**Problem Set I: Poisson distribution, fitting positions** *due 1 Mar 2007*

In this problem set, you will use numerical simulations to test some ideas presented in class. You can use any numerical package you happen to know, but should consider writing your own code, especially if you have never done this. To help you get started, you can find the code I wrote myself for question (1) via the course web site. It is a C programme using the GNU Scientific Library (GSL; my first semi-serious C programme and first use of GSL; this so that I could check the problem set remained reasonable...); if you would like to use GSL too, but do not know how to install this (or do not want to), I can provide a guest account on my machine. In general, if you need help with writing code, do ask! An important goal of this problem set is to get you to a point where you don't hesitate to do a Monte-Carlo simulation if it seems useful.

*Note: evaluation will be based on plots and text you hand in, as well as on a quick description of how you did the simulations. The latter is probably easiest done at your computer.*

1. In class, I mentioned that one has to be careful assigning weights $\sigma = \sqrt{N}$ for Poisson-distributed data. Verify that statement, by doing a simulation in which you generate $n_{\rm trial} = 100$ sets of $n_{\rm sample} = 100$ points each for Poisson-distributed data with mean $\mu = 100$, and calculate both the straight average and a weighted average. Make a histogram of these averages, and discuss in a few words why the two give inconsistent results. Also try one or two different mean values $\mu$ and look for consistencies in the deviations. (*For more discussion, see Mighell, 1999, ApJ 518, 380.*)
   **Solution.** *From my simulation, I found that the straight average was 100, as expected, while the weighted average was 99. The reason for the bias is that values that fluctuate low are given higher weight than values that fluctuate high. Increasing $\mu$ gives fractionally smaller errors. Looking in detail, one finds that weighted average is always 1 smaller than the straight average (see reference).*
2. Now write a routine that simulates a one-dimensional image of a star, with model data $M(x)$ given by,

$$M(x) = C + A\frac{1}{\sqrt{2\pi}\sigma_{\rm w}} \exp\left[-\frac{1}{2}\left(\frac{x - x_{\rm c}}{\sigma_{\rm w}}\right)^2\right]. \qquad \text{(I.1)}$$

   For definiteness, use pixels $x = 0 \ldots 99$, sky $C = 100$ photons, amplitude $A = 5000$ photons, and width $\sigma_{\rm w} = 2$ pixels. Now add Poisson noise to all these data points, and do a non-linear least-squares fit with a Gaussian (assuming uncertainties of $\sqrt{N}$, i.e., ignoring the minor deficiency arising from that assumption that you found above). Repeat 100 times and compare the typical uncertainty on $x_{\rm c}$ inferred from the covariance matrix with the scatter between different trials. How does the scatter relate to the width $\sigma_{\rm w}$ and the signal-to-noise ratio?
   **Solution.** *See web site for my programme and output; I fitted for sky, amplitude, and centre position, but not for width (taking this as a case where we fit a star with a known PSF). I find $\langle\chi^2\rangle \simeq 100$, roughly consistent with the expected $100 - 3 = 97$ degrees of freedom. As expected, the uncertainties from the covariance matrix for $x_{\rm c}$ are consistent with the scatter around the input value. This is true also for the amplitude, but not for the sky, where the scatter is higher: 1.50 instead of 1.05 from the covariance matrix.*
   *Looking in detail, I find that the mean sky and mean amplitude derived from the fits are smaller by 1 from the input values (as could be expected from part 1); for the amplitude, this effect is small and does not increase the scatter much, but for the sky it is important. If I calculate the scatter relative to the mean best-fit sky of 99, I find 1.06, which is consistent with what is inferred from the covariance matrix.*

*For the uncertainty on $x_c$, one sees that a reasonable estimate is $\sigma_x = \sigma_w/(S/N)$, where $S/N \simeq 5000/\sqrt{5000} \simeq 70$ for the case here (where the source is substantially brighter than the sky). For a fainter source, the same would hold, but one would need to take account the contribution of the sky to the noise.*

3. Now, on purpose, make stars that do not conform to a simple Gaussian, e.g., by replacing the above model with,

$$M(x) = C \;+\; A(1-\alpha)\frac{1}{\sqrt{2\pi}\sigma_{\rm w}}\exp\left[-\frac{1}{2}\left(\frac{x-x_{\rm c}}{\sigma_{\rm w}}\right)^2\right] + \;A\alpha\frac{1}{\sqrt{2\pi}\sigma_{\rm w,2}}\exp\left[-\frac{1}{2}\left(\frac{x-x_{\rm c}-\delta_x}{\sigma_{\rm w,2}}\right)^2\right]. \tag{I.2}$$

This mimics a poor PSF, in which a fraction $\alpha = 0.2$ of the light is in, say, a wider PSF with width $\sigma_{\rm w,2} = 4$ pixels, which is offset from the good "core" by $\delta_x = 1$ pixel. Again, add Poisson noise and fit the model star, but fit it with the simple model from the previous question (i.e., ignore the second Gaussian in the fit). Compared to the good fits, what happens to $\chi^2$? What happens to the uncertainty on $x_{\rm c}$ inferred from the covariance matter? How does the latter compare to the offset from the true position? And how does it compare to the scatter from the different trials? If you wanted to measure differences in positions on an image, what would you suggest is a good estimate for the uncertainties on these position differences?

**Solution.** *Since the model is not good any more, one find a larger $\langle\chi^2\rangle = 167$, but the errors estimated from the covariance matrix do not change. These errors are now substantially smaller than the mean deviations from the input values for the amplitude ($\langle A - A_{\rm in}\rangle = -300$, while $\sigma_A = 76$) and the centre position ($\langle x_{\rm c} - x_{\rm c,in}\rangle = 0.074$ while $\sigma_x = 0.033$); for the sky, the effect is less big ($\langle C - C_{\rm in}\rangle = 1.3$ while $\sigma_C = 1.06$; but note that in part 2 we had $\langle C - C_{\rm in}\rangle = -1$, so the sky did take up some of the flux of the star).*

*One could hope that by rescaling the errors with $\sqrt{\chi^2_{\rm red}} \simeq 1.3$, the error estimates would be better, but this clearly is not the case here: the mean deviations are still larger than these rescaled errors.*

*Again calculating the scatter relative to the mean deviations, I find numbers that are much closer to the values inferred from the covariance matrix. Thus, if one is determining relative positions, where systematic effects cancel, it is better to use the covariance matrix.*

*The lesson to take, of course, is that relative measurements are much more secure.*