10.5 Surface density changes and Counts-in-Cells

(a) Magnitude of offset in $w(\theta)$. You will need to start with the uniformly random data set over the area $0° < \alpha < 60°, -20° < \delta < 20°$ specifically provided for this example.

The example suggests going to circles as small as 0.03°in radius; a rough calculation shows that for a surface density of $100000/(40 \times 60) \approx 40$ per square degree, the average number is 0.03 points, an unrewarding number on which to check Poisson statistics. Stick with the range 0.3° to 3° radii, with mean numbers thus running from 3 to 300 per cell.

Setting up the grid of circles can consist simply of the grid of circle *centres*; each point in the sky distribution can be tested to see whether it falls within the given cell radius of each centre. There are efficient and other ways of doing this.

You will find the distributions of cell numbers for a given cell size to be indistinguishable from Poisson distributions, as they should be. Figure 1 shows the distributions for 10 logarithmically-spaced cell sizes, diameters 0.3° to 3°, on a log-log plot. The Poisson distributions computed from the mean number of points per cell-size are shown superposed.

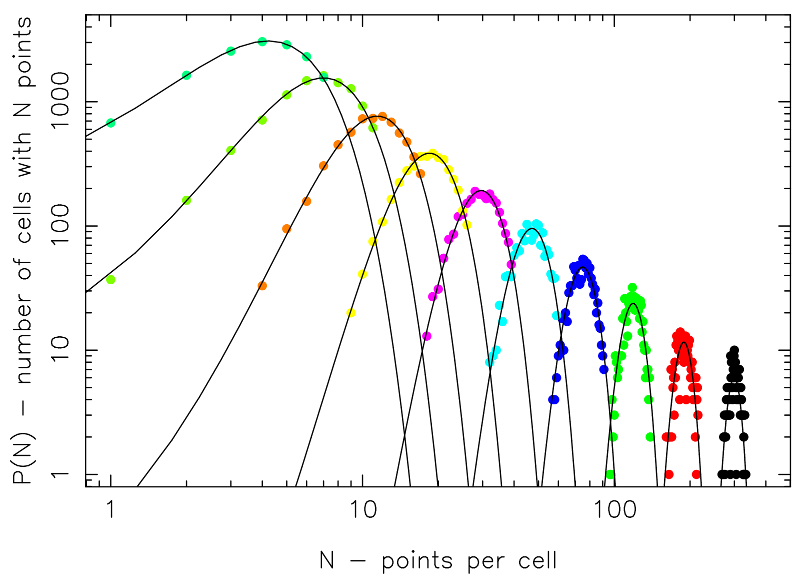

Figure 1: Number of cells versus number of objects per cell, the counts-in-cells for the uniform toy sky with 100000 points provided in the data set for this example. There are 10 cell sets, with cell diameters evenly distributed logarithmically from 0.3° (the left-most distribution) to 3° (far right). Cells are independent. The curves represent Poisson distributions with means given by the average number of points per cell in each cell set.

(As an aside, note the difficulty of a straight-out computation of the Poisson distribution for observed number k and mean number μ:

$$f(k, \mu) = \frac{e^{-\mu} \mu^k}{k!} \tag{1}$$

For values of $\mu > 10$ this innocuous-looking sum soon runs any computer out of significant digits. You need recourse to an identity: that of the cumulative Poisson distribution with the second incomplete Gamma function Q, the complement of the first incomplete Gamma function P:

$$F_\mu(< k) = Q(k,\mu) = 1 - P(k,\mu),\ \text{where} P(k,\mu) = \frac{1}{\Gamma(k)} \int_\mu^\infty e^{-t} t^{(k-1)} dt \tag{2}$$

and the (complete) Gamma function is

$$\Gamma(k) = \int_0^\infty t^{(k-1)} e^{-t} dt \tag{3}$$

See Numerical Recipes; routines are available for incomplete Gamma functions in Numerical Recipes and elsewhere.)

The variance statistic is

$$y(L) = \frac{\mu_2(L) - \overline{N}(L)}{\overline{N}^2} \tag{4}$$

with the second moment $\mu_2 = \overline{(N - \overline{N})^2}$. When this statistic is computed and plotted for the random-sky data set of this example, the result is shown by the crosses in Figure 2.

