Astro 121, Fall 2005
Week 4 (September 21 [happy equinox!])

Topic: Data modeling, and CCDs
Break: Blair (the following week, 9/28: Michael)

For next week we'll deal with two different topics. The first is comparing data and models, and assessing "goodness of fit", how well your model matches your data; the second is starting to understand CCDs, the detector of choice for many modern astronomical observations. As I mentioned last week, remember to keep thinking probabilistically. When evaluating a model and/or set of observations, try to phrase a hypothesis that you're testing, and a quantitative probability that the hypothesis is correct.

## I. Data modeling

**Reading:** All of these books are on reserve in Cornell unless otherwise noted.

- *Astronomy Methods,* Bradt, pp. 165–172. (Pages 151–165 cover material we discussed last week, if you'd like to have a look at Bradt's treatment of this.)
- *Data Reduction and Error Analysis for the Physical Sciences,* Bevington and Robinson, Section 4.3.
- *An Introduction to Error Analysis*, by John R. Taylor. Chapter 12.
- "Practical Statistics for Astronomers – II. Correlation, Data-modeling, and Sample Comparison", J.V. Wall 1996, *Quarterly Journal of the Royal Astronomical Society*, **37**, 519–563. As I mentioned last week, this contains less theory than some of the other sources, but it is more closely aligned with our goals as astronomers, and it is pretty readable. Don't read the whole thing, but it's good to have on your shelf as a starting place if you want to find references to other works on statistics. For our purposes this week, look at Section 3. It covers chi-square (with a slightly different definition than the other references) but has the advantage of comparing it to a number of other ways of estimating parameters of a fit. In particular it's good to be aware of Bayesian techniques for analyzing data.

**Important terms and concepts:** $\chi^2$; reduced $\chi^2$;

## Problems

1. $\chi^2$ fitting (Bevington and Robinson Sec. 4.3) is often used in astronomy (and other sciences) to find the set of coefficients of a given function that give a curve that best fits the data. The basic idea is to find the set of coefficients that minimizes $\chi^2$ ; this provides "best fit" parameters for a function that we specify to fit the data, but it also allows us to assess the probability that it is a "good" fit, i.e. that the remaining differences between the curve and the data arise just due to experimental uncertainty. Examples:

   a. You fit a straight line to your set of 20 data points, determining a slope and an intercept. The total (not reduced) $\chi^2$ of the fit is 15. What is the value of $\chi_\nu^2$ (reduced $\chi^2$)? What is the probability of this level of disagreement between model and data arising by chance? What do you conclude about whether a straight line is a good fit to the data? (You may find it helpful to look at Sec. 11.1, and Table C.4 on p. 258; Taylor's Chapter 12 also provides a nice discussion of $\chi^2$.)

   b. In the paper by Herbst *et al.* 1997 (AJ 114:744; see pp. 748-749), the authors fit a model (with 6 free parameters) to their observations (11 data points) and obtain a $\chi^2$ value of 0.4. They conclude that the model is a good fit to the data. What do you conclude? (It's not clear in the paper whether they are quoting $\chi^2$ or reduced $\chi^2$, so you may want to work the problem both ways.)

2. Find a paper in the astronomical literature (one you're already reading for some other purpose is fine) that does some statistical analysis, and, using the theory and resources that we've gone through in class, try to understand what the authors are doing. (An example using chi-square, such as those above, is fine, but feel free to branch out and look at other techniques.) Come to class prepared to give a short presentation on what the authors were trying to do, what technique they used, what conclusions they reached, and whether or not you agree with those conclusions. This sequence of reasoning is exactly what you need to do as a scientist when you are reading the literature. If you find an example that interests you but are having a hard time understanding what the authors are doing, feel free to come talk to me about it.

## II. CCDs

Since I'd like you to get started working with real data very soon (like for the assignment after this one), for the moment we're going to skip over the process by which a telescope collects light and forms an image, and go straight to understanding the most important detector for optical astronomy, the CCD.

**Reading:** Bradt gives fairly terse coverage of the details of working with CCD images, given how important they are in astronomy. I've recommended some additional sources for more detail. Chromey is a good place to start; it is fairly comprehensive without being overwhelming. Berry and Burnell give a very clear, practical approach to understanding CCDs and their calibration; Howell is a little more detailed and technical, and more oriented toward professional astronomy.

