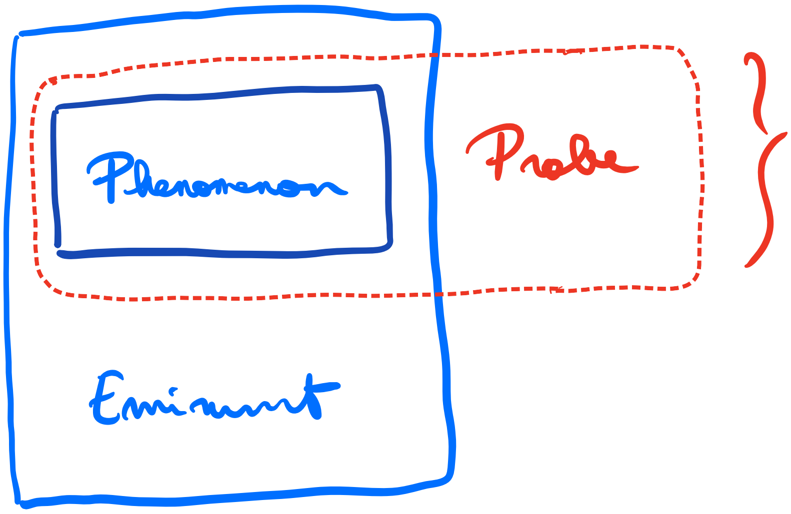
LEC04 Probabilistic modeling

I. Data generating process



Observational process

Eg.
- landing position of a ball
- images of galaxies taken by a telescope
- user preferences collected on a movie platform

$\Downarrow$?

Observations $x \in X$
:= Random outcomes of the probe

X ↓ Observational space

Often unknown or uncertain!

