Tech Talk: Bayesian Analysis

Overview on the next coolest thing in town

Kenneth Lim

04 May, 2017

Contents

Why Bayesian Analysis?

Reason

1. Provides a framework for continuous learning.
2. Enable us to quantify uncertainty.
3. Provides a different paradigm for solving problems, more creative.

Prior Probability

Likelihood of the evidence 'E' if the Hypothesis 'H' is true

$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

Posterior Probability of 'H' given the evidence

Priori probability that the evidence itself is true

Simple AB Testing

Background

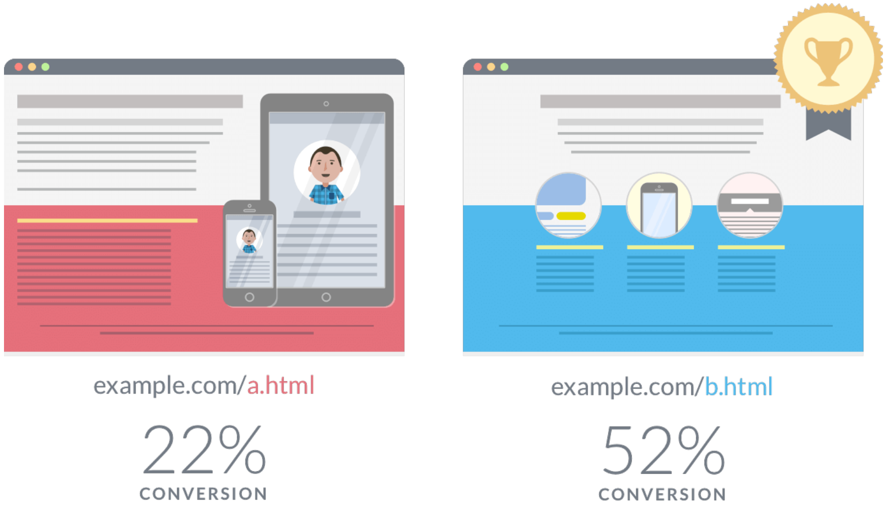
Given 2 versions of web pages, A and B.

Page A has 500 visits and 110 click-through.

Page B has 1000 visits and 520 click-through.

We would like to know:

- Which version works better for conversion
- How much uplift
- How certain are we about the analysis



Simple AB Testing

Frequentist Way

Establish the null hypothesis

Version A and version B have the same conversion rate.

H_0: P_A = P_B

Establish the alternative hypothesis

Version B's conversion rate is not equal to Version A.

H_1: P_A != P_B

Perform t-test and look at p-value

Reject null hypothesis if p-value is < 0.05

0.0001 < 0.05

p-value **Significance Level**

Probability that we would observe a value more extreme by chance is 0.0001.

Hence, reject the null hypothesis, we prefer Version B to Version A.

Simple AB Testing

Bayesian Way

<u>Establish the prior belief</u>

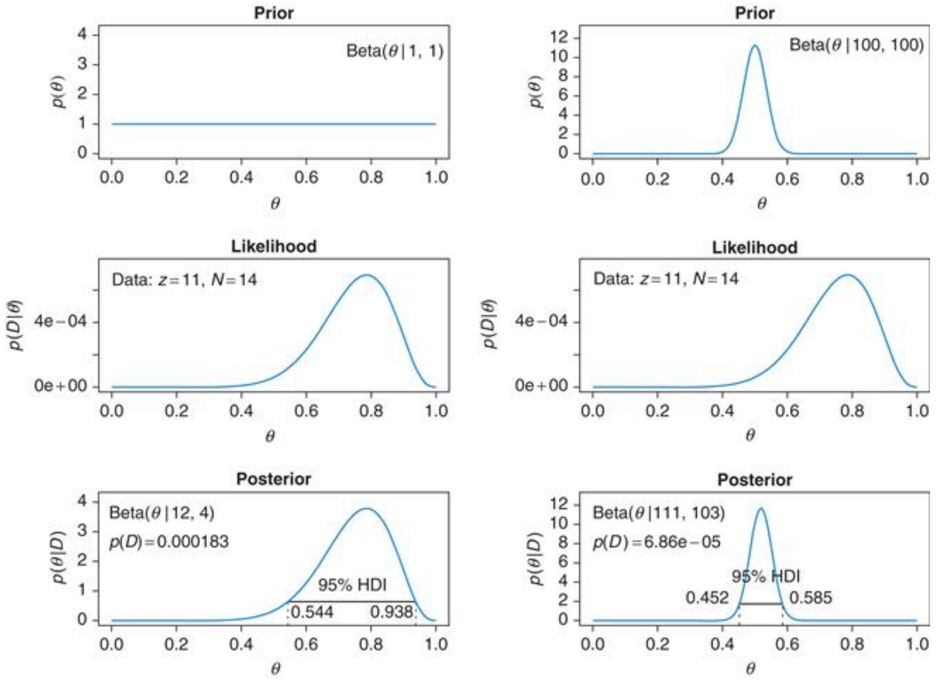
We have some weak belief that both Version A and Version B have uniform Beta(1,1).

