Simple Coin Tossing Experiment and Binomial Beta Model

Investigating Deep the effect of prior when data do agree/not-agree with prior

Overview

Here we are going to investigate deeply the **Bernoulli - Beta Model** Where Data our the Outcome of the simple coin tossing experiment follow Binomial and the Prior we are going to take is Beta,our primary goal here is to investigate the influence by priors on the posterior ,and we will do this by taking different data for same priors, and see how the posterior is affected

Model

We are going to use binomial model for this, Lets us assume X_i , a bernoulli random variable such as

$$X_i = 1 \qquad When\ i^{th}\ Toss\ Head$$
$$X_i = 0 \qquad Otherwise$$

We are assuming Head as **Success** and Tail as **Failure** Now,

$$X = \sum_{i=1}^{n} X_i$$

Where

$$X \sim Bin(n, \theta)$$

Now We have to take Prior for our proportion θ , Now we are going to use beta - distribution for prior , and we want the parameters ,We want our prior which is biased towards more success, so we are taking the distribution for the population parameters such distribution is negatively skewed and we are also assuming there is highest possibility that the proportion lies near to 80% that is $\theta = 0.8$, So

$$\theta \sim Beta(a, b)$$

Where Mode will be given by

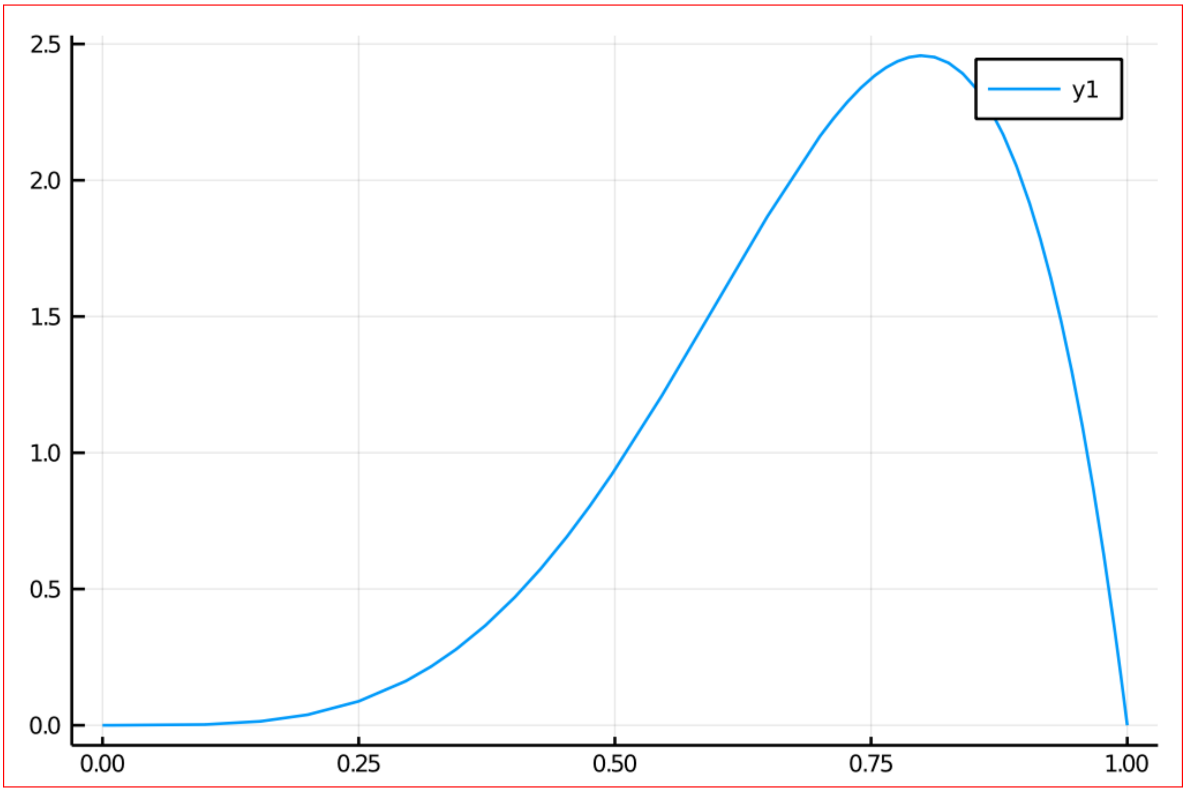
$$Mode = \frac{a-1}{a+b-2}\ for \quad a > 1, b > 1$$