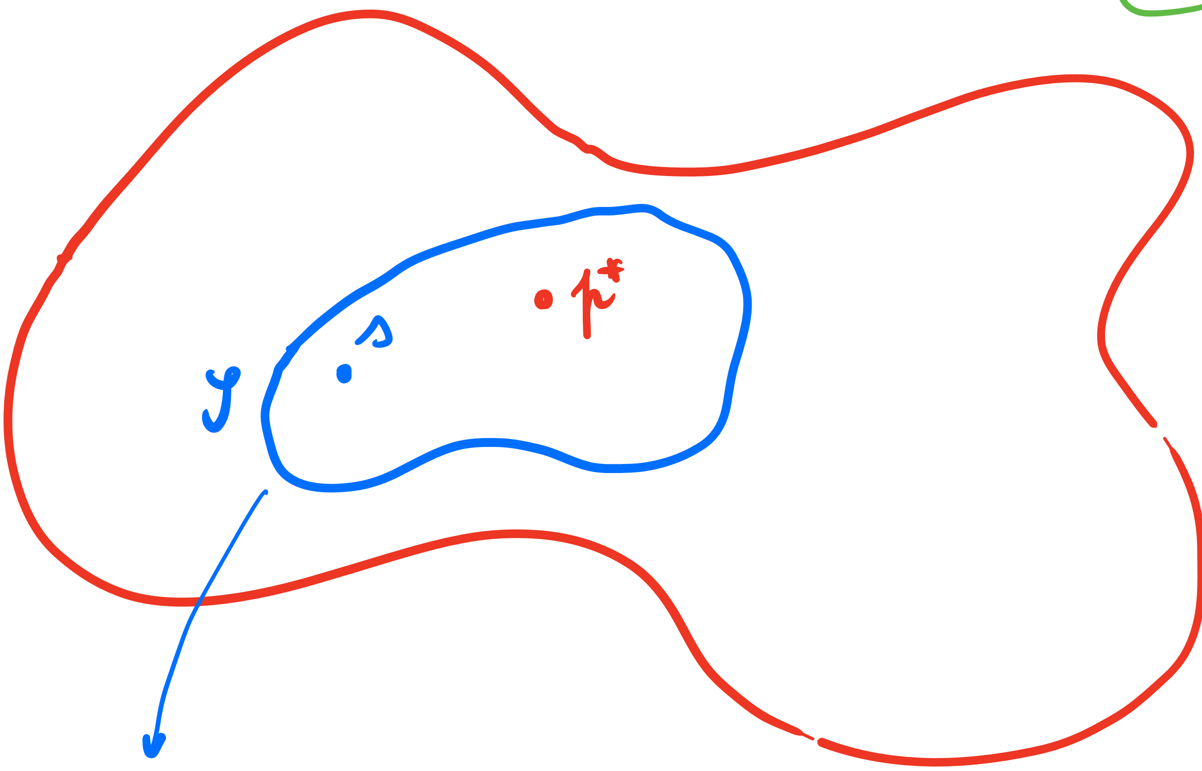
Data generating process = $p(x)$

True data generating process = $p^*(x)$

Abuse of notation, should be π

II. The observational model

What if $p^* \notin S$?



Observational model / Model configuration space

= set of assumed distributions s, each defining a possible mathematical narrative of how the data could be generated.

$\mathcal{P}$ set of all distributions over X.

