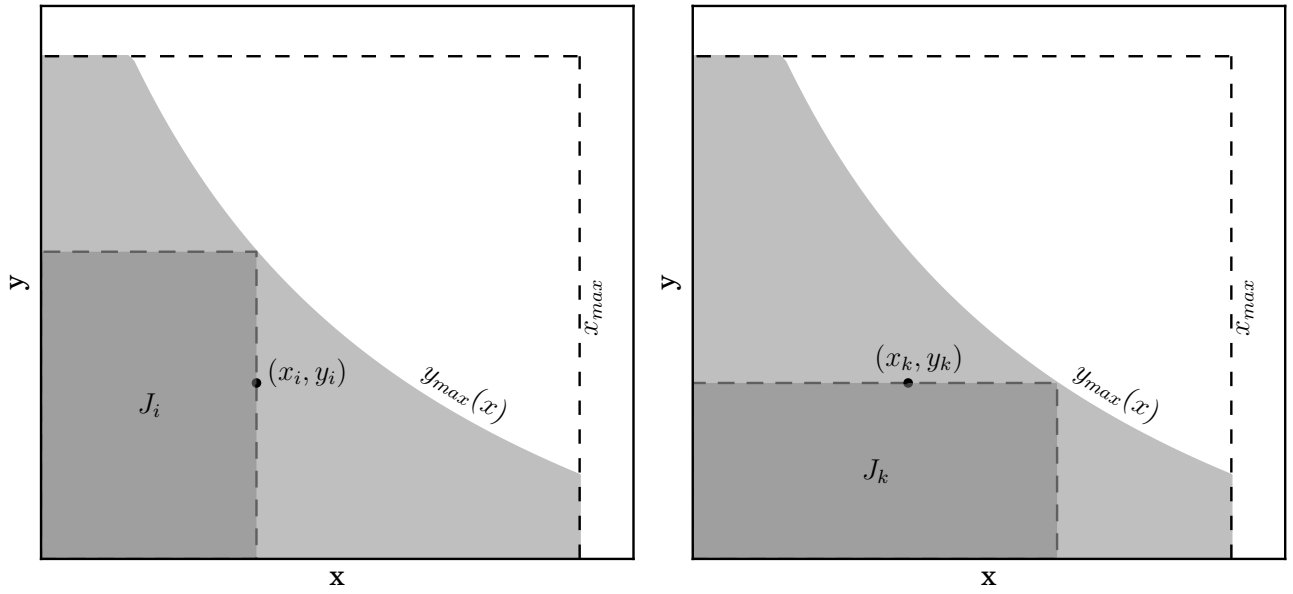
- parametric methods are usually based on maximizing likelihood (e.g. Marshall 1985)

- the simplest and most famous method, the V_{max} method (Schmidt 1968), requires binning in two axes simultaneously, while with the C^- method data is binned only one axis at a time (e.g. Fan et al. 2001)

- How do we know that separable LF is a good guess for our data?

C^- method

- Given a set of measured pairs (x_i, y_i), with $i = 1 \ldots N$, and *known* relation $y_{max}(x)$, estimate the two-dimensional distribution, $n(x, y)$, from which the sample was drawn. Assume that measurement errors for both x and y are negligible compared to their observed ranges, that x is measured within a range defined by x_{min} and x_{max}, and that the selection function is 1 for $0 \leq y \leq y_{max}(x)$ and $x_{min} \leq x \leq x_{max}$, and 0 otherwise.

