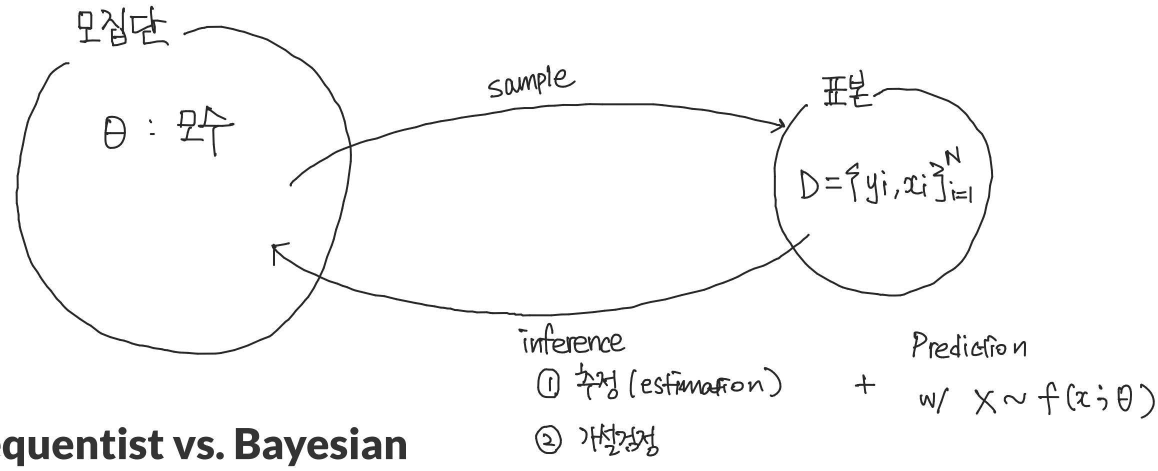
ESC-21WINTER

Bayesian Statistics

Purpose of Statistics : Inference and Prediction



Frequentist vs. Bayesian

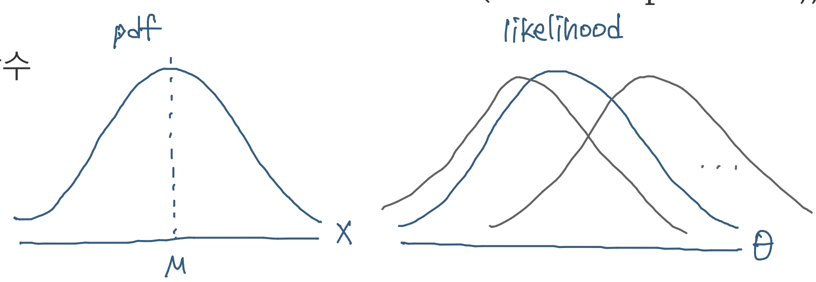
위와 같은 통계학의 문제들을 해결하는 데 두 가지 접근법이 Frequentist (빈도론적 통계학)와 Bayesian (베이지안 통계학)이다.

들어가기 전에 : PDF (probability density function)과 Likelihood Function

$$y_i \sim \underbrace{f(x_i, \theta)}_{known\ \theta} \Longrightarrow \underbrace{\mathbf{L}(\theta|x, y)}_{unknown\ \theta}$$

density function : 전적으로 모수 θ에 의해 결정되는 함수 (θ를 알아야 pdf 아니냐,,)

likelihood : data의 함수



Frequentist : Fixed Parameter, Random Data

빈도론적 관점에서 추론(Inference)은 **알지는 못하지만 고정된 상수로 존재하는 모수** θ^* **를 찾는 것** 이다. 그리고 이런 'θ가 뭐냐' 라는 물음에 대해서 확률적으로 답한다.

$\mathcal{D} = \{y_i, x_i\}$: data

$\hat{\theta} = s(\mathcal{D})$ + [,] error bound w.r.t. θ

%

Estimation : How to answer about error bound?

1. Solve optimization problem (e.g. MLE) : $\hat{\theta} = s(\mathcal{D}) = \max_{\theta} l(\theta|\mathcal{D})$
2. Error bound : Confidence Interval
   - (Asymptotic) Sampling distribution of $s(\mathcal{D})$: 극한분포 구하기
   - Bootstrapping : 시간과 비용이 넘 비쌈ㅠ

$D \longrightarrow D_1^*, D_2^*, D_3^*, \cdots, D_B^*$

$s(D_1^*) \quad s(D_2^*) \quad \cdots \quad s(D_B^*)$ ⇒ empirical dist.

Hypothesis Testing : 너무 이분법적이야!

그리고 여러모로 p-value가 불러오는 오해에 빠지기 쉬움...

Prediction ; 어떤 "분포" 에 대해서 예측

$\tilde{y}_i \sim f(\tilde{x}_i, \hat{\theta})$

Uncertainty of the prediction comes from

1. Uncertainty of parameter estimate $\hat{\theta}$
2. Inherent variability in $\tilde{y}_i$

Bayesian : Fixed Data, Random Parameter

$\hat{\theta}$ = a distribution of θ given data

빈도론적 관점과의 가장 큰 차이를 모수 θ를 random variable로 본다는 것이다! 베이지안 관점에서 모르는 건 (uncertainty) 다 확률변수라고 생각한다.

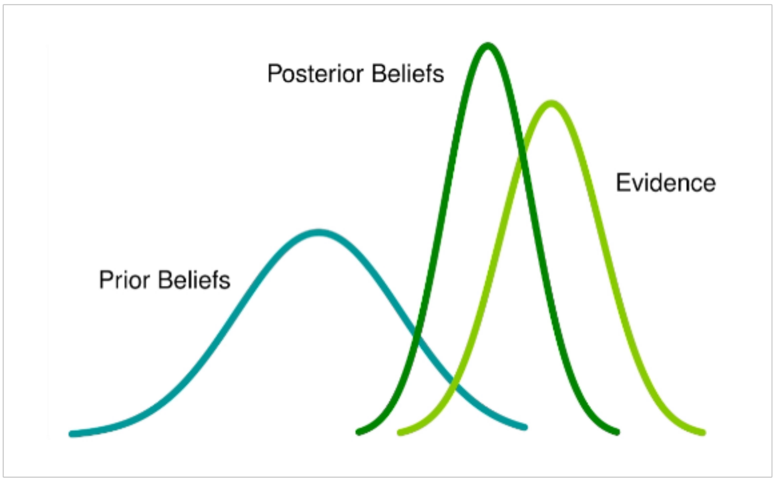
Estimation : full probability model

Bayes Rule : $p(\theta|\mathcal{D}) = \dfrac{p(\mathcal{D}, \theta)}{p(\mathcal{D})} = \dfrac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta}$

$p(\theta)$: Prior

$p(\mathcal{D}|\theta)$: Likelihood (observed data per each value of θ)

$p(\theta|\mathcal{D})$: Posterior



Bayesian process is all about applying Bayes Rule!

Prediction : weighted average over all possible θ

베이지안 관점에서는 모수에 대한 답을 '분포'의 형태로 내린다!

$$\tilde{y}_i \ \sim \ p(\tilde{y}_i|\mathcal{D}) = \int \underbrace{p(y_i|\theta)}_{\text{likelihood}}\underbrace{p(\theta|\mathcal{D})}_{\text{posterior}} d\theta$$

w.r.t. θ
모든 θ에 대한 weighted avg.

Problems

1. Prior on θ
   - 모수에 대한 사전적인 믿음/지식이 0이라면 어떻게 하죠...?
   - 어떤 scale에 대해서 uniform한 prior를 줄건데...?
   - prior를 변수변환에 대해서 invariant하게 줄 수 있어...?

   No clear answer...

2. Normalizing constant : 보통 적분이 어려움

   Solution ? Given that $p(\theta|\mathcal{D}) \propto p(\theta, \mathcal{D})$ ...
   - Use conjugacy
   - MCMC (Markov Chain Monte Carlo) : sample from posterior
   - Approximate posterior : normal approx, variational inference... etc.