Intro to Bayesian Modeling

Often times when running click experiments with ad content, analysts will want to run smaller sample sizes in order to minimize experimental costs. Other times, the ad hasn't been on the marketplace long enough in order to gather a large number of clicks in order to make interpretations. Is it fair to say that an ad that has been clicked 4 times in 5 impressions has a click rate of 80%? This can create a few problems. When there are thousands of clicks we can use a simple ratio, but for small data we might want to find a better way to think about this.

Binomial Distribution

When I think about ad clicks or video plays in a scrolling feed, I think about millions of coin flip experiments that are being conduted with coins of varying probabilities of heads or tails. For example: A visitor to a site sees an ad. He/she tosses a coin to determine whether he/she will click the ad. If it lands heads, the ad is clicked and if it lands tails the user will pass by. The probability of success inherent in each toss we're assuming is the click rate. Imagine we would like to simulate the number of successes(clicks) in a certain number of trials? We could use rbinom in r to simulate this where:

- **n**: number of trials in the experiment
- **size**: number of trials in each observations. We're only concerned with 1 between click and no click
- **prob**: the probabilty of success

For example, if we would like to simulate the number of clicks we would see in 100 impressions of an ad with a probability (click through rate) of 10% we could write:

R Source

```
rbinom(100, 1, .10)
```

R Output

```
##   [1] 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0
##  [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
##  [71] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0
```

Given a random variable, the number of successes in a series of Bernoulli Trials will follow the Binomial Distribution, which can be written:

$$B(k, n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where k is the number of success, n is the number of trials and p is the probabily of success. As an example, if 100 people saw an ad, with a click through rate of 20%, the probabiliy distribution looks like:

R Source

```
# Plot the results of the trials
df <- data.frame(x = 1:100, Prob = dbinom(1:100, 100, prob= .20))
p1 <- ggplot(data=df, aes(x=x,y=Prob)) +
    theme(axis.title.x = element_blank(),
        axis.title.y = element_text(size = 12, margin = margin(0, 20, 0, 0))) +
    geom_line()
p1
```
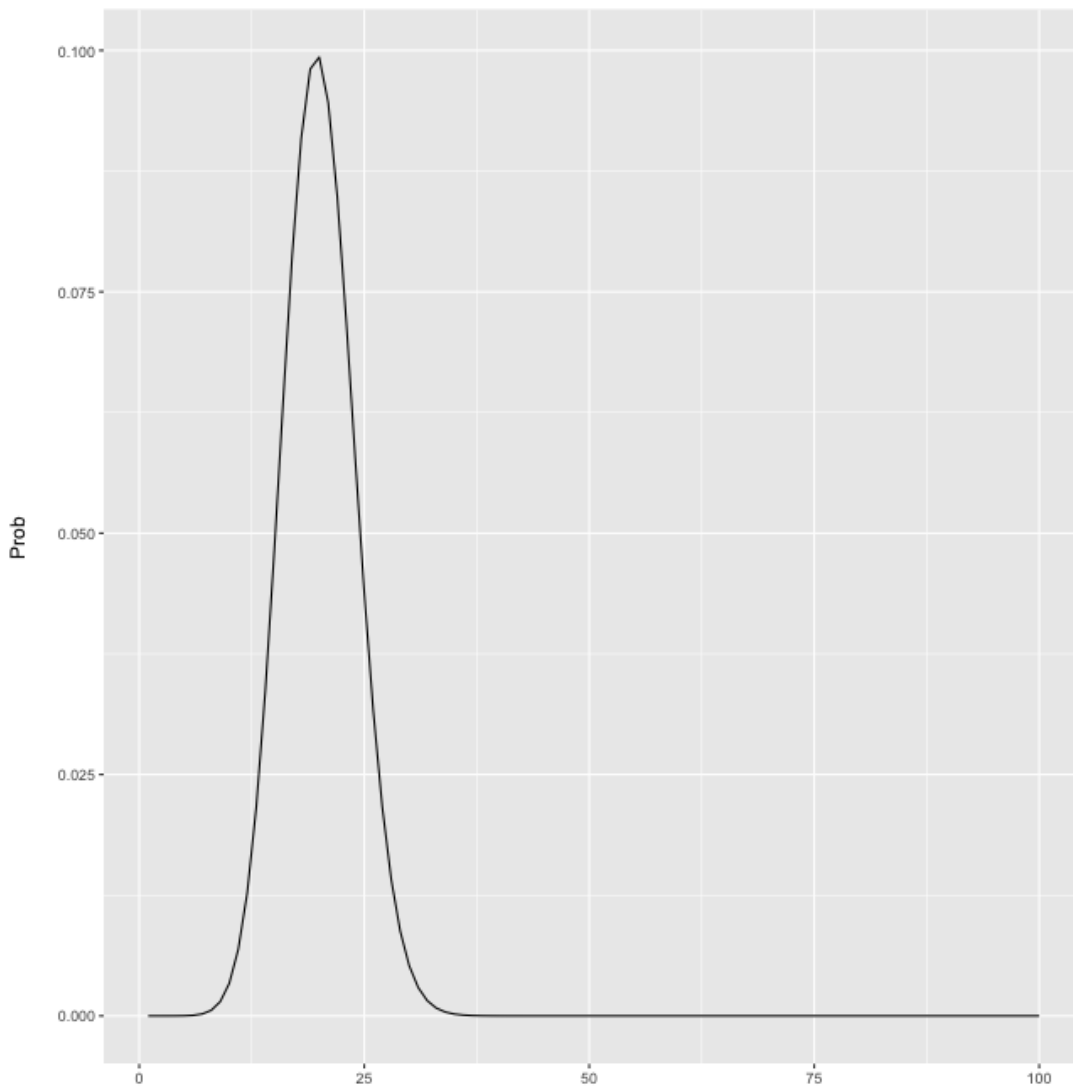
Beta Distribution

