Astro 121, Fall 2005
Week 3 (September 14)

Topic: Uncertainty and statistics
Break: Victoria

Reading: All of these books are on reserve in Cornell unless otherwise noted. You don't have to read all of these; read what you need to understand the essential concepts and do the problems. I'd recommend Chromey as a good place to start.

- *Data Reduction and Error Analysis for the Physical Sciences,* Bevington and Robinson, Chapters 1–4.

- *To Measure the Sky* by Fred Chromey, Chapter 2. Note that the equation at the bottom of p. 2-17 has a mistake. It should read:

$$\text{Fractional uncertainty in counting N events} = \frac{\sigma}{\mu} \approx \frac{1}{\sqrt{N}}$$

- *An Introduction to Error Analysis*, by John R. Taylor. Taylor covers much the same ground as the references above, but more slowly. You might want to skim the intro material and see Chapters 3–5 and 11–12.
- "Practical Statistics for Astronomers – I. Definitions, the Normal Distribution, Detection of Signal", J.V. Wall 1979, *Quarterly Journal of the Royal Astronomical Society*, **20**, 138–152. 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. You probably wouldn't want it to be the only thing you ever read about statistics, but it's a good complement to other sources. You can find it through ADS.

Important terms and concepts:
- Parent distribution vs. sample distribution; mean; standard deviation; standard deviation of the mean; Poisson distribution; Gaussian distribution; propagation of uncertainty

Problems and questions

1. Why is the Poisson distribution so important for astronomy? (Hint: observational astronomy is fundamentally about waiting for photons to arrive.)

2. You take a CCD image of a nearby star. On examining the image, you find that the background level (due to emission from the sky) is 200 counts per pixel, and that the pixel centered on the star contains 1000 counts. (Assume that all the light from the star falls in this one pixel.) What is the signal-to-noise ratio of your measurement of the flux from the star? What would the signal-to-noise ratio be if you observed the same star with *no* background emission? (This problem assumes a perfect CCD with no intrinsic noise; we'll deal with real CCDs next week.)

3. Rules of thumb, round 2. A useful rule of thumb in astronomy is that an uncertainty of 0.01 magnitudes is roughly equal to a fractional uncertainty of 0.01 (i.e. 1%) in flux. Use the formalism of error propagation discussed in any of the readings to show that this is indeed true. (It turns out that it simply is the result of a numerical coincidence—what is that coincidence?) At

what magnitude uncertainty does this approximation break down? (In other words, for what x does a 0.0x magnitude uncertainty *not* roughly equal a 0.0x relative flux uncertainty?)

4. In a 10-pc-radius volume around the star TW Hya, there are three stars that lie well above the main sequence and have space velocities of less than 10 km/s relative to the Local Standard of Rest. In 28 other fields selected in the same way, there are 22 fields with no stars that meet these criteria, 5 fields with one such star, and 1 field with two. Is the TW Hya field unusual? (This is a problem I really had to solve. If you want, see Jensen, Cohen, & Neuhäuser 1998, AJ [you can get the exact reference from ADS!] to check your answer against mine *after* you've worked the problem out for yourself.)

5. In Preibisch et al. (2001; AJ 121:1040), the authors primarily examine the low-mass population of the Upper Scorpius OB association, but in Section 5.2.2, they briefly discuss the high-mass population. Based on previous studies, they say that the expected number of high-mass stars in a field that is 2° in diameter is 1.3 ± 0.1 stars. In the 2°-field that they study in this work, they find two massive stars, which they say is "consistent with the expected number within the uncertainties." Given that one could *never* find a field with 1.2–1.4 stars (the range that naively appears to be allowed by the uncertainties), what does it mean to be consistent with the expected number? You might want to find the probability that the authors would find what they did in a random field, as well as finding the most likely number of stars in a field.

6. Uncertainty on the mean. Read Bevington and Robinson, Section 4.1, paying close attention to the important result in Eq. 4.14. What happens to the sample standard deviation s (Eq. 4.13) as the sample size increases? What happens to the uncertainty on the mean σ_μ? Can you give an intuitive explanation for this difference?

7. The Central Limit Theorem. Why do astronomers use Gaussian distributions so much? Follow one of the links from the course web page http://astro.swarthmore.edu/astro121/, or do your own web search, to find a good discussion of the Central Limit Theorem, and come to class prepared to explain it. Also want to look at the demo at http://astro.swarthmore.edu/astro121/distributions.html that illustrates the special case of the Poisson distribution converging to a Gaussian distribution.

8. Look at the paper "Infrared radiation from an extrasolar planet", by Deming et al., *Nature* 434:740 (2005), which presents the first detection of photons from an extrasolar planet. On the top of page 742, the authors say "Because about half of the 1,696 points are out of eclipse, and half are in eclipse, and the SNR≈111 per point, the error on the eclipse depth should be $0.009 * 2^{0.5}/848^{0.5} = 0.044\%$ of the stellar continuum." Derive that expression for the uncertainty on the eclipse depth.

9. Survival analysis. Chromey, problem 1 from Chapter 2. The type of statistical analysis necessary to deal with questions like this formally is called "survival analysis". But astronomy doesn't generally deal with survival times—so why should we care about questions like this? (As I'm sure you have guessed, we *should* care.)

10. You are reading the ApJ, and you find one paper that gives a value of 32 ± 2 km/s for the radial velocity of a star, and another paper that gives a value of 28 ± 2 km/s for the radial velocity of the same star. Do these numbers agree with each other? Is the radial velocity of the star variable? Explain your answer carefully, defining exactly what you mean by "agree".

11. Explain why the factor of N – 1 (rather than N) is present in the denominator of the expressions for the sample variance and sample standard deviation (e.g. Chromey Eq. 2.4 or Taylor Eq. 4.9). The Wall article is helpful for understanding this intuitively.