<u>Update our belief with data</u>

Use Bayes rule to update our prior beliefs to get the posterior belief.

<u>Calculate the High Density Interval (HDI)</u>

Calculate the HDI which gives 95% range of conversion rate.



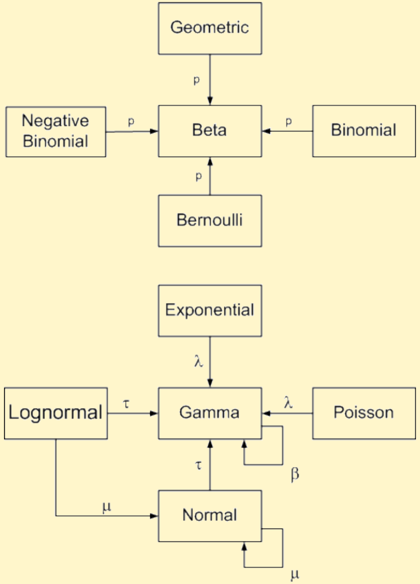
Simple AB Testing

	Frequentist Way	Bayesian Way
Which version works better for conversion?	Reject the null hypothesis that conversion rate of A and B are equal.	We are 96% confident that B is better than A.
How certain are we?	Probability of observing the data as extreme as the p-value is very unlikely.	95% probability that the conversion rate for A and B lies between 19%-26%, 49%-55% respectively.
How much uplift?	We don't really know	95% probability that the conversion rate is 25%-35% higher for B.

Methods to Perform Bayesian Updates

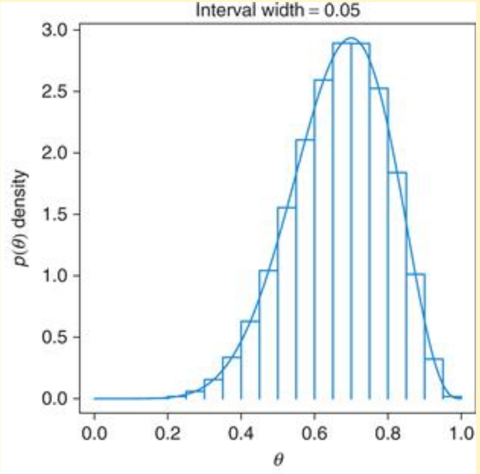
Analytical

Using conjugate priors. Mathematically tractable.



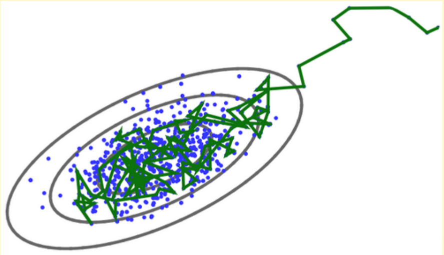
Approximation

Approximate posterior using grid method. More freedom in specifying prior distribution.



MCMC

Markov Chain Monte Carlo. Metropolis-Hastings algorithm draws samples from the posterior distribution. Samples are only from those high probability regions.