The problem that we're confronted with in the above scenario is the choice of a probability of success, which we've defined as the click through rate. To account for the variability in the click through rate we're going to model it from a list of values in a distribution. There are a variety of distributions that could be used to model a click through rate, but a common choice is the beta distribution. The [beta distribution](#) is always positive and continuous on [0,1], which makes it more appropriate than the normal distribution for example. The normal distribution can take negative values, which wouldn't make sense for a click through rate.

The shape of the beta distribution is defined by two parameters, a & b. To define a and b we're going to use a mean and sample size parametrization. Let's assume we have 5 clicks in 20 impressions. We'll set v as the sample size ($v = 20$) and let μ be the CTR, $\mu = (5/20) = 0.25$. After some simplification of the [underlying formulas,](#) the result is: $\alpha = \mu v, \beta = (1 - \mu)v$.

Imputing the values leaves us with: $\alpha = 0.25 * 20 = 5, \beta = (1 - 0.25)*20 = 15$

The derived distribution for the click through rate can be found below.
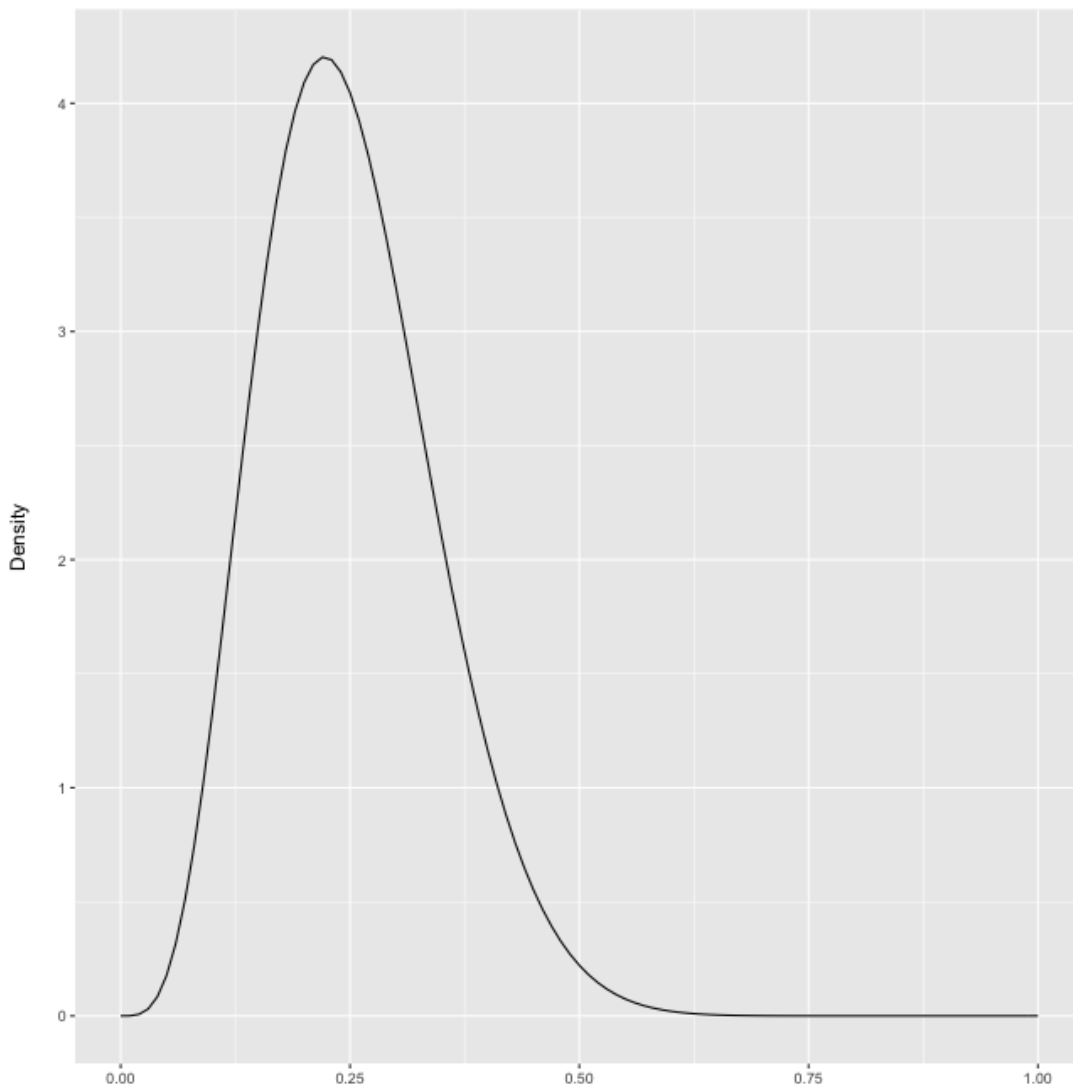
Bayesian Modeling

Oftentimes you may have multiple ads running at the same time: as time passes some ads will have many views, while others will have fewer. We can use Bayesian inference in order create a new click through rate given the prior observation of the click through