

## 2. Maximum Likelihood Estimation(MLE)

### Likelihood Function

입력으로 주어진 확률 분포(파라미터)가 데이터를 얼마나 잘 설명하는지를 나타내는 점수 (Likelihood)를 출력하는 함수= $P(\text{Data}|\text{parameter})=\text{Likelihood}(\text{parameter})$

입력 : 확률 분포를 표현하는 파라미터

출력 : 데이터를 설명하는 정도  $\rightarrow$  높은 확률값을 가질수록 데이터를 잘 설명하는 것

### Log Likelihood

Likelihood는 확률 값의 곱으로 표현되어 있기 때문에 underflow의 가능성과 cost가 큼  
따라서 log를 취해줌으로써 덧셈으로 바꾸어 해결해줌

$$\prod_{i=1}^n P_{\theta}(x=x_i) \quad \rightarrow \quad \sum_{i=1}^n \log P_{\theta}(x=x_i)$$

### MLE via Gradient Ascent

랜덤 생성 대신, gradient Ascent를 통해 likelihood 값을 최대로 만드는 파라미터를 찾자!

$$\theta \leftarrow \theta + \alpha \cdot \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$$

### Summary

1. 가상의 확률분포를 모사하는 확률 분포의 파라미터를 찾고 싶음
2. 목표 확률 분포로부터 데이터를 수집한 후, 데이터를 잘 설명하는 파라미터를 MLE를 통해 찾자.

3. 이 때, Gradient Ascent를 통해 log likelihood를 최대로 하는 파라미터를 찾을 수 있다.