

# MLE / MAP

Jeonghun Yoon

# Terms

Bayes rule

Maximum Likelihood Estimate (MLE)

Maximum A Posteriori Estimate (MAP)

# Bayes rule

The diagram shows the Bayes' rule formula  $p(\theta|\mathbb{x}) = \frac{p(\mathbb{x}|\theta)p(\theta)}{\sum p(\mathbb{x}|\theta)p(\theta)}$  enclosed in a blue rectangular box. Three labels with arrows point to parts of the formula: 'likelihood (우도 값)' points to the numerator's first term  $p(\mathbb{x}|\theta)$ ; 'prior (사전 확률)' points to the numerator's second term  $p(\theta)$ ; and 'posteriori (사후 확률)' points to the entire left side of the equation  $p(\theta|\mathbb{x})$ .

$$p(\theta|\mathbb{x}) = \frac{p(\mathbb{x}|\theta)p(\theta)}{\sum p(\mathbb{x}|\theta)p(\theta)}$$

- 사후 확률 : 관찰 값들이 관찰 된 후에 모수(parameter)의 발생 확률을 구한다.
- 사전 확률 : 관찰 값들이 관찰 되기 전에 모수의 발생 확률을 구한다.
- 우도 값 : 모수의 값이 주어졌을 때 관찰 값들이 발생할 확률

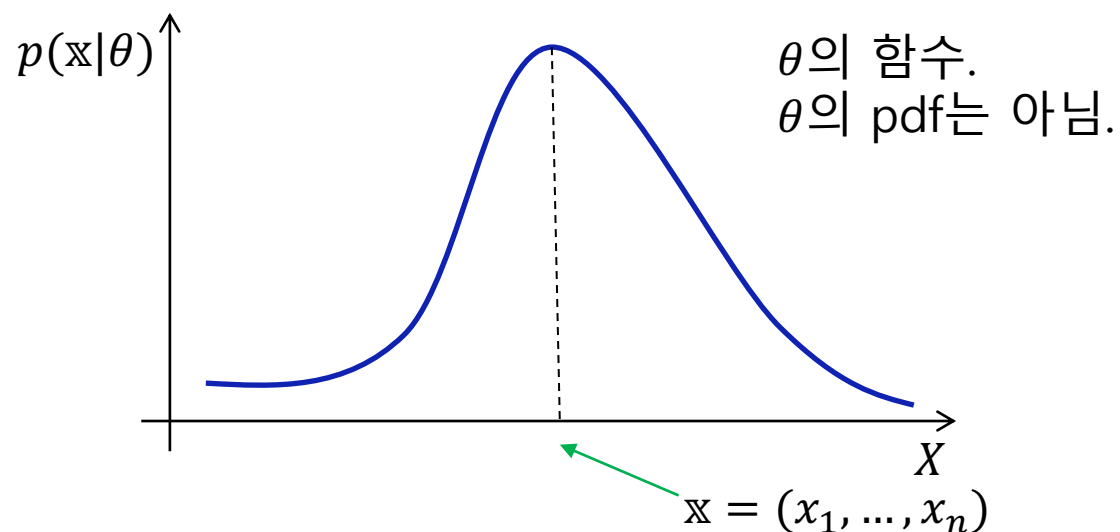
# Maximum Likelihood Estimate

$$\mathbb{x} = (x_1, \dots, x_n)$$

우도(likelihood)는 다음과 같이 정의 된다.

$$\mathcal{L}(\theta) = p(\mathbb{x}|\theta)$$

변수(parameter)  $\theta$ 가 주어졌을 때, data set  $\mathbb{x} = (x_1, \dots, x_n)$  (관찰 된, observed) 를 얻을 수 있는(obtaining) 확률

