NAME: Maxfield Green
DATE: 1/17/2019
Homework: Assignment 1
STAT/CS 387

Problem 1

1.1

What probability distribution (or type of random variable) characterizes these data and why?

The data are a collection of battles between the warriors and two different models of the Bayesian Hunter Killers, the Mk1s and the Mk2s. Each battle is unique, take place one at a time, result in a maximum of one death and there is a fixed probability that the warrior will die that is constant for all battles p_{death}. We see that each battle is a Bernoulli trial with a binary outcome of success or failure. The data as a whole, a collection of Bernoulli trials belongs to the binomial distribution.

1.2

Using Bayesian Inference, as the frequentest have been hunted to extinction, what is the probability that the Mk2 model type is deadlier than the Mk11?

We are interested in whether the Mk2s are more deadly than the Mk1s. i.e, we are interested in whether more warriors die from one model than the other. This can inform battle protocol and reduce uncertainty in tactical decision making processes. We know that 0.68% of battles so far with the Mk1s have resulted in warrior death. Alternatively brushes with the Mk2s have yielded 0.81% death for the warriors.

We hold the prior assumption that every battle has an equal chance of success or failure for the warriors.As stated above, our prior distribution is a binomial Bernoulli distribution. Thus our likelihood function for the Bayes rule will be the pmf of the binomial distribution given by:

$$P(Success|\theta) = \theta^{\#Successes}(1-\theta^{\#failures}) \quad (1)$$

In order to determine the probability that the Mk2 model is deadlier than the Mk1 model, we will sample from the posterior distribution given a prior uniform distribution and our observations of battles with both models and the number of resulting deaths of the warriors.

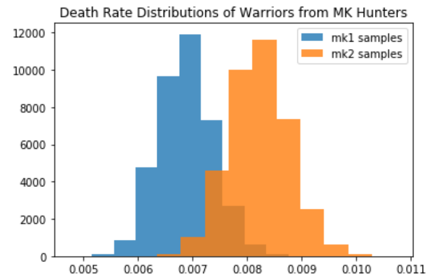

Figure 1: Posterior Samples

In the above histogram of the posterior distributions for deaths from MK1s and MK2s, we see that the Mk2 sample distribution is centered higher than the Mk1. To confirm our intuitions, I calculate the difference between the two samples, titled "diff" in line 33 of the attached code. Taking the mean of this new sample difference confirms that the Mk2s are more deadly than the Mk1s with a probability of 96.5% Additionally, this is confirmed by the trace-plot outputs in Figure 2 and the Gelman-Rubin statistic for convergence which is explained later in the assignment.

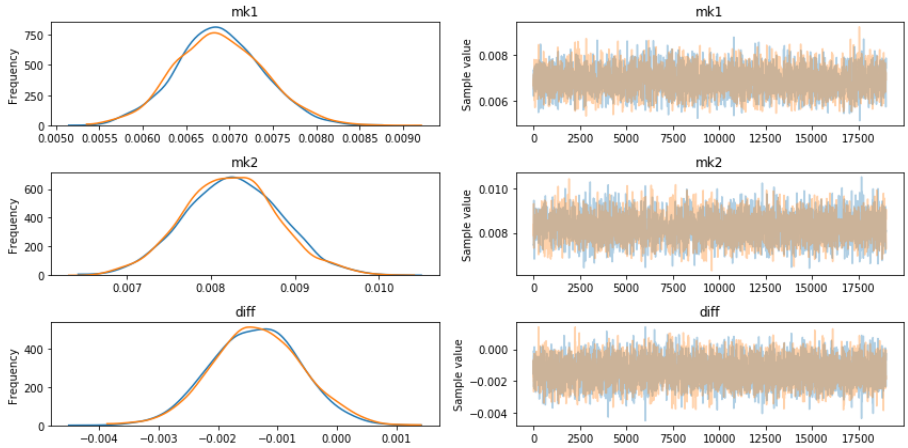

Figure 2: Posterior Samples of Mk1, Mk2 and their difference

2 Problem Two: Txt - Message Sampling Analysis

2.1 Problem 2.1

Implement this inference problem yourself (taking PyMC code from Data Science 1 lecture notes if you wish). The count data (Ct = of messages received on day t) is provided in the le txtdata.csv. Construct an argument with plots as needed to demonstrate that your sample has converged in distribution to the underlying posterior. What steps did you take to ensure you have converged? What did you calculate to show convergence and why?