rate. From the mean and sample size model above, we can derive an estimate of the prior probabilty, the click through rate. Next, we need a liklihood function, the liklihood of a certain number clicks given the click through rate. This function is given by the binomial distribution. The relationship between the beta prior and a binomial distribution allows us to use a conjugate prior, which provides a solution to the posterior probability. There is a great derivation from [Carnegie Mellon](#) that you can review here on beta bernoulli/binomial models that derives the posterior probability. The posterior probability is given by: $\beta(a + x, b + N - x)$ To see this as an example in r:

```
R Source
# Median Approximation of beta distribution
MedianBeta <- function(a,b) {
  (a - (1/3)) / (a + b -(2/3))
}

# Definition of a,b parametrization
clicks <- 2
impressions <- 10
mu <- clicks/impressions
a <- mu * impressions
b <- (1 - mu) * impressions

# Set PriorA and PriorB to the median of the Beta Distribution
priorA = MedianBeta(a,b)
priorB = MedianBeta(a,b)

# Plot the results of the posterior probability
plot(seq(0, 1,0.01), dbeta(seq(0,1,0.01), priorA + clicks, priorB + impressions - clicks), type="l", ylab="Density")
```
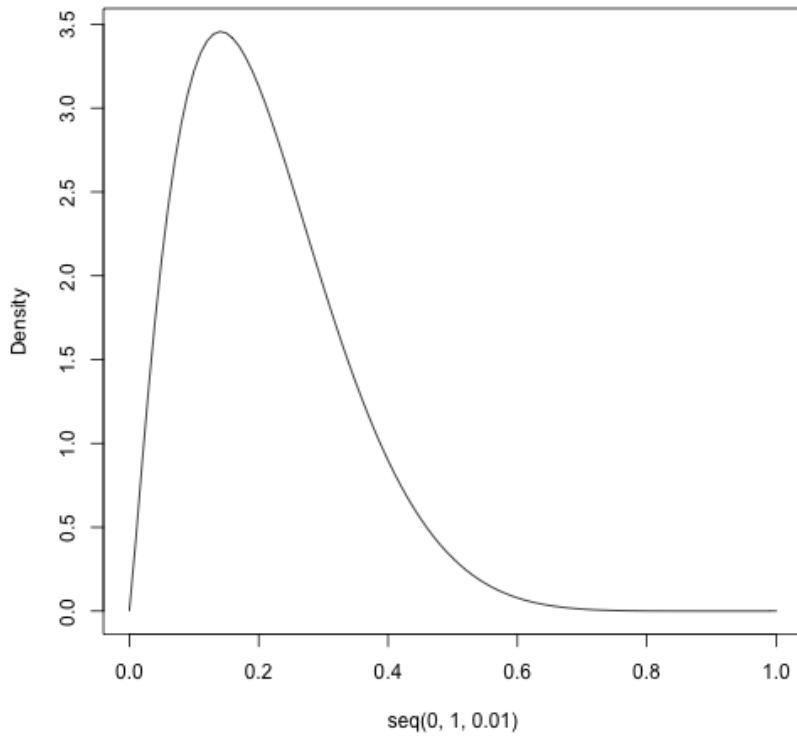
The code above uses an estimate of the median of the beta distribution as the prior probability, the click through the rate. This is used in conjunction with the definition of the posterior probability to calculate a distribution of the click through the rate.