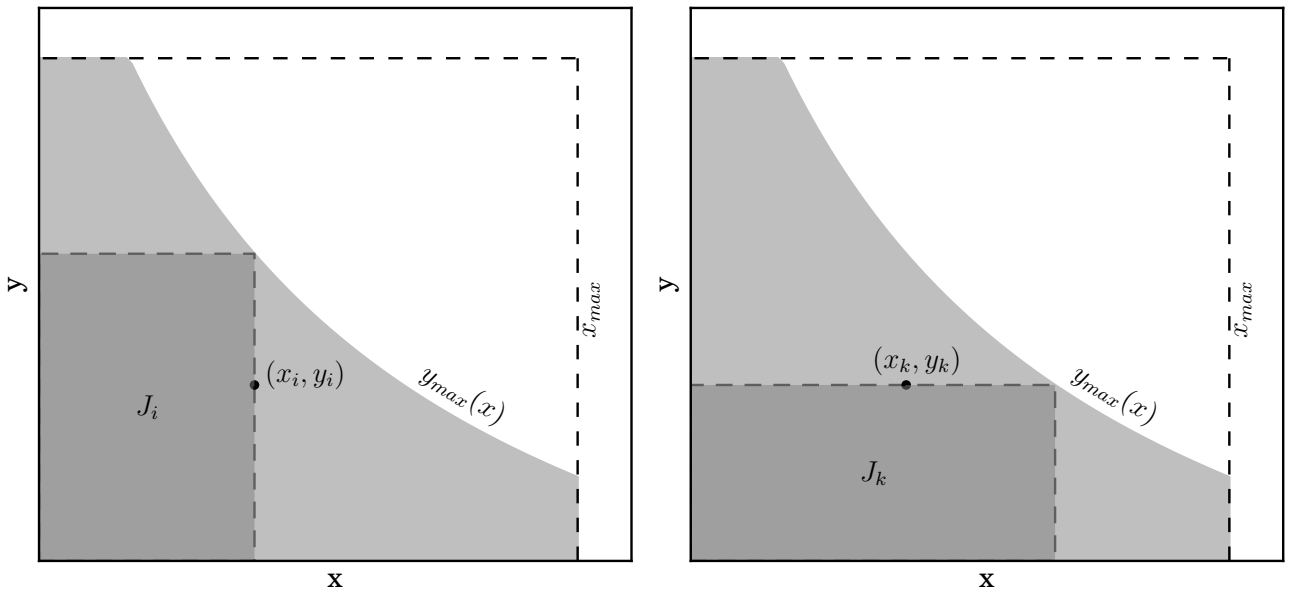


C^- method

- C^- method is applicable when the distributions along the two coordinates x and y are uncorrelated, that is, when we can assume that the bivariate distribution $n(x, y)$ is separable

$$n(x, y) = \Psi(x)\, \rho(y). \tag{1}$$

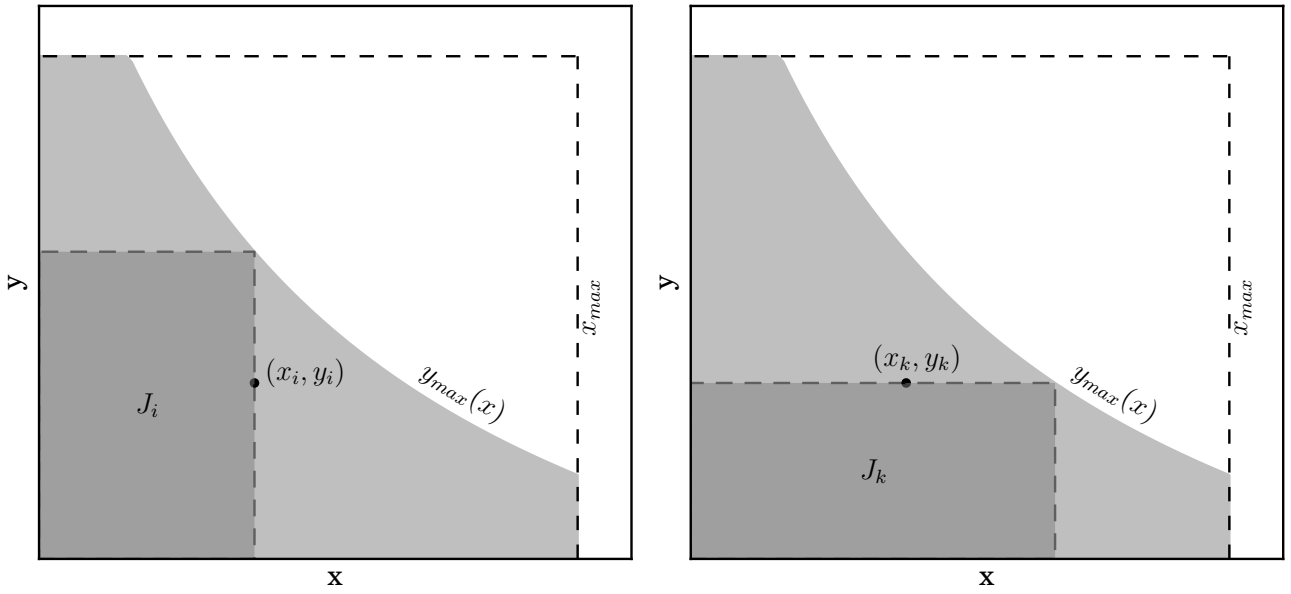
Therefore, before using the C^- method we need to demonstrate that this assumption is valid.



