Bertinoro 5 (JVW)

Surveys: Detection: Luminosity Functions

Detection – what do we mean?

- **preliminary** to much else that happens in astronomy, whether it means locating a spectral line, a faint star or a gamma-ray burst.

- we take it here as **location**, and **confident measurement**, of some sort of **feature** in a fixed region of an image or spectrum.

- elusive objects or **features at the limit of detectability** tend to become the focus of interest in any branch of astronomy.

- then concept of detection (and non-detection) requires careful examination.

- **non-detections important** because they define how representative any catalogue of objects may be. This set of non-detections can represent **vital information in deducing the properties of a population of objects.**

- if something is **never detected**, it's a datum, and can be exploited statistically; **every observation potentially contains information, constrains emission energy.**

- possible to deduce **distributions of parameters**, e.g. luminosity function from the detections/catalogue approach, or directly from **fairly raw sky data**.

- we consider both – but detections first.

Detection – a model-fitting process

- When we say "We've got a detection" we generally mean **"We have found what we were looking for"** – OK at reasonable signal-to-noise.
- e.g comparison of model **point-spread functions** with the data.
- **extended objects**? – **wider range of models** needed.

A clear statistical model is required. The noise level (residuals from the model) may follow **Poisson (√N) statistics → Gaussian**. The statistics depend on more than the physical and instrumental model. **How were the data selected** for fitting in the first place? Picking out the **brightest** spot in a spectrum means that we have a **special** set of data. The **peak pixel**, in this case, will follow the distribution appropriate to the maximum value of a set of, say, Gaussian variables. **Adjacent pixels** will follow a less well-defined distribution.

Simulation – "Model sources" are strewn, and the reduction software is given the job of telling us **what fraction is detected**. These essential large-scale techniques are very necessary for handling the detail of how the observation was made.

Example: radio astronomy synthesis images: noise level at any point depends on:
- gains of all antennas,
- noise of each receiver,
- sidelobes from whatever sources happen to be in the field of view,
- map size, tapering parameters, ionosphere, cloud…..**input data not known!**

Detection – the simpler, classical approach

What do we really want from the survey?

Are we more concerned with **detecting as much as possible (completeness)** ?

Are we more worried about **false detections (reliability)** ?

What are we going to do with the detections ?

- publish complete set of posterior probabilities of observed params everywhere?
- or just the covariance matrix, as an approx?
- or marginalized signal-to-noise, integrating away nuisance params?

From the classical point of view, if we are trying to measure a parameter ***α***, the **likelihood** sums up what we have achieved: $\mathcal{L} = \mathrm{prob}(\mathrm{data} \mid \alpha).$

Suppose that ***α*** is a ***flux density*** and we wish to set a **flux limit for a survey**. We only catalogue detections when our data exceed this limit s_{lim}. Then two properties of the survey are useful to know:

1. The false-alarm rate is the chance that pure noise will produce data above the flux limit:

$$\mathcal{F}(\mathrm{data}, s_{\mathrm{lim}}) = \mathrm{prob}(\mathrm{data} > s_{\mathrm{lim}} \mid \alpha = 0).$$

   The *reliability* is $1 - \mathcal{F}$, *i.e.* $\mathcal{F} = 5/100$ gives 95 per cent reliability. The may sound good, but note that it is the infamous 2σ result.

2. The *completeness* is the chance that a measurement of a real source will be above the flux limit:

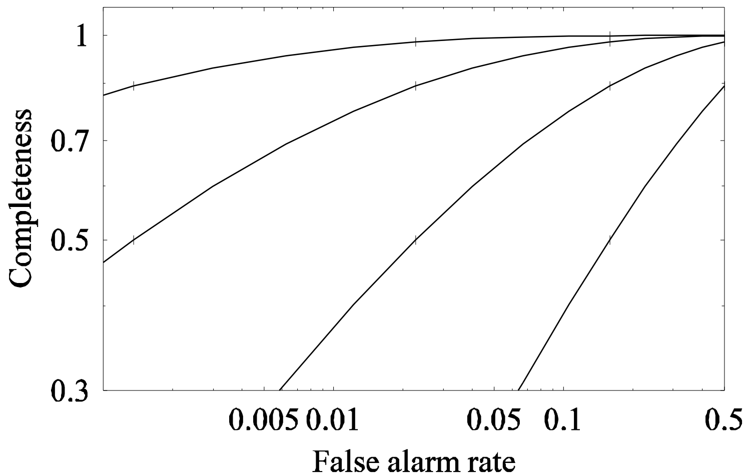
$$\mathcal{C}(\mathrm{data}, s_{\mathrm{lim}}, S) = \mathrm{prob}(\mathrm{data} > s_{\mathrm{lim}} \mid \alpha = S).$$

Classical detection - Example

Suppose our measurement is of ***s*** and the noise on the measurement is Gaussian, of unit std dev. The source has a "true" flux density s_0, measured in units of the std dev. We then have repectively for the **pd of the data given the source**, and for the **pd of the data when there is no source:**

$$\mathrm{prob}(s \mid s_0) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(s-s_0)^2}{2}\right]$$

$$\mathrm{prob}(s \mid s_0 = 0) = \frac{1}{\sqrt{2\pi}} \exp\frac{-s^2}{2}$$



Integrating these functions from ***0*** to ***s*** makes it easy to plot up the **completeness against the false-alarm rate**, taking the flux limit as a parameter.

For source flux densities in units of σ_{noise} ranging from 1 unit (lower right) to 4 units (upper left). Flux limits are indicated by dots, from 0 on right to 3 on far left. A 4σ source and a 2σ flux limit give a false-alarm rate of 2% and a completeness of 99%.

High completeness does indeed go hand in hand with a high false alarm rate. However there are quite **satisfactory combinations** for flux limits and source intensities of a few standard deviations.

In real life? The problem - outliers not described by Gaussians.

Detection – the Bayesian way

Conditional probabilities! => the Rev Thomas Bayes is calling…..
We have

prob(data | a source is present, flux density *s*) **(1)**

and **prob(data | no source is present).** **(2)**

(1) Take the prior probability that a source, intensity ***s***, is present in the measured area to be ***εN(s)***, where ***N(s)*** is a normalized distribution, the probability that a single source will have a flux density ***s***.

(2) The prior probability of no source is ***(1- ε) δ(s)***; ***δ*** is a Dirac delta function.

