

Bayesian A/B Testing

Conjugate Priors

Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- Likelihood
- Posterior
- Conjugate Prior

Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- Likelihood
- Posterior
- Conjugate Prior

Frequentist - Hypothesis Testing

Define a Metric	Declare null and alternative hypothesis.
Set Parameters	Significance, number of observations, etc.
Run Experiment	Make sure you follow it to a tee.
Compute Test Statistic	Make sure it's the appropriate one.
Calculate P-Value	
Draw Conclusions	Reject H_0 in favor of H_A , or fail to reject.

Frequentist A/B Testing - Limitations?

- If one of the pages your testing appears to be obviously better, can you scrap the experiment and declare it the winner?
- At the end of an experiments, can you quantify how much better the winning pages is than the losing page?

Example

"It's 95% likely that page A is better than page B."

- If, after you've begin your test, your boss comes to you with another version of the page and asks you to test it it too, can you update the test?

Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- Likelihood
- Posterior
- Conjugate Prior

Bayesian - Overview

Define a Metric

To quantify what "better" means.

Run Test

Collect data.

Continually Monitor
Results

Can decide to stop at any time.

Suggest Next Step

Based on probabilities
calculated.

Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- Likelihood
- Posterior
- Conjugate Prior

Bayes' Theorem

$$P(\theta|y) = \frac{P(y|\theta) \times P(\theta)}{P(y)}$$

$P(y|\theta)$ Likelihood

$P(\theta)$ Prior

$P(y)$ Normalizing
Constant

Remember: $P(y) = \int P(y, \theta) d\theta$

This equation captures the essence of Bayesian modeling: that the uncertainty about some unknown parameter can be quantified.

Bayes' Theorem Distilled

All that mathiness can be boiled down to the straightforward idea that:

$$Posterior \propto Likelihood \times Prior$$

Back to A/B Testing

How, then, do we apply this supposedly straightforward idea of Bayes' Theorem to our statistics based question of A/B testing?

Let's first consider what we are really trying to do in the process of A/B testing.

Trying to determine which of options A & B is "better".

For example, trying to determine which page layout has a higher click-through rate.

Moving Towards Bayes

Considering that distilled version of Bayes' Theorem we saw earlier:

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior}$$

Let's attach some tangible ideas to these abstract notions of a "likelihood" and a "posterior" for the click-through rate example. For the next two slides let's consider the data associated with each of our pages separately.

Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- **Likelihood**
- Posterior
- Conjugate Prior

Likelihood

What are the possible outcomes for each observation in our data for a single page?

A click-through occurs or not, binary result.

How, then, are our data distributed; aka, what is the **likelihood**?

Likelihood

What are the possible outcomes for each observation in our data for a single page?

A click-through occurs or not, binary result.

How, then, are our data distributed; aka, what is the **likelihood**?

Binomial

$$P(X=k) = \binom{n}{k} p^k \times (1-p)^{n-k}$$

→ k successes in n trials.

Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- Likelihood
- **Posterior**
- Conjugate Prior

Posterior

Again, with respect to the data from a single page, what are the possible values that a click-through rate can take on?

Any real value in the open interval $[0, 1]$.

What distribution, then, should we use to model this click-through rate parameter; aka, what is the **posterior**?

Posterior

Again, with respect to the data from a single page, what are the possible values that a click-through rate can take on?

Any real value in the open interval $[0, 1]$.

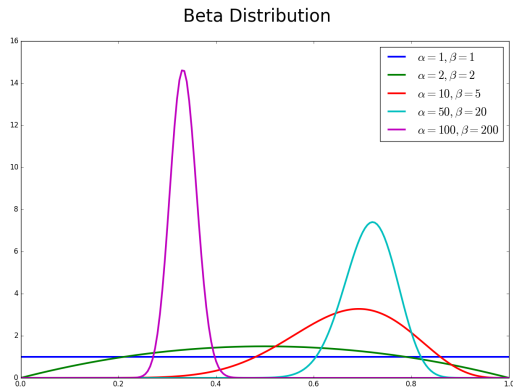
What distribution, then, should we use to model this click-through rate parameter; aka, what is the **posterior**?

