7.7 Luminosity-distance correlation

The data file contains two luminosities and a distance for each "galaxy". We will suppose that the set of galaxies has been chosen by selection on some other luminosity, uncorrelated with the other two. Of course in real life this is an assumption we would want to check somehow.

We now want to set up the case where there are two flux limits in operation. For each "galaxy" we have a triple (luminosity 1, luminosity 2, distance). For definiteness, let us call the first luminosity in the triple the radio luminosity, and the second the x-ray luminosity.

Our first task in the simulation is to select on the first luminosity, marking objects as detected if $L_R/R^2 \geq 5$ and otherwise ascribing a limit $5R^2$. We then select on the second luminosity similarly, also using $L_X/R^2 \geq 5$; the resulting sample, flux-limited on two quantities, has 87 detections. If we make the usual mistake of plotting luminosity against luminosity for the detected galaxies, we get Figure 1.

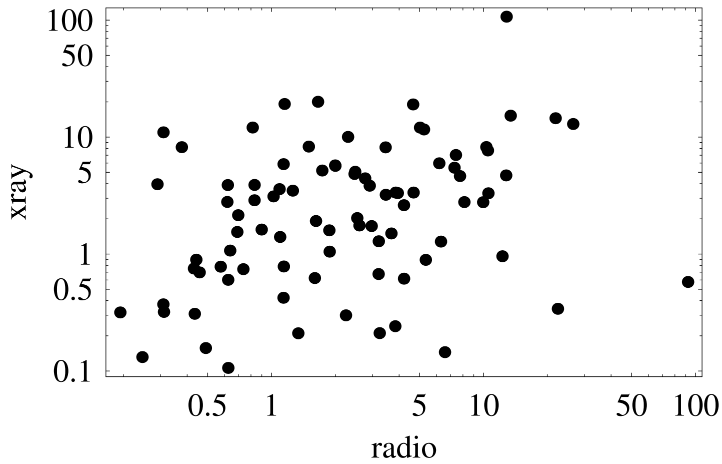

Figure 1: A luminosity - luminosity plot for the jointly flux-limited sample.

The correlation looks convincing, but let us now put the upper limits on the plot.

Now we work out the Kendall statistic. To do this, it's convenient to set up a data structure for each galaxy which contains the luminosity (if detected), the luminosity limit (flux times distance squared) if not, and a flag indicating whether there is a detection or not. The heart of the calculation is working out the a and b matrices. The pseudo-code for this is in the book:

`create a square matrix` a `of size` $n+m \times n+m$
`initialize it to zero`
`for each` X_i
`if` $X_j > X_i$ `and` X_j `is detected, set` $a_{ij} = 1$
`if` $X_j < X_i$ `and` X_i `is detected, set` $a_{ij} = -1$

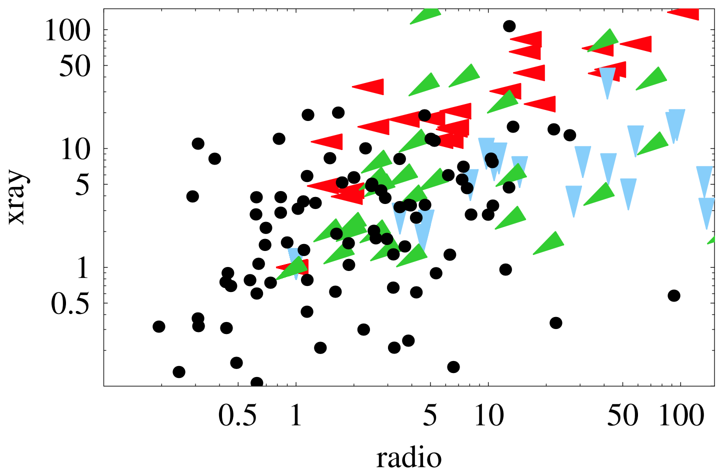

Figure 2: The same plot, but now showing the cases where there are only limits for either or both of the luminosities.

Applying the formulae for the Kendall statistic, we then find it has a value of 2612, with a standard deviation of 19023. This is a one-tenth sigma result, and shows the power of using the upper limits to rule out bogus correlations. It also shows how important it is to be sure that the censoring of this type of data is indeed random, as described in the book.

Now let us see what happens when there is a real correlation. A convenient toy model of correlated luminosities is provided by a bivariate Gaussian, of mean zero, variance 100 units and correlation coefficient 0.7. We chose 500 of these luminosity pairs and deleted any that contained a negative luminosity. The distances were obtained in the usual way for these examples, by taking the cube root of a random number uniformly distributed between zero and two. This gives around 400 "galaxies" and flux limits of 5 units (in both variables) yields a sample consisting of around 100 detections. The correlation between detections is comparable to the previous case, as seen in Figure 2. However the Kendall test now consistently gives a two-sigma result.

The take-home message of this example is that two similar diagrams yield very different significances of correlation because the distribution of upper limits is different.

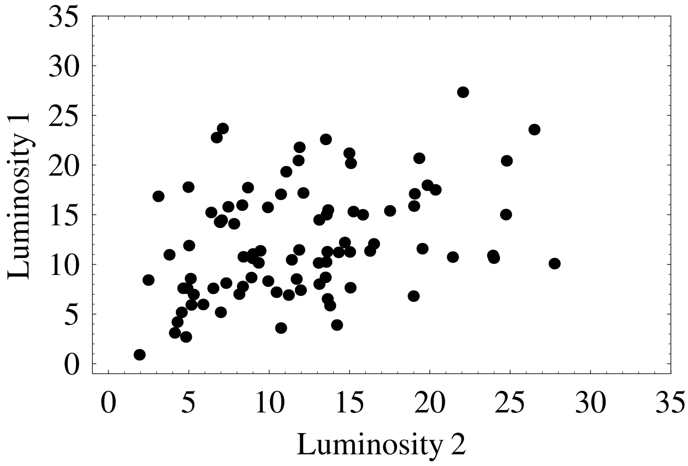

Figure 3: An example of the correlation between detections for the intrinsically-correlated data.

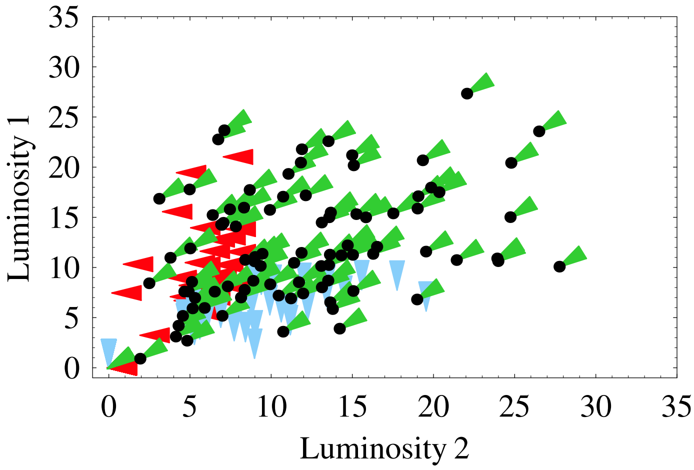

Figure 4: The same data but also plotting the upper limits. Notice that they are differently distributed to Figure 2.