So,

$$
\begin{aligned}
\frac{a-1}{a+b-2} &= 0.8 \\
\Rightarrow a-1 &= 0.8(a+b-2) \\
\Rightarrow a-1 &= 0.8a+0.8b-1.6) \\
\Rightarrow 0.2a &= 0.8b-0.6 \\
\Rightarrow a &= 4b-3
\end{aligned}
$$

Taking a = 5 we get b = 2 , so

$$\theta \sim \beta(5,2)$$



Then the posterior will be given by

$$p(\theta|x) = \frac{p(\theta)L(x;\theta)}{\int p(\theta)L(x;\theta)}$$

where

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} = \frac{\Gamma(5+2)}{\Gamma(5)\Gamma(2)}\theta^{5-1}(1-\theta)^{2-1} = 30\theta^4(1-\theta)^1$$

and

$$L(x;\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Then

$$p(\theta|x) = \frac{30\theta^4(1-\theta)^1\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\int 30\theta^4(1-\theta)^1\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,d\theta}$$

The Denominator of the above equation is a normalizing constant and that evaluates to

$$
\begin{aligned}
&\int 30\theta^4(1-\theta)^1\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,d\theta \\
&\Rightarrow 30\int \theta^{4+\sum x_i}(1-\theta)^{n-\sum x_i+1}\,d\theta \\
&\Rightarrow 30\cdot\beta(4+\sum x_i+1, n-\sum x_i+2)
\end{aligned}
$$

Now,

$$p(\theta|x) = \frac{1}{\beta(4+\sum x_i+1, n-\sum x_i+2)}\theta^{(4+\sum x_i+1)-1}(1-\theta)^{(n-\sum x_i+2)-1}$$

Here

$$p(\theta|x) \sim Beta(4+\sum x_i+1, n-\sum x_i+2)$$

And we will take posterior mean as a bayes estimator that is

$$\hat{\theta}_B = \frac{5+\sum x_i}{7\ +\ n}$$

Analysis

We can look at the posterior as it have same form as our prior and it is known as conjugate prior , for this situation where Random Variable X_i denote success and failure the $\sum x_i$ will denote the total number of success observed and $n-\sum x_i$ is the total number of failures then the distribution of posterior can be given by

$$p(\theta|x) \sim Beta(5+total\ success\ observed, 2+total\ failures\ observed)$$

Where our prior is

$$p(\theta) \sim Beta(5,2)$$

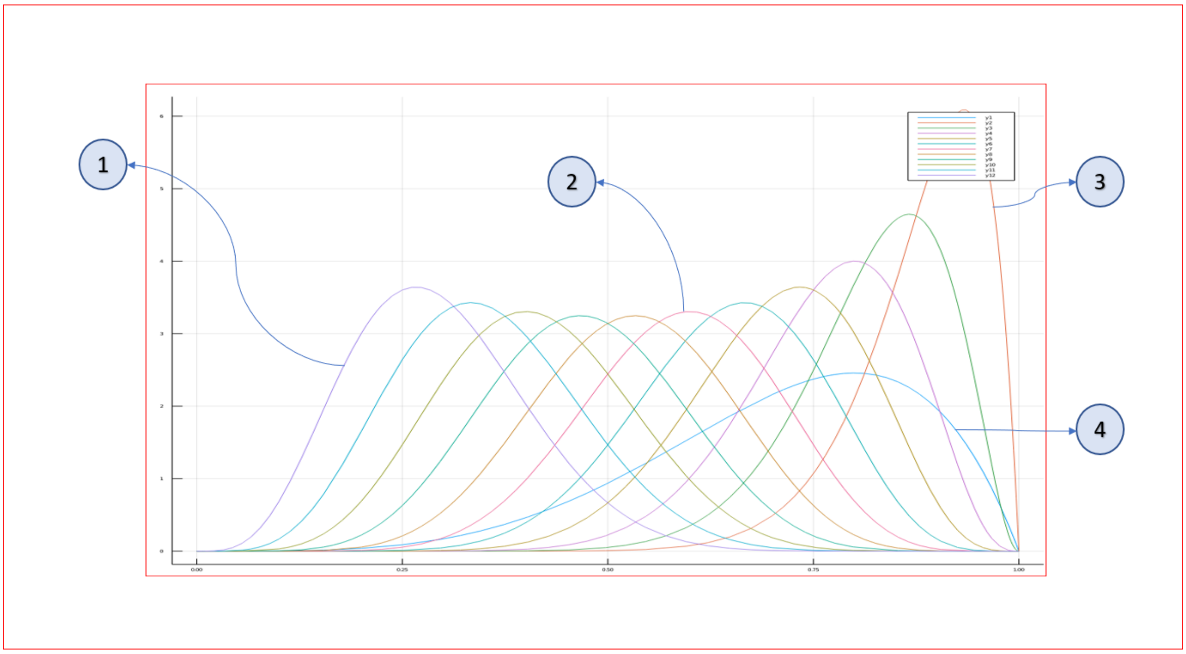
So we must analyze the What will happen if we add number to the parameters of beta distribution , here we must consider

