$$\vec{x} = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^N \end{bmatrix}$$

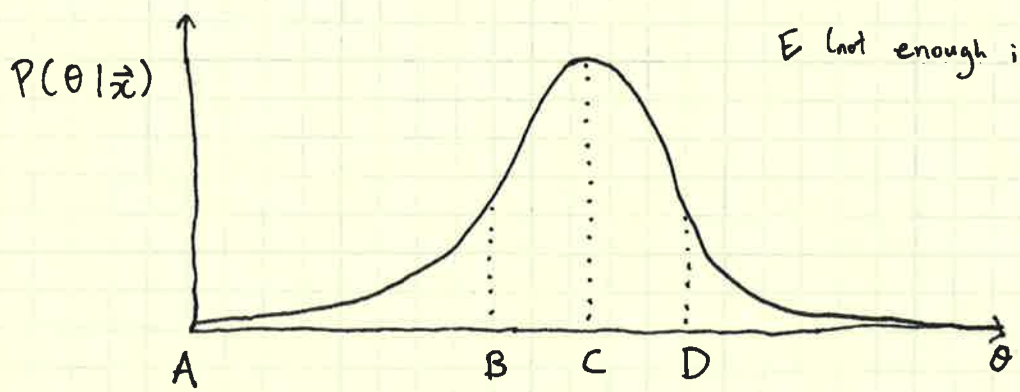
x^i is data (trial i)

hypothesis: $P(\vec{x} \mid \theta)$ → model parameter

$$\underbrace{P(\theta \mid \vec{x})}_{\text{Posterior}} = \frac{\overbrace{P(\vec{x} \mid \theta)}^{\text{Likelihood}}\ \overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{P(\vec{x})}_{\text{marginal probability}}}$$

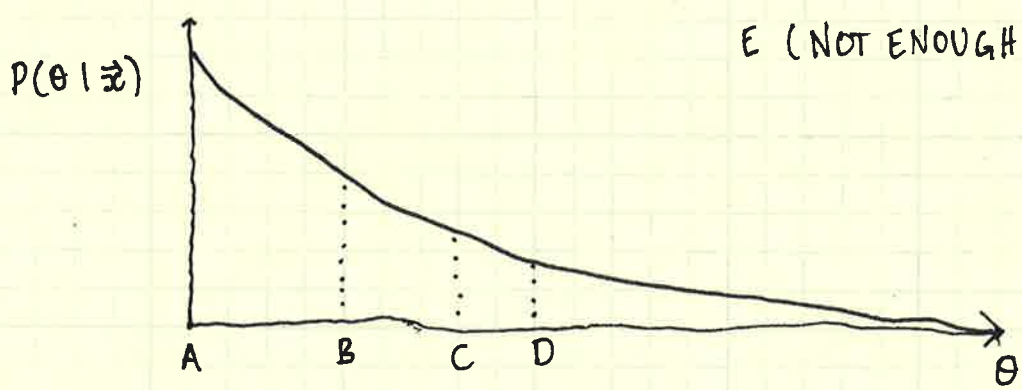
Which parameter to choose?

Ex. 1:



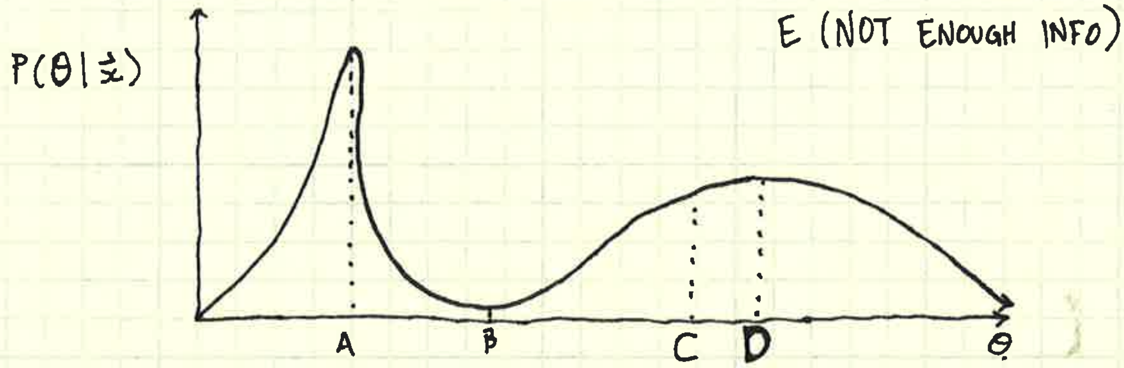
E (not enough info)