First, we conduct a basic analysis of the time series data. We see that over time, the count of text oscillates with certain peeks and troughs. Visually, its pretty hard to tell if there are two regimes or not. There does certainly look to be a decent amount of variance.

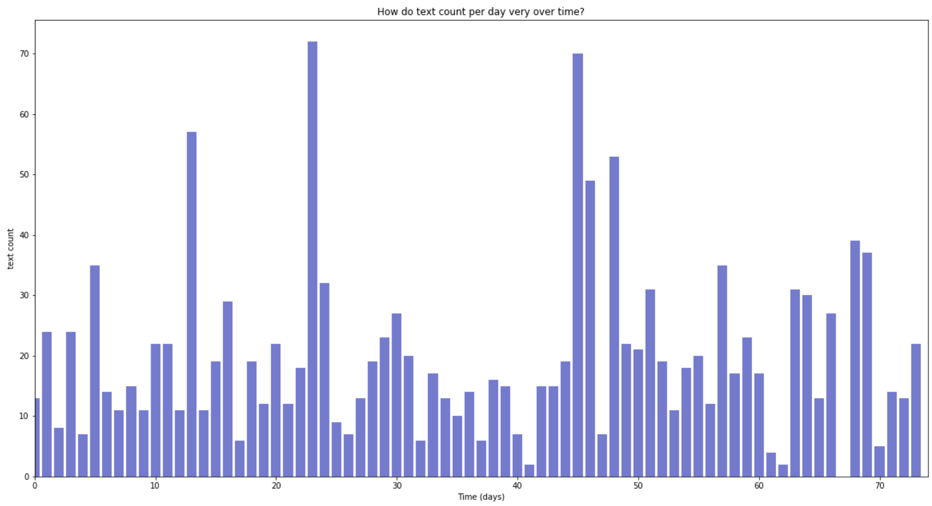

Figure 3: Txt Count per Day

To answer this question, we will generate the prior distributions and then update our priors with the information we are given in the txt data. We then can sample the posterior distribution using Markov Chain Monte Carlo(MCMC). If our our has converged, than we are seeing the true posterior distribution of the two paramters λ_1 and λ_2. As these data are continuous, it makes sense to use the exponential distribution to model λ. In this case, we will need to define the hyper parameter α. It was suggested by [1] that this could be the "inverse of the average of the count data". We will assign our parameter τ a uniform distribution as we want to assume a uniform prior for every day in the data range. See lines 72-84 for parameter creation using the pymc3 package.

After sampling the posteriors, we can take a peak via the histograms in table 4The histograms of the sampling are shown below. Again we are looking at the prior distributions informed by our observations. We see that the two λ distributions are set apart from one another, centered at different values. Additionally, the pdf for τ tau distribution show monotonically increasing probability of a switch point approaching day 45.