$$total\ success\ observed + total\ failure\ observed = total\ number\ of\ observations = n$$

For Simplicity let us consider a coin tossing experiment experiment with 10 observations the we may end up in any of the following situations

	Total Success Observed	**Total Failure Observed**
1	10	0
2	9	1
3	8	2
4	7	3
5	6	4
6	5	5
7	4	6
8	3	7
9	2	8
10	1	9
11	0	10

So let us plot posterior for these situations



1. This represent the 0 Success and 10 Failures
2. This represent the 5 Success and 5 Failures
3. This represent the 10 success and 0 Failures
4. Out prior $\beta_{5,2}$

As we can see prior is effecting the estimate for the proportion, as the number of successes increasing, the data is giving evidences that the proportion is biased towards 1 and our prior is agreeing with the data the probability increases , but one thing that we should we notice even when success 9 and failure 1 , data is more agreeing with the prior but its since our prior have the mode at 0.8 , since it have the maximum probability in the Neighborhood of 0.8, and 9 success with 1 failure give good evidence for 0.9 and 10 success and 0 failure is giving good evidence for proportion to be 1 , whereas 0.9 is more nearer to our prior belief about the proportion than 1 but still the posterior for 10 success and 0 failure have higher probability,Let us see it analytically

10 Success and 0 Failure

When we have 10 Success and 0 Failure,our posterior will be given by

$$p(\theta|x) \sim \beta_{15,2}$$

Whose mean will be given by $\frac{15}{17}$ and Mode of the posterior is given by $\frac{14}{15}$ and

$$\beta_{15,2}(\frac{15}{17}) \approx 4.89$$
$$\beta_{15,2}(\frac{14}{15}) \approx 6.09$$

9 Success and 1 Failure

When we have 9 Success and 1 Failure , our posterior will be given by

$$p(\theta|x) \sim \beta_{14,3}$$

Mean will be given by $\frac{14}{17} = 0.82$ and Mode by $\frac{13}{15} = 0.86$

$$\beta_{14,3}(0.82) \approx 4.19$$
$$\beta_{14,3}(0.86) \approx 4.64$$

Intuition

Assumptions

1. Every Priors have have influence factor
2. No matter what information data have the posterior always get influenced by priors