Then the posterior probability

prob(a source is present, brightness s | data)

is given by

$$\frac{\epsilon \, \mathrm{prob}(\mathrm{data} \mid s) N(s)}{\epsilon \int \mathrm{prob}(\mathrm{data} \mid s) N(s) \, ds + (1-\epsilon) \int \mathrm{prob}(\mathrm{data} \mid s=0)}.$$

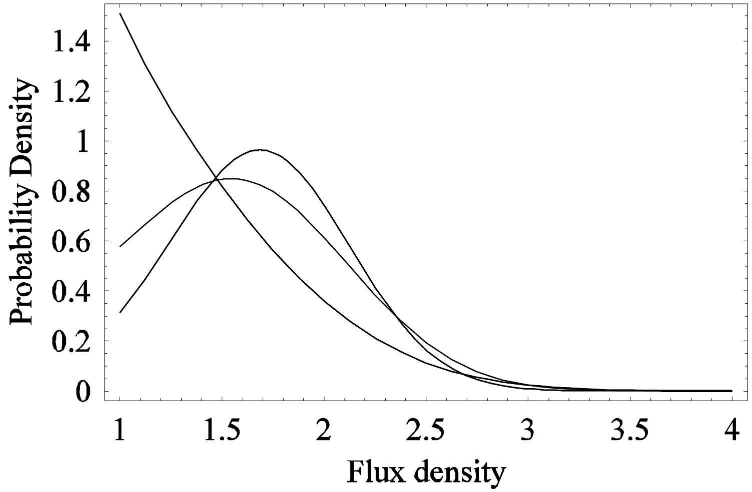
Integrating this expression over ***s*** gives the probability that a source is present, for given data.

Bayesian detection – Example 1

1. Radio telescope, randomly-selected position in the sky.
2. Model of the data ***D***, the single measured flux density ***f*** : Gaussian distribution about true flux density ***S*** with a variance σ^2.
3. The extensive body of radio source counts also tells us the **prior for *S***; approximate this by the power law ***prob(S) = K*** $S^{-5/2}$. (***K*** normalizes the counts to unity; there is presumed to be one source in the beam at some level.)
4. Probability of observing ***f*** when the true value is ***S*** : $\exp[-\frac{1}{2\sigma^2}(f-S)^2]$.
5. From Bayes : with ***n*** independent flux measurements f_i then

$$\text{prob}(S \mid D) = K'' \exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(f_i - S)^2]S^{-5/2}.$$

Suppose that the source counts were known to extend from ***1*** to ***100*** units, the noise level was ***σ = 1***, and the data were ***2, 1.3, 3, 1.5, 2*** and ***1.8***. The Fig. shows posterior probabilities for the first 2, then 4, then 6 measurements. The increase in data gradually overwhelms the prior but the prior affects conclusions markedly (as it should) when there are few measurements.



Power-law prior, Gaussian error dist. Posterior probability distribution for 2, 4, 6 measurements shows that the form approaches Gaussian as numbers increase.

Detection - summary

1. **Bayesian treatment of detection gives a direct result**; we may read off a suitable flux limit that will give the desired probability of detection.

2. But – the **confusion, and confusion limit** - images or spectral lines **crowd together**, **overlap** as we reach fainter. **Several** different objects may contribute to to the total flux at any coordinate. Even if only one 'object' is present, with a steep ***N(s)*** it will be more likely that the flux results from a **faint source plus a large upward noise excursion**, rather than vice-versa.

3. Then we only expect to measure population properties -- parameters of the flux-density distribution ***N(s)*** or its spectral equivalent. We are getting in to the **confusion regime**, a concept we'll consider later. Beware of the return of **hyperparameters.**

4. **Detection is a modelling process**:
   - it depends on what we are looking for,
   - how the answer is expressed depends on what we want to do with it next.

5. The **simple idea of a detection**, making a measurement of something that is really there, **applies when signal-to-noise is high** and individual objects can be isolated from the general signal. At low s/n, measurements can constrain population properties, with the notion of "detection" disappearing, in two senses: we drop into the **confusion level**, and/or we deal with **censored data**.

Catalogues, selection effects

Typically, a body of astronomical detections is published in a **catalogue**. Objects are in this catalogue – or not - on the basis of **some clear criteria**. If not, upper limit is info.

Most astronomical measurements are **affected by the distance to the object**.
- **proper motion**, fixed velocity, becomes a smaller angle **inversely** as the distance.
- **Apparent intensity** (flux density) drops off as the **square** of the distance.
- **ellipticity** of a galaxy becomes harder to detect (seeing blur).
- so-called **k-corrections** can make objects relatively brighter or fainter.

We measure an **apparent** quantity X and infer an **intrinsic** quantity by a relationship $Y=f(X,R)$ where R is the **distance** to the object in question. The function f may be complicated, for reasons of both **observation and relativistic geometry.**

Take a simple case (the principles carry over). Observe a flux density S and infer a luminosity L given by $L=S\,R^2$. The smallest value of S we are prepared to believe is S_{lim}; if a measurement is below this limit, the object is not in our catalogue or sample.

Our objects (="galaxies") are assumed to be drawn from a **luminosity function** $\rho(l)$, **the number of objects within** Δl **about** l **per unit volume**. Using only our catalogue set of measurements $L_1, L_2, \ldots$ however, we will **not** be able to reproduce ρ. Instead, we will get the luminosity distribution η, where

$$\eta(l) \propto \rho(l) V(l).$$

....and Malmquist Bias + Example

Crucially, ***V(l)*** is the volume within which sources of intrinsic brightness ***l*** will be near enough to find their way into our catalogue. We get

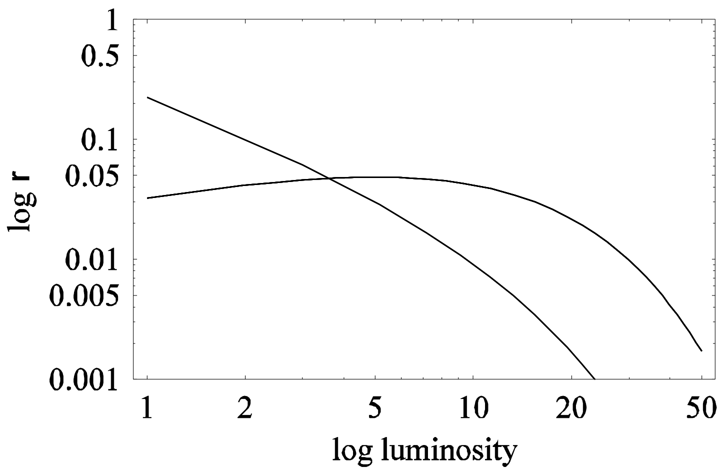
$$\eta(l) \propto \rho(l)(\frac{l}{S_{lim}})^{3/2}.$$

Obviously ***η*** will be biased to higher values of luminosity than ***ρ***. This sort of bias occurs in a multitude of cases in astronomy, and is often called **Malmquist bias**.

Example: The luminosity function of field galaxies is well approximated by the Schechter function

$$\rho(l) \propto \left(\frac{l}{l_*}\right)^{\gamma} \exp(-\frac{l}{l_*}),$$

in which we take ***γ = 1*** and ***l*~*~ = 10** for illustration. To obtain the form of the luminosity distribution in a flux-limited survey, we multiply the Schechter function by $l^{3/2}$. The differences between the luminosity function and luminosity distribution are shown in the Fig.