C^- method

- Define a *comparable* or *associated* set for each object i such that $J_i = \{j : x_j < x_i, y_j < y_{max}(x_i)\}$; this is the largest x-limited and y-limited data subset for object i, with N_i elements (see the left panel).

- Sort the set J_i by y_j; this gives us the rank R_j for each object (ranging from 1 to N_i)



C^- method

- Define a *comparable* or *associated* set for each object i such that $J_i = \{j : x_j < x_i, y_j < y_{max}(x_i)\}$; this is the largest x-limited and y-limited data subset for object i, with N_i elements.

- Sort the set J_i by y_j; this gives us the rank R_j for each object (ranging from 1 to N_i)

- Define the rank R_i for object i in *its* associated set: this is essentially the number of objects with $y < y_i$ in set J_i.

- If x and y are truly independent, R_i must be distributed *uniformly* between 0 and N_i.

C^- method

- If x and y are truly independent, R_i must be distributed *uniformly* between 0 and N_i.

- In this case, it is trivial to determine the expectation value and variance for R_i: $E(R_i) = E_i = N_i/2$ and $V(R_i) = V_i = N_i^2/12$. We can define the statistic

$$\tau = \frac{\sum_i (R_i - E_i)}{\sqrt{\sum_i V_i}} \tag{2}$$

  If $\tau < 1$, then x and y are uncorrelated at $\sim 1\sigma$ level.

- A quiz question for in-class discussion: how would the above expression for τ look like if the uniform distribution were replaced by a Gaussian (normal) distribution?

C^- method

Assuming that $\tau < 1$, it is straightforward to show using relatively simple probability integral analysis (e.g., see Appendix in Fan et al. 2001), as well as the original Lynden-Bell's paper, how to determine cumulative distribution functions. The cumulative distributions are defined as

$$\Phi(x) = \int_{-\infty}^{x} \Psi(x')dx', \tag{3}$$

and

$$\Sigma(y) = \int_{-\infty}^{y} \rho(y')dy'. \tag{4}$$

Then,

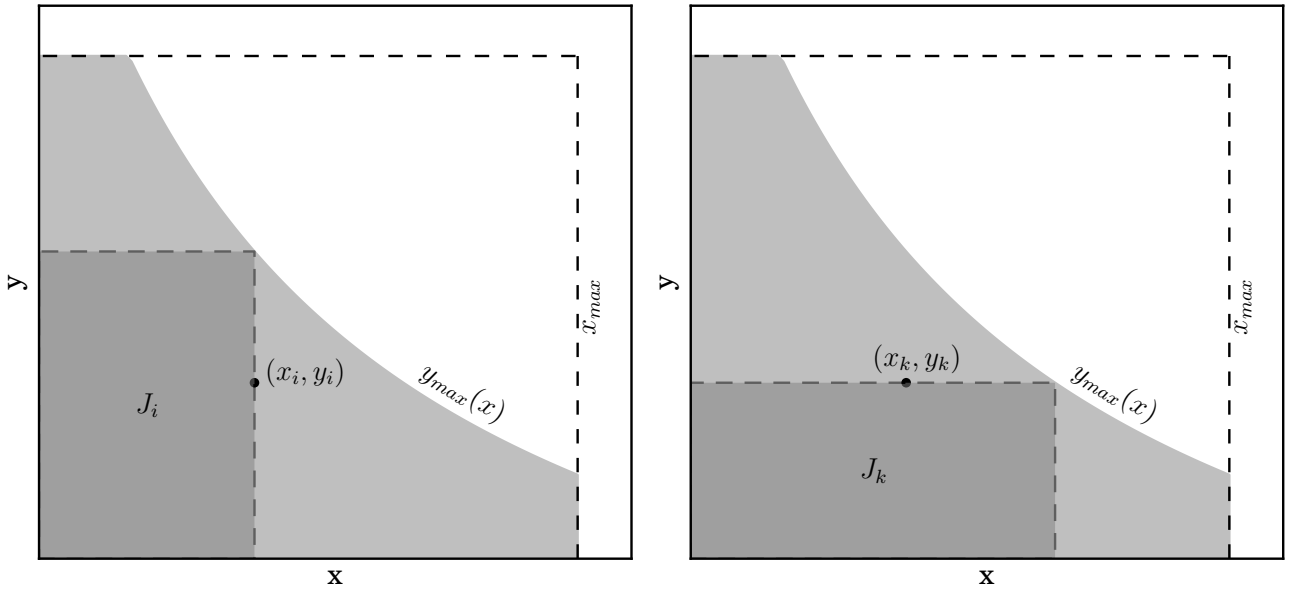
$$\Phi(x_i) = \Phi(x_1)\,\Pi_{k=2}^{i}(1 + 1/N_k) \tag{5}$$