Our aim here is to find a metric to measure the influence of priors

Proposition

1. As number of data increases and tends to infinty the influencing power of prior will tend to zero.

   Let us denote our influencing power by $I_{Beta(5,2)}$, then

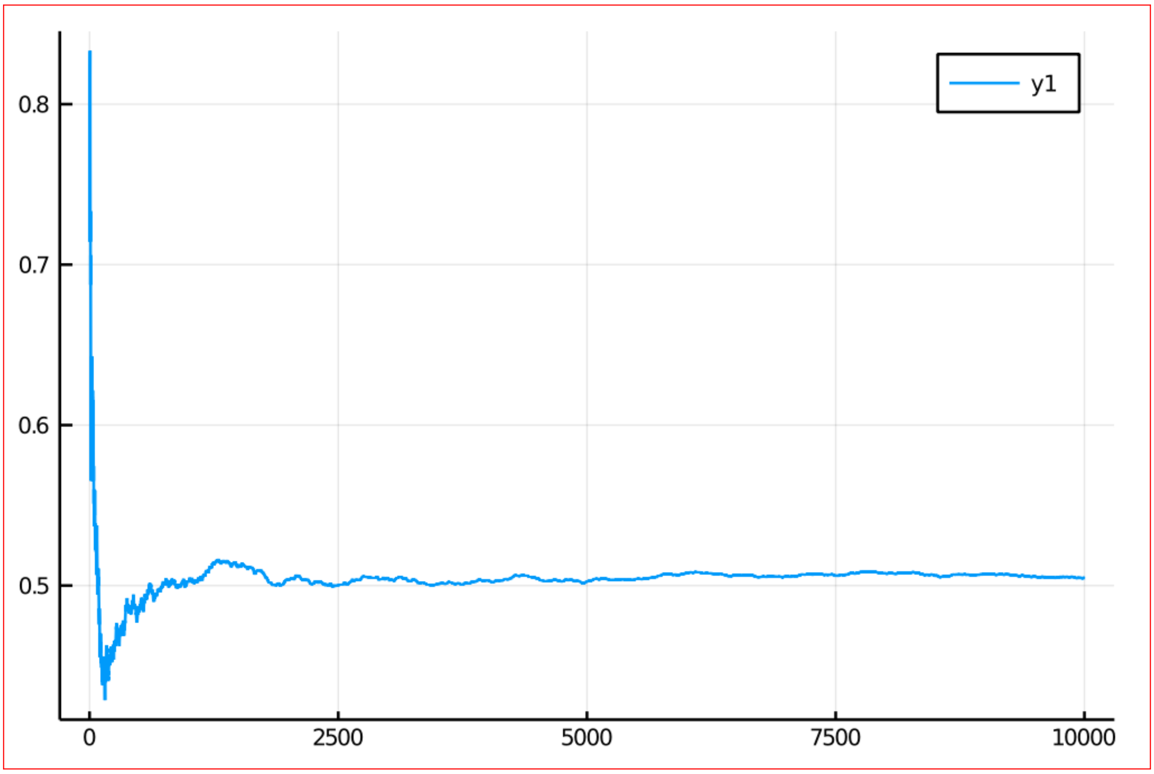
$$I_{Beta(5,2)} \to 0 \quad as \quad n \to \infty$$

   Let we have n observations and out of n observations $\frac{n}{2}$ are success means head then without any prior we will estimate proportion to be 0.5, so as we increase number of observation if our estimate through posterior tend to 0.5 the we can say $I_{\beta(5,2)} \to 0$ so

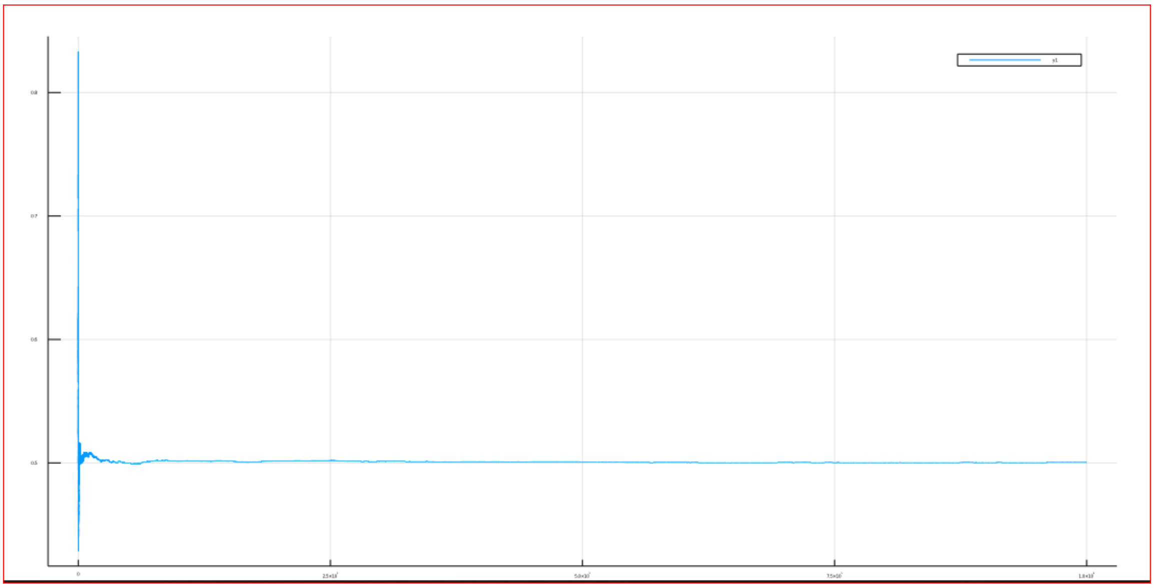
$$\hat{\theta}_B = \frac{5 + \sum x_i}{7 \, + \, n}$$