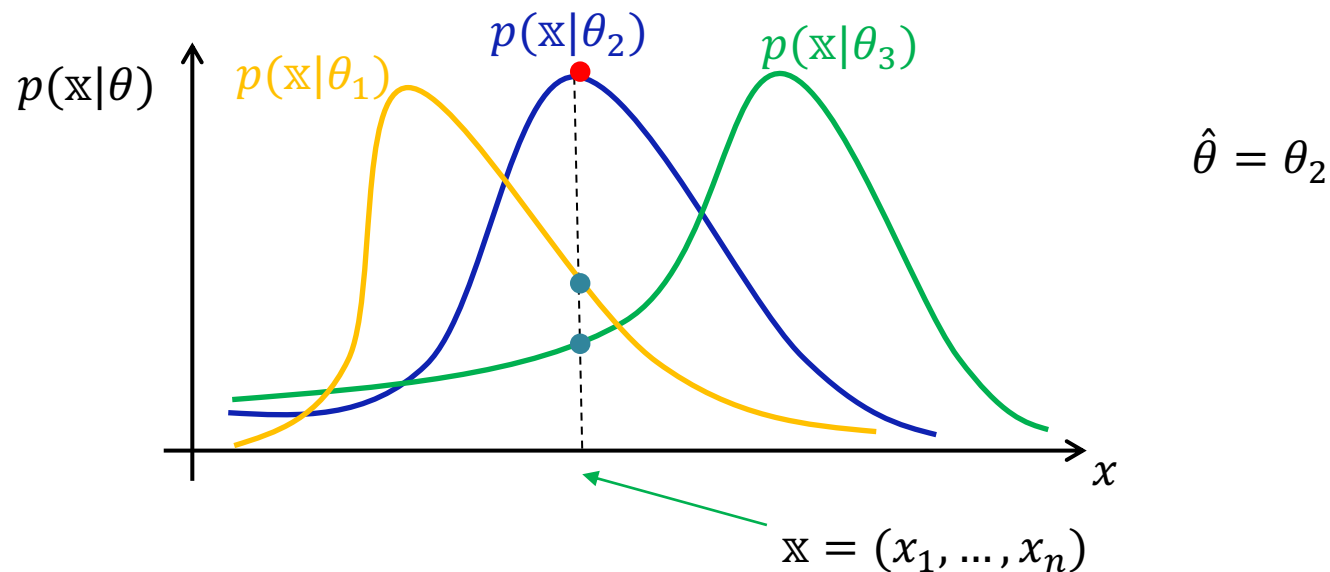


# Maximum Likelihood Estimate

Maximum Likelihood Estimate는 다음과 같이 정의 된다.

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta) = \arg \max_{\theta} p(\mathbb{x}|\theta)$$

관찰 된 data set  $\mathbb{x} = (x_1, \dots, x_n)$ 을 얻을 수 있는 확률이 가장 큰  $\theta$ 가 MLE이다.



# Maximum A Posteriori Estimate

우리가 likelihood function  $p(\mathbf{x}|\theta)$ 와 prior  $p(\theta)$ 를 알 때, Bayes rule에 의하여 posteriori function의 값을 구할 수 있다.

A diagram enclosed in a blue rectangular box. It shows the relationship between likelihood, prior, and posterior. At the top left, the text 'likelihood (우도 값)' has an arrow pointing to the numerator of a fraction. At the top right, the text 'prior (사전 확률)' has an arrow pointing to the same numerator. The fraction is  $p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\sum p(\mathbf{x}|\theta)p(\theta)}$ . At the bottom left, the text 'posteriori (사후 확률)' has an arrow pointing to the entire fraction. The label 'likelihood (우도 값)' is positioned above the fraction, and 'prior (사전 확률)' is positioned to the right of the fraction. The label 'posteriori (사후 확률)' is positioned to the left of the fraction.

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\sum p(\mathbf{x}|\theta)p(\theta)}$$