Parametrization

$$\mathcal{Y} = \underbrace{\{ p_{\cancel{X}}(x;\theta) \}}_{\text{Family of PDFs (s) indexed by } \theta \in \Theta.}$$

⚠ Not unique!

Example:

$\mathcal{X} = \mathbb{R}$

$\mathcal{P}$ = set of distribution Gaussian density.

a) $\Theta = \mathbb{R} \times \mathbb{R}_0^+$

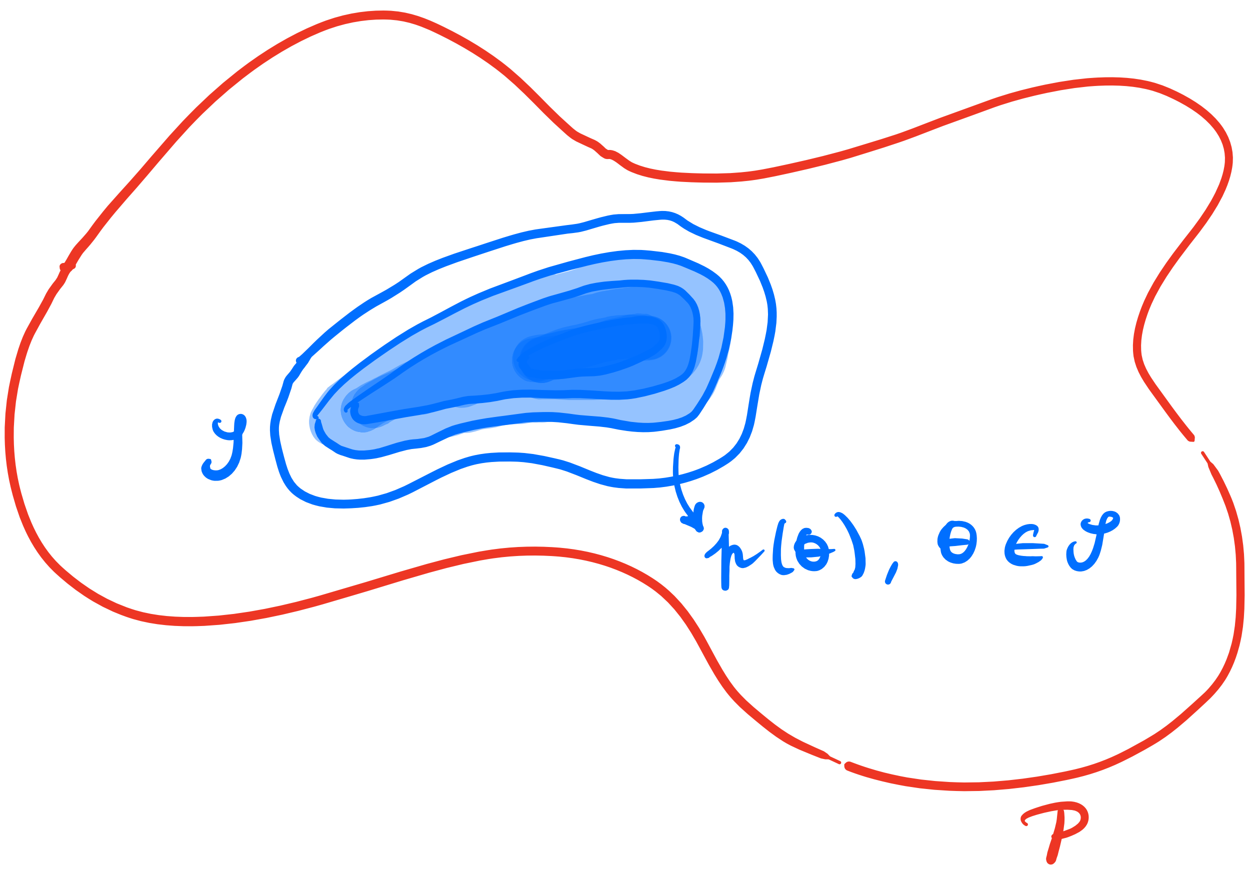
$$p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

