Monte Carlo Markov Chain

↳ gives us a way to draw a sample from any PDF

$$P(H|D) = \frac{P(D|H)\; P(H)}{P(D)}$$

posterior; likelihood of D if H ✓; prior; P(D) likelihood of evidence.

↓

we sample from P(H|D) when P(D) is a constant

we use it for posterior distribution.

↓

Based on constructing a markov chain that has this distribution as a stationary distribution

only if

— x —

Frequentist Regression

data generated from:

$$y = \beta^T x + \varepsilon$$

response; model param; input; random error.

OLS = ordinary least square

- single point estimate.
- calculate model parameters by minimizing sum of square errors.

$$L(y,\hat{y}) = |y-\hat{y}|^2$$

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

Bayesian Regression

data generated/sampled from

$$y \sim N(\beta^T x,\ \sigma^2)$$

normal dist; mean of gaussian; SD of gaussian

\+ parameters from posterior dist.

↓

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

$$\propto \text{likelihood} \times \text{prior}.$$

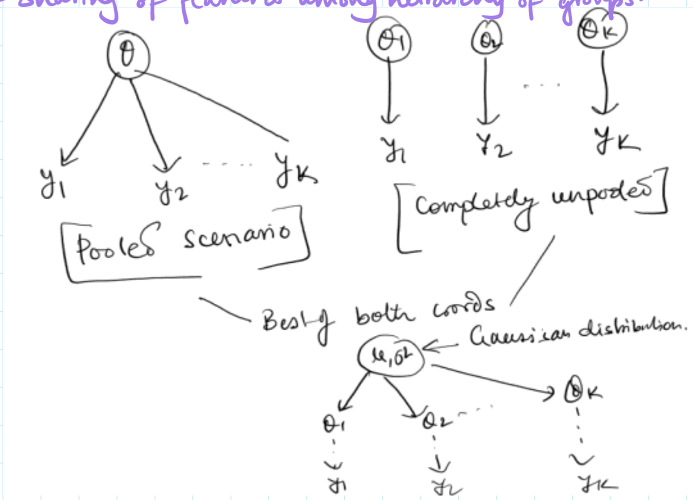
$$P(\beta|\epsilon) \propto P(\epsilon|\beta) \cdot P(\beta)$$

posterior of β given residuals; prob. density of residuals given params; prob. density of params

Gaussian prior ⟶ Gaussian Posterior

Heirarchical Bayesian Regression:

↳ sharing of features among heirarchy of groups.



When to do HBR:

- tiny Dataset, OLS impractical
- online learning
- band of predⁿ.
- regression problem
- dataset has nested relⁿ

When not to do it:

- lots of data
- likelihood dominates prior
- lines overlap.

Namita Achyuthan