where it is assumed that x_i are sorted ($x_1 \leq x_k \leq x_N$).

C^- method

Analogously, if M_k is the number of objects in a set defined by $J_k = \{j : y_j < y_k, y_{max}(x_j) > y_k\}$ (see the right panel of figure below), then

$$\Sigma(y_j) = \Sigma(y_1)\,\Pi_{k=2}^{j}(1 + 1/M_k). \tag{6}$$



astroML implementation of the C^- method

- Note that both $\Phi(x_j)$ and $\Sigma(y_j)$ are defined on non-uniform grids with N values, corresponding to the N measured values.

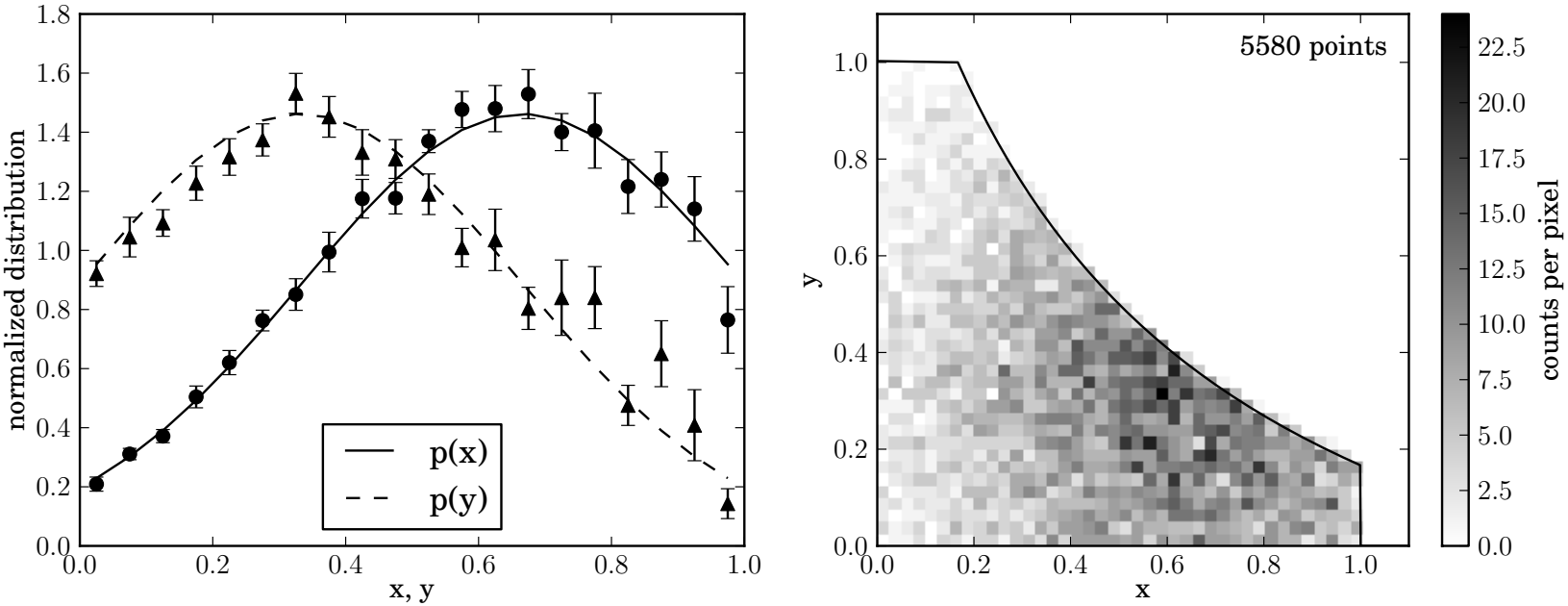
- Essentially, the C^- method assumes a piece-wise constant model for $\Phi(x)$ and $\Sigma(y)$ between data points (equivalently, differential distributions are modeled as Dirac's δ functions at the position of each data point).