b) $\Theta = \mathbb{R} \times \mathbb{R}$

$$p(x;\mu,\underbrace{\lambda}_{=\log\sigma}) = \frac{1}{\sqrt{2\pi}} e^{-\lambda} \exp\left(-\frac{1}{2}(x-\mu)^2 e^{-2\lambda}\right)$$

III. Bayesian models

All uncertainties are captured in a single unified framework, including epistemic uncertainty summarizing what we know (and don't) about the model configurations. $\Leftrightarrow$ Θ becomes a R.V.



Ingredients

* Prior model $p(\theta)$
* Observational model $p(x|\theta)$

The complete Bayesian model is the joint

$$\boxed{p(\theta, x) = p(\theta)\, p(x|\theta)}$$

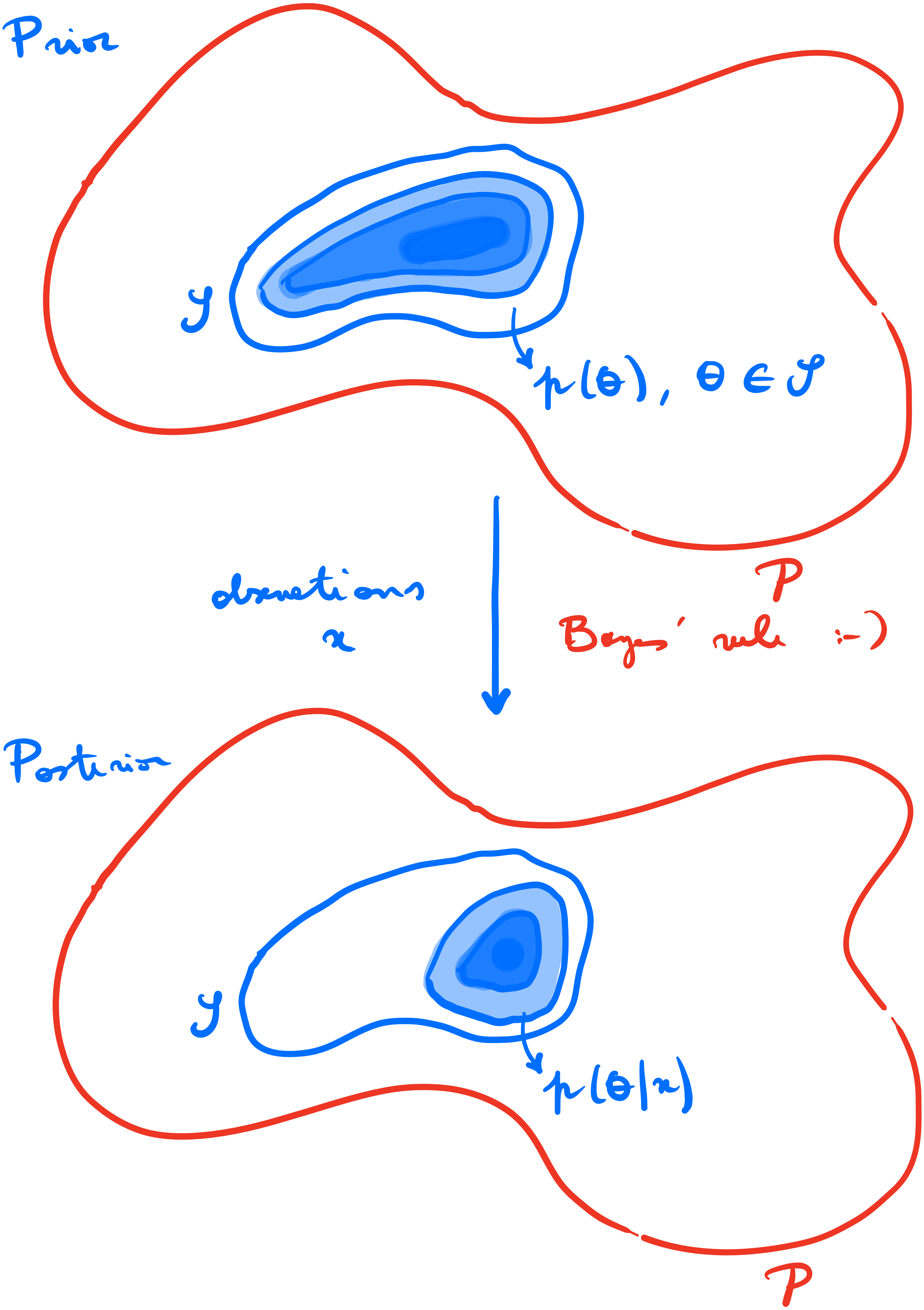
The joint probability perspective makes it easy to develop <u>generative</u> models that follow the structure of the true data generating process.