Beta! Has support over the continuous interval $[0, 1]$.

Beta

$$X \sim Be(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

Beta Distribution



$$E[X] = \frac{\alpha}{\alpha + \beta}$$

$$Mode = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Piecing it Together

Recalling the form of Bayes' Theorem:

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior};$$

we now know what the likelihood and the posterior look like:

$$\text{Likelihood} \quad \binom{n}{k} p^k \times (1 - p)^{n-k}$$

$$\text{Posterior} \quad \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1 - p)^{\beta-1}$$

What do you notice about these two distributions?

Piecing it Together

Recalling the form of Bayes' Theorem:

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior};$$

we now know what the likelihood and the posterior look like:

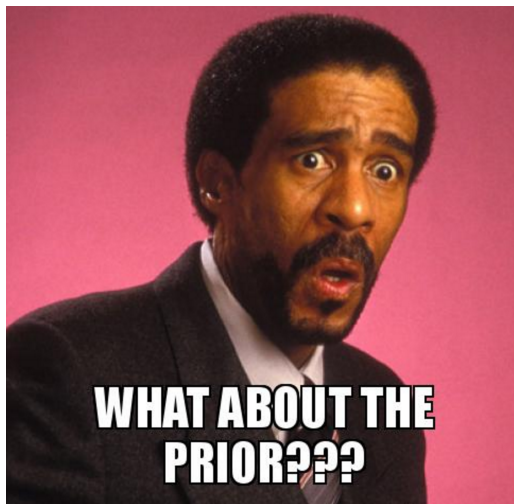
$$\text{Likelihood} \quad \binom{n}{k} p^k \times (1 - p)^{n-k}$$

$$\text{Posterior} \quad \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1 - p)^{\beta-1}$$

What do you notice about these two distributions?

They have the same form!

Interlude



Overview

A/B Testing

- Frequentist - Review
- Bayesian - Overview

Bayesian A/B Testing

- Bayes' Theorem
- Likelihood
- Posterior
- Conjugate Prior

Priors

Before we reason about what our prior, $P(\theta)$, should look like, let's discuss what a prior actually is.

Prior

The prior is a distribution that encodes our beliefs about the possible values that the parameter in question, θ , can take on before we collect any data.

Choosing a Prior

How do we choose our prior, then, since technically, the only requirement is that it is a distribution, aka $\int P(\theta) = 1$?

More specific to our current situation though, we might ask the question, what distribution would we like to use to encode our prior beliefs about θ ?

To answer this question let's consider the functional form of the likelihood and posterior that we've visited already:

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior}$$

$$\textit{Beta} \propto \textit{Binomial} \times \textit{Prior}$$

$$p^{\alpha-1}(1-p)^{\beta-1} \propto p^k(1-p)^{n-k} \times \textit{Prior}$$

Conjugate Prior

In this case, the obvious distribution we should choose to encode our prior beliefs about $P(\theta)$ is the **beta**. Let's look at why this makes sense.

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior}$$

$$\textit{Posterior} \propto \textit{Binomial} \times \textit{Beta}$$

$$\textit{Posterior} \propto p^k (1-p)^{n-k} \times p^{\alpha-1} (1-p)^{\beta-1}$$

$$\textit{Posterior} \propto p^{k+\alpha-1} (1-p)^{n-k+\beta-1}$$

Beta Conjugate Prior

This is the beta distribution! Just as we wanted our posterior to be.

$$p^{k+\alpha-1}(1-p)^{n-k+\beta-1}$$

What are the parameters of this distribution?

Beta Conjugate Prior

This is the beta distribution! Just as we wanted our posterior to be.

$$p^{k+\alpha-1}(1-p)^{n-k+\beta-1}$$

What are the parameters of this distribution?

$$\alpha_1 = k + \alpha$$

$$\beta_1 = n - k + \beta$$

A.k.a.

$$\textit{Posterior} \sim \textit{Be}(k + \alpha, n - k + \beta)$$

Bayesian A/B Testing in Practice

$$\textit{Posterior} \sim \text{Be}(k + \alpha, n - k + \beta)$$

How do we use this new-found tool?

