Seoul
National
University

Bayesian Inference

Name: Chong-kwon Kim

SCONE
Lab.

Classical vs. Bayesian

- Notations
  - Θ (θ): Random variable of parameter that we will predict, instance of the r.v.
    - Pr(Θ = θ), $f_{\Theta}(\theta)$
  - X: (Big data) observations
    - Pr(X=x | Θ = θ), $f_{X|\Theta}(x \mid \theta)$

How to obtain A Priori?
- Conjecture
- Uniform, Normal
- Long observations

- Before: Classical (Frequentists)
  - Ignore distribution of priors
    - Consider θ as an unknown parameter

- Now: Bayesian
  - Big data
  - Consider a priori distributions as well as the likelihood of observed data
    - Pr
    - Pr(X=x | Θ = θ), $f_{X|\Theta}(x \mid \theta)$
    - (Θ) OR $f_{\Theta}(\theta)$

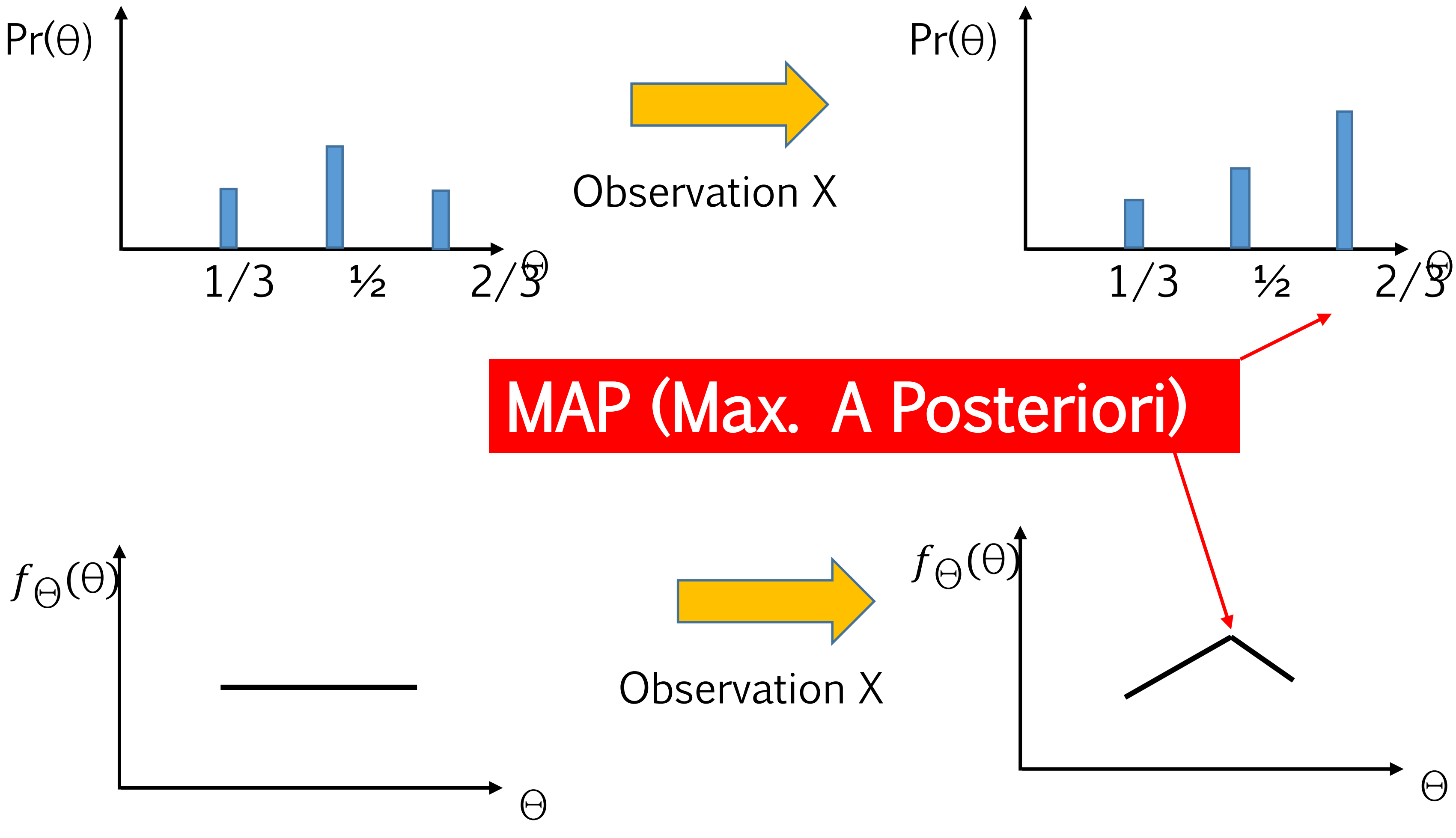
Bayes' Rule

Laplace(1749~1827) was a prominent French mathematician and scientist. Called Newton of France, he invented Laplace transform, predicted black holes and extended Bayes' theorem.
One of his student at Ecole Militaire was Napoleon who was good at Mathematics.