$$\Theta = (\Theta_{phenom}, \Theta_{env}, \Theta_{probe})$$

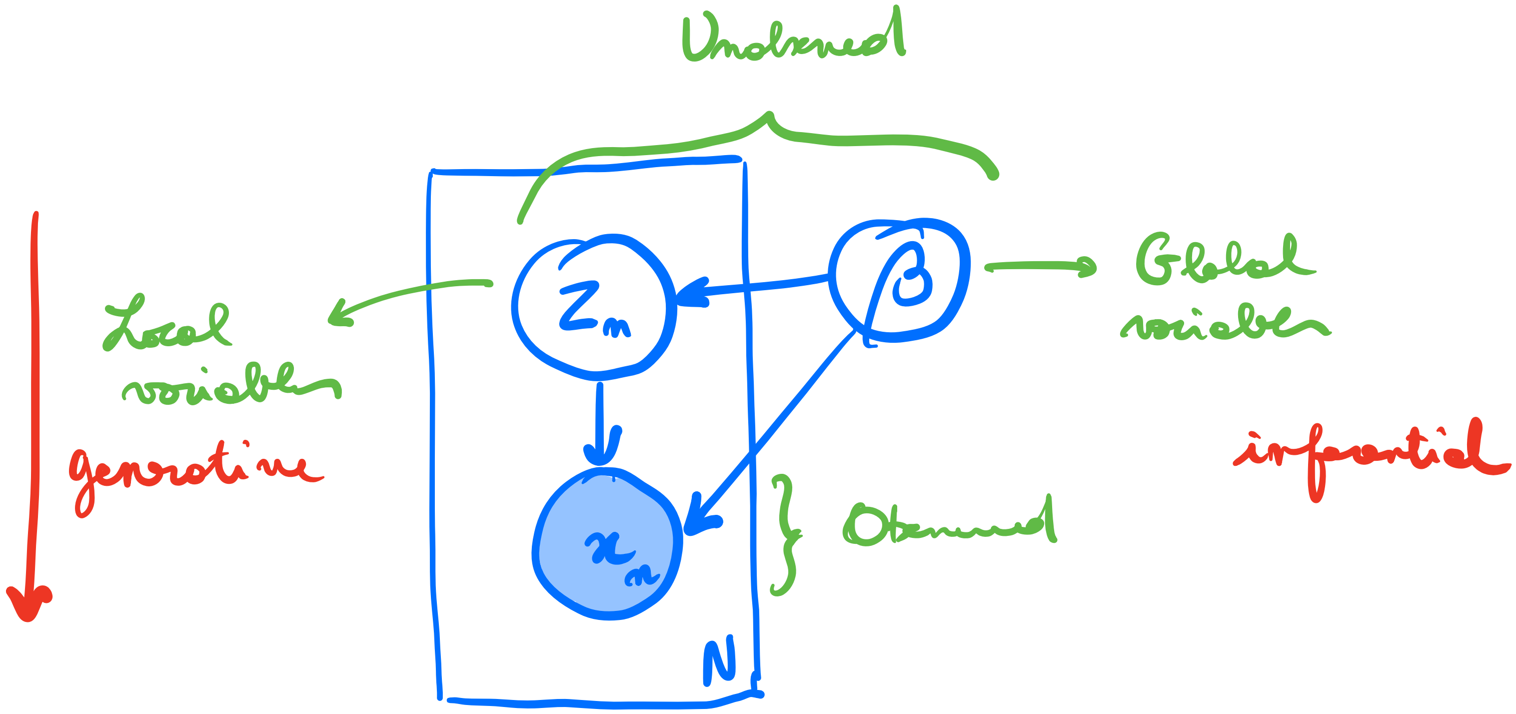
$$p(x, \Theta) = p(x, \Theta_{phenom}, \Theta_{env}, \Theta_{probe})$$
$$= p(x \mid \Theta_{phenom}, \Theta_{env}, \Theta_{probe})$$
$$\times\ p(\Theta_{probe} \mid \Theta_{phenom}, \Theta_{env})$$
$$\times\ p(\Theta_{env} \mid \Theta_{phenom})$$
$$\times\ p(\Theta_{phenom})$$