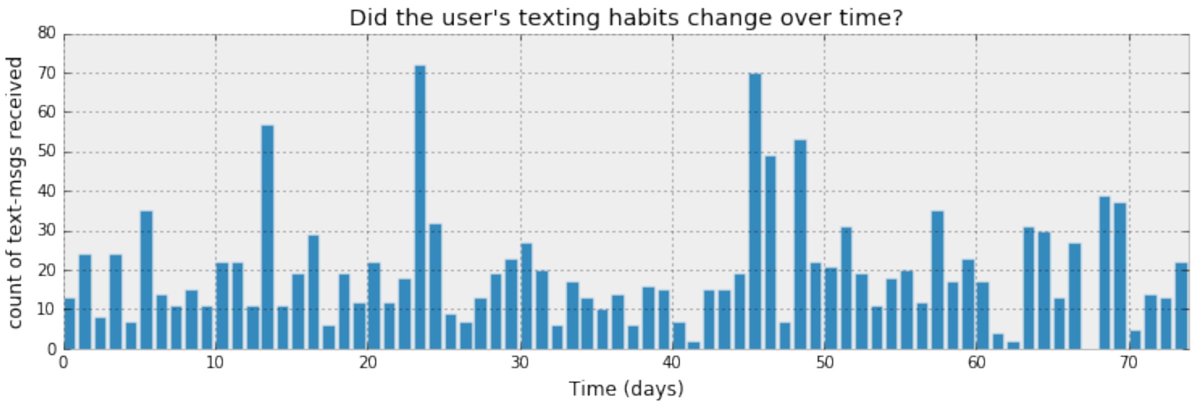


Hierarchical Models

Background

We have some daily text-message counts from a user of our system. We suspect that there's some change in user's usage rate at some time. We would like to know:

- Did user text message habit change?
- When did it change?



Hierarchical Models

The Model

Priors

$$\tau \sim \text{DiscreteUniform}(1,70)$$

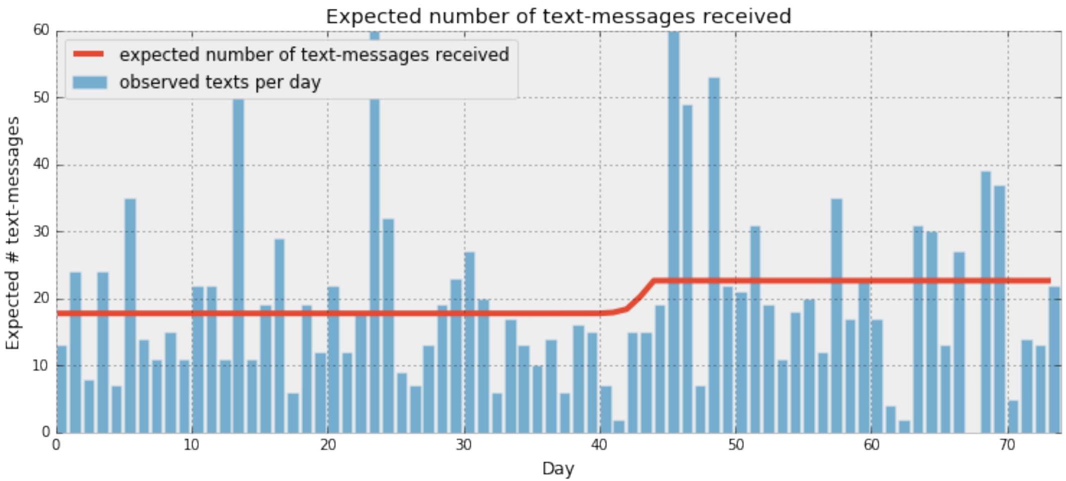
$$\Rightarrow P(\tau = k) = \frac{1}{70}$$

$$\lambda_1 \sim \text{Exp}(\alpha)$$
$$\lambda_2 \sim \text{Exp}(\alpha)$$

$$\lambda = \begin{cases} \lambda_1 & \text{if } t < \tau \\ \lambda_2 & \text{if } t \geq \tau \end{cases}$$

