Alexander Booth
Found Visualize
November 01, 2017

Analyzing conversion rates with Bayes Rule

Purpose:
This quick tutorial will outline how to analyze a success metric in the Bayesian sense.

Scenario:
McMaster-Carr has just launched a new app, ScrewUp. It's a hip little app matching people who love screws with specialty screws in our catalog. Our marketing team has gotten some early publicity by advertising on Google - the result was 794 unique visitors of whom 12 created an account. Doing some division I have computed an empirical conversion rate of 12/794=1.5%.

To begin with, this seems promising. A 1.5% conversion rate isn't great, but it's certainly enough to get started. Investors have suggested that they will probably invest if the conversion rate exceeds 1%.

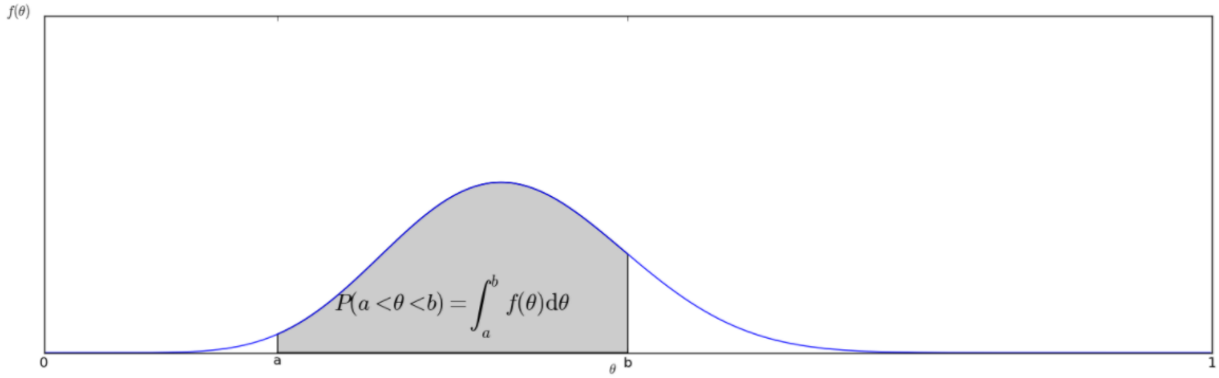
Now, suppose the marketer has the ability to get a lot more publicity. She can expose the ScrewUp app to approximately 10,000 visitors just by talking to some people at Google. Suppose we make the assumption that these 10,000 visitors will convert at the same rate as the 794 early visitors. How many people can I reasonably expect to sign up? This isn't a trick question - the expectation is about 150 signups. But how confident are we that we will really see 150 signups? How confident are we that the conversion rate is higher than 1%?

Bayesian Basics:
The first important concept in Bayesian reasoning is the underlying model. In our case, we take a very simple model. We assume there exists an (unknown) parameter $\theta \in [0,1]$. A unique visitor to the site will create an account with probability θ - i.e., θ is our true conversion rate.

In Bayesian reasoning, the fundamental goal is to compute a posterior distribution on θ. This means we want to find a function f(x) with the property that:

$$P(a < \theta < b) = \int_a^b f(\theta) d\theta$$

In graphical terms, the probability that $a<\theta<b$ can be interpreted as the area under the curve of $f(\theta)$:



Bayesian analysis provides us with an objective method of altering f(x), our estimator of f(θ), based on the evidence we have about it.

Updating our beliefs with Bayes rule:

As you might expect, Bayes rule plays a crucial part in changing our beliefs based on evidence. As a refresher, Bayes rule states that:

$$P(fact|evidence) = \frac{P(evidence|fact)P(fact)}{P(evidence)}$$

To use Bayes rule in our context, we simply need to plug our model into this formula. In our context, the fact we want to compute the probability of the true conversion rate being θ.

Unfortunately, we cannot compute P(fact) directly. P(fact) is our prior distribution, and is purely a subjective choice. It represents our beliefs before we have gathered any evidence. The need to choose a prior is one of the two major sources of subjectivity in Bayesian reasoning - the other source of subjectivity is the underlying choice of model.