$\Leftrightarrow$ Formulate the observational model as a series of generative steps, from the phenomenon of interest down to the measurements.

Posterior inference

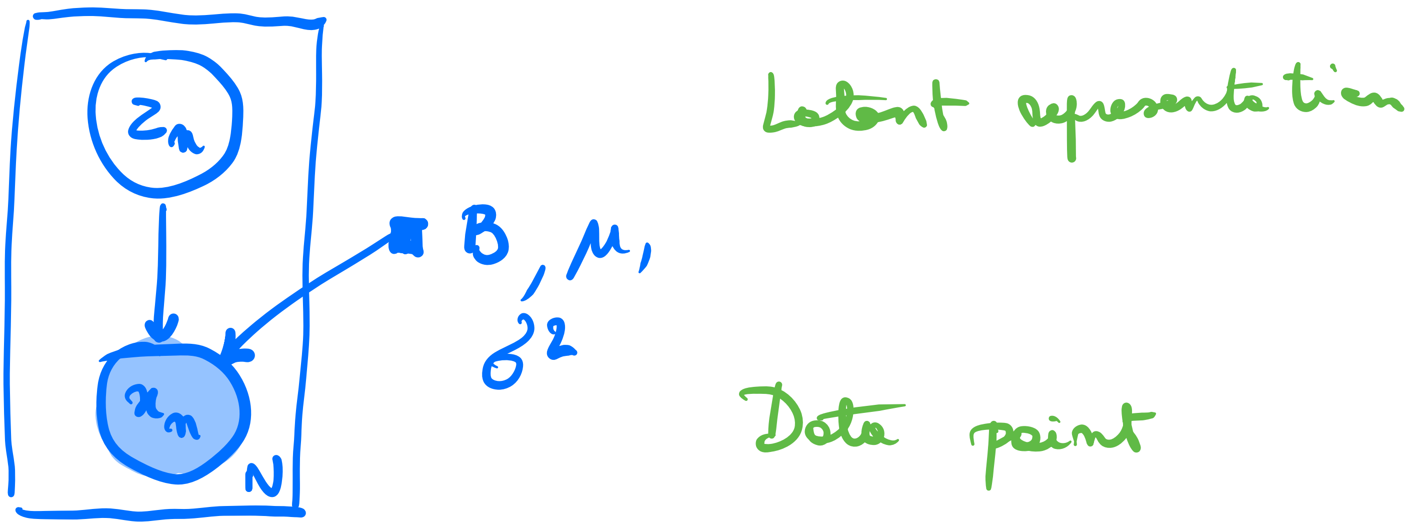


IV. Latent variable models



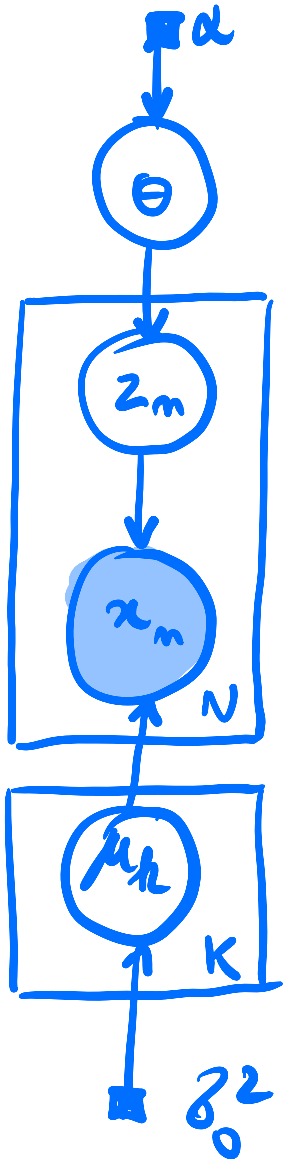
$$p(\beta, z, x) = p(\beta) \prod_{n=1}^{N} p(z_n|\beta)\, p(x_n|z_n, \beta)$$

a) PCA



$$p(z, x \mid B, \mu, \sigma^2)$$
$$= p(z)\, p(x \mid z, B, \mu, \sigma^2)$$
$$= \mathcal{N}(z \mid 0, I)\, \mathcal{N}(x \mid Bz + \mu, \sigma^2 I)$$

2) Mixture models



$\theta \in \Delta^K$

Mixture proportion — $\text{Dirichlet}_K(\alpha)$

Mixture component assigned to n — $\text{Discrete}(\theta)$ Multinomial

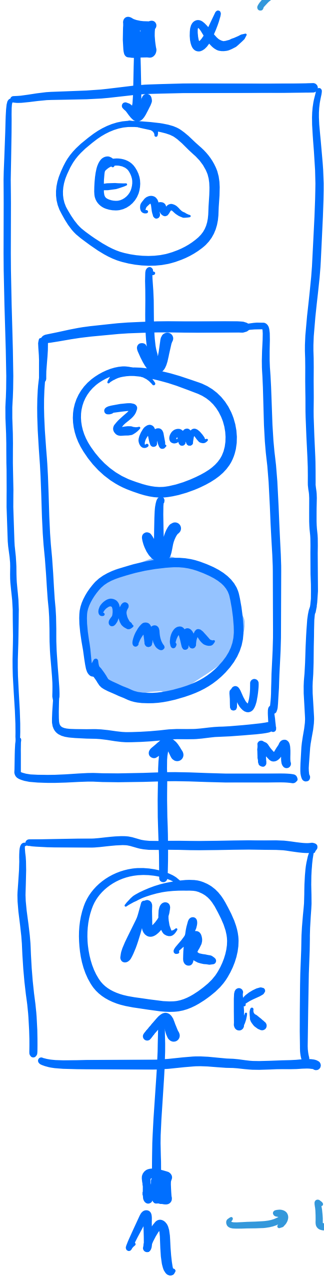
Data point — $\mathcal{N}(\mu_{z_n}, \sigma^2)$