The luminosity function ρ (steep curve) and the (flat-space) luminosity distribution, plotted for the Schechter form of the luminosity function.

Malmquist bias – a serious issue in survey astronomy. The bias depends on the (probably unknown) **form of the luminosity function**. It passes on all kinds of unwelcome information.

Malmquist Bias and the luminosity-distance correlation

Malmquist bias arises because intrinsically bright objects can be seen within proportionately much greater volumes than small ones. Because most of the volume of a sphere is at its periphery, it follows that in a flux-limited sample the bright objects will tend to be further away than the faint ones - **there is an in-built distance-luminosity correlation.**

The **luminosity - distance correlation is widespread, insidious and very difficult to unravel**. It means that **for flux-limited samples, intrinsic properties correlate with distance**:

- **two unrelated intrinsic properties will appear to correlate** because of their mutual correlation with distance. Plotting intrinsic properties -- say, X-ray and radio luminosity -- against each other will be very misleading.
- much further analysis is necessary to establish statistical dependence. Such analyses may require **detailed modelling of the detection process.**
- consider measuring the ellipticity of galaxies - distant ones may well look rounder because of the effects of seeing. More distant galaxies appear more luminous thanks to Malmquist. We are on course for deducing that round galaxies are more luminous or vice versa. A
- detailed model will be necessary to establish the relationship between true ellipticity, measured ellipticity, and the size of the galaxy relative to the seeing disc.

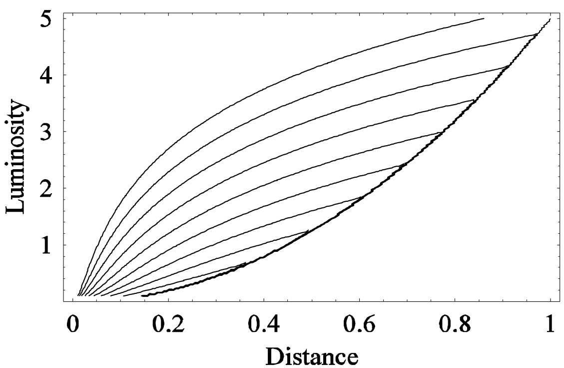
The luminosity – distance correlation - Example

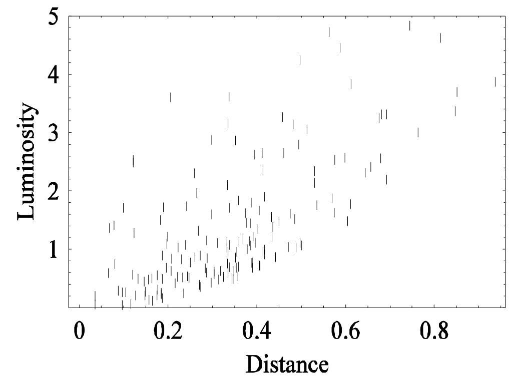
Adopt a Schechter function with $\gamma = 1$ and $l_* = 10$ for the purposes of illustration.

The **probability** of a galaxy being at distance ***R*** is proportional to R^2, in flat space.
The **probability** of it being of brightness ***l*** is proportional to the **Schechter function**.
The **probability** of of a galaxy of luminosity ***L*** at distance ***R*** being in our sample is

$$\begin{aligned} \text{prob(in sample)} &= 1 \quad L < S_{lim} R^2 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

The product of these three probability terms is **the bivariate distribution *prob(l,r)***, the probability of a galaxy of brightness ***l*** and distance ***r*** being in our sample.





Left: Contour plots of the bivariate prob(l,r). The contours are at logarithmic intervals; galaxies tend to bunch up against the selection line, leading to a bogus correlation between luminosity and distance. Right: Results of a simulation of a flux-limited survey of galaxies drawn from a Schechter function.

There is a clear **correlation between distance and luminosity**. Look familiar? +++>
A direct check of this is to **simulate a large spherical region** filled with galaxies whose **luminosities are drawn from a Schechter function**, and select a flux-limited sample.

Example – remember 'The fishing trip'?

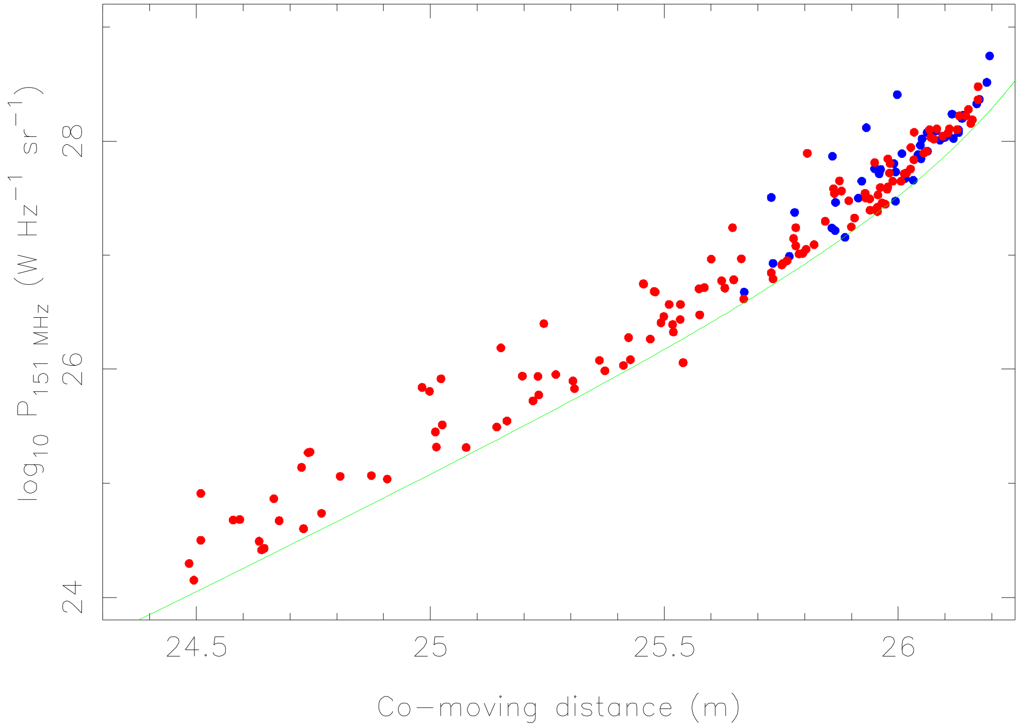
Suppose that we have plotted something against something, on a Fishing Expedition.

