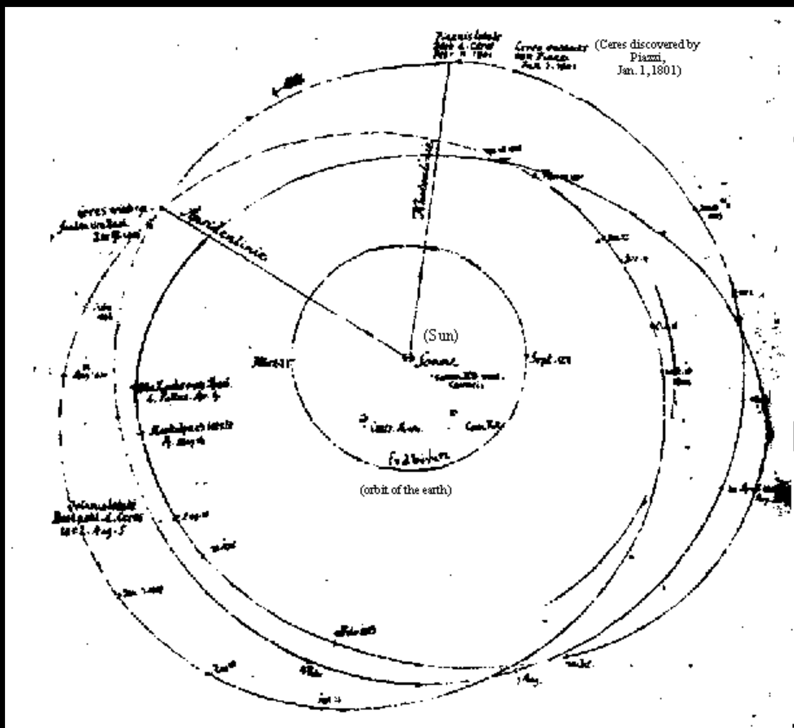
Bayesian A/B Testing

Beta Distribution and Multi-Arm Bandit

A history lesson

- Understanding the frequentist world view



Sketch of the orbits of Ceres and Pallas (nachlaß Gauß, Handb. 4). Courtesy of Universitätsbibliothek Göttingen.

Astronomy , Gauss and confidence intervals

https://www.math.rutgers.edu/~cherlin/History/Papers1999/weiss.html

What about Bayes Ways?

Posterior Probability

Likelihood of Observations

Prior Probability

$$\Pr(\theta|y) = \frac{\Pr(y|\theta)\Pr(\theta)}{\Pr(y)}$$

Normalizing Constant

prior: initial belief

likelihood: likelihood of data given outcome

posterior: updated belief

Bayes Theorem

posterior ∝ prior x likelihood

Binomial (Likelihood)

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

- p: conversion rate (between 0 and 1)
- n: number of visitors
- k: number of conversions

Beta Distribution

$$\boxed{\frac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}}$$

- p: conversion rate (between 0 and 1)
- α, $\boldsymbol{\beta}$: shape parameters
  - α = 1 + number of conversions
  - $\boldsymbol{\beta}$ = 1 + number of non conversions
- Beta Function (B) is a normalizing constant
- $\alpha = \boldsymbol{\beta} = 1$ gives the *uniform distribution*

Conjugate Priors

posterior ∝ prior × likelihood

beta ∝ beta × binomial

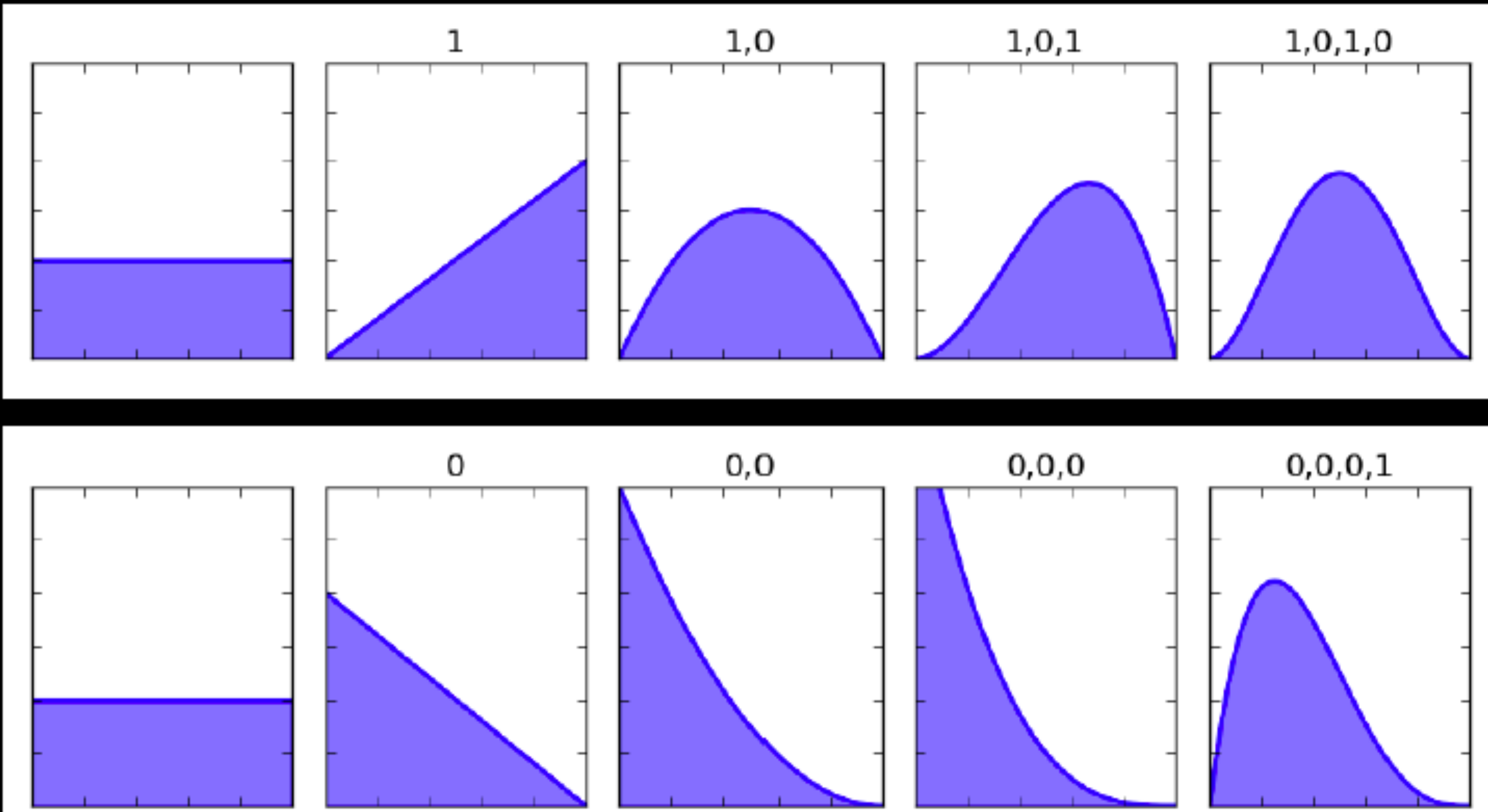
THE MATH:

$$
\begin{aligned}
\text{posterior} &\propto \text{prior} \times \text{likelihood} \\
&= \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(a,b)} \times \binom{n}{k} p^k (1-p)^{n-k} \\
&\propto p^{\alpha-1}(1-p)^{\beta-1} \times p^k (1-p)^{n-k} \\
&\propto p^{\alpha+k-1}(1-p)^{\beta+n-k-1}
\end{aligned}
$$

The result is a Beta Distribution with these shape parameters:

$\alpha + k$ and $\boldsymbol{\beta} + n$ - k

The Distribution



1 = conversion 0 = non conversion

A/B Testing

(frequentist)

- Define a metric (CTR, for example)
- Determine parameters of interest for study (number of observations, power, significance threshold, and so on)
- Run test, without checking results, until number of observations has been achieved
- Calculate p-value associated with your hypothesis test
- report p-value and suggestion for action

A/B Testing

(frequentist)

- Can you say “it is 95% likely that site A is better than site B”?
- Can you stop test early based on surprising data?
- Can you update the parameters of your test while it is running?

A/B Testing

(bayesian)

- Define a metric (CTR, for example)
- Run test, continually monitor results
- At any time calculate a probability that site A has better results on the defined metric than site B
- Suggest courses of action based on this probability

A/B Testing

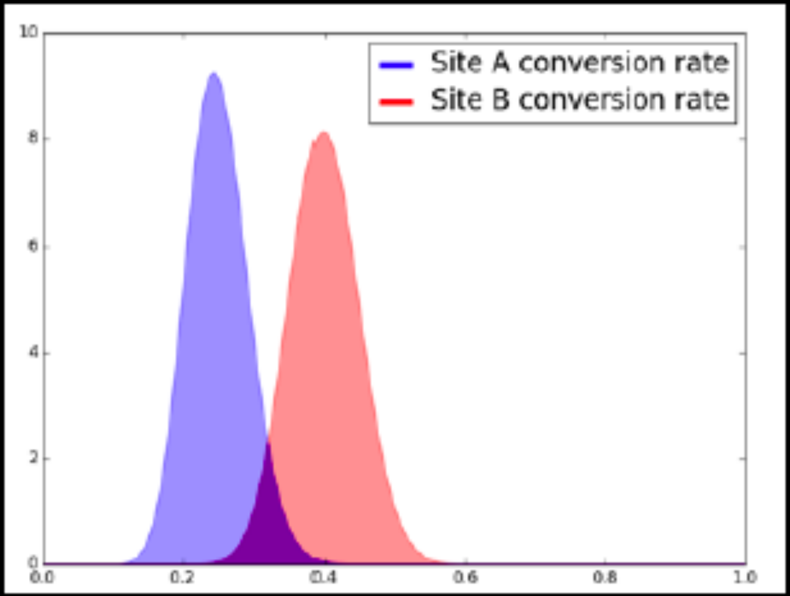
(bayesian)

- Can you say “it is 95% likely that site A is better than site B”?
- Can you stop test early based on surprising data?
- Can you update the parameters of your test while it is running?

A/B Testing

- We want to know if this is true:
  conversion rate of site A > conversion rate of site B

- We can also answer if this is true:
  conversion rate of site A > conversion rate of site B + 5%



Method:

- Sample a large number from both distributions

- Count the percent of times site A wins

The code

```
num_samples = 10000

A = np.random.beta(1 + num_clicks_A,
                   1 + num_views_A - num_clicks_A,
                   size=num_samples)

B = np.random.beta(1 + num_clicks_B,
                   1 + num_views_B - num_clicks_B,
                   size=num_samples)

### The probability that A wins:
print np.sum(A > B) / float(num_samples)

### The probability that A > B + 0.5%:
print np.sum(A > (B + 0.05)) / float(num_samples)
```

In Summary:

- Explain the difference between a frequentist A/B test & a Bayesian A/B test.
- Define & explain prior, likelihood, & posterior.
- How are conjugate priors useful for A/B testing? Explain what a conjugate prior is and how it applies to A/B testing.
- Analyze an A/B test with the Bayesian approach.