- Four cases
  - DD Case
    - $\Pr(\Theta = \theta \mid X=x) = \frac{\Pr(\Theta = \theta)\cdot Pr(x|\theta)}{\sum_{\theta} \Pr(\Theta = \theta)\cdot Pr(x|\theta)}$
  - DC Case
    - $\Pr(\Theta = \theta \mid X=x) = \frac{\Pr(\Theta = \theta)\cdot f_{X|\Theta = \theta}(x|\theta)}{\sum_{\theta} \Pr(\Theta = \theta)\cdot f_{X|\Theta = \theta}(x|\theta)}$
  - CD Case
    - $f_{\Theta|X}(\theta|X = x) = \frac{f_{\Theta}(\theta)\cdot \Pr(X=x|\Theta = \theta)}{\int_{-\infty}^{\infty} f_{\Theta}(\theta)\cdot \Pr(X=x|\Theta = \theta)d\theta}$
  - CC Case
    - $f_{\Theta|X}(\theta|x) = \frac{f_{\Theta}(\theta)\cdot f_{X|\Theta}(x|\theta)}{\int_{-\infty}^{\infty} f_{\Theta}(\theta)\cdot f_{X|\Theta}(x|\theta)\, d\theta}$

Bayesian Inference

- A priori ➔ A posteriori



Bayesian Inference

- **MAP**
  - $\hat{\theta} = \arg max_\theta \Pr(\theta \mid x)$
  - $\hat{\theta} = \arg max_\theta f_{\Theta|X}(\theta \mid x)$

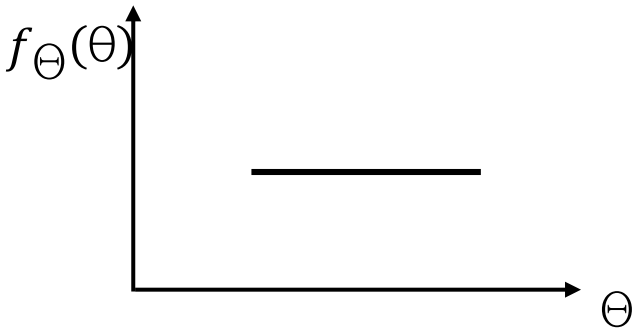
  ➔ Select $\hat{\theta}$ that maximizes
    - $\Pr(\Theta = \theta) \cdot Pr\ (x|\theta)$
    - $\Pr(\Theta = \theta) \cdot f_{X|\Theta = \theta}\ (x|\theta)$
    - $f_\Theta(\theta) \cdot \Pr(X = x|\Theta = \theta)$
    - $f_\Theta(\theta) \cdot f_{X|\Theta}(x|\theta)$

- **LMS (Least Mean Squares)**
  - Estimation error: $\Theta - \hat{\theta}$
  - Mean Squared Error: $E[(\Theta - \hat{\theta})^2]$
  - $\hat{\theta} = \arg min_\theta E[(\Theta - \theta)^2 \mid X = x]$ ➔ $\hat{\theta} = E[\Theta \mid X=x]$

MAP: Example

- A coin factory (Mint) manufactures coins that are known to have probability of Heads ~ U[1/3, 2/3]. Consider a one particular coin. To estimate Pr(H) of the coin, we toss the coin five times and the results are HHTHT. Estimate Pr(H) of the coin.

  – Random Variable: Θ
    - $f_{\Theta}(\theta) = 3,\ 1/3 \le \theta \le 2/3$

  – Observation: X
    - $\Pr(\mathrm{H} \mid \theta) = \theta = 1 - \Pr(\mathrm{T} \mid \theta)$
    - x = 'HHTHT'
    - $\Pr(\mathrm{HHTHT}) = \theta \cdot \theta \cdot (1-\theta) \cdot \theta \cdot (1-\theta) = \theta^3(1-\theta)^2$