3. Does the eye see much correlation? If not, formal testing for correlation is probably a waste of time.

2. Could the apparent correlation be due to selection effects? Consider for instance the beautiful correlation obtained by Sandage (1972): 3CR radio luminosities vs distance.

Nb (1) Sandage recognized the problem.

Nb (2) Hasn't stopped anybody plotting luminosity vs redshift and 'proving' that quasars and radio galaxies are the same because they fit the line…so do baseballs.



Radio luminosities of 3CR radio sources versus distance modulus.

Survey bias – last warning

Note an effect that competes with Malmquist bias due to observational error. The number of objects as a function of apparent intensity ***N(s)***, the **number counts or source counts**, usually rises steeply to smaller values ***s*** - there are many more faint objects than bright ones.

In compiling a catalogue we in effect **draw samples from the number count distribution**, forget those below S_{lim}, and convert the retained fluxes to luminosities. The effect of observational error is to convolve the number counts with the noise distribution. Because of the steep rise in the number counts at the faint end the effect is to **contaminate the final sample with an excess of faint objects**. (An object of observed apparent flux density is much more likely to be a faint source with a positive noise excursion than a bright source with a negative excursion. Observe with zero noise!) This can severely bias the deduced luminosity function towards less luminous objects.

This effect **does not occur if the observational error is a constant fraction of the flux density**, and the source counts are close to a power law.

Many if not most types of astronomical observations suffer from the **range of problems due to Malmquist bias, parameter-distance correlation and source-count bias.** Think that yours does not? **I am prepared to take a bet.**

The luminosity function and V_{max}

- assume a catalogue of objects, high reliability and well-defined limits.
- examination of an intrinsic variable l (e.g. luminosity) requires $\rho(l)$.
- measure L_i for all of the objects in some (large) volume? How?

One of the best methods is the $1/V_{max}$ estimator.

- $V_{max}(L_i)$ are the maximum volumes within which the i^{th} object could lie, and still be in the catalogue.
- V_{max} thus depends on the survey limits, distribution of the objects in space, and the way in which detectability depends on distance.
- simplest case: a uniform distribution in space is assumed. Given the $V_{max}(L_i)$, an estimate of the luminosity function is

$$\hat{\rho}(B_{j-1} < l \leq B_j) = \sum_{B_{j-1} < i \leq B_j} \frac{1}{V\mathrm{max}(L_i)}$$

in which its value is computed in bins of luminosity, bounded by the B_j .

- a maximum-likelihood estimator, and so has minimum variance for any estimate based on its statistical model.
- errors are uncorrelated from bin to bin and can easily be estimated – the fractional error in each bin is ~ $1/\sqrt{N_j}$, where N_j is the number of objects in each bin. Better error estimates - bootstrap.

The luminosity function and V_{max} cont…

- determination of V_{max} is the crunch; choosing the flux limit of a survey affects:
  (1) the number of sources that are missed,
  (2) the number of bogus ones included, and
  (3) the extent to which faint sources are over-represented.

=> MC simulations? Rough idea of the luminosity function enables this.

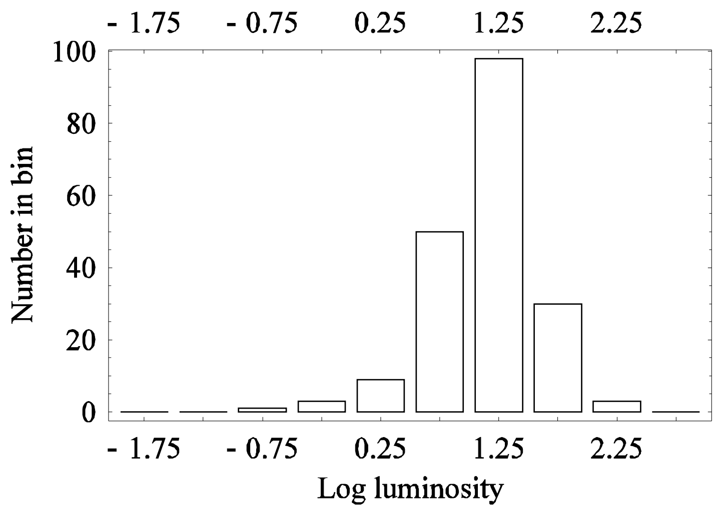
- with V the volume defined by the distance to the source as its radius, the distribution of V/V_{max} is very useful in estimating the actual limit of a survey.
- if the correct flux limit adopted then we expect V/V_{max} to be uniformly distributed between zero and one. This can be checked by, e.g. A K-S test, and such a process is a model-fitting procedure to estimate survey limit.

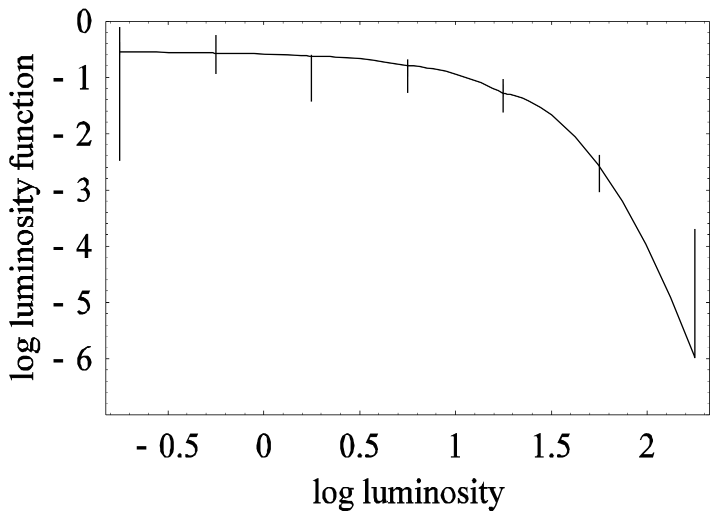
- this procedure fails horribly at large ($z > 0.2$) cosmological distances where cosmic evolution dominates.
- the derivation of cosmic evolution led Schmidt (1968) to derive the technique.

- there's a vast literature; summary: Willmer 1997.

$1/V_{max}$ luminosity function - Example

From the previous simulation with a Schechter lum fn, a flux limit of 20 units gives a sample of ~200 objects. Left Fig.: the **luminosity distribution** ***n(l)*** shows a strong peak at ~5 units, related to the characteristic luminosity ***l* = 10***. Bins are 0.5 dex wide. (Faint sources are greatly under-represented, because they are above the flux limit for only small distances.)





Right Fig: the result of applying the V_{max} method and **bootstrapping** to derive errors. The input lum fn is the solid line and the estimate from the ~200 sample via $1/V_{max}$ is shown as dots and error bars. Because V_{max} is so small for the faint sources, the few faint sources in the sample give a large contribution to $\underline{\rho}$ although the errors are correspondingly large. For simplicity the luminosity functions were normalized, so giving **luminosity probability distributions**.