- *Astronomy Methods,* Bradt, pp. 137–143. Don't worry too much about the details of the different semiconductor layers in the CCD.

- Chromey, pp. 7-1 to 7-24. Don't worry about the difference between QE and DQE.

- *Handbook of CCD Astronomy*, Howell. Chapters 1 through 3, and through Sec. 4.3 of Chapter 4. In Chapter 2, some of the descriptions of different types of CCDs (e.g. in Sec. 2.2.2) are a bit confusing since much of the terminology is not discussed thoroughly until later chapters; just get through it as best you can, and some of it will be clarified in Chapter 3. Also, note that the sentence that runs from the bottom of p. 31 to the top of p. 32 is incorrect. Read noise is an *uncertainty* in the number of electrons detected, and thus it can increase or decrease the readout value compared to the true value; it is not a number simply *added* to each pixel.

- *The Handbook of Astronomical Image Processing*, Berry and Burnell. Chapter 1, pp. 1–17; Chapter 4. In Chapter 1, you can skim pp. 1–8; we'll come back to some of those concepts in a few weeks.

**Important terms and concepts:** charge transfer efficiency, gain, bias, overscan, dark current, flat field, ADU, quantum efficiency

**Problems:**

3. What is the difference between subtracting overscan strips from CCD images, and subtracting separate bias frames from the images? When might you choose to do both? When might you choose *not* to do both?

4. A particular CCD has read noise of 3 $e^-$/pixel and a gain of 4 $e^-$/ADU. If each pixel in a CCD bias image contains (on average) 100 ADUs, what is the typical noise per pixel in this bias frame?

5. Go to the Kitt Peak National Observatory web site and follow the links for CCD information to find the characteristics of the CCD named T2KB.

   a. If you were using this CCD, what value of gain would you choose if you were observing an object that yielded a flux at the detector of 1,000 photons / second / pixel? What is the longest exposure you could take and still obtain a reliable flux measurement from your image at the 0.1% level?

   b. Would the answer to the above questions be the same for observing a source that yielded 10 photon / sec / pixel? If not, how would your answer(s) differ?

   c. If a typical exposure during your observing run was 30 minutes long, would you subtract a dark frame from your images? (Think carefully about [and calculate] the noise in the final image in the case where you subtract the dark frame and the case where you don't.)

   d. What is the peak quantum efficiency of this CCD? At what wavelength does it occur?

6. At the bottom of p. 10, Howell gives a typical value for charge transfer efficiency (CTE) of a CCD as 0.99999 (sometimes called "five nines").

   a. If you are using a 2048x2048 CCD with this charge transfer efficiency, what fraction of the charge would you expect to remain at the readout of the last pixel?

   b. Comment on how the fact that electric charge is quantized affects your answer to part a.

   c. Geoff Marcy (discoverer of many extrasolar planets using the radial velocity method) told me a few years ago that poor charge transfer efficiency in their CCD was one of the major problems giving systematic errors in their radial velocity measurements at Lick Observatory. Explain how poor CTE could give a systematic radial velocity error.

7. As part of a proposal I wrote to NSF to fund a new telescope for the new science center, I had to calculate what signal-to-noise ratio we might be able to achieve in taking spectra with the proposed instrumentation. Part of that calculation involved determining noise sources in the CCD. The camera I proposed that we buy is made by Finger Lakes Instrumentation, and it is based on the Marconi CCD 42-40 chip. If we run this chip at a temperature of –20° C, what dark current should it have? (The chip is very similar to the Marconi CCD 47-10 chip.) For what exposure times would dark current dominate over read noise as the primary background noise source?

8. Suppose you want to observe a white dwarf at 4000 Å to look for variability on a 1 second timescale (i.e., you will make many 1 second integrations). Consider the CCDs T2KB and S2KB, both available at Kitt Peak National Observatory; in your proposal, which CCD would you request? Presume that with the telescope and instrumentation available to you at Kitt Peak, 500 white-dwarf-4000 Å-photons fall on a pixel each second, and ignore any source of background other than readout noise.