Homework 4 Problem 4

William Matzko

March 4, 2021

4.4.1

The figure showing the probability of k solar flares in one day has the theoretical binomial and Poisson distribution shown with the data. For the probability of each distribution, I used the average number of flares in a day divided by 24. We see that both theoretical distributions capture the rise/fall of the distribution fairly well, but both overestimate the peak of the distribution. Considering that the Poisson distribution 'hugs' the data curve better, I would say that the probability distribution is most similar to the Poisson distribution.

4.4.2

The plot showing the probability distribution of the time between flares has an exponential curve overlaid. On line 134 of the code (or around that), one can make the y-axis a log scale. In the former case, we see that the exponential distribution approximates the time between flares reasonably well. In the latter, we see that the theoretical curve fits well out to ≈ 23 hours. Afterwards, the fit becomes quite horrible. However, at that point, we suffer from a limited amount of data, which will skew the statistics. Considering both the regular and log plots, I would say that the probability distribution of the time between flares is reasonably similar to that of the exponential distribution.

4.4.3

The assumptions for the binomial distribution (Bernoulli trial) are

- There are two possible outcomes (i.e. flares or no flares)
- The probability of 'success' is p and the probability of 'failure' is $1 - p$
- The probability p does not change

In contrast, the Poisson distribution (which may be derived from the binomial distribution in the correct limits) makes the assumptions:

- k events may occur in some time interval ($k = 0,\ 1,\ 2$)
- Events occur independently
- The rate at which events occur is constant
- Two events cannot be simultaneous

A cursory glance of the data reveals that it is possible for two events to occur in the same hour, but two events may not occur in the same minute. Unless one is modeling the number of flares on a time scale of minutes, then a Poisson distribution would be more appropriate. However, if we look at the shape of the probability curve of k flares in one day, then we notice that the flares do not necessarily occur independently of each other. If they did occur independently, we would expect more of a Gaussian distribution (I think...). I'm not sure how to check if the solar flare rate is a constant, but given the nature of solar flares, I have a hunch it isn't (though on small enough time scales we can perhaps assume p is constant).

Hence, strictly speaking, I do not think either the binomial or Poisson distributions fit the data *exactly*. However, I would say that the Poisson distribution fits better. This is reflected in the theoretical probability curves plotted in the chart, as discussed in 4.4.1.