   Let us simulate the X_i and plot the estimate with the number of samples on x-axis

   For 10000 observations



   For 1000000 observations



Its time to define what I mean by influence

Def : Influence of a prior $p(\theta)$ can be defined $I_{p(\theta)}$ as a change in estimate due to the change in prior

It can be mathematically written as

$$I_{p(\theta)} = \frac{d\hat{\theta}}{d(p(\theta))}$$

Let us take a case where

$$\hat{\theta} = E_{p(\theta|x)}(\theta|x) = \frac{\int \theta L(x;\theta)p(\theta)d\theta}{\int L(x;\theta)p(\theta)d\theta}$$

So,

$$I_{p(\theta)} = \frac{d\hat{\theta}}{d(p(\theta))} = \frac{d\left(\frac{\int \theta L(x;\theta)p(\theta)d\theta}{\int L(x;\theta)p(\theta)d\theta}\right)}{d(p(\theta))}$$

It can also the written as

$$\begin{aligned} I_{p(\theta)} &= \frac{d\left(\frac{\int \theta L(x;\theta)p(\theta)d\theta}{\int L(x;\theta)p(\theta)d\theta}\right)}{d\theta} \Bigg/ \frac{d(p(\theta))}{d(\theta)} \\ \Rightarrow & \frac{\theta L(x;\theta)p(\theta)\int L(x;\theta)p(\theta)d\theta - L(x;\theta)p(\theta)\int \theta L(x;\theta)p(\theta)d\theta}{\left(\int L(x;\theta)p(\theta)d\theta\right)^2} \Bigg/ \frac{d(p(\theta))}{d(\theta)} \\ \Rightarrow & \frac{\theta p(\theta|x) - p(\theta|x)E_{p(\theta|x)}(\theta)}{\frac{d(p(\theta))}{d(\theta)}} \\ \Rightarrow & \frac{p(\theta|x)(\theta - E_{p(\theta|x)}(\theta))}{(p(\theta))'} \end{aligned}$$

If we are right till now , then the above function must give us Influence but we need to put any value for θ to calculate , so we may want to calculate influence from the parameter calculated using only data without priors and one of that estimate will be MLE. Let us see through our above coin tossing experiments , so MLE for the coin tossing experiment will be given by