All we have to do is take the number of conversions, k , and the total number of views, n , for each of our pages and plug them into our posterior distribution.

Bayesian A/B Testing in Practice

$$\textit{Posterior} \sim \textit{Be}(k + \alpha, n - k + \beta)$$

How do we use this new-found tool?

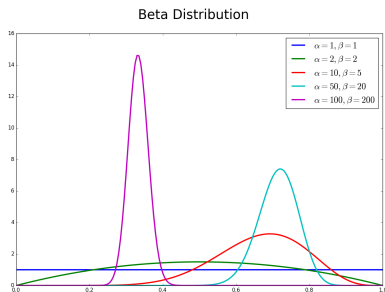
All we have to do is take the number of conversions, k , and the total number of views, n , for each of our pages and plug them into our posterior distribution.

What about α and β ? What are the values of those parameters?

Back to the Prior

Those values α and β are the parameters for the beta distribution that we cleverly chose to describe our prior beliefs.

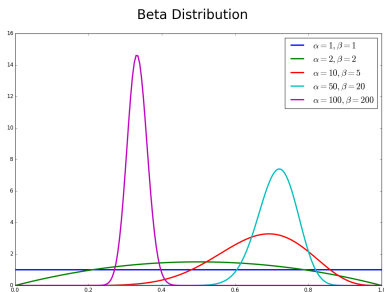
The values of α and β literally encode our prior beliefs about the values that θ could possibly take on. With this in mind, there are a couple of ways that we can choose these values:



Back to the Prior

Those values α and β are the parameters for the beta distribution that we cleverly chose to describe our prior beliefs.

The values of α and β literally encode our prior beliefs about the values that θ could possibly take on. With this in mind, there are a couple of ways that we can choose these values:

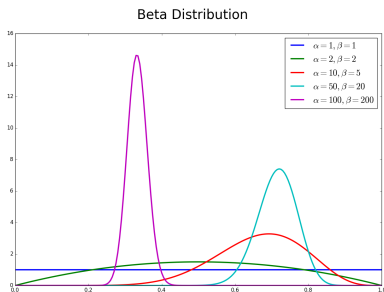


- Choose an uninformative prior
→ $\alpha = 1$ and $\beta = 1$.

Back to the Prior

Those values α and β are the parameters for the beta distribution that we cleverly chose to describe our prior beliefs.

The values of α and β literally encode our prior beliefs about the values that θ could possibly take on. With this in mind, there are a couple of ways that we can choose these values:



- Choose an uninformative prior
→ $\alpha = 1$ and $\beta = 1$.
- Act like you have observed some data before that represent data your experiment has to overcome.

Bayesian A/B Testing in Code

How do we actually perform a test with code, then? By simulation.

```
1  from numpy.random import beta
2
3  num_samples = 10000
4  alpha = beta = 1
5  site_a_simulation = beta(num_conv_a + alpha,
6                          num_views_a - num_conv_a + beta,
7                          size=num_samples)
8  site_b_simulation = beta(num_conv_b + alpha,
9                          num_views_b - num_conv_b + beta,
10                         size=num_samples)
```

Bayesian A/B Testing in Code Cont.

What's the probability that site A has a higher conversion rate than site B?

```
11 np.mean(site_a_simulation > site_b_simulation)
```

Bayesian A/B Testing in Code Cont.

What's the probability that site A has a higher conversion rate than site B?

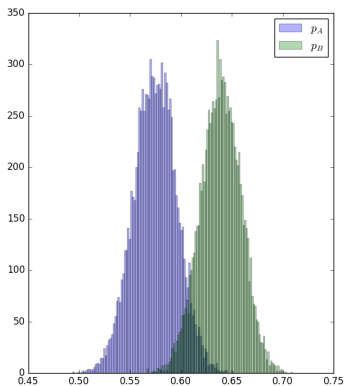
```
11 np.mean(site_a_simulation > site_b_simulation)
```

What's the probability that site A has a 5% higher conversion rate than site B?

```
12 np.mean(site_a_simulation > (site_b_simulation + 0.05))
```

Visualizing the Simulation

Histogram of posterior p 's for site A & B



Histogram of posterior p 's for site A & B

