R Class Homework 6: Poisson distributions

Background

The Poisson distribution: This is a data distribution defined by discrete observations (like pollinator visits) made against a continuous sampling space (like time), and where the variance of the distribution is equal to the mean. Given a mean of 'lambda', the function below calculates the probability of a given value (k) of X. E.g., if the mean number of pollinator visits to a flower in a day is 20 (lambda=20), what is the probability of observing exactly 13 visits (k=13)?

$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Plotting multiple k against a given lambda gives you a probability distribution. As lambda increases, the Poisson distribution approximates a normal distribution.

The assignment

Make a neatly formatted R script with a header and clearly identified subsections for the following problems. Include clear comments about what the code is doing. Submit the file with a name like this example: "R Class HW06_T Taylor.R"

1) Translate the probability mass function of the Poisson distribution into R. That means make a line of code equal to this equation $\frac{\lambda^k e^{-\lambda}}{k!}$ where lambda is an object named 'l' and k is an object named 'k'. e is Euler's constant. Use the internet to find out how to use e in R. k! is "k factorial". Look up how to do factorials in R. Keep track of parentheses!

2) Make a for loop that will:

- Calculate the probability (p) of X=k given lambda=1, for k=1:20. That is, the loop should keep lambda fixed at 1, but calculate p 20 times, for k=1, k=2, … k=20
- Grow a vector called 'p' containing each calculated p. See the section on "growing a vector" in the for loop tutorial.

3) Deposit the vector 'p' into a data frame called 'pois'. Also include a column 'k' equal to 1 through 20. Use ggplot2 to plot 'p' (y axis) against 'k' from your data frame with both points and lines. Label the axes. The plot should look like the one here for lamba=1: http://en.wikipedia.org/wiki/Poisson_distribution .

4) Do that again for lambda=5. Look at the two graphs to see how the Poisson distribution looks more normal with higher lambda.

5) Make three data frames from the above code, for lambda=1, lambda=4, and lambda=10. Add a 'lambda' column to each data frame, containing the lambda value used to make that data frame. Use rbind() to get those three data frames into a single data frame. Then repeat your ggplot() code, but referring to this new data frame, and adding the argument to aes (fill=lambda). This should re-create the figure from the Wikipedia page (linked in problem 3).