As a way to precisely define the prior, a common choice is the beta distribution. The use of Beta distributions in Bayesian inference is due to the fact that they provide a family of conjugate prior probability distributions for binomial (including Bernoulli) and geometric distributions. The domain of the beta distribution can be viewed as a probability, and in fact the beta distribution is often used to describe the distribution of a probability value p

$$f_{\alpha,\beta}(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

The uniform Beta Prior is the beta distribution where alpha=beta=1

More generally, suppose that for any problem of this nature we choose the prior $f_{\alpha,\beta}(\theta)$. Then suppose we gather evidence by running N trials and observe K successes. The posterior, proved by math, is then:

$$\text{posterior} = f_{\alpha+K,\beta+N-K}(\theta)$$

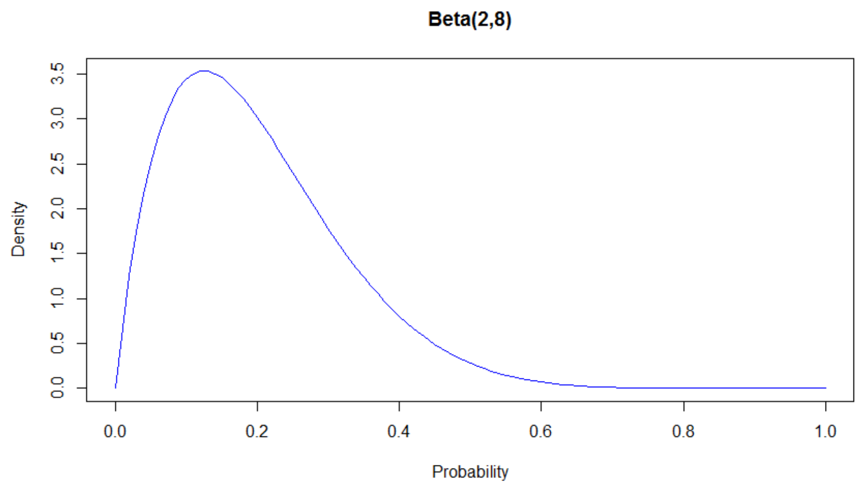
For our app then, P(θ | 12 signups and 794 visitors) = $f_{1+12,\ 1+794-12}(\theta)$. This can be integrated to determine the specific possibility of a conversion rate between two values

For instance, the probability that our true conversion rate is greater than 1% (.01) is 94%. Nice.

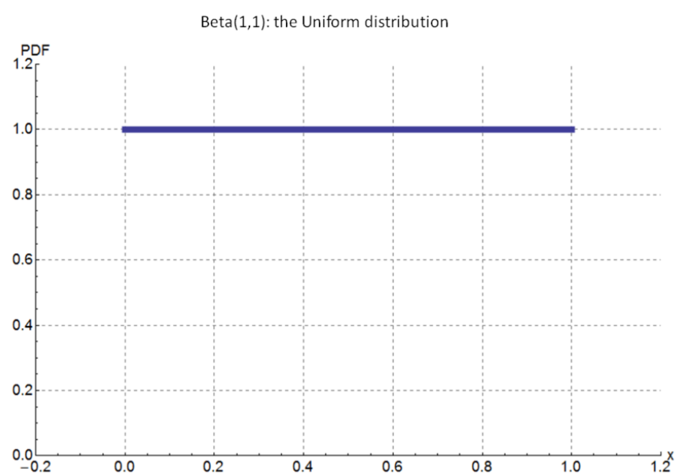
The Beta Distribution
The Beta distribution can be understood as representing a distribution *of probabilities*. It represents all the possible values of a probability when we don't know what that probability is.

It has two parameters:

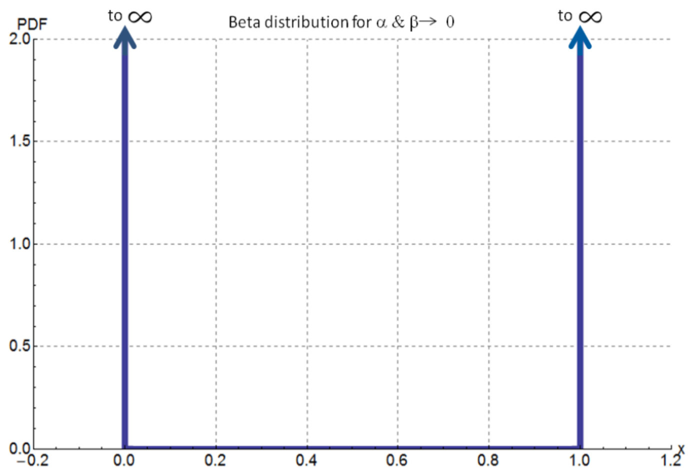
- Alpha = # of successes
- Beta = # of failures



Popular Uniformed Priors
$f_{1,1}(\theta)$, Bayes' Prior. This essentially states that all potential probabilities are equal



$f_{0,0}(\theta)$, Haldane's Prior. This essentially states that it either happens always, or never. This is as uninformed as you can get.



