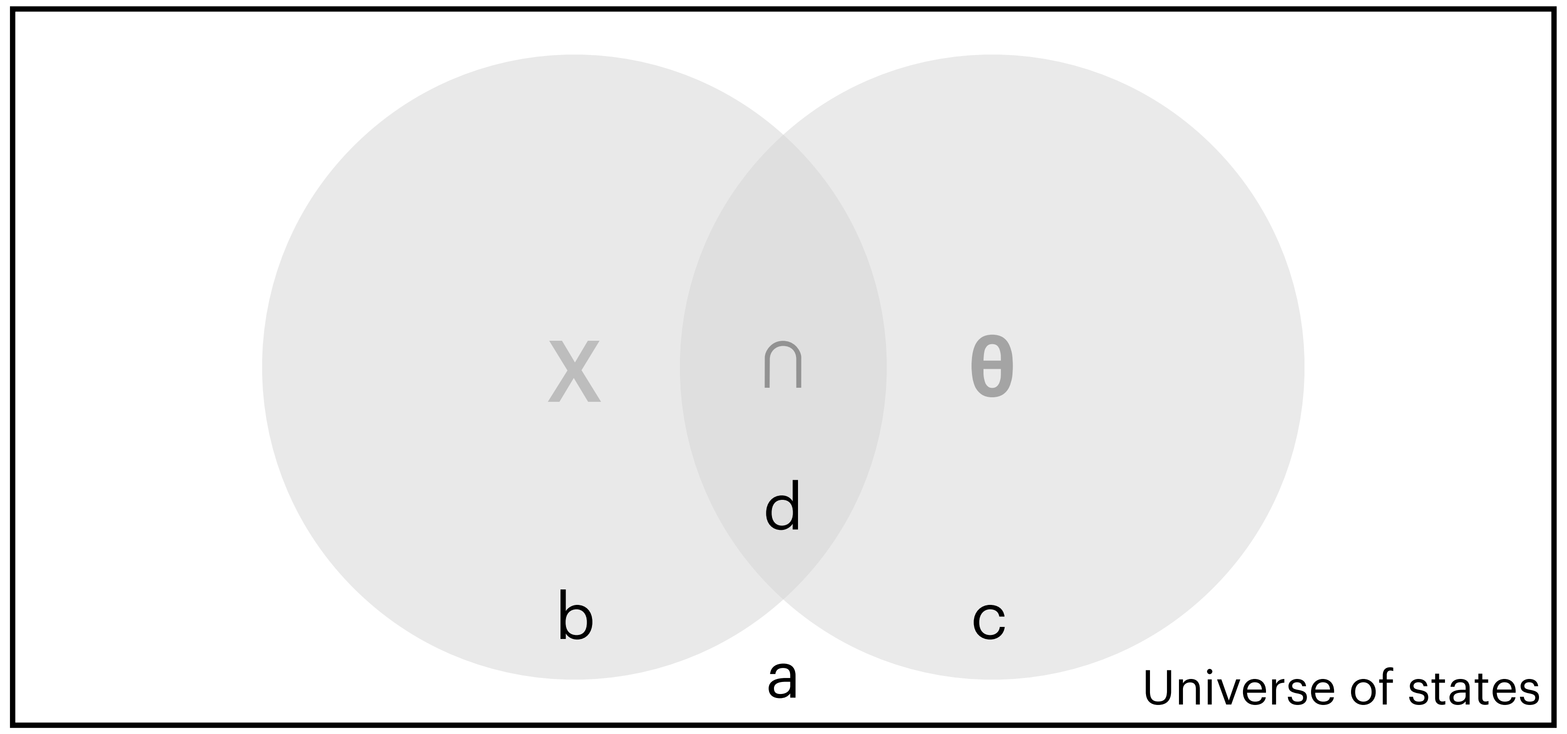
Bayes' Rule

- Intuitive statement: $P(x|\theta)P(\theta) = P(\theta|x)P(x)$

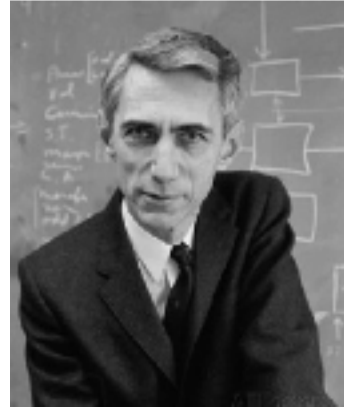
  - Proof $\frac{d}{\cancel{c}} \cdot \frac{\cancel{c}}{a} = \frac{d}{\cancel{b}} \cdot \frac{\cancel{b}}{a} = \frac{d}{a}$



- Bayes' Rule follows by deduction: $P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$ $\quad posterior = \frac{likelihood \times prior}{evidence}$

- Prescription for how to update a model given new evidence (i.e., new data)

Similarity measures between distributions



Claude Shannon

- Shannon postulated that any measure of the informativeness of an event, x, should satisfy three conditions:
  1. An event with probability 1 yields no information
  2. The probability of an event and the information it yields vary inversely with each other
  3. The total information coming from independent events is purely additive

  - Which he used to define *self-information*: $I(x) = -\log P(x)$

- Shannon entropy: $H(P) = \mathbb{E}_{x\sim P}[I(x)] = -\mathbb{E}_{x\sim P}[\log P(x)] = -\sum_{x\sim P} P(x)\log P(x)$

- Kullback-Leibler (KL) divergence: $D_{KL}(P||Q) = \mathbb{E}_{x\sim P}\left[\log \frac{P(x)}{Q(x)}\right]$

- Cross entropy: $H(P, Q) = H(P) + D_{KL}(P||Q) = -\mathbb{E}_{x\sim P}[\log Q(x)] = -\sum_{x\sim P} P(x)\log Q(x)$