BAYESIAN ESTIMATION

UNSUPERVISED

WHY? When there's a small amount of data, Maximum Likelihood estimation might be poor → Bayesian Learning.

$\hat{\theta}$: random variables with a distribution.

PROPERTIES:

- Exploit the fact that there is uncertainty in estimating θ
- NOT a single value of θ
  ↳ Average over the uncertainty to estimate θ
- Use Bayes Theorem

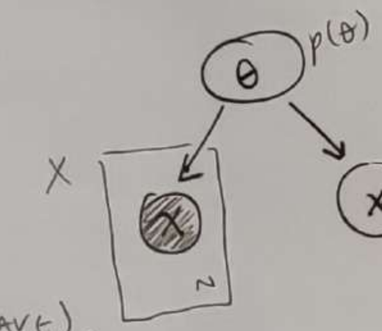
$$p(\theta|X) = \frac{p(\theta)\,p(X|\theta)}{p(X)} = \frac{\text{prior} \cdot \text{likelihood}}{\text{marginalization.}}$$

$p(\theta)$: prior distribution: belief of parameters before the data is shown.

$p(X|\theta)$: sample likelihood: how likely is the sample X is if θ takes a specific value.

$p(X)$: It's there to make sure that the posterior $p(\theta|X) \in [0,1]$

OBJECTIVE: create a generative model representing how data is created.



WHAT WE HAVE:

- ~~[illegible]~~ observed data $X = \{\vec{x}_t\}_{t=1}^{N}$
-

WHAT WE DON'T KNOW: (BUT WE WANT TO HAVE).

- DISTRIBUTION WITH UNKNOWN PARAMETER θ
- DRAWING OF ONE INSTANCE x'

OBJECTIVE

- CALCULATE THE PROBABILITY DISTRIBUTION $p(x'|X)$

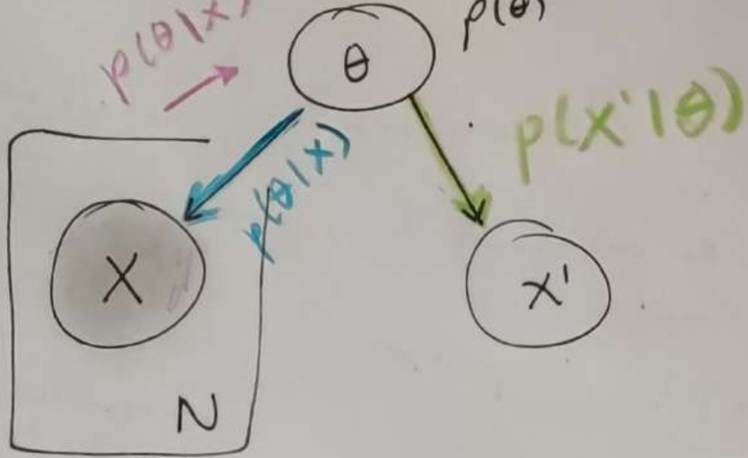
JOINT DISTRIBUTION OF BAYESIAN NETWORK:

$$p(x', X, \theta) = p(\theta)\,p(X|\theta)\,p(x'|\theta)$$

$$p(x'|X) = \frac{p(x', X)}{p(X)} = \frac{\sum_\theta p(x', X, \theta)}{p(X)} = \sum_\theta \underbrace{\left[\frac{p(\theta)\,p(X|\theta)}{p(X)}\right]}^{\text{BAYES}} p(x'|\theta)$$

$$p(x'|X) = \sum_\theta p(\theta|X)\,p(x'|\theta) \quad \text{or} = \int p(\theta|X)\,p(x'|\theta)\,d\theta$$

CALCULATING THE POSTERIOR $p(\theta|X)$, BAYES RULE INVERTS THE DIRECTION OF THE ARC AND MAKES A DIAGNOSTIC INFERENCE.



$p(\theta|X)$: POSTERIOR $\hat{=}$ DIAGNOSTIC INFERENCE.

$p(\theta|X)$: LIKELIHOOD

$p(x'|\theta)$: ESTIMATE

$$p(x'|X) = \int d\theta \cdot \underline{p(\theta|X)} \cdot \underline{p(x'|\theta)}$$