Make tractable with MC sampling, importance sampling
(lower variance)

Big picture :-

We understand the data generation process

However, we only observe X.

Sample

$$z \sim P(z)$$

$$x \sim P_\theta(x|z)$$

Train

$$\max_\theta \sum_i \log P_\theta(x^{(i)}) = \sum_i \log \sum_z P_z(z) P_\theta(x^{(i)}|z)$$

max log likelihood

$$= \sum_x \log \sum_z P_\theta(x,z)$$

intractable

Representation $x \to z$

Next we will look at strategies to compute maximizing the log likelihood