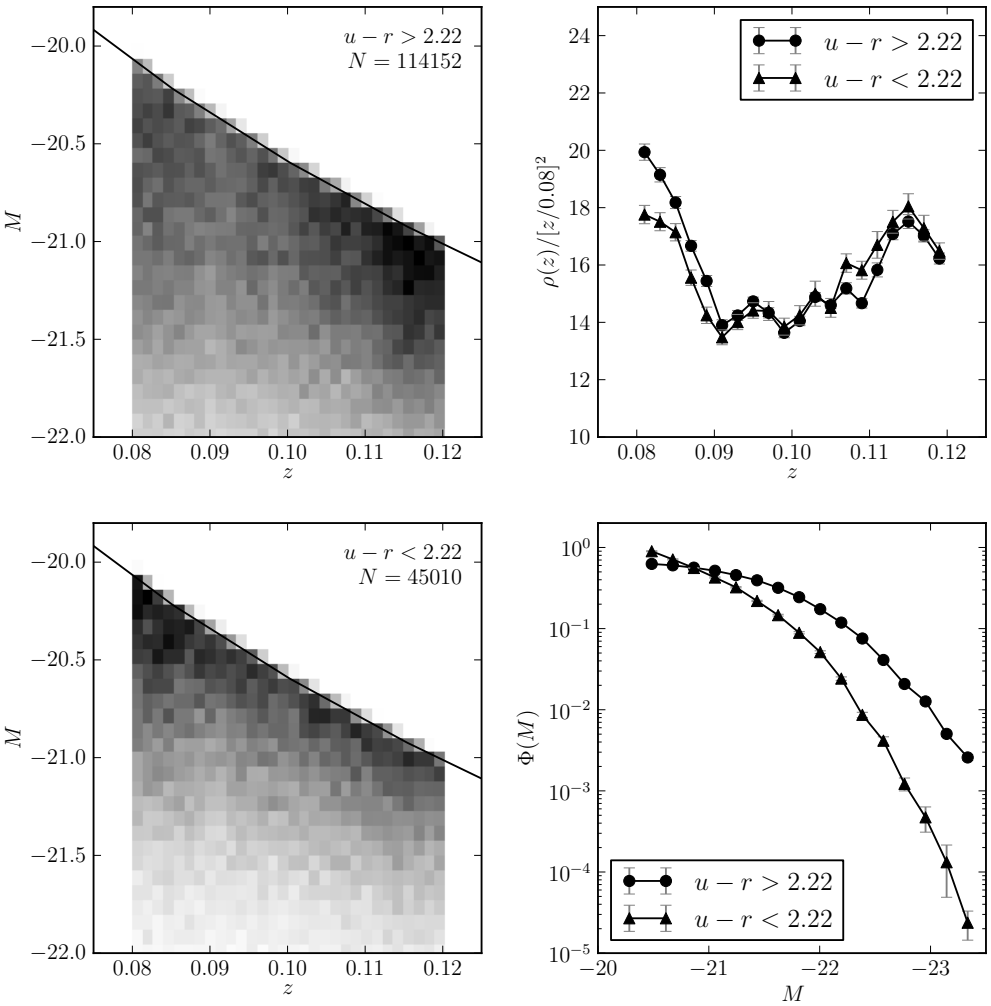
- As shown by Petrosian (1992), $\Phi(x)$ and $\Sigma(y)$ represent an optimal data summary.