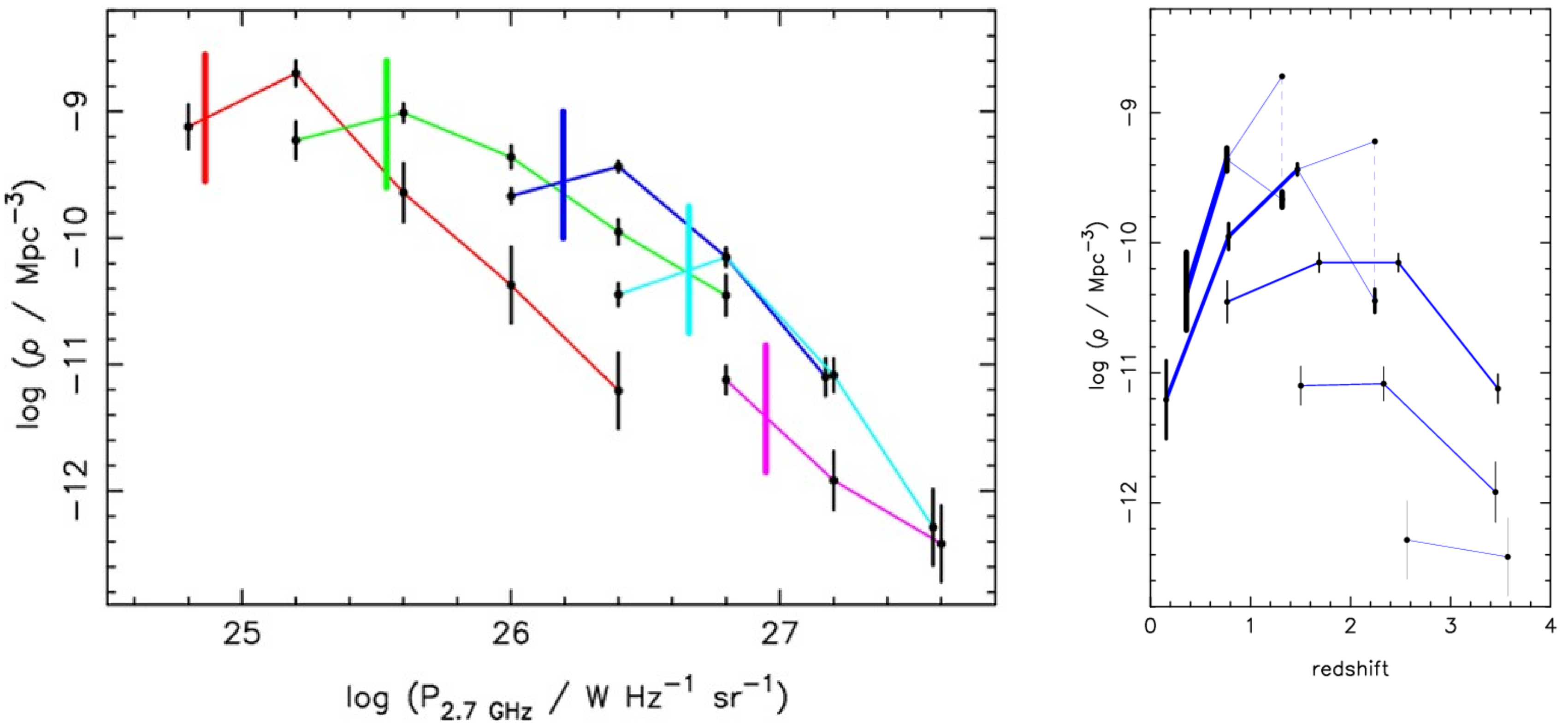
V_{max} – non-uniform distribution

- If uniform distribution is **not** a good assumption (the case for ~ all cosmological
- projects) then there are three ways to a better estimate.
-
- **(1) Bin the data** into narrow ranges of distance, and estimate the lum fn within each bin. But at large distances we will know nothing about low luminosities. The approach is further limited by the decreasing numbers of objects as the number of distance bins is increased.

(2) We can make progress on spatial dependence with stronger assumptions. Our data is **the set of pairs** (R_i, L_i); we want the bivariate distribution $\rho(r,l)$.
Usual assumption: **the distribution factorizes**, so that the **form of the luminosity function does not change with distance**, but the normalization can. This method will allow extrapolation of info from small distances and low luminosities. The **standard estimator of** ρ is the C^- **method** (Lynden-Bell 1971), important because **it is also a MLE**. If **parametric forms** are known for ρ and φ then normal modelling methods can be used.

(3) "Free-form" methods are intermediate; they fit quite general functions to the data populating the ***r - l plane.*** Examples of these are $\sum_{i,j} a_i (\log r)^i (\log l)^j$, where the cross terms break the factorizability.
If we use small number of terms in the expansion => extrapolation into areas of the ***r - l plane*** unpopulated by observations.

Vmax – non-uniform distribution - Example



A complete sample of ~ 400 radio quasars. Left: radio luminosity function in redshift ranges 0 – 0.5 (red), 0.5 – 1.0 green, 1.0 – 2.0 (blue), 2.0 – 3.0 (light blue) and 3.0 – 5.0 (disgusting pink). Vertical bars – lower limits in completeness for each redshift range. Right: space density as a function of redshift for 5 power ranges: log $P_{2.7}$ = 25.8-26.2, 26.2-26.6, 26.6-27.0, 27.0-27.4, 27.4-27.8, denoted by decreasing line weight.

Survey data - Censored and Confused

Peter A G Scheuer 1930 - 2001

Outstanding theoretical astrophysicist, mentor and friend, intensely modest and private man of very few words and very many brilliant ideas.

"In 1955 Ryle's group compiled the 2C catalogue of radio sources using interferometric techniques. The catalogue contained vastly more faint radio sources than would be expected in any of the standard cosmological models. In particular, these observations were in fatal conflict with the predictions of the steady state cosmological model of Bondi, Gold and Hoyle. These results were hotly contested, not only by the proponents of steady state theory, but also by the Australian radio astronomers at Sydney, led by Bernard Mills, who found far fewer faint sources and called into question Ryle's result. The Cambridge radio astronomers realised that the flux densities of the faintest sources were systematically overestimated because of the presence of faint sources within the beam of the radio interferometer. In a paper rightly considered a masterpiece, published in 1957 in the Proceedings of the Cambridge Philosophical Society, Scheuer showed how the correct form of the number counts could be found by studying the interferometer traces and found that there was indeed an excess of faint sources, but less than originally thought. This discovery was the first really convincing evidence for the evolution of the radio source population over cosmological time- scales." M S Longair obit, 2001.

Survival Analysis – censored data

A primary sample of objects - a series of measurements from which we pick out 'detections'. The results often find their way into catalogues, e.g. the New General Catalogue (NGC) or the 3CR Catalogue.