Random position of the mixture components — $\mathcal{N}(0, \sigma_0^2)$

$$p(\theta, \mu, z, x \mid \alpha, \sigma^2)$$

$$= p(\theta \mid \alpha) \prod_{k=1}^{K} p(\mu_k \mid \sigma_0^2) \prod_{n=1}^{N} p(z_n \mid \theta)\, p(x_n \mid z_n, \mu)$$

c) Mixed membership model



→ LDA: proportion parameter

→ LDA: per-document topic proportion

Mixture proportion

$\text{Dirichlet}_K(\alpha)$

→ LDA: per-word topic assignment

Mixture component for x_{mn}

$\text{Discrete}_K(\theta_m)$

→ LDA: Observed word

Data point

$p(\cdot \mid \mu_{z_{mn}})$

$(\text{Discrete}_W(\mu_{z_{mn}}))$

→ LDA: Topics

Mixture component parameter

$p(\cdot \mid \eta)$

$(\text{Dirichlet}_W(\beta))$

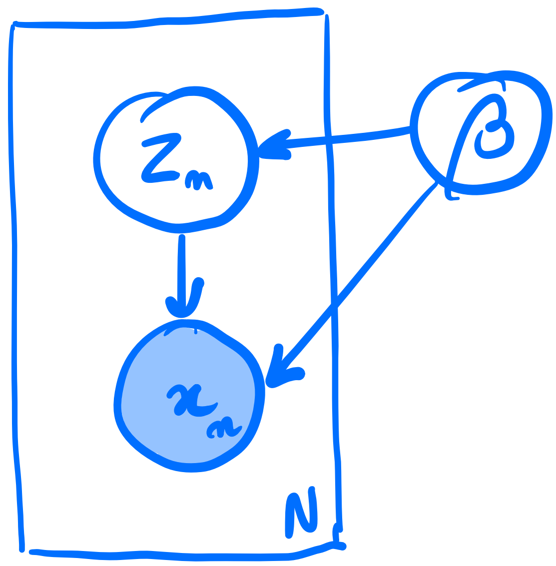
↓ number of words

→ LDA: topic parameter

$$p(\theta, \mu, z, x \mid \alpha, \eta)$$

$$= \prod_{k=1}^{K} p(\mu_k \mid \eta) \prod_{m=1}^{M} p(\theta_m \mid \alpha) \prod_{n=1}^{N} p(z_{mn} \mid \theta_m)\, p(x_{mn} \mid \mu_{z_{mn}})$$

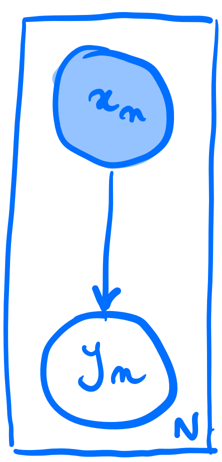
d) Probabilistic programs

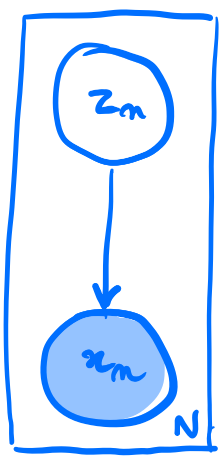


Show Naman's
Molavi Kan's

$$p(\beta, z, x) = p(\beta) \prod_{n=1}^{N} \underbrace{p(z_n|\beta)\, p(x_n|z_n, \beta)}_{\text{forward generation model / computer simulations}}$$

e) Discriminative / generative ML models





Also...

linear factor models

Time series models

Matrix factorization

Multilevel regression

...

V. Prior predictive checks

Often, the observational model is understood much better than the prior model.

$\Rightarrow$ We need to investigate the consequences of the prior model in the context of the observational model.

Summary functions

$t : X \longrightarrow T$ $\qquad\qquad$ $t : \Theta \longrightarrow T$

$\Leftarrow$ "statistics" $\qquad\qquad$ some intermediate quantity

Prior predictive checks

$$x \sim p(x) \longrightarrow \theta \sim p(\theta), \; x \sim p(x \mid \theta)$$

Then plot $t(x)$ to evaluate the assumptions

Summary

* Bayesian data analysis provides a principled framework for modeling and inference.

* Generative models capture the generative structure of the true data generating process.
  observed

They make all assumptions explicit.