- The differential distributions $\Psi(x)$ and $\rho(y)$ can be obtained by differentiating cumulative distributions in the relevant axis; an approximate normalization can be obtained by requiring that the total predicted number of objects is equal to their observed number.

- **astroML Book Figure 4.9:** The right panel shows a realization of truncated separable two-dimensional Gaussian distribution (with the truncation given by the solid line). The lines in the left panel show the true one-dimensional distributions of x and y, and the points are computed from the truncated data set using the C^- method (with error bars from 20 bootstrap resamples).

https://www.astroml.org/book_figures/chapter4/fig_lyndenbell_toy.html





- **astroML Book Figure 4.10:** The luminosity function for two $u-r$ color-selected subsamples of SDSS galaxies from the spectroscopic sample, with redshift in the range $0.08 < z < 0.12$ and flux limited to $r < 17.7$.
- The left panels show the distribution of sources as a function of redshift and absolute magnitude. The distribution $p(z, M) = \rho(z)\Phi(m)$ is obtained using Lynden-Bell's method, with errors determined by 20 bootstrap resamples, and shown in the right panels.

https://www.astroml.org/book_figures/chapter4/fig_lyndenbell_gals.html

Test of L-z Correlation. II

- In reality, the selection function is typically complex: $s(L, z|SED, ...)$ (no sharp faint limit!)

- First define a generalized comparable set (Fan et al. 2001; AJ 121, 54) $J_i = \{j : L_j > L_i\}$; this is a luminosity limited data subset for object i

