L6: The Central Limit Theorem

→ Motivate Normal Dist

→ why we want a lot of data.

Jan 26, 2026

FoDA

Jeff M. Phillips

Review

Bayes' Thm

posterior $\Pr(\underline{M} \mid D) \propto \underbrace{\boxed{\Pr(D \mid M)}}_{\text{likelihood}} \cdot \underbrace{\Pr(M)}_{\text{prior}}$

proportional to

if likelihood + prior are <u>Normal</u>

$\hookrightarrow$ log(posterior) $\Longrightarrow$

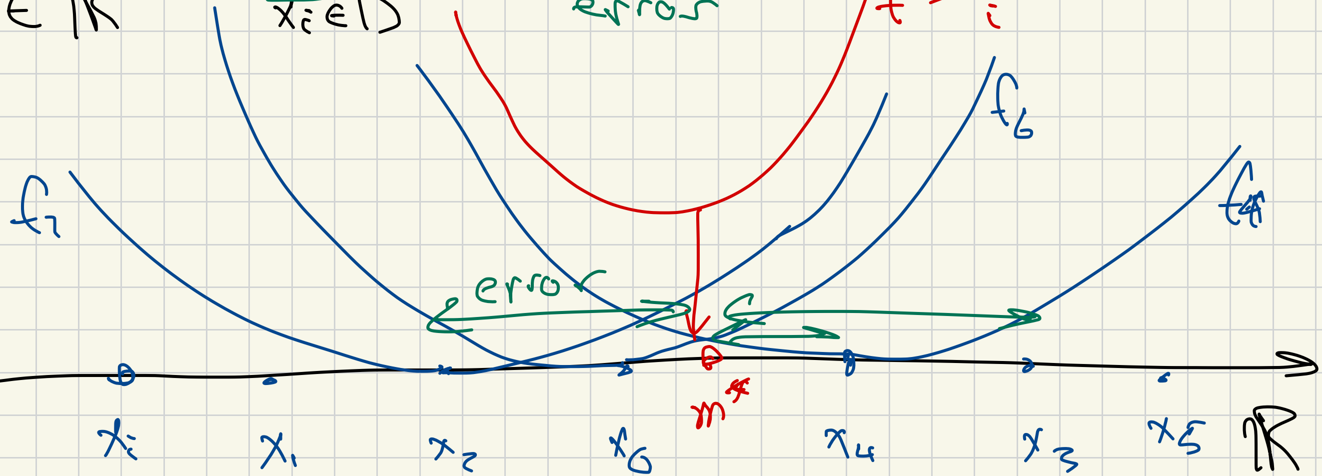
$\underset{m}{\text{argmax}}$ log(likelihood) + log(prior)

$\hookrightarrow$ $\underset{m \in \mathbb{R}}{\text{argmin}} \sum_{x_i \in D} (x_i - m)^2 + (\text{Prior} - m)^2$

Sum of Squared Errors (SSE)

$$\underset{m \in \mathbb{R}}{\operatorname{argmin}} \ \underbrace{\boxed{\sum_{x_i \in D}}}_{\text{sum}} \underbrace{(x_i - m)}_{\text{error}}^{\boxed{2}\ \text{square}}$$

$$f = \sum_i f_i$$



Algo

$$m^* = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \text{average}(\{x_u\}_i)$$

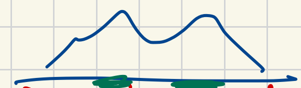
$$f_i(m) = (x_i - m)^2$$

Data and where it comes from?

now **Observations** set $P = \{p_1, p_2 \dots p_n\}$

fixed, given as input.

drawn iid from an unknown distribution pdf f.



iid : Independently and Identically Distributed

↳ probabilities are independent.
one p_i does not affect another $p_{i'}$

↳ from the same distribution

Random Variable $X \sim f$

RVs $X_1, X_2 \dots X_n \underset{iid}{\sim} f$

- X_1 independent of X_2 — independent
- $X_i \sim f$ — identical

Observations $P = \{p_1 \dots p_n\}$

Algo $\bar{P} = \frac{1}{n} \sum_i p_i$

$p_i \xleftarrow[\text{realize}]{} X_i$

data

$\bar{P} \underset{\frac{1}{n}\sum_i}{\Leftarrow} \{p_i\} \xleftarrow[\text{realize}]{} \{X_i\} \underset{iid}{\sim} f$

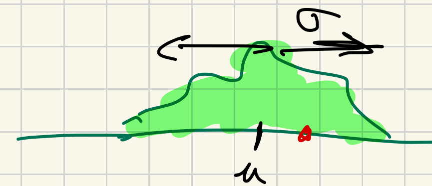
deterministic — probabilistic

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \Rightarrow$ Random Variable

Central Limit Theorem

n iid RVs $X_1, X_2, \dots X_n$ each $X_i \underset{iid}{\sim} f$

- $E[X_i] = \mu$ $\quad X_i \sim f$
- $Var[X_i] = \sigma^2 < \infty$ $\quad X_i \sim f$



$\hookrightarrow \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

① Converges to Normal Distribution

② $E[\bar{X}] = E[X_i] = \mu$

③ $Var[\bar{X}] = \frac{Var[X_i]}{n} = \frac{\sigma^2}{n}$

$\text{stddev}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

Probably Approximately Correct (PAC)

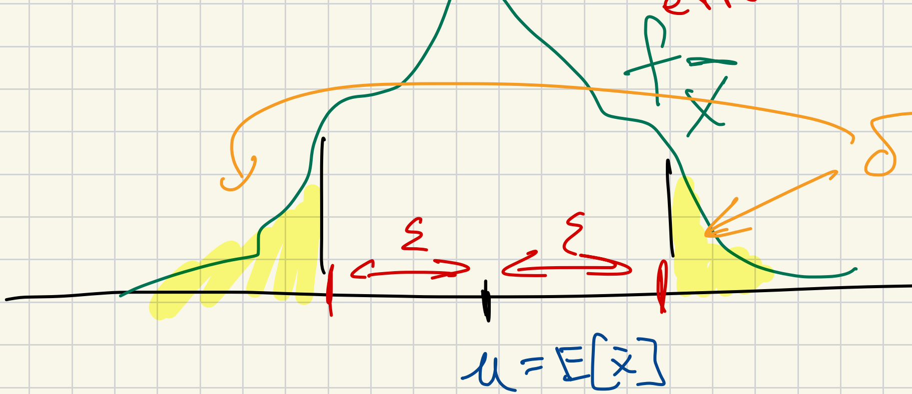
random process outcome

$\mu \equiv$ target value = true optimal model.

$$\Pr\left[\,\left|\bar{x} - E[\bar{x}]\right| > \varepsilon\,\right] \leq \delta$$

tolerable error

probability of failure



example

$\varepsilon = 10$

$\delta = 0{,}05$

Write equation ← nope
↳ inequality
↓ bounds

\# samples needed n
to get ε, δ-PAC bound

or

Given n samples, acceptable prob. failure δ

↳ write error guarantee
upto δ-prob. failure.