=> complementary data? Upper limit at every pixel? Not retained.

'Resurvey' is different, e.g. measure $H\alpha$ luminosity of NGC galaxies.
- **now very useful to quote upper limits; real objects are there.**
- **sometimes a resurvey may yield lower limits, e.g. measurement of X-ray and radio flux densities for the NGC galaxies would yield both upper limits and lower limits for the radio to X-ray spectral index.**

Statistics dealing with limits is called 'survival analysis', from medical statistics: at end of a study some subjects have survived, some not.
- **measurements which are only limits are called 'censored'.**
- **introduced into astronomy by Avni, Feigelson, co-workers 1980 >.**

Survival analysis offers (i) estimation of intrinsic distributions (like luminosity functions), (ii) modelling and parameter estimation, (iii) hypothesis testing and (iv) tests for correlation and statistical independence, for cases in which some of the available measurements are limits.

Survival Analysis – Censored Data

The key assumption is that the **censoring is random**; this means that the chance of only an upper limit being available for some property is independent of the true value of that property.

Assumption often met for flux-limited samples. For an object of true luminosity ***L*** and distance ***R***, the condition for censoring is that $\frac{L}{R^2} < S_{\rm lim}$

the flux limit for the survey. If ***R*** is a random variable, independent of ***L***, and S_{lim} is fixed, then the chance of censoring is independent of ***L***.

Careful examination of how a sample was selected is necessary to determine that survival analysis is applicable.

Let's be definite. Survey at **wavelength A**; resurvey the sample at **wavelength B.** So we get some **values of** L_B and some **upper limits** L^U_B.
Our aim – **to construct the normalized distribution** of L_B , which will be $\alpha\ \rho_B$

Two estimators available, a **recursive relation due to Avni** and the **Kaplan-Meier estimator of the cumulative distribution**. Both are **maximum likelihood**. Both work for upper and lower limits. The former requires binning; the latter does not rely on binning, but being cumulative, **errors are highly correlated** from one point on the estimate to the next.

Survival Analysis – Censored Data

Kaplan-Meier estimator:

$$\hat{\mathcal{K}}(L_k) = 1 - \prod_{i=1}^{k-1} (1 - d_i/n_i)^{\delta_i}$$

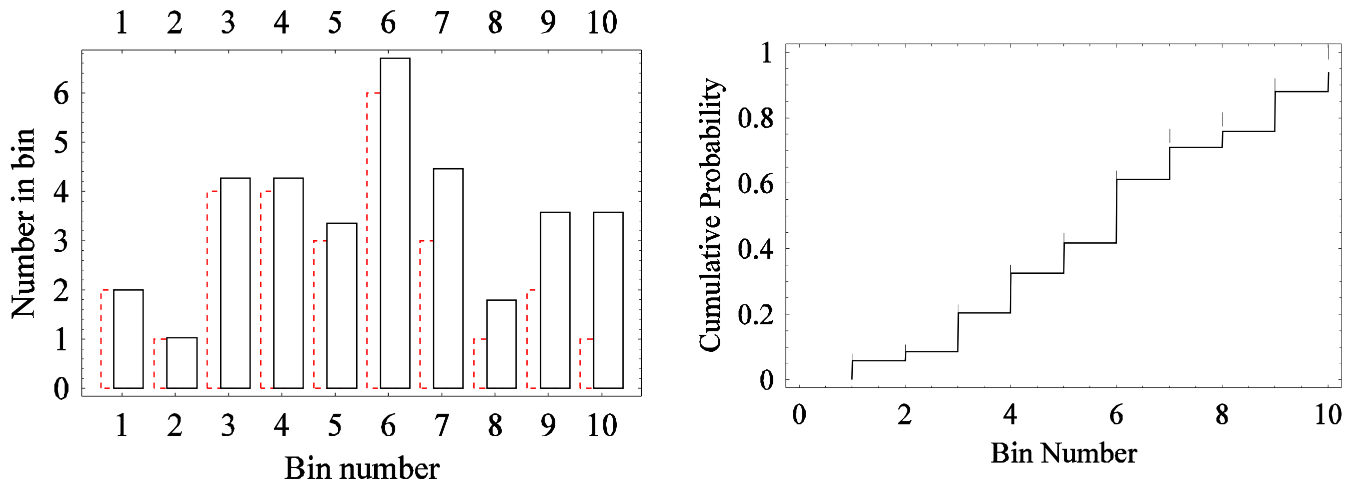
Lower limits: arrange data in **increasing** order.

- d_i is the number of observations of L_i
- n_i is the number of observations $=/> L_i$
- $\delta_i = 1$ for detection, 0 for lower limit

Upper limits: arrange data in **decreasing** order

- n_i is the number of observations $</= L_i$
- $\delta_i = 1$ for detection, 0 for upper limit

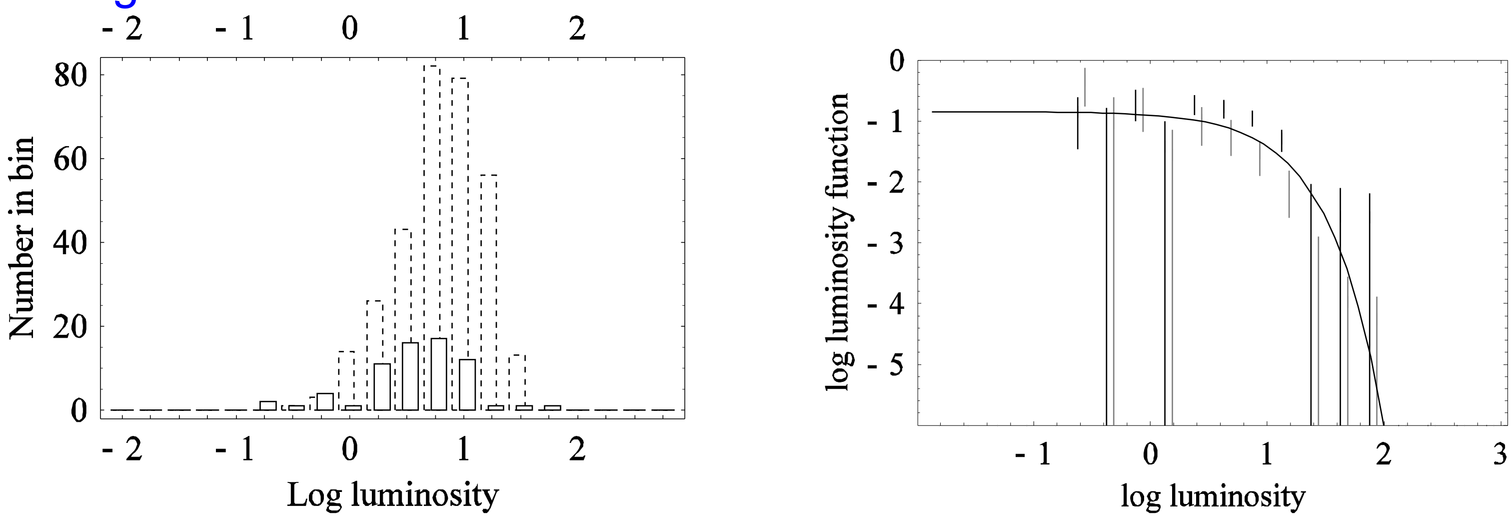
Example 1 :



Left: Avni estimator, binned data. Distribution of spectral indices (optical to X-ray) for a sample of optically-selected QSOs; observed distribution (red dashed), and true distribution after including lower limits. Right: same data, shown cumulatively as dots. Kaplan-Meier estimator is the solid line, following the bins used in the Avni formulation.