- Then generalize N_i to

$$T_i = \sum_{j=1}^{N_i} \frac{s(L_i, z_j|SED_j)}{s(L_j, z_j|SED_j)}, \tag{7}$$

  and redefine the rank accordingly

$$R_i = \sum_{j=1}^{N_i} \frac{s(L_i, z_j|SED_j)}{s(L_j, z_j|SED_j)}, \tag{8}$$

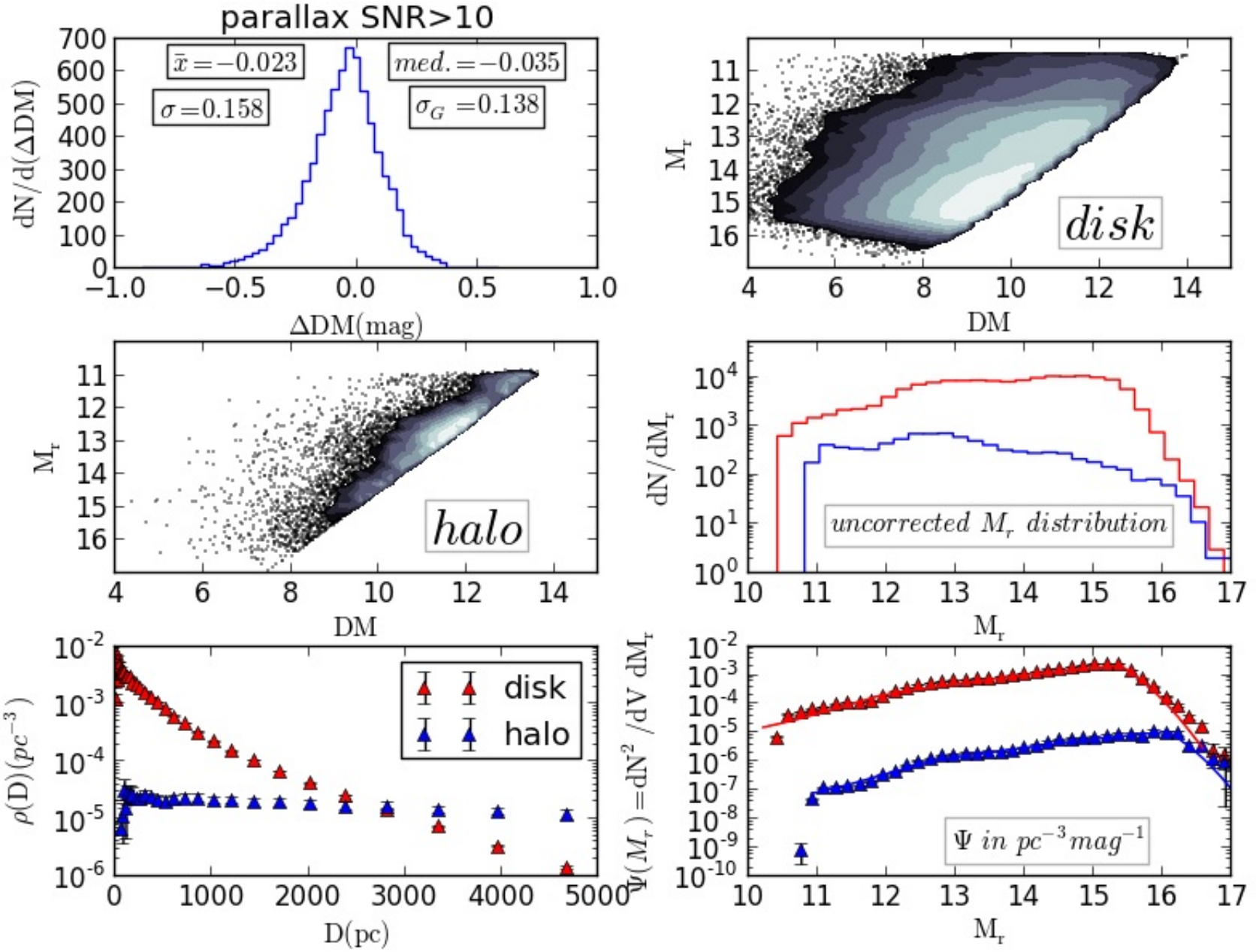
  for $z_j < z_i$. It follows that $E(R_i) = T_i/2$ and $V(R_i) = T_i^2/12$.

Evolving Luminosity Function

- The C^- method is simple and optimal, but it is valid only for uncorrelated variables (separable luminosity function). What do we do when the τ test suggests correlated variables?

- Kelly, Fan and Vestergaard (2008, ApJ 682, 874) described a powerful and completely general Bayesian approach (see their Appendix A for a nice introduction to Bayesian methodology). While too complex for homework, this is a fantastic method – if you ever come again across the problem of estimating a general multi-dimensional distribution that is sampled with non-negligible and possibly complex selection function, remember that paper!

- . . . and let's now say a few words to help you with HW # 2


parallax SNR>10
$\bar{x} = -0.023$
$med. = -0.035$
$\sigma = 0.158$
$\sigma_G = 0.138$
dN/d(ΔDM)
ΔDM(mag)
disk
halo
DM
M_r
uncorrected M_r distribution
dN/dM$_r$
disk
halo
$\rho(D)(pc^{-3})$
D(pc)
$\Psi(M_r) = dN^2/dV\,dM_r$
Ψ in $pc^{-3}mag^{-1}$

LF normalization

- The C^- method does not know (or need) details about our sample; in particular, it cannot give us the overall LF normalization!

- We will use HW#2 problem to discuss normalization in more detail; we can talk about three levels of normalization in this case:

  1. The sample normalization: if we didn't have the selection effects, how many objects would our sample contain?

  2. Normalization to the full sky: we need to know the sky coverage for our sample (and have arguments why we can extrapolate to the whole sky).

  3. Extrapolation from the volume probed by the sample to some other position; here, we want to know LF at $z = 0$.

LF normalization: the sample normalization

If we didn't have the selection effects, how many objects would our sample contain?

- To recap, the cumulative luminosity (absolute magnitude) function is $\Phi_c(M_j)$ and the cumulative distance distribution is $n_c(D_j)$ where $j = 1...N$.

- Both $\Phi_c(M_j)$ and $n_c(D_j)$ are direct outputs from the C^- method; let us renormalize them as $\Phi_c(M_N) = 1$ and $n_c(D_N) = 1$, where it is assumed that M_j and D_j arrays are sorted so that M_N and D_N are their maxima (btw, C^- would return $\Phi_c(M_N) = N$ and $n_c(D_N) = N$).