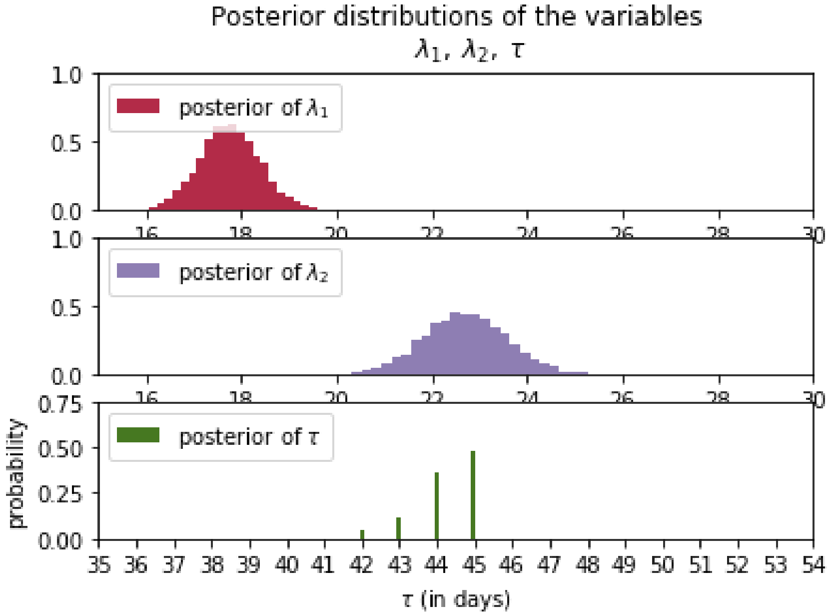

Figure 4: Posterior Histograms

To confirm that the sampling algorithm has converged, I use the Gelman Rubin test. The test examines the variance between the multiple chains from the MCMC trace. A low Gelman Rubin statistic indicates convergence. The Gelman Rubin statistics were all low on the scale, thus we have no reason to assume that convergence has not occurred. Additionally, we can examine the output trace plots in Figure four. We see that the models look highly uncorrelated, this is exactly what we want.

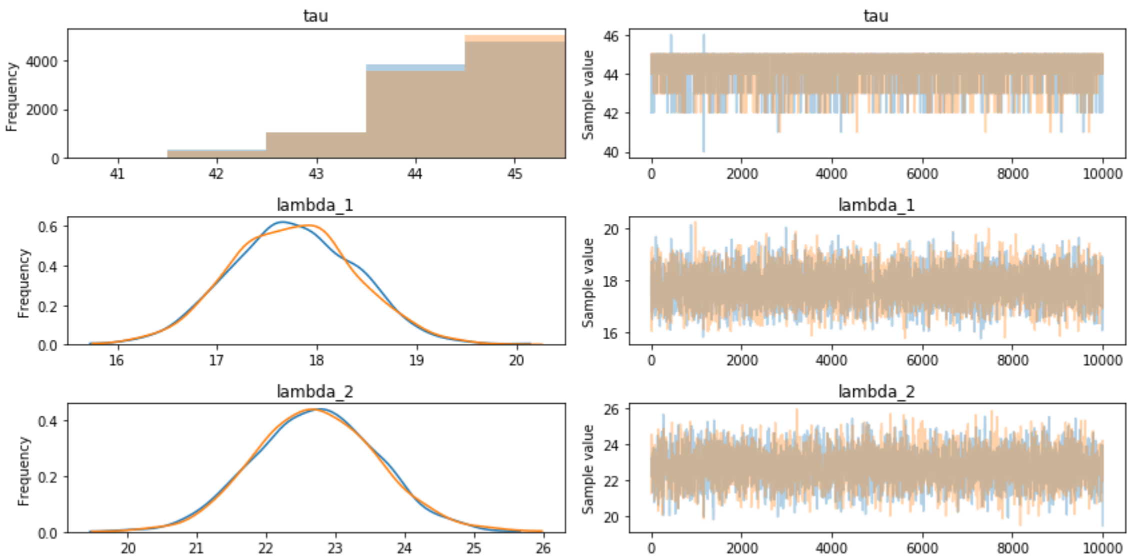

Figure 5: Trace Plots

2.2 Problem 2.2

Propose a function of time that smoothly changes the Poisson rate λ(t) = $f(t)$ from λ_1 to λ_2. How do you interpret these new parameters? What are the appropriate prior distributions and why? Demonstrate convergence of your sample.

We now aim to remove the τ parameter and smooth the transition from the distinct Poisson rates $_1$ to λ_2. We introduce ϕ_1 and ϕ_2. We will define our function f(t) as the sigmoid with parameters ϕ_1 and ϕ_2 and time (t).

$$\frac{\lambda_1}{1 + e^{-(\phi_1 + \phi_2 * t)}} + \lambda_2 \tag{2}$$

Here ϕ_2, controls for the rate at which λ_1 transitions to λ_2 and ϕ_1 controls at which time (t) the transition will occur. In the logit model, these parameters are the slope and intercept of the linear model ϕ_1+ϕ_2*t. Both ϕ will be parameters of the Normal distribution, centered around zero with a standard deviation of 3 to preserve uncertainty. The sigmoid is then translated by λ_2 and transformed by λ_1
We can create a boolean mask and loop over the samples to determine which sample is apart of each distribution. See lines 144-152 in the attached code. Now we can calculate the expected Poisson rate, or the expected number of text received on a given day. A plot of this overlaid on the observed data is provided below. 95% Confidence intervals are also provided. However, the confidence bands are extremely thin and are fairly invisible.