Survival Analysis- Censored Data

Example 2 : back to the simulated field-galaxy sample. This was selected at one wavelength. We 'resurvey' at another, and get 67 detections and 317 upper limits, in a simulation allocating luminosites from independent Schecter functions at both wavelengths.



Left: luminosity distribution and upper limits for the field galaxy simulation; there are 67 detections and 317 upper limits (dashed bins). Right: luminosity probability distribution (black dots) from the Avni estimator, together with bootstrap error estimates. Solid line – theoretical distribution. Lighter dots are a $1/V_{max}$ estimate, displaced slightly in luminosity for clarity.

Survival analysis and 1/Vmax results agree – both are MLEs and are based on similar models. The survival analysis gives better estimates here of the luminosity function, using more data – but the real advantage comes in correlation analyses, or reconstructions of non-distance-dependent distributions (e.g. spectral indices).

The confusion limit

In many cases of astronomical interest, we find that faint objects are much more numerous than bright ones.

Faint objects crowd together; ultimately they start to be unresolved and our signal becomes a mixture of objects of various intensities, **blended together by the point spread function of our instrument.**

Examples include radio sources in deep surveys, spectral lines in the Lyman-α forest, stars at the cores of globular clusters, and faint galaxies observed in the optical.

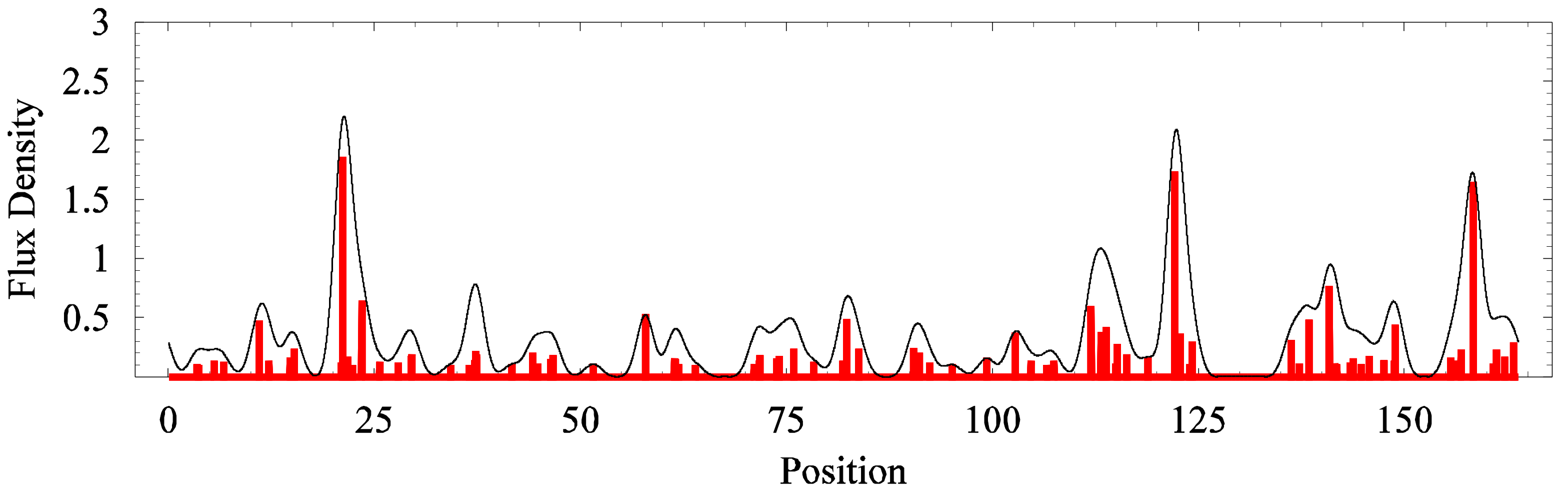
The **confusion limit** concept was developed during a bitter controversy amongst radio astronomers and cosmologists in the 1950s, the **source-count/Big Bang/Steady State** controversy.

The root of the problem was **instrumental**, wildly different source counts being obtained at Sydney (Mills Cross; essentially filled aperture) and Cambridge (interferometer).

Scheuer (1957) analyzed the statistics of the source counts and showed that the Cambridge results were seriously affected by **confusion**. The wide beam of the interferometer meant that **many radio sources were contributing to each peak** in the record; these had erroneously been interpreted as discrete sources.

The confusion limit - Example

To show the pronounced effect of confusion, here is a simulation of a 1D scan of sources obeying a Euclidean source count ***N(>S) α S***$^{-3/2}$. The beam is a simple Gaussian and there is, on average, one source per beam.



Confusion simulation at a level of one source per beam area. Input sources – red verticals; solid line is the response when convolved with a Gaussian beam.

A simple count of the peaks in the record gives a maximum-likelihood slope for the source count of -1.8 with standard deviation 0.3, very different from the true value of -1.5. In the case of an interferometer, the presence of sidelobes biases the faint counts to much steeper than true values; the apparent cosmological evolution this implied was the subject of the original controversy.

The confusion limit – the p(D) technique

The technique developed by Scheuer is known to astronomers as **"p(D)"** , or **"probability of Deflection"** - deflections of the needle of a chart recorder.

Technique has been used in the radio, the X-ray, infrared, and Lyman-α. The method **derives the probability distribution of measurements in terms of the underlying source count**, which may be recovered by a model-fitting process. Its benefit is that information is obtained from sources which are much too faint to be "detected" as individuals.

The derivation of the distribution requires conditional probabilities, Poisson statistics, characteristic functions, Fourier transforms. It is a beautiful bit of math by Scheuer.