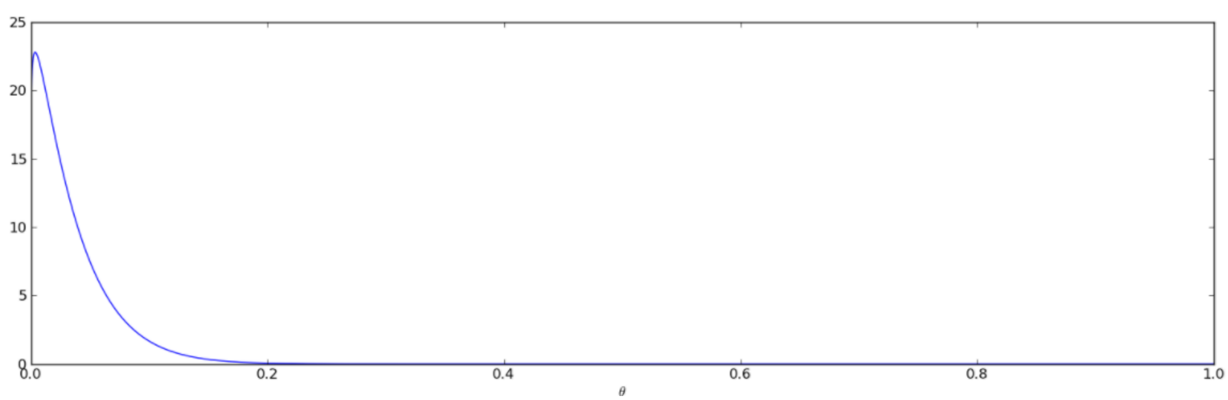
Calculating the Posterior

Confused? Don't be! Here are 4 Easy Steps:

- Step 1: Pick a prior. f(1,1) or f(0,0) are good choices
- Step 2: Observe N trials, with K success and F failures
  - E.g 794 people saw the ad, 12 signed up, 782 didn't
- Step 3: Do simple addition to calculate new alpha or beta. This is optional if you pick the f(0,0) prior
- Step 4. New Alpha = Prior Alpha + K
  New Beta = Prior Beta + F

Using a Prior from Intuition or Business Domain Knowledge.

The other prior we could choose is based on our intuition about website conversion rates, in general, they tend to be low. A conversion rate larger than 10% is extremely unlikely. It is much more likely we made a mistake than we actually observe a 50% conversion rate. In this case, we would want to choose a prior that looks something like this:

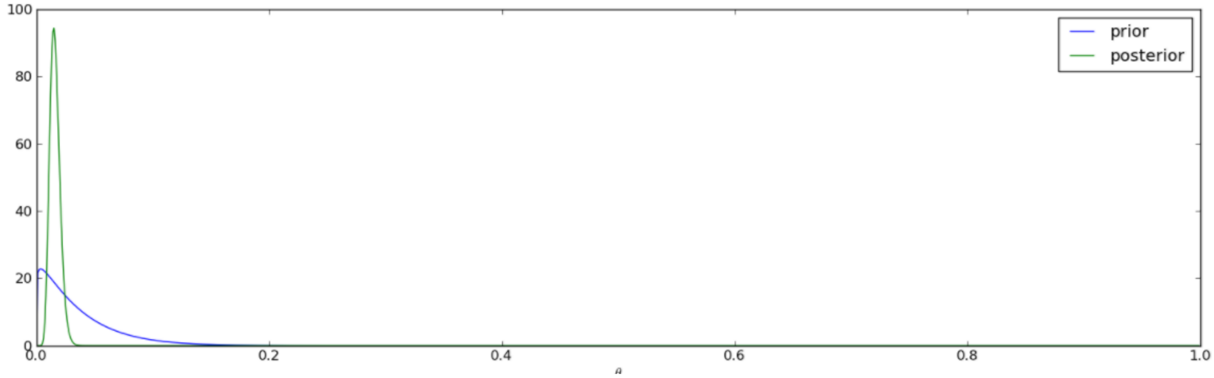


This prior says that although the true value of θ is unknown to us, we are almost certain it is smaller than 0.1. This distribution is the beta function: $f_{1.1,30}(\theta)$.