Likelihood

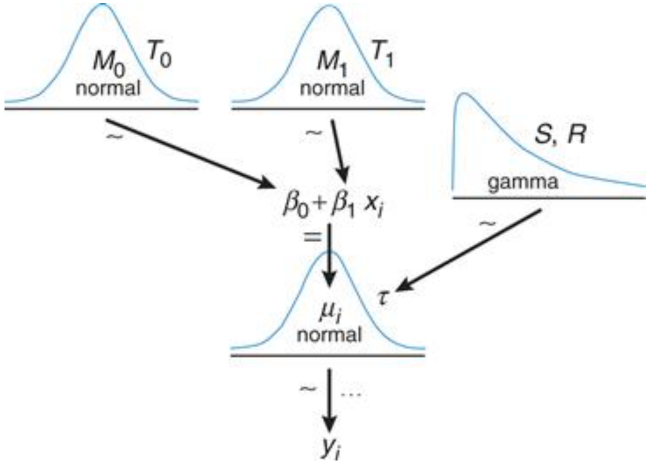
$$C_i \sim \text{Poi}(\lambda)$$

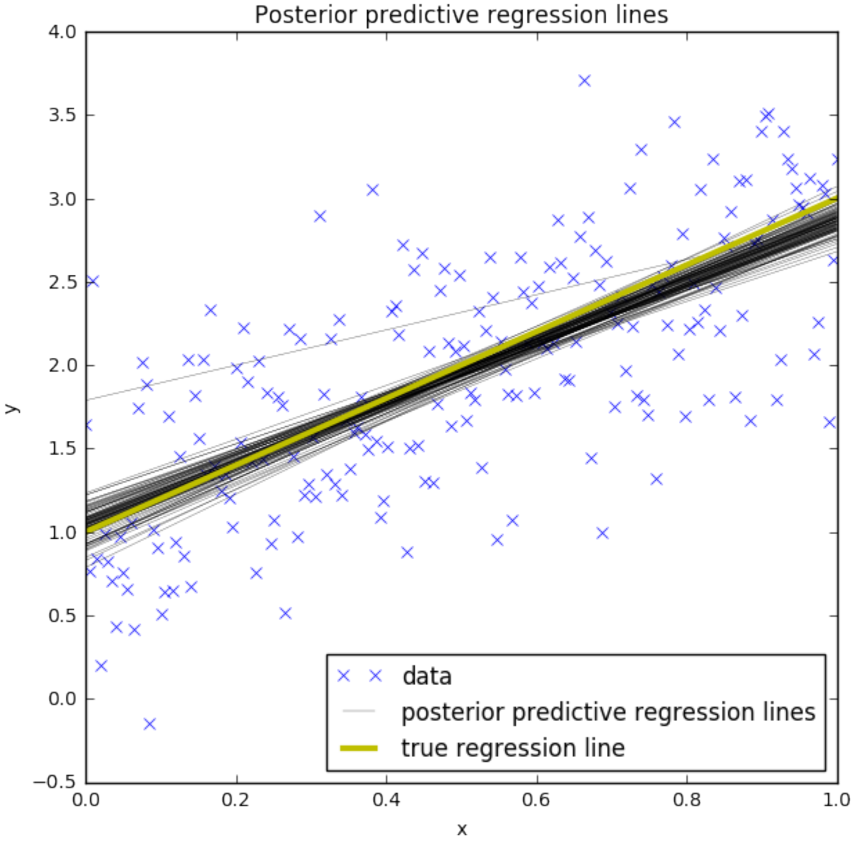


Bayesian Machine Learning

Linear Regression

Estimate coefficients with MCMC. MCMC allows to sample for multiple regression lines to estimate uncertainly of regression line.