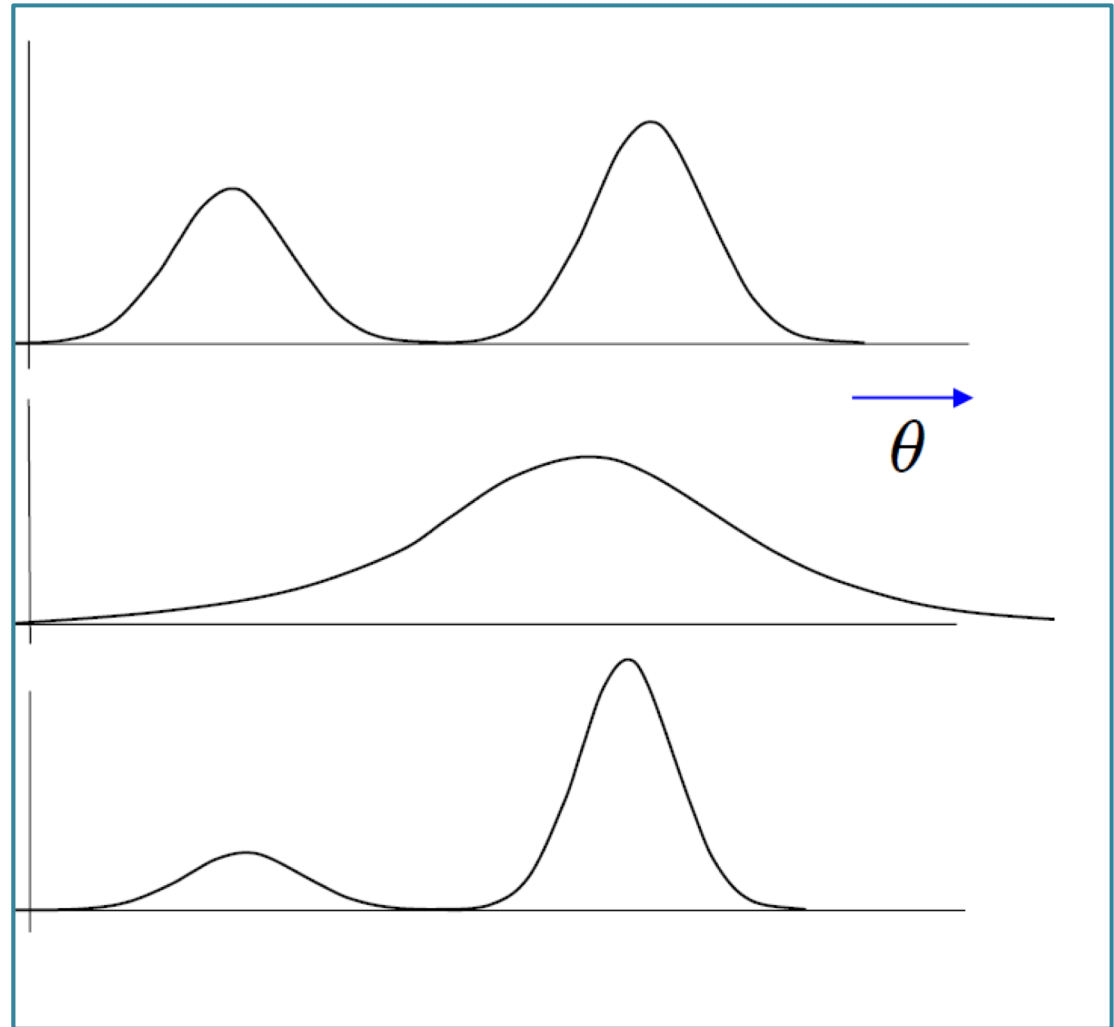
$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$$