The result: the FT of the p(D) distribution is

$$P(\omega) = \exp(R(\omega) - R(0)) \;\; \text{in which} \;\; r(s) = \int N\left(\frac{s}{\Omega(x)}\right) \frac{dx}{\Omega(x)}$$

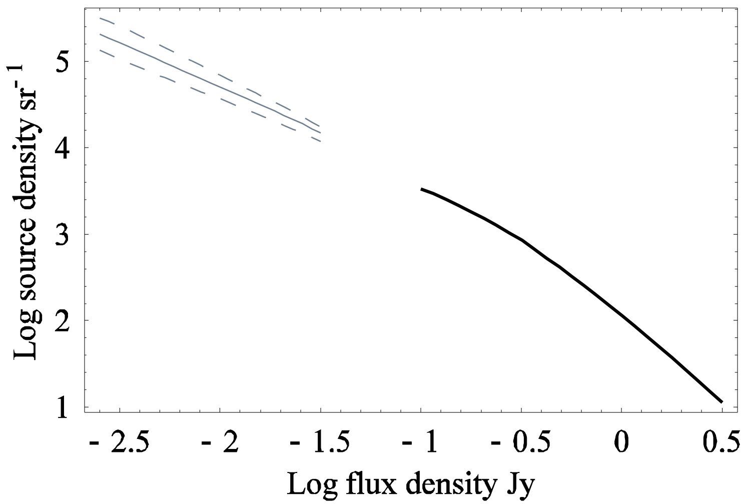
contains the source count *N*. *R* is the FT of *r*. Analytic solutions are available when ***N(S)*** is a power law, but the inverse transform to get ***p(D)*** has to be done numerically.

In real life we need to take account of **differential measurement** techniques in which measurements from two positions are subtracted to avoid baseline errors.
And there's always **noise**. A modelling process is needed to recover N(S).

The confusion limit and p(D) – Example 1

The derivation of source counts from p(D) is another technique in which population characteristics are derived from observations of discrete objects or features without forming an object list or catalogue.

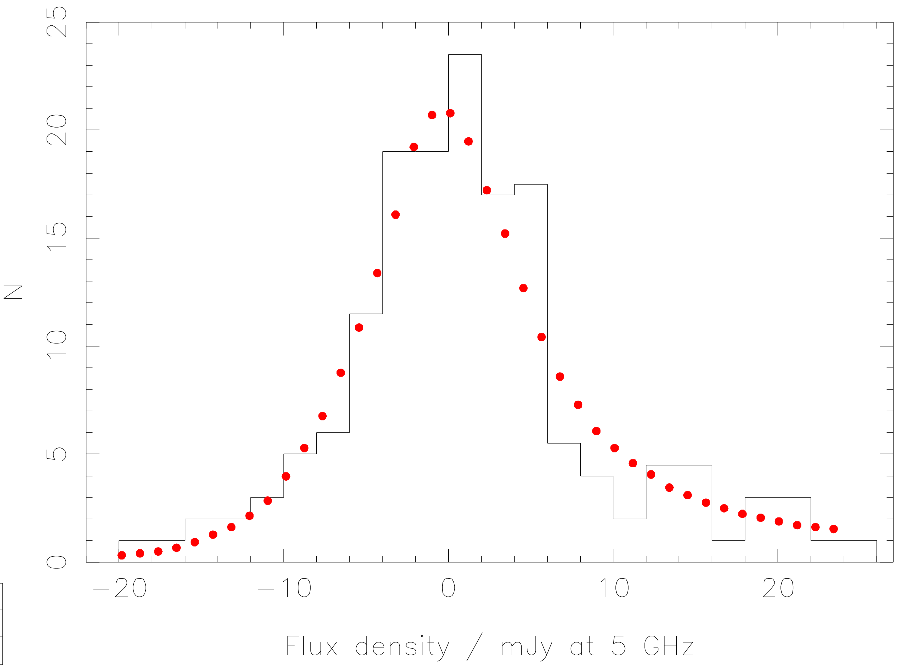
Wall & Cooke (1975) applied the p(D) technique for filled aperture telescopes to extend the 2.7-GHz radio source counts to much fainter levels than could be achieved by identifying individual sources:



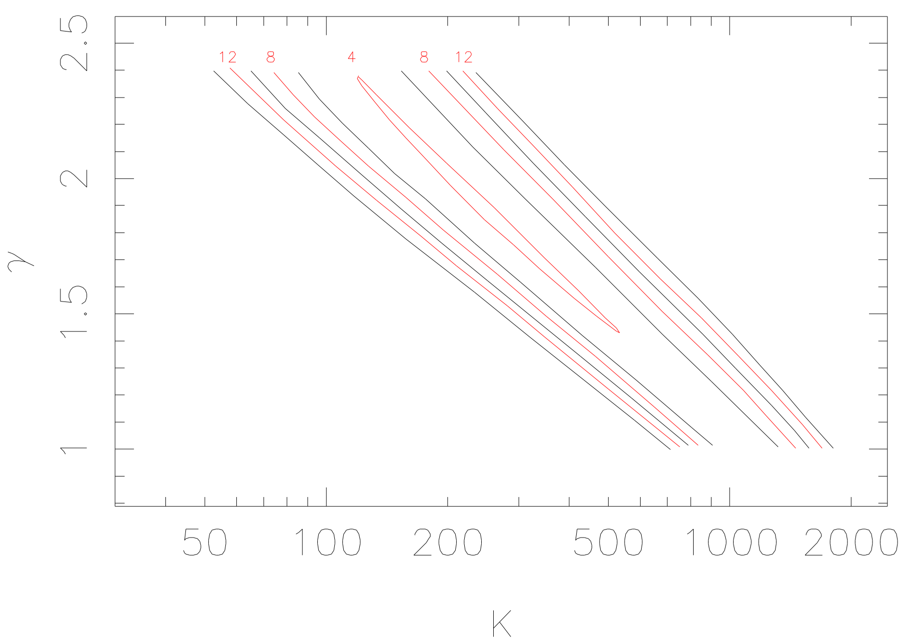
The 2.7-GHz counts from Wall & Cooke (1975): the darker line is derived from ordinary source counts with error bars not much wider than the line, while the p(D) results are shown in grey, the dashed lines representing one standard deviation of the fitted parameters.

(Minimum chi-square method) / p(D) – Example 2

Chi-square testing/modelling: the object of the experiment was to estimate the surface-density count (the ***N(S)*** relation) of faint radio sources at 5 GHz, assuming a power-law $N(>S) = KS^{-(\gamma-1)}$, γ and K to be determined from the distribution of background deflections, the **p(D) method**. The histogram of measured deflections is shown right.



The dotted red curve above represents the optimum model from minimizing χ^2. Contours of χ^2 in the **γ - K plane** are shown left.



With the best-fit model, $\chi^2 = 4$ for 7 bins, 2 parameters; thus dof = 4. **Right on.**

Final challenge:

Use $1/V_{max}$ estimator of the luminosity function to recover each of the two power-law luminosity functions from the 'surveys' of your universe of 1000000 objects.

End JVW Bertinoro 5