Using the equation for the posterior calculated above, the posterior for the informative prior is:

P(θ | 12 signups and 794 visitors) = $f_{1.1+12,\ 30+794-12}(\theta)$

Plotting them both:



Integrating gives us:

$$\int_{0.01}^{1} f_{1.1+12,30+794-12}(\theta)d\theta = 0.93127$$

The probability that our true conversion rate is greater than 1% (.01) is 93%. This is less than the previous calculation because this prior emphasizes conversion rates less than .01%. We would need more data for the probability of the true conversion rate to be 94%. Consequently, the inverse is true.

Let us say we only saw 5 conversions out of the original 794 trials, instead of 12. Quick Math:
Posterior Prob < .013 with uninformed prior: 0.944413078014
Posterior Prob < .013 with informed prior: 0.951749074075

See the appendix for the Python code

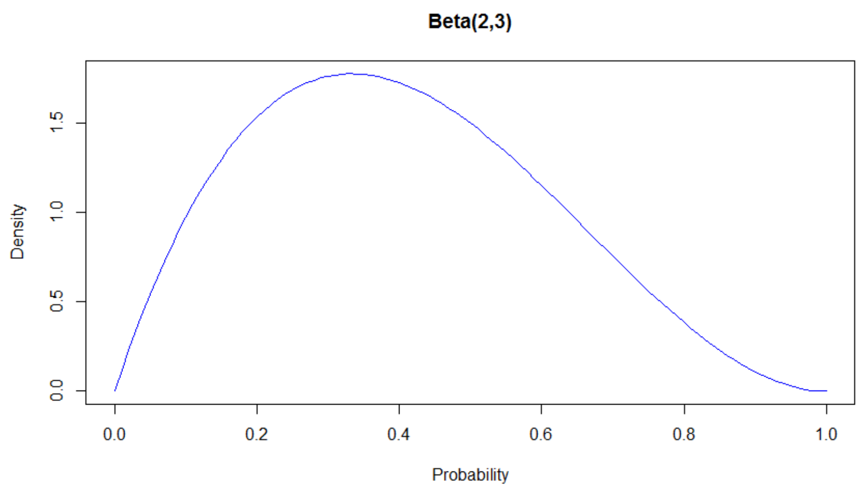
A Deeper Dive into Using a Prior from Intuition or Business Domain Knowledge:

When does this matter most?
An informed prior is most important when dealing with small sample sizes

- The astute among you may have already noticed the Beta(2,8) graph was a lot wider, without a narrow spike when compared to Beta(13,783). Good work.

Examples:
Of the first 5 people to see our ad, 2 created an account! Do we really have a 40% conversion rate though?

Using uninformed prior of Beta(0,0) gives a posterior function = Beta(2+0, 3+0) = **Beta(2,3)**

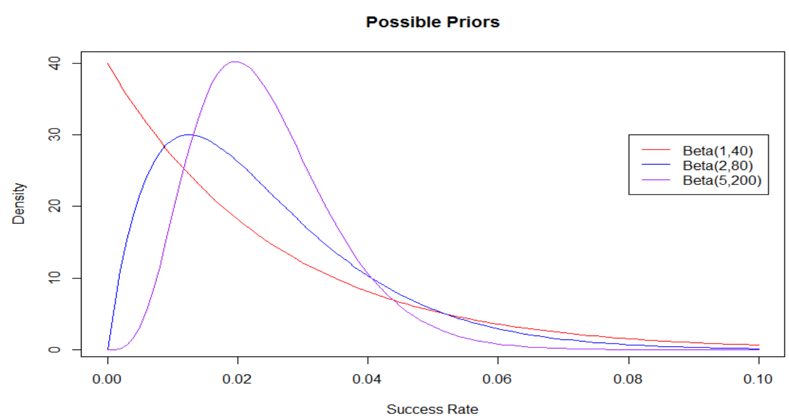


Unlike the last example, where we had a pretty narrow spike, we have this huge range of possible values. As such, our true conversion rate could be pretty much anything between 0.05 and 0.8! This reflects the very little information that we actually have acquired so far.

Given that we have had 2 conversions we know the true rate can't be 0, and we also know that it can't be 1 since we've had 3 people not convert. Everything else is pretty much fair game.