Ex 2:



E (NOT ENOUGH INFO)

Ex 3:



Maximum a posteriori (MAP) and Maximum Likelihood (ML)

MAXIMIZE $P(\theta|\vec{x})$, $P(\vec{x}|\theta)$

$$P(\theta|\vec{x}) = \frac{P(\vec{x}|\theta)\,P(\theta)}{P(\vec{x})}$$

Use MAP:

1) Prior is KNOWN and IMPORTANT
2) To show $P(\theta)$ UNIMPORTANT

USE ML:

1) Prior does not exist, or UNIMPORTANT
2) Convenience, tractability

ML estimate is asymptotically ($N \to \infty$) [if model is correct]

- UNBIASED ($\hat{\theta} \to \theta$)
- MINIMUM VARIANCE

TOPS 35502

Example: Fitting a Gaussian

$$P(x \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Likelihood: $P(\vec{x} \mid \mu,\sigma) = \prod_i P(x^i \mid \mu,\sigma)$ (Independence)

$$= \prod_i \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x^i-\mu)^2}{2\sigma^2}\right]$$

$$\log P(\vec{x} \mid \mu,\sigma) = -\sum_i \left(\log\sigma + \frac{(x^i-\mu)^2}{2\sigma^2} + \tfrac{1}{2}\log 2\pi \right)$$

$\left\{ \begin{array}{l} \log \text{ monotonic,} \\ \text{argmax}_\theta P(\vec{x} \mid \theta) = \text{argmax}_\theta \log P(\vec{x} \mid \theta) \end{array} \right.$

$$\frac{\partial}{\partial\mu}\left[\log P(\vec{x} \mid \mu,\sigma)\right] = \sum_i \frac{x^i-\mu}{\sigma^2} = \frac{1}{\sigma^2}\sum_i x^i - \frac{N}{\sigma^2}\mu$$

$$= 0 \underset{\text{extremum condition}}{\Longrightarrow} \hat{\mu} = \frac{1}{N}\sum_i x^i \qquad \boxed{\text{sample mean}}$$

$$\frac{\partial}{\partial\sigma}\left[\log P(\vec{x} \mid \mu,\sigma)\right] = -\sum_i \left(\frac{1}{\sigma} - \frac{(x^i-\mu)^2}{\sigma^3} \right)$$

$$= 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N}\sum_i (x^i - \hat{\mu})^2$$ sample variance...

$$\langle \hat{\mu} \rangle = \mu_0$$ (if data is from the model)

$\hat{\mu}$ unbiased

$$\langle \hat{\sigma}^2 \rangle = \left\langle \frac{1}{N}\sum_i (x^i - \hat{\mu})^2 \right\rangle = \left\langle \frac{1}{N}\sum_i \left((x^i-\mu) - (\hat{\mu}-\mu) \right)^2 \right\rangle$$

$$= \underset{①}{\frac{1}{N}\sum_i \langle (x^i-\mu)^2 \rangle} \; \underset{③}{- \frac{2}{N}\langle (\hat{\mu}-\mu) \sum_i (x^i-\mu) \rangle} \; \underset{②}{+ \sum_i \frac{1}{N}\langle (\hat{\mu}-\mu)^2 \rangle}$$

$$= \underset{①}{\frac{1}{N}(N\sigma^2)} + \underset{②}{\frac{N}{N}\left(\frac{\sigma^2}{N}\right)} - \underset{③}{\frac{2}{N^2}\sum_{i,j} \langle (x^i-\mu)(x^j-\mu) \rangle}$$

Since trials are independent,

$$\langle (x^i-\mu)(x^j-\mu) \rangle = \begin{cases} \sigma^2 & i=j \\ 0 & i \neq j \end{cases}$$

$$= \sigma^2 + \frac{1}{N}\sigma^2 - \frac{2}{N}\sigma^2 = \frac{N-1}{N}\sigma^2 \rightarrow \boxed{\text{BIASED}}$$