ML AND MAP

Shingchern D. You

ML concept

- Recall the basic definition: Likelihood of θ given the sample $\mathcal{X}$

$$l\,(\theta|\mathcal{X}) = p\,(\mathcal{X}\,|\theta) = \prod_t p\,(x^t|\theta)$$

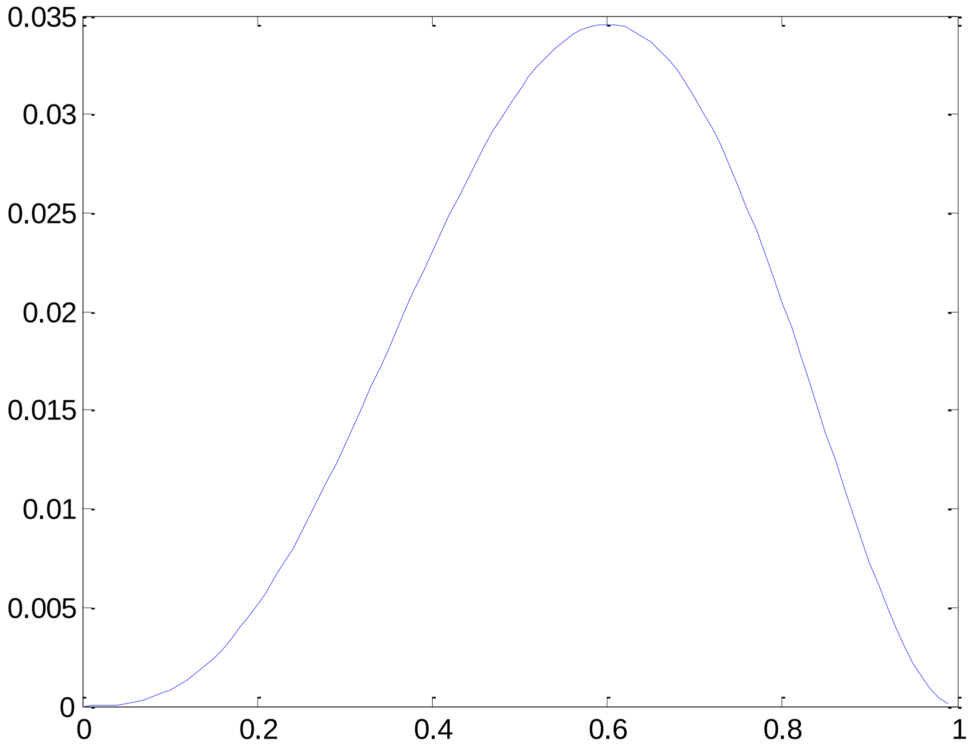
- Want to find θ such that l is max
- Simple example: Tossing a coin 5 times and the results being H, H, T, T, H.

ML concept

- If $P(\{H\}) = \theta = 0.1$, we have
  $P(\{H,H,T,T,H\}) = 0.1*0.1*0.9*0.9*0.1 = 8.1\text{x}10^{-4}$
- If $P(\{H\}) = \theta = 0.2$, we have
  $P(\{H,H,T,T,H\}) = 0.2*0.2*0.8*0.8*0.2 = 5.1\text{x}10^{-3}$
- We can repeat this computation many times with different values of θ to find the max value
- Theoretic answer: $\theta_{ML} = \frac{N_H}{N}$, where$N$: total tossing, N_H: tossing with head shown

ML concept

- With a program, this can be done easily. Observe that $\theta = 0.6$ yields highest probability, $\theta_{ML} = 0.6$



ML concept

- This result is the same as theoretical derivation
- But, it seems counter-intuitive…If tossing a coin 100 times and head shows 100 times, what is the probability head shown on next tossing?

MAP

- ML says 1.0, but we know it is likely 0.5 because tossing coins is modeled as "repeated" (and/or independent) trials
- That is the difference between ML and MAP
- ML is derived ONLY based on observation
- MAP incorporates ***a priori*** probability to ML
- Recall Bayes theorem

$$\mathrm{P}(\theta|\chi) = \frac{P(\chi|\theta)P(\theta)}{P(\chi)}$$

MAP

- MAP estimator wants to find

$$\theta_{MAP} = \arg\max_{\theta} \mathrm{P}(\theta|\boldsymbol{\chi})$$

- As $P(\chi)$ does not affect the max operation, we need to consider only

$P(\boldsymbol{\chi}|\theta)P(\theta)$

- The first term is equal to ML, the second term is the *a priori* probability

MAP

- The probability $P(\{\mathrm{H}\}) = \theta$ is typically modeled as outcome from beta distribution, whose pdf is

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{(a-1)}(1-x)^{(b-1)}$$

- Therefore, we need to determine values of a and b (*a priori* knowledge) in order to use MAP

MAP

- In terms of simulation: (1) Generate many uniformly spaced values for θ, say, 100 numbers between 0 and 1; (2) Compute likelihood for each θ, OK to use pdf in place of $P(\theta)$; (3) Find θ_{max}
- With lots of math, we have

$$\theta_{MAP} = \frac{N_H + a - 1}{N + a + b - 2}$$

Source: www.mi.fu-berlin.de/wiki/pub/ABI/Genomics12/MLvsMAP.pdf

MAP

- To have a "mean" $\theta = 0.5$, we set a = b
- If a = b = 1, we have $\theta_{MAP} = 0.6$ (same as ML)
- If we are more confident about the *a priori* knowledge, we can set larger values of a and b, such as a = b = 10

MAP

- If a = b = 10, we have $\theta_{MAP} = 0.522$