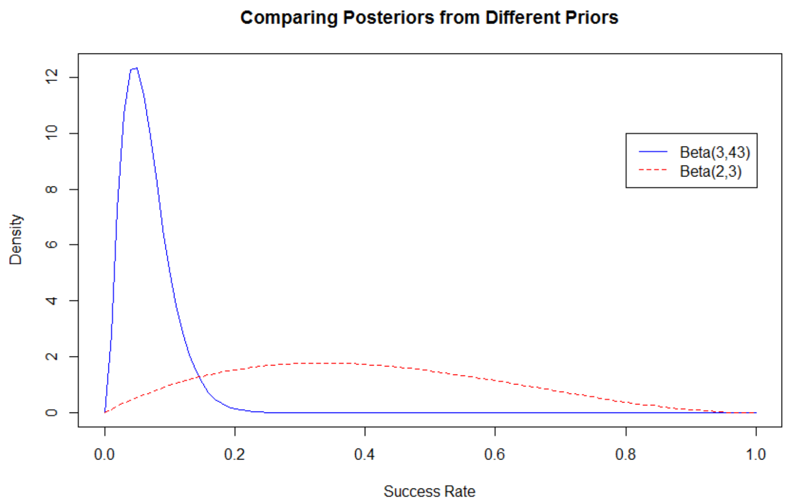
Hold your horses though. We may be new to the app advertising department, but an 80% sign up rate still sounds really high. Doing some quick research, we know that similar ads for mobile apps garner a 2.5% conversion rate. How do we map that to a Beta function?

We want alpha/(alpha+beta) = 2.5%. But which to pick? There are many number combinations to get that. For instance: Beta(1,40), Beta(2, 80), Beta(5,200)



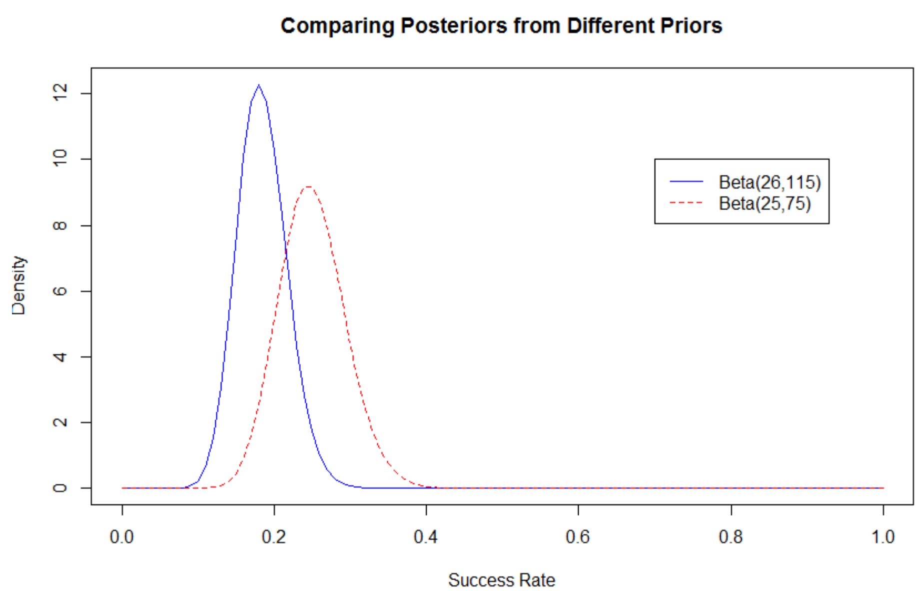
Notice that the lower the combined alpha + beta, the wider our distribution is. Thus, different Beta distributions can be used to represent different strengths of belief in the same data.

Our beliefs aren't too strong, so we'll just take the most pessimistic prior. The world is a cold, dark place after all. This is the prior Beta(1,40) which gives a new posterior of Beta(3,43). Let's compare this to the other posterior of Beta(2,3) calculated from an uninformed prior of Beta(0,0)

