# 4. MLE & MAP

From name, we know MLE is "Maximum Likelihood".
MAP is "Maximum A Posterior". From Bayes' Law

$$P(\theta|x) = \frac{P(x|\theta) \cdot P(\theta)}{P(x)}$$

Prior $P(\theta)$ is needed for MAP.

Let's use an example:
We flip a coin (not fair) $N$ times and wish to estimate the probability of seeing head $\theta$ after observing $X$ heads.

Assume: $\theta \sim \text{Beta}(a, b)$

$X \sim \text{Binomial}(N, \theta)$

① MLE approach.

1) likelihood

$$p(x|\theta) = \binom{N}{x} \theta^{x} (1-\theta)^{N-x}$$

2) maximum likelihood

$$\begin{aligned}\hat{\theta}_{ML} &= \arg\max_{\theta} p(x|\theta) \\ &= \arg\max_{\theta} \log p(x|\theta) \\ &= \arg\max_{\theta} \underbrace{\log \binom{N}{x} + x \log\theta + (N-x)\log(1-\theta)}_{(*)}\end{aligned}$$

$$\frac{\partial}{\partial\theta}(*) = \frac{x}{\theta} - \frac{N-X}{1-\theta} = 0 \Rightarrow \hat{\theta}_{ML} = \frac{x}{N}$$

② MAP approach

1) Prior

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}$$

2) Posterior

$$p(\theta|x) = \frac{P(x|\theta) \cdot P(\theta)}{P(x)}$$

$$= \frac{\binom{N}{x} \theta^{x} (1-\theta)^{N-x} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}}{P(x)}$$

$$\hat{\theta}_{MAP} = \arg\max_{\theta} P(\theta|x)$$

$$= \arg\max_{\theta} P(x|\theta) P(\theta)$$

$$= \arg\max_{\theta} \binom{N}{x} \theta^{x} (1-\theta)^{N-x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}$$

$$= \arg\max_{\theta} \left[\binom{N}{x} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right] \cdot \theta^{x+a-1} (1-\theta)^{N-x+b-1}$$

log(·)

$$= \arg\max_{\theta} \underbrace{\log\left[\binom{N}{x} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right] + (x+a-1)\log\theta + (N-x+b-1)\log(1-\theta)}_{(*)}$$

$$\frac{\partial}{\partial\theta}(*) = \frac{x+a-1}{\theta} - \frac{N-x+b-1}{1-\theta} = 0$$

$$\Rightarrow \hat{\theta}_{MAP} = \frac{x+a-1}{N+a+b-2}$$