# MLE vs MAP

Likelihood  
 $p(\mathbf{x}|\theta)$

Prior  
 $p(\theta)$

Posterior  
 $p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$

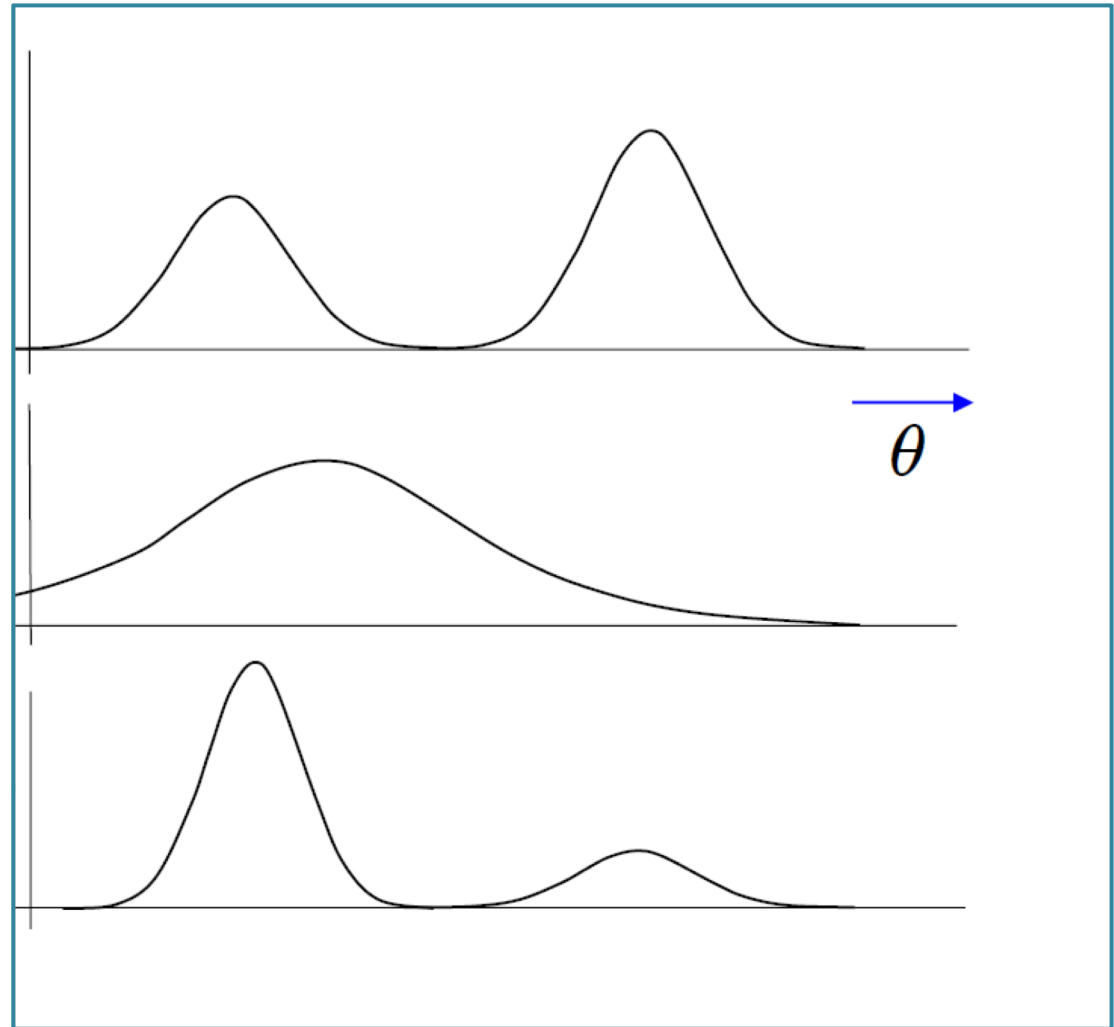


# MLE vs MAP

Likelihood  
 $p(\mathbf{x}|\theta)$

Prior  
 $p(\theta)$

Posterior  
 $p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$



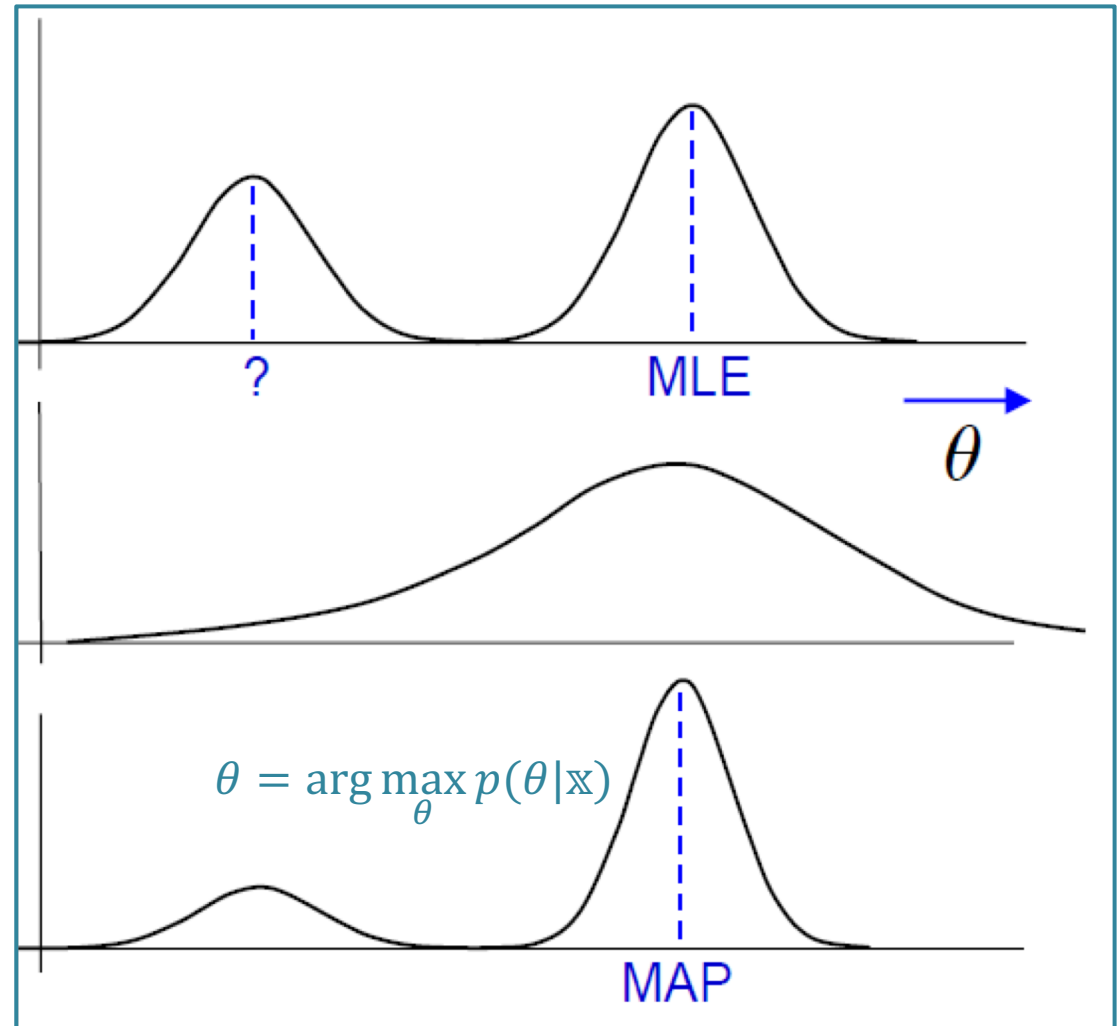


# MLE vs MAP

Likelihood  
 $p(\mathbf{x}|\theta)$

Prior  
 $p(\theta)$

Posterior  
 $p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$



# MLE 예제

최대 우도 추정은 관측 값이 주어졌을 때 모델의 매개변수들 중 그 관측 값을 발생시켰을 가능성이 가장 높은 매개변수들의 값을 추정하는 것이다.

이항 분포를 예로 들어보자. 표본 (0, 1, 0, 0, 1, 0)이 이항 분포로부터 추출되었다.  $P(X = 1) = \mu$ ,  $P(X = 0) = 1 - \mu$  이다. 매개변수  $\mu$ 에 대한 최대 우도 추정(MLE)은 무엇일까?

# MLE 예제

$$\begin{aligned}\mu\text{의 우도} &= L(\mu) \\ &= P(x=0) \times P(x=1) \times P(x=0) \times P(x=0) \times P(x=1) \times P(x=0) \\ &= (1-\mu) \times \mu \times (1-\mu) \times (1-\mu) \times \mu \times (1-\mu) \\ &= (1-\mu)^4 \times \mu^2\end{aligned}$$

우도를 최대화하는 것과 로그 우도를 최대화하는 것은 동일한 과제이므로 수학적 편의를 위해 양변에 로그를 취한다.