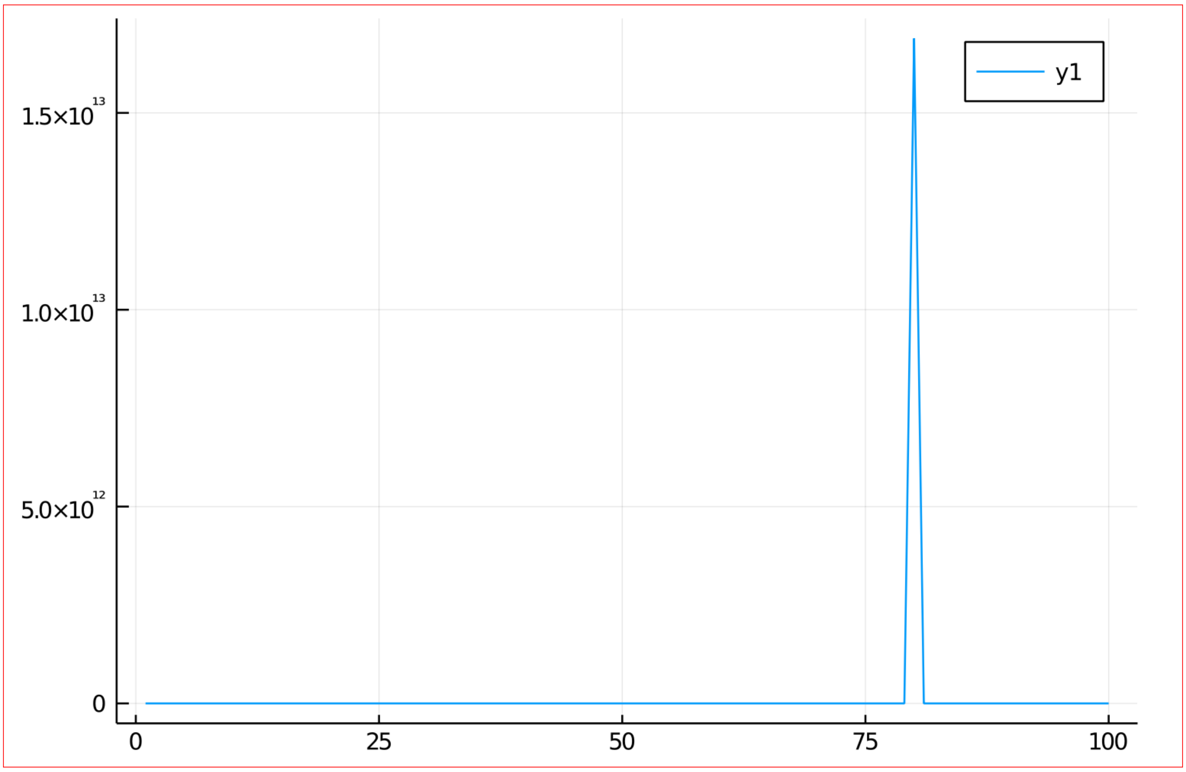
$$\theta_{MLE} = \frac{\sum x_i}{n}$$

So it also agree with our first proposition , when we have a lots of data points influence tends to zero, the term $(\theta - E_{p(\theta|x)}(\theta))$ will tend to zero as our data increases because $E_{p(\theta|x)}(\theta) \to \theta_{MLE}$ as $n \to \infty$. Now let us take a prior $\beta(a, b)$, so then the influence will be given by

$$
\begin{aligned}
I_{\beta(a,b)} &= \frac{\left[\frac{S_n}{n} - \frac{a+S_n}{a+b+n}\right] \frac{1}{\beta(a+S_n,b+n-S_n)} \left(\frac{S_n}{n}\right)^{a+S_n-1} \left(1-\frac{S_n}{n}\right)^{b+n-S_n-1}}{\frac{1}{\beta(a,b)} \left(\frac{S_n}{n}\right)^{a-1} \left(1-\frac{S_n}{n}\right)^{b-1}} \\
&\Rightarrow \frac{\left[\frac{S_n}{n} - \frac{a+S_n}{a+b+n}\right] \frac{\beta(a,b)}{\beta(a+S_n,b+n-S_n)} \left(\frac{S_n}{n}\right)^{S_n+1} \left(1-\frac{S_n}{n}\right)^{n-S_n+1}}{(a-1)(1-\frac{S_n}{n}) - (b-1)\frac{S_n}{n}}
\end{aligned}
$$