- The number of points, n, brighter than some arbitrary M^* and closer than some arbitrary D^* is then

$$n(M < M^* \text{ and } D < D^*) = C\,\Phi_c(M^*)\,n_c(D^*). \qquad (9)$$

where we (still) don't know C (n.b. C is dimensionless).

LF normalization: the sample normalization

- Now, if make sure that M^* and D^* are within our selection volume (the implied apparent mag must be above our cutoff) and thus unaffected by selection effects, then we get C from

$$N^o(M < M^o \text{ and } D < D^o) = C\,\Phi_c(M^o)\,n_c(D^o), \qquad (10)$$

  which is almost the same expression as on the previous page, except that here $n(M < M^*$ and $D < D^*)$ is replaced by $N^o(M < M^o$ and $D < D^o)$: the actual number of objects in our sample that satisfy this condition.

- This is not mathematically optimal solution for C because N^o is a random variable, but with modern large samples this is nit-picking; the optimal procedure would integrate over the full sample, but nevertheless would still need to adopt an interpolation procedure for $\Phi_c(M)$ and $n_c(D)$...

- Given the real sample size, N, that is affected by selection effects, the "corrected" sample size is C!

LF normalization: the sample normalization

- The number of points per unit two-dimensional area, $dA = dM\,dD$, is then

$$\frac{d^2N}{dMdD} = C\left(\frac{d\Phi_c(M)}{dM}\right)\left(\frac{dn_c(D)}{dD}\right), \qquad (11)$$

  where we now know C and can easily take (numerical) derivatives $d\Phi_c(M)/dM$ and $dn_c(D)/dD$ (where $\Phi_c(M)$ and $n_c(D)$ came from C^- and are normalized to 1).

- The quantities in parenthesis are differential distribution functions.

- When normalizing to the full sky (step #2), we need to know the fraction of sky, f_{sky}, covered by our sample; if justified, we need to multiply C by $1/f_{sky}$: $C_{sky} = C/f_{sky}$.

LF normalization: extrapolation

How do we go from $d^2N/(dMdD)$ to volume density?

$$\Phi(M,D) \equiv \frac{d^2N}{dMdV} = \frac{d^2N}{dMdD}\frac{dD}{dV}, \tag{12}$$

where $dV = 4\pi D^2 dD$. We have two cases of interest:

- **Case 1:** We seek the volume density vs. D, $\rho(D)$, and we don't care about M distribution:

$$\rho(D) = \int \Phi(M,D)dM = \frac{C_{sky}}{4\pi D^2}\left(\frac{dn_c(D)}{dD}\right), \tag{13}$$

  where we used the fact that $\int_{-\infty}^{\infty}\left(\frac{d\Phi_c(M)}{dM}\right)dM = \Phi_c(M_N) = 1$.

- Unit for $\rho(D)$ is the number of objects per (distance unit)^3 (remember that $n_c(D)$ was dimensionless and normalized to unity at $D = D_N$; the unit comes from taking derivative with respect to D, $dn_c(D)/dD$, and from the $1/D^2$ term).

LF normalization: extrapolation

- Given $\rho(D)$, we can fit some function to it and **extrapolate** to get $\rho(D = D_0)$ (and thus the ratio $\rho(D)/\rho(D = D_0)$ for any D_0, including $\rho(D_0)/\rho(D_N)$).

- **Case 2:** We want to know the M distribution at some $D = D_0$, call it $\psi(M|D = D_0)$ (e.g. $D_0 = 0$ corresponding to solar neighborhood, as discussed in this HW). First, at $D = D_N$ (recall $n_c(D_N) = 1$)

$$\psi(M|D = D_N) = \int \Phi(M, D) dD = C_{sky} \left(\frac{d\Phi_c(M)}{dM} \right). \quad (14)$$

- Then, extrapolating to D_0 (unit for ψ is the number of objects per mag; this is what we compare to the "true" LF)

$$\psi(M|D = D_0) = \psi(M|D = D_N) \frac{\rho(D_0)}{\rho(D_N)}. \quad (15)$$