$$\log(L(\mu)) = \log((1-\mu)^4 \mu^2) = 4 \times \log(1-\mu) + 2 \times \log(\mu)$$

미분을 하여 0이 되는 지점을 찾는다.

$$\frac{\partial(\log L(\mu))}{\partial \mu} = 0 \Rightarrow 4 \times \frac{1}{1-\mu} \times (-1) + 2 \times \frac{1}{\mu} = 0 \Rightarrow -4 \times \mu + 2(1-\mu) = 0 \Rightarrow \mu = \frac{1}{3}$$

# MLE 예제

하지만 미분을 0으로 하는 식에서 찾은 안장값(saddle point)이 최대값인지 최솟값인지 확인하기 위해 한 번 더 미분을 해야 한다.  $\mu$ 값이 최대값이라면  $\log(L(\mu))$ 를 두 번 미분한 값은 음수여야 한다.

$$\frac{\partial^2}{\partial \mu^2} \log(L(\mu)) = -4 \times \frac{1}{(1-\mu)^2} - \frac{2}{\mu^2}$$

$\frac{1}{3}$  을 대입하자.

$$\frac{\partial^2}{\partial \mu^2} \log(L(\mu)) = -4 \times \frac{1}{\left(1 - \frac{1}{3}\right)^2} - \frac{2}{\left(\frac{1}{3}\right)^2} = -9 - 18 = -27$$

따라서 우도를 최대화하는  $\mu$ 는  $\frac{1}{3}$  이라는 것이 증명된다.

# MLE 예제

따라서

- Likelihood :  $L(\mu) = \left(1 - \frac{1}{3}\right)^4 \left(\frac{1}{3}\right)^2 = 0.021948$

- Log Likelihood :  $\ln(L(\mu)) = \ln(0.021948) = -3.819$

정리해보면 관측값이 (0, 1, 0, 0, 1, 0)과 같을 때, 앞면이 나올 확률은  $\frac{1}{3}$ 인 것이다. 표본에서 앞면이 2번, 뒷면이 4번이다. 관측값만 살펴보았을 때 동전의 앞면이 나올 확률은 직관적으로 보아도  $\frac{1}{3}$ 이다. 아주 그럴 듯한, 모수의 예측값이다.