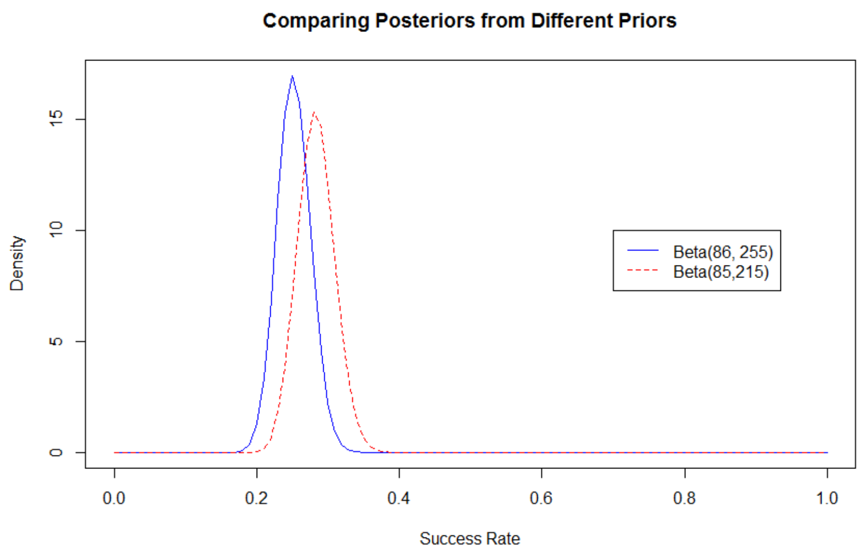


Adding a Prior Probability can drastically improve estimates from small sample sizes! For instance, we can definitely throw out the chance that we have a conversion rate between 40% and 80%. But how can we know for sure? We need to gather more data!

After a few more hours, we notice that 25 of 100 people have created an account after seeing our ad.



The next morning, we notice that 85 of 300 people have created an account after seeing our ad.



What have we learned then?

1) When we had no evidence our Likelihood proposed some things we know are absurd.
2) In light of little evidence, our Prior beliefs squashed any data we had.
3) As we continue to gather data that disagrees with our Prior our Posterior beliefs shift towards what the data tells us and away from what our initial thoughts were

Another important thing to realize is that we started with a pretty pessimistic prior. Even after just a small time period of collecting a relatively small set of information we were able to find a Posterior that seems much, much more reasonable.

The most important thing to realize is that the more data we gather, the more our Prior beliefs become diminished by evidence! Thus, with large amounts of data collected, it can be fine to use an uninformed prior.

Appendix

Sources:

Bayesian Theory for Conversion Rates

https://en.wikipedia.org/wiki/Beta_distribution#Bayes.27_prior_probability_.28Beta.281.2C1.29.29
https://www.chrisstucchio.com/blog/2013/bayesian_analysis_conversion_rates.html
https://www.countbayesie.com/blog/2015/4/4/parameter-estimation-adding-bayesian-priors

Python Code:

```
#imports
from pylab import arange
from scipy.stats import beta

#set integral dx
dx = 0.0001
```

```
#get x vals from .01 to 1
x = arange(0.01,1.0,dx)

#uniformed prior
result = beta(1+12, 1+794-12).pdf(x).sum()*dx
print(result)

#informed prior
result1 = beta(1.1+12, 30+794-12).pdf(x).sum()*dx
print(result1)

#get x vals from 0 to .013
x = arange(0.0,.013,dx)

#uniformed prior
result2 = beta(1+5, 1+794-5).pdf(x).sum()*dx
print(result2)

#informed prior
result3 = beta(1.1+5, 30+794-5).pdf(x).sum()*dx
print(result3)
```

R Code:

```
curve(dbeta(x, 2, 8), col = "blue", xlab = "Probability", ylab = "Density", main= "Beta(2,8)")

curve(dbeta(x, 13, 783), col = "blue", xlab = "Success Rate", ylab = "Density", main= "Posterior = Beta(13,783)")

curve(dbeta(x, 2, 3), col = "blue", xlab = "Probability", ylab = "Density", main= "Beta(2,3)")

#Potential Priors
curve(dbeta(x, 1, 40), col = "red", xlim=c(0,0.10), xlab = "Success Rate", ylab = "Density", main= "Possible Priors")
curve(dbeta(x, 2,80), add = TRUE, col = "Blue")
curve(dbeta(x, 5,200), add = TRUE, col = "Purple")
legend(0.08, 30, legend=c("Beta(1,40)", "Beta(2,80)", "Beta(5,200)"),
      col=c("red", "blue", "purple"), lty=1)


#Final Model
curve(dbeta(x, 86, 255), col = "blue", xlim=c(0,1), xlab = "Success Rate", ylab = "Density",
     main= "Comparing Posteriors from Different Priors", lty=1)
curve(dbeta(x, 85,215), add = TRUE, col = "red", lty=2)
```

```
legend(0.7, 10, legend=c("Beta(86, 255)", "Beta(85,215)"),
    col=c("blue", "red"), lty=1:2)
```