Bayesian Statistics, Assignment for Tuesday, Sept. 17

From *Statistical Treatment of Experimental Data*

This assignment is meant to go along with Section 9 of Young, pp. 64-76. You have learned enough about integrals that the results in Section 9 should make sense, but the derivations — for example, how he gets from Eq. 9.10 to Eq. 9.12 — you don't have near enough training in integral calculus to follow. Again though, the result, Eq. 9.12, should make sense. It just says that the mean value in a Gaussian distribution $\overline{x}$ is the parameter m. Ditto for Eq. 9.15 which says that the variance σ^2 is $1/2\,h^2$. Using the latter result, Young gets rid of h^2 in Eq. 9.16, which is his final form for the normalized Gaussian.

For Problem Set 5

Gaussian Distributions

1. In Problem Set 3, you calculated the mean and standard deviation of the cocker spaniel puppy litters. When I did it I got that the mean was about 8 and the standard deviation was about 2.

(a) Make a nice new graph of the cocker spaniel puppy litters, but instead of it being a histogram with all the histogram bars adding up to 200, normalize the distribution. In other words, divide every bar by 200 so all the bars add up to 1.

(b) On top of the same graph that you made in (a) plot a Gaussian with mean 8 and standard deviation 2.

2. If so many events happen that it is no longer obvious that they are discrete events, then a Poisson distribution starts to look like a Gaussian distribution. Let's imagine we break 16 mugs a week on average.

(a) Make a table of the distribution that shows the odds of breaking 8 mugs, 9, mugs, 10 mugs, etc. all the way out to 24 mugs in a week. The formula you are looking for is Eq. 8.5 with $a = 16$, and you are plugging in n from 8 to 24. (Yah, I know, you have to bust out your calculator and punch in the formula 17 times, sorry). Note: Because this part has huge numbers like 16^{24} and 24!, some calculators might barf, but hopefully yours can handle it.

THE PROBLEMS ARE CONTINUED ON REVERSE

(b) Graph the table you made in (a).

(c) According to Eqs. 8.6 and 8.9 what is the mean and standard deviation of the distribution you graphed in (b)?

(d) Write out Eq. 9.16 for a Gaussian function with the mean and standard deviation you found in (c).

(e) If you fuzz your eyes out or hold the graph paper a ways away, maybe the graph you made in (b) is already looking like a Gaussian. But now graph the Gaussian function you found in (d) on top of the bar chart data you graphed in (a).

DISCUSSION: I hope this is enough to start sensing a deep relation between the Poisson and Gaussian distributions. (Remember that we derived Poisson from binomial, so actually all the distributions we have been studying are deeply related.) It takes Young a whole appendix (Appendix C) of nasty algebra to establish the relation.

The bottom line is if a random event doesn't occur discretely, but instead is the sum of of a very large number of small random events, so many small events that it might as well be infinity, then we can plausibly transition from discrete to continuous distributions. All of p. 65 was devoted to general discussion of when this is expected to be true, including some amusing commentary.

In the problem you just did, you saw that even 20 small events (on average) per week is a large enough number to start seeing a Gaussian distribution of mug breakage emerge.

3. Consider the function $f(m) = M_2\, m^2 + 2\, M_1\, m + M_0$. This is a parabola in the variable m. M_2, M_1, and M_0 are just unknown constants. If M_2 is negative the parabola opens downward. We don't have to care whether it is negative or positive. If it is negative it has a maximum. If it is positive it has a minimum.

(a) Use the "complete the square" trick to find m value where the minimum or maximum occurs. I put the 2 in to make the algebra tidier.

(b) Now that you have found the m value where the minimum or maximum occurs, stick it back in to $f(m)$ to get the minimum or maximum value (or just read the formula off from the complete the square algebra you did in (a)).

4. Graph and indicate the location of the maximum and the value of the maximum for the function,

$$f(m) = -m^2 + 2\, m - 1$$

HINT/CROSS-CHECK: If you put in $M_2 = -1$, $M_1 = 1$, and $M_0 = 1$ into the formulas you got in Problem 3 they better agree with the location and value of the maximum you indicated in this problem.