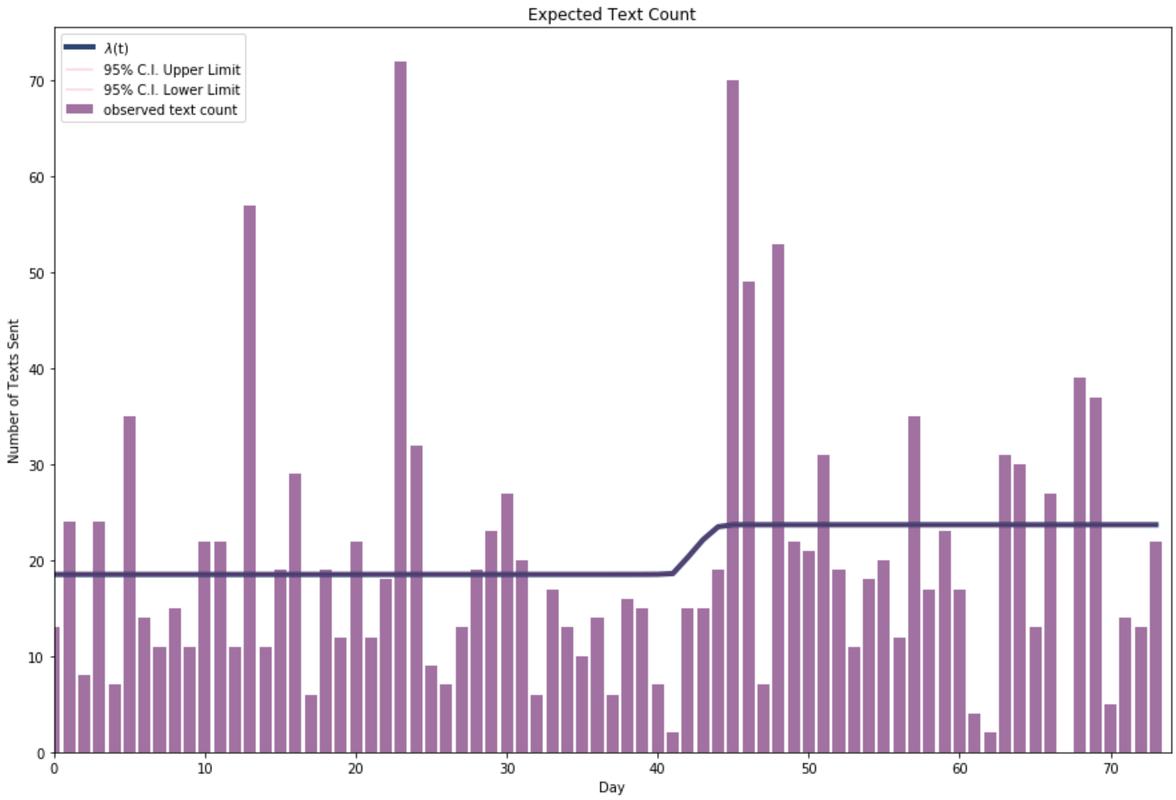

Figure 6: Trace Plots

Now we can see that this new model supports that old switch point model according to the tau parameter of the uniform distribution. The switch-point from txt frequency regimes occurs at around the same place, day 43.

3 Problem Three

Which model is better justified and why?
In this particular situation with this particular data set, the first model is superior because of its simplicity. It is much more computationally efficient as we only need to sample from 3 random variables as opposed to four. However, given a different data set, it is quite possible that the more complex model would be robust to different types of frequency regimes. Perhaps more than one switch point. Why would there only be one in a given period. Given more time coupled data, I would expect to see periodic cycles with txt count phase changes that occur.It would be more interesting to compare our two models under more observations. However, in any model, more data can always be used. In this case the initial model is better due to its simplicity and confirmed result.

4 Sources

[1]Davidson,Cam , "Probabilistic Programming and Bayesian Methods for Hackers", Chapters 1 and 2,Jupyter Notebook