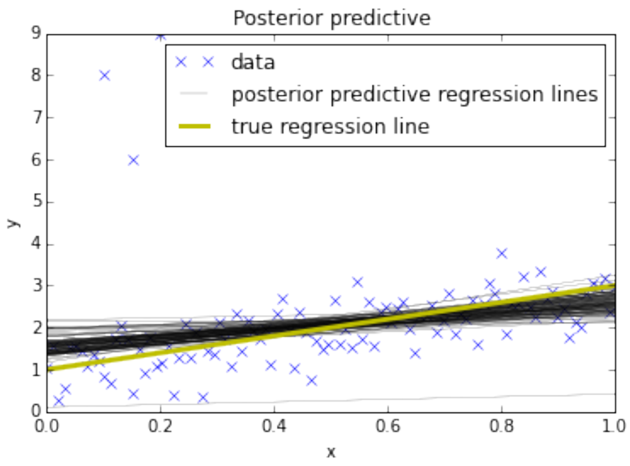


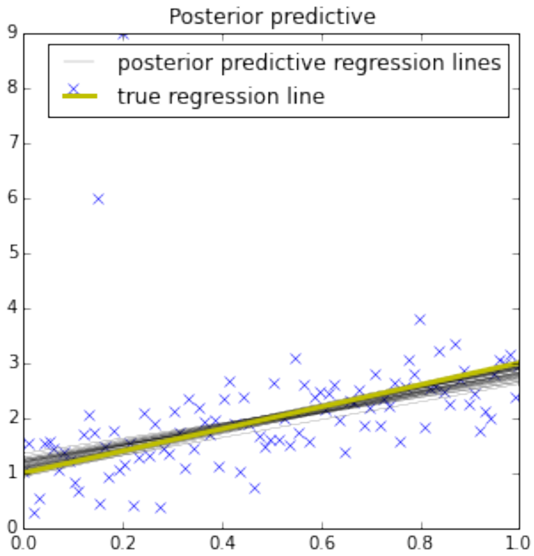


Bayesian Machine Learning

Robust Linear Regression

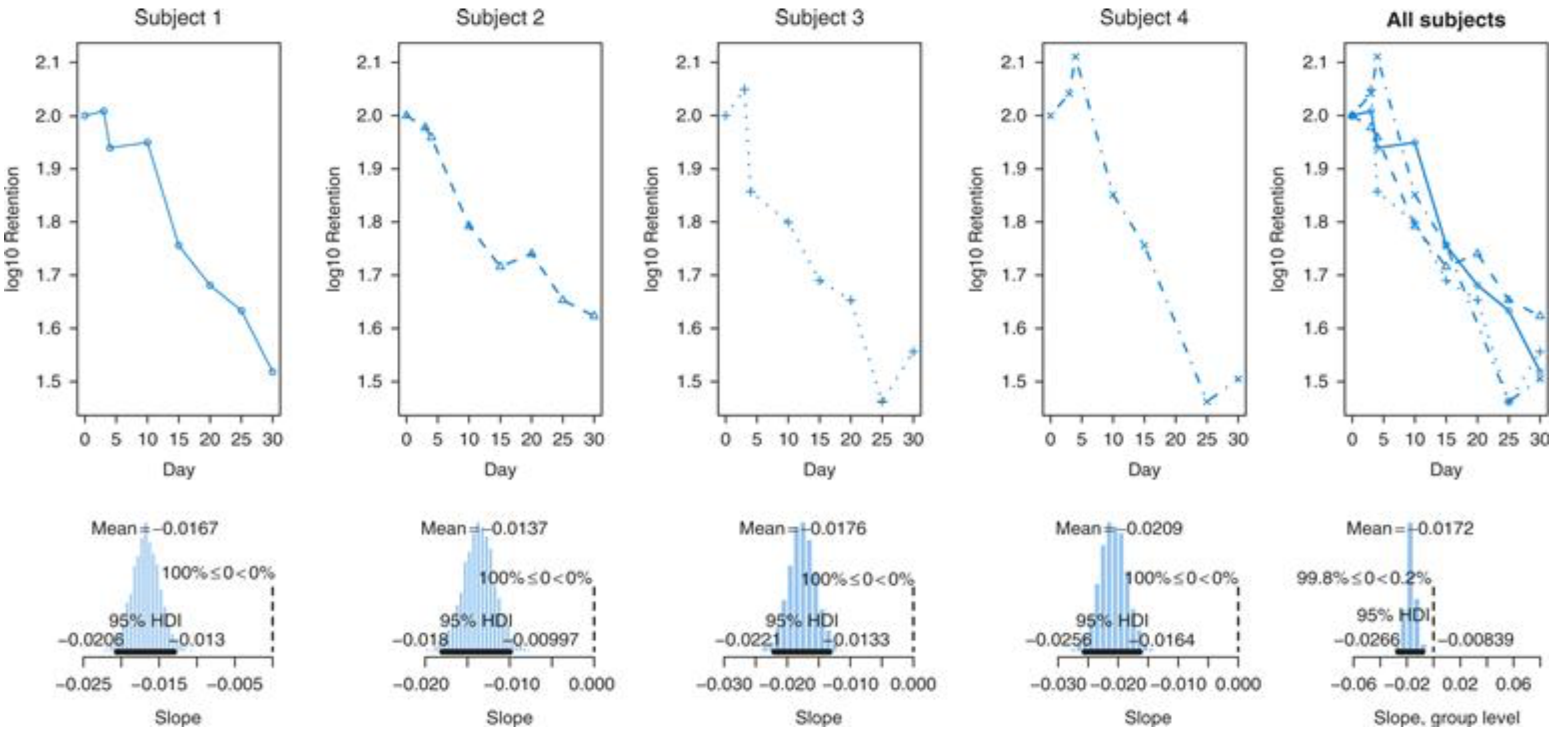
Use Student-t distribution instead of normal Gaussian distribution. Having more mass at tails allows regression line not be heavily influenced by outliers.





Bayesian Machine Learning

Hierarchical Linear Regression



Why Bayesian Analysis?

Reason

1. Provides a framework for continuous learning.
   - Beta shape parameters can be continuously updated
2. Enable us to quantify uncertainty.
   - Credible intervals, 95% High Density Intervals
3. Provides a different paradigm for solving problems, more creative
   - Hierarchical Modeling

Thank You.