In Data Analysis

-Data가 어떤 확률변수로부터 실현된 표본 (가정)

- 분포 / 모수를 가진다.

1) 특정 분포를 가지는가 ex) 가우시안, 베르누이 ----------- test(분포 검정)
2) (그렇다면) parameter(모수)의 값은 얼마인가 -----------모수검정(parameter Test) H0 = `\( mu\)`

3) 모수추정 -> 모수가 실제로 어떤값인지 (어떤값일 확률이 높은지)
MSE, MLE, 베이지안 추정

- 모수 추정
  - 모멘트 방법
  - MLE :

$$\hat{\theta} = \arg\max_{\theta} L\left(\theta; \{x_i\}\right)$$

  - 베이지안

1. MLE 모수추정(ex)

$$P(x;\theta) = Bern(x;\theta) = \theta^{x_i} \times (1-\theta)^{(1-x_i)}$$

$$\log^{L} = N_1 / \log^{\theta} + (N - N_1)\log(1-\theta)$$

2. 베이지안 모수추정 - 모수적 방법
   - 비모수적 방법

- 모수적 방법

https://m.blog.naver.com/PostView.nhn?blogId=muzzincys&logNo=60204777552&proxyReferer=https%3A%2F%2Fwww.google.co.kr%2F

베이지안 추론은 설명하기 이전에는 먼저 사후확률(posterior)에 대한 설명이 필요하다. 베이즈 정리에서도 설명 했듯 사후확률은 사전확률(prior)와 주어진 데이터(정보)를 결합하여 모수에 대한 정보를 업데이트하여 추론하는 확률이다.

즉, MLE에서는 사전 정보를 전혀 반영하지 않고, 주어진 정보만으로 확률을 구하였다면, 베이지안의 경우, 사후확률을 구할 때, 우리가 알고 있는 사전 정보를 이용하여 구한다.

예를 들면 앞의 MLE 예에서 우리는 주어진 정보, 즉 동전을 던졌을 때 100번 중 앞면이 63번이 나왔다라는 정보만을 이용했다. 하지만 베이지안은 우리가 알고 있는 앞면이 나올 확률이 0.5(공정한 동전이라는 조건)라는 사전 정보를 이용하여 사후확률을 구한다.

- 비모수적 방법

https://statisfaction.wordpress.com/2018/07/19/different-ways-of-using-mcmc-algorithms/

EM 혼합모형

likelihood 함수

- 확률 밀도 함수를 랜덤변수의 값 xx의 함수가 아닌 파라미터 θθ의 함수로 보는 것
- 확률 분포로부터 특정한 샘플 값 xx가 발생하였을 때, 이 샘플 값 xx가 나오게 하는 파라미터 θθ의 가능성
- 확률 분포로부터 특정한 샘플 값 xx가 발생하였을 때, 샘플 값 xx와 변수 θθ에서의는 확률(밀도함수)

Q1) EM 에서 loglikelihood 와 resposibility 관계
Q2) K-means 와 EM의 관계