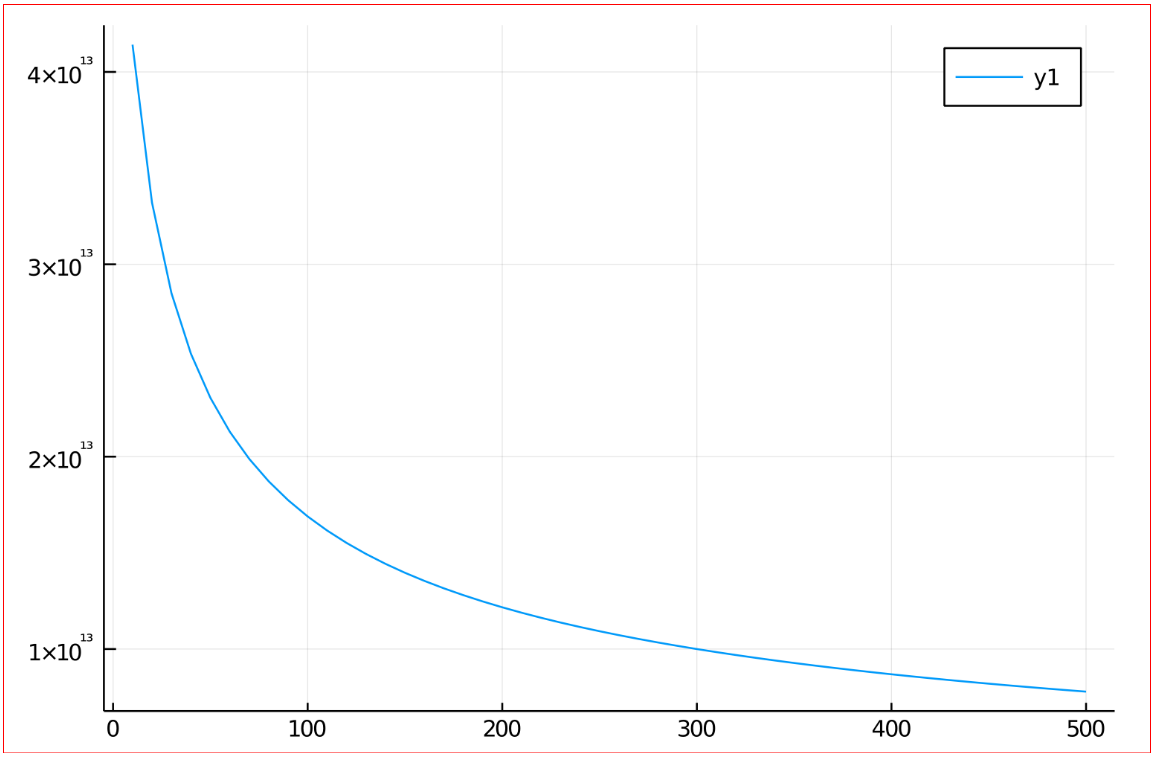
Here $\sum x_i = S_n$

Let us plot for a=5 , b=2 so that the prior have mode at 0.8, Let us plot the modulus of influence with different number of outcomes so let us flip a coin and and plot the influence for different success