We can use this distribution in tamdem with information about costs or revenue to derive information about the performance of an ad. For example how much might we spend on a given ad after 10,000 impressions, if the cost per click is $0.35:
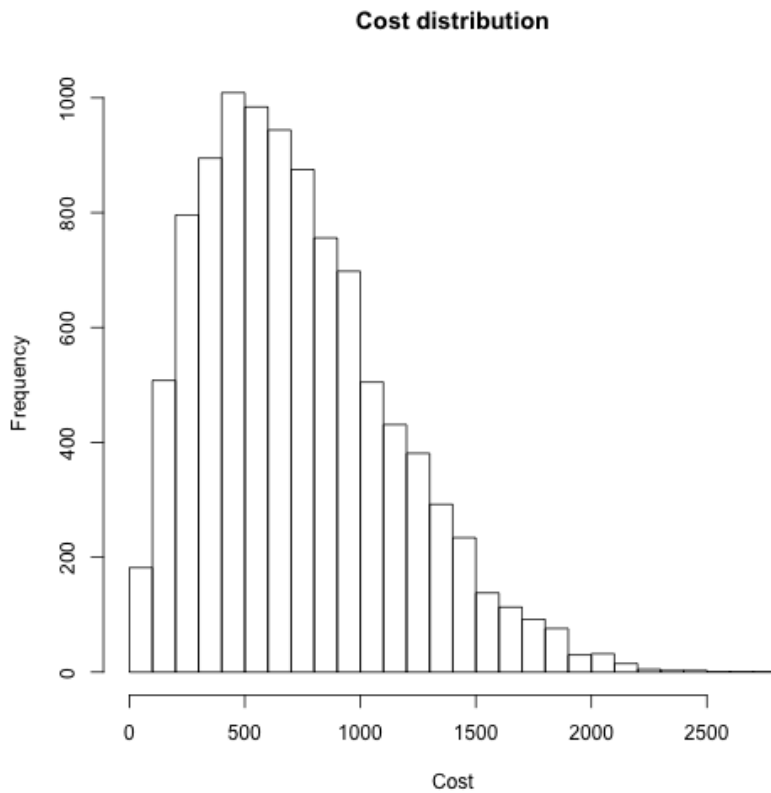
R Source

```
cpc <- 0.35

clicks <- 2
impressions <- 10

hist(rbeta(10000, priorA + clicks, priorB + impressions - clicks) * 10000 * cpc, 20, main="Cost distribution", xlab="Cost")
```

Cost distribution

This was just a sample of the types of statistical modeling that can be done using click data. There are many more applications of bayesian statistics to experimentations that I'll take some time to go over in the next few projects.

- **Intro to Bayesian Modeling**

- Toggle
- Code
    - Languages
    - R
    - All
- Output
    - Type
    - Output
    - All
- Figures

Styled with knitrBootstrap