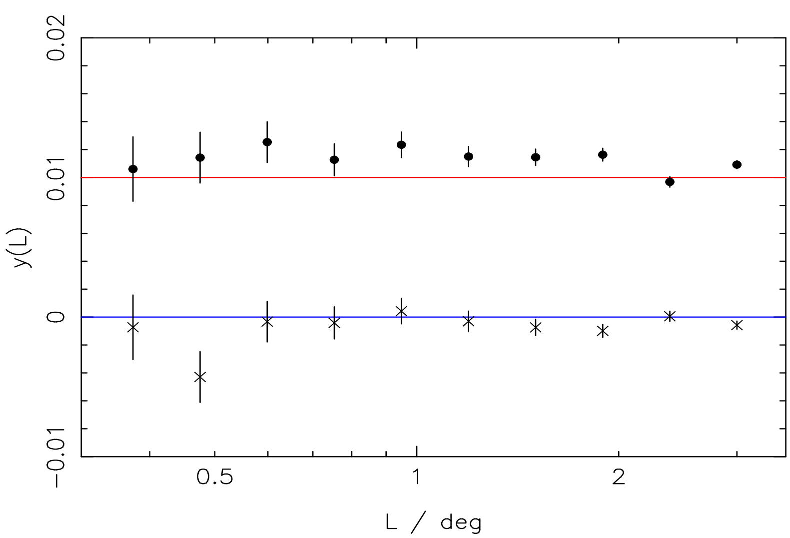

Figure 2: The variance statistic $y(L)$ as a function of cell diameter L, for a 10 cell sets with diameters logarithmically spaced from 0.3° to 3°. The crosses are $y(L)$ determinations for the toy random sky data set of this example. The dots are for the toy sky of example 10.4, in which there is a step of 20 per cent in surface density between the northern and southern halves of the field.

The errors on $y(L)$ are given by

$$\sigma_y = \sqrt{\frac{2}{N_c (\overline{N})^2}} \tag{5}$$

where N_c is the number of independent cells. (Verify this from standard error analysis - see section 3.3.) As should be the case for a sky of uniformly-distributed independent random points, no significant offset in the set of values of $y(L)$ is evident in Figure 2.

(b) Carrying out the same exercise for the sky data-set of example 10.4 gives the points in Figure 2 designated by solid dots. Now there is an offset consistent with the prediction of $\Delta y(L) = 0.01$.

From the text: 'Systematic surface density gradients spuriously offset the counts-in-cells variance: a spread in the mean surface density across the cells will inevitably broaden the overall probability distribution $P(N)$, which is constructed from fluctuations about those means. For a cell of area S at local surface density ς, $< N >= \varsigma S$ and $< N^2 >= \varsigma S + \varsigma^2 S^2$ for no clustering; averaging over many cells produces $< \overline{N} >= \overline{\varsigma} S$ and $< \overline{N^2} >= \overline{\varsigma} S + \overline{\varsigma^2} S^2$. It follows that the variance statistic y (equation 4) is offset by

$$< \Delta y >= \frac{\overline{\varsigma^2}}{(\overline{\varsigma})^2} - 1, \qquad (6)$$

precisely the same offset as that experienced by $w(\theta)$ in the presence of surface gradients.'

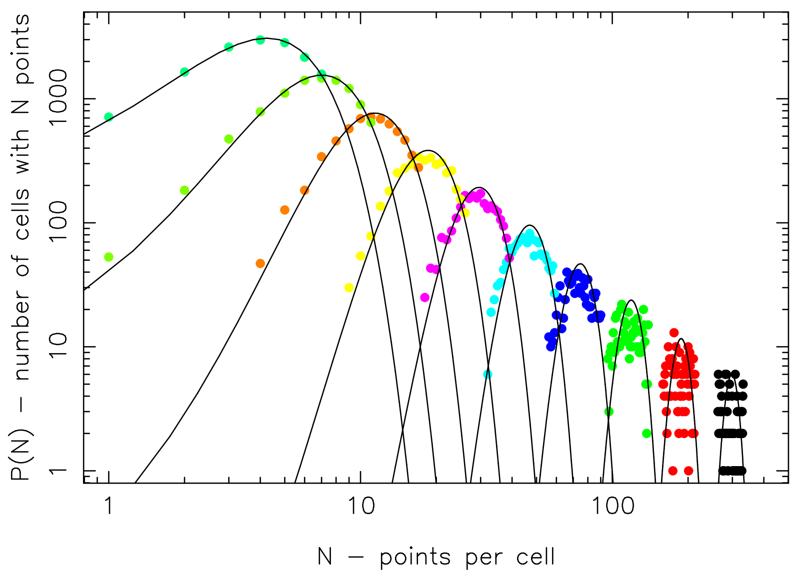

Figure 3: Number of cells versus number of objects per cell, the counts-in-cells for the toy sky with 100000 points having a discontinuity in surface density, the data set of example 10.4.) The same 10 cell-sets were used as for the uniform sky calculation of Figure 1, cell diameters evenly distributed logarithmically from 0.3° (the left-most distribution) to 3° (far right). The curves again represent Poisson distributions with means given by the average number of points per cell in each cell-set.

Thus as for the offset in $w(\theta)$ of example 10.4, with a fractional offset in surface density of ϵ, we get for the overdensity $\overline{\delta^2}$

$$\overline{\delta^2} = \frac{1}{2}(\epsilon/2)^2 + \frac{1}{2}(-\epsilon/2)^2 = \epsilon^2/4 \qquad (7)$$

so that $\epsilon = 0.2$ gives the $\Delta y(L) = 0.01$ shown as the red line in Figure 2.

Figure 3 shows the distributions of counts-in-cells for this non-uniform sky using the same sets of independent cells as for the uniform sky c-in-c. The distributions still look Poissonian for the most part, but are they are noticeably more ragged with greater scatter (=> increased variance) than those of Figure 1. This is quantified by the offset in $y(L)$ apparent in Figure 2.

As in example 10.4, investigate the effects of surface-density changes on scales smaller than that of the example, as well as the effects of large-scale gradients rather than steps. Does c-in-c methodology offer any obvious advantage in these issues?