# Inference

## An Overview of ML & MAP Approaches

By Chris Tufts
@devlintufts
miningthedetails.com

# What is Statistical Inference?

"Learning characteristics of the population from ~~the~~ a sample"

\- PSU STAT 504

onlinecourses.science.psu.edu/stat504/node/16

# Example: Is a coin fair?


Before we get too far...

## Necessary Background

☐ Bernoulli distribution

☐ Beta distribution

# Bernoulli Distribution

- 2 possible outcomes

  ON/OFF

  Heads/Tails

- Single parameter 'p'

  $p(\text{heads}) = p$

  $p(\text{tails}) = 1 - p$

p is between 0 & 1

# Beta Distribution

- Distribution of probabilities

- 2 parameters

  $\alpha$ - Alpha

  $\beta$ - Beta

- We will use it to estimate a prior

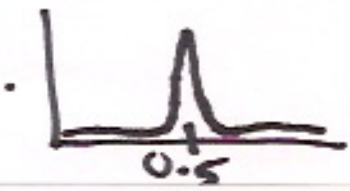
# Beta

- $\alpha$ $\beta$ → Integers

<u>Ex</u>
$\alpha$ → # heads
$\beta$ → # tails

$\alpha = 2$ After 4 flips
$\beta = 2$ how confident are
you the coin is fair?


$\alpha = 490$ After 1000
$\beta = 510$ flips, maybe
we can make a better guess



# For More Info

## Bernoulli:

PSU STAT 504

onlinecourses.science.psu.edu/stat504/node/27

## Beta:

Dave Robinson's post on beta distributions

varianceexplained.org/statistics/beta_distribution_and_baseball/

# Bayes Theorem

Likelihood

Prior

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{p(D)}$$

Posterior

Evidence

**Prior**
$p(\theta)$

Estimated probability of a value, $\theta$, prior to observing the data, $D$.

Do I have a strong prior belief my coin is fair???

**Evidence**
$p(D)$

Probability of observing data

**Likelihood**
$p(D|\theta)$

How likely is the data given $\theta$?

**Posterior**
$p(\theta|D)$

How credible is $\theta$ given the data?

# Maximum Likelihood

- Find $\theta$ to Maximize the Likelihood

- $\theta_{ML} = \frac{\#\text{ heads}}{\#\text{ flips}}$

Note:
$\theta$ refers to model parameter(s) in our case it is: $p_{ML} = \theta_{ML}$

# Maximum a Posteriori

# MAP

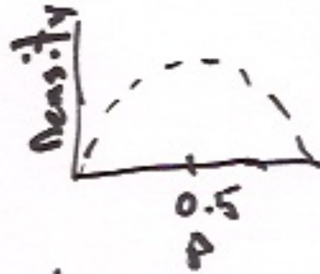
- Find $\theta$ that Maximizes the posterior

- Uses a 'prior' ... in our case a beta distribution is used

$$\theta_{MAP} = \frac{\#\text{heads} + \alpha - 1}{\#\text{flips} + \alpha + \beta - 2}$$

# Example 1:

20 flips
$\alpha = \beta = 2$




- I don't know how much I can trust this coin so I set $\alpha$ & $\beta$ to 2.

$$p_{MAP} = \frac{\#\text{heads} + \alpha - 1}{\#\text{heads} + \#\text{tails} + \alpha + \beta - 2}$$

$\#\text{heads} = 12$
$\#\text{tails} = 8$

$$p_{MAP} = \frac{12+2-1}{20+4-2}$$

$$\boxed{p_{MAP} = 0.59}$$

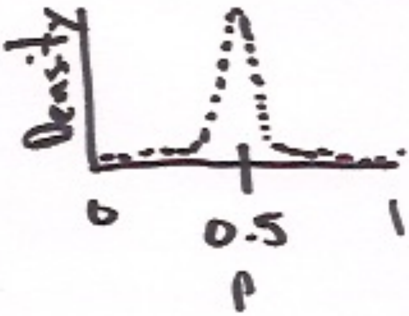
$$p_{ML} = \frac{\#\text{heads}}{\#\text{flips}}$$

$$\boxed{p_{ML} = 0.6}$$

# Example 2:

20 flips
$\alpha = \beta = 100$

\# heads = 12
\# tails = 8




- This time I got my coin from a trusted source, so I use higher $\alpha$ + $\beta$ values for my prior...
- We already know $\boxed{p_{ML} = 0.6}$
- $p_{MAP} = \dfrac{12 + 100 - 1}{20 + 200 - 2}$

$$\boxed{p_{MAP} = 0.51}$$

Example 3:

$\#$ flips $= 1000$

$\#$ heads $= 723$

$\#$ tails $= 277$

$\alpha = \beta = 2$

$$\boxed{p_{ML} = \frac{723}{1000} = 0.72}$$

$$p_{MAP} = \frac{723+2-1}{1000+4-2}$$

$$\boxed{p_{MAP} = 0.72}$$

# Example 4:

\# heads = 723

\# tails = 277

$\alpha = \beta = 100$

$$\boxed{p_{ML} = 0.72}$$

$$p_{MAP} = \frac{723 + 100 - 1}{1000 + 200 - 2}$$

$$\boxed{p_{MAP} = 0.68}$$

# Some Notes...

- MAP:
  - selection of prior can have a strong impact
  - the <u>stronger</u> the prior (i.e. higher $\alpha$ & $\beta$) the more samples (# flips) are needed to overcome an incorrectly estimated prior [or prior that differs from observations]
- see/compare $p_{MAP}$ in examples

| | Bayes Components | Pros | Cons |
|---|---|---|---|
| ML | Likelihood | Fast/ Easy to implement | Doesn't take prior into account |
| MAP | Likelihood + Prior | Allows use of prior | If prior is selected incorrectly requires lots of observation to overcome |

# References

Parameter Estimation for Text Analysis

https://faculty.cs.byu.edu/~ringger/CS679/papers/Heinrich-GibbsLDA.pdf

Author: Gregor Heinrich

provides derivations of MAP + ML approaches/formulas used in examples!

# UP NEXT!

Gibbs Sampling