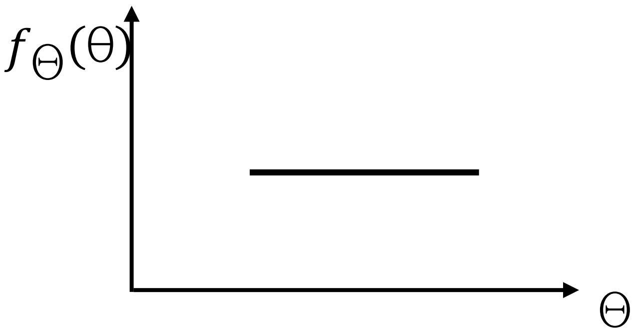
  – A Posteriori probability $\propto f_{\Theta}(\Theta) \cdot \Pr(X{=}x \mid \Theta = \Theta) = 3 \cdot \theta^3 (1-\theta)^2$

  ➔ $\hat{\theta} = 3/5$

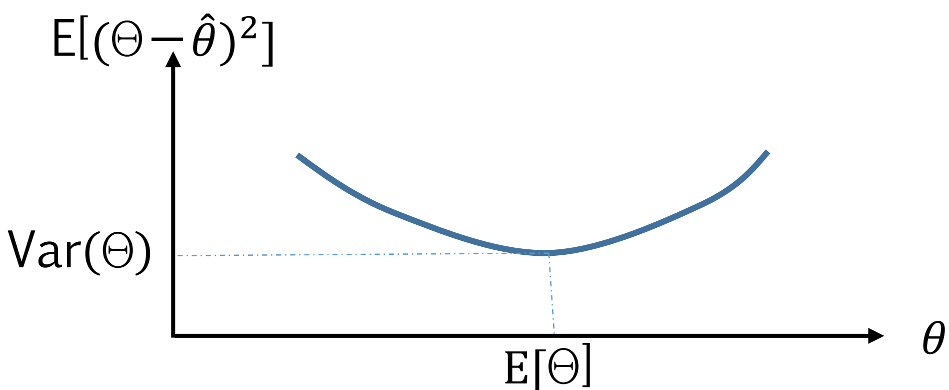
Note: $\int_{-\infty}^{\infty} f_{\Theta}(\theta) \cdot \Pr(X = x \mid \Theta = \Theta)\, d\Theta$
Is constant

LMS (Least Mean Squared)

- **Note**
  - ➔ Estimation error, $\Theta - \hat{\theta}$, is a R.V.
  - ➔ MSE(Mean Squared Error), $E[(\Theta - \hat{\theta})^2]$, depends on $\hat{\theta}$
    - Example:
      - $f_\Theta(\theta) = 3,\ 1/3 \le \theta \le 2/3$
      - $E[(\Theta - \hat{\theta})^2] = \int_{1/3}^{2/3} (t - \hat{\theta})^2 \cdot 3\, dt$
      - At $\hat{\theta}=1/3$ ➔ $E[(\Theta - \hat{\theta})^2] = (\frac{1}{3})^3$
      - At $\hat{\theta}=1/2$ ➔ $E[(\Theta - \hat{\theta})^2] = 2 \cdot (\frac{1}{6})^3$



    - $E[(\Theta - \hat{\theta})^2] = Var(\Theta - \hat{\theta}) + (E[\Theta - \hat{\theta}])^2 = Var(\Theta) + (E[\Theta - \hat{\theta}])^2$



LMS

- Observe X ➔ Conditional Mean Squared Error
- $E[(\Theta-\hat{\theta})^2]$ ➔ $E[(\Theta-\hat{\theta})^2 \mid X = x]$

$$E[(\Theta-\hat{\theta})^2]$$
$$\hat{\theta} = \arg min_{\theta} E[(\Theta-\theta)^2]$$
$$\hat{\theta} = E[\Theta]$$

➔

$$E[(\Theta-\hat{\theta})^2 \mid X = x]$$
$$\hat{\theta} = \arg min_{\theta} E[(\Theta-\theta)^2 \mid X = x]$$
$$\hat{\theta} = E[\Theta \mid X=x]$$

c

X

LMS Example

- Example:
  - Random Variable: $\Theta \sim U[4, 10]$
    - $f_\Theta(\theta) = 1/6,\ 4 \le \theta \le 10$
  - Observation $X = \Theta + W$ where $W \sim U[-1, 1]$
    - $f_W(w) = \frac{1}{2},\ -1 \le w \le 1$
  - ➔ For a given $\Theta = \theta$, $X = \theta + W$ and $X \sim U[-1, 1]$ as W
  - $f_{X|\Theta}(x|\theta) = \frac{1}{2}, \theta - 1 \le x \le \theta + 1$
  - $f_{\Theta,X}(\theta, x) = f_\Theta(\theta) \cdot f_{X|\Theta}(x \mid \theta) = (1/6)(1/2) = 1/12$