Bayesian Machine Learning for Scientific Research

Maximizing information from data

Ezequiel Alvarez
Organizer: Rogerio Rosenfeld
sequi@unsam.edu.ar
ICTP-SAIFR October 2024