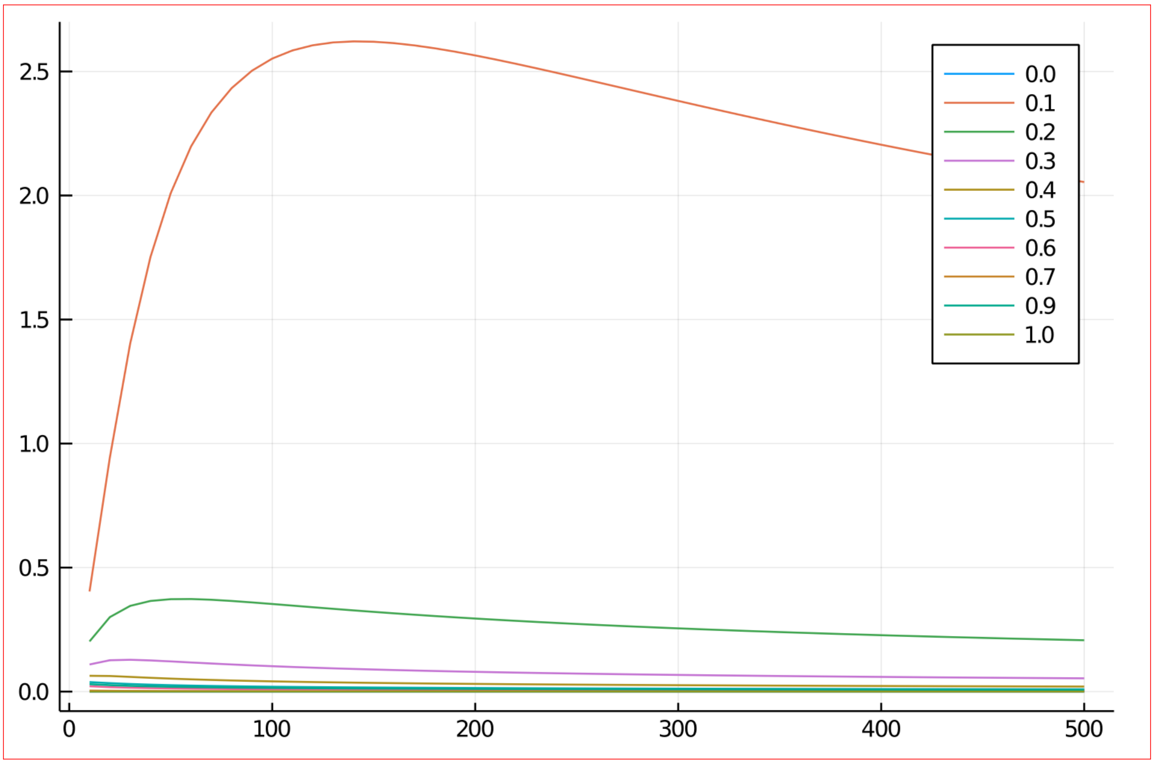


As we can see there is highest influence for 80 success and 20 failure so we can say the prior in influence higher when data agree with prior highest Now let us plot influence for different number coin flipping leaving the θ_{MLE} constant

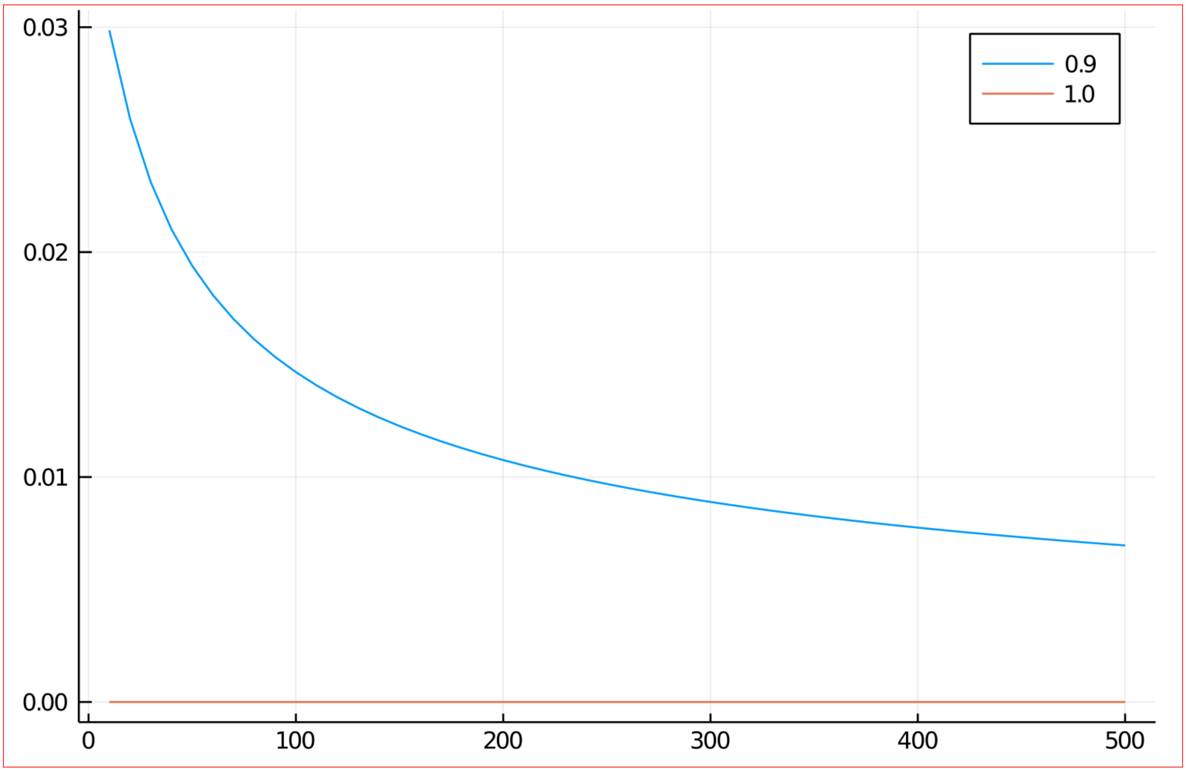
1. for $\theta_{MLE} = 0.8$



2. For different θ



Now let us plot for $\theta = 0.9$ and $\theta = 1.0$



Now we can see influence is greater for 0.9 than 1.0 hence if we toss a coin 10 times and we got 10 times success(head) that means data is suggesting to be proportion equal to 1.0 whereas our prior belief is that it should be 0.8 and the influence on our data is approximately equal to 0 hence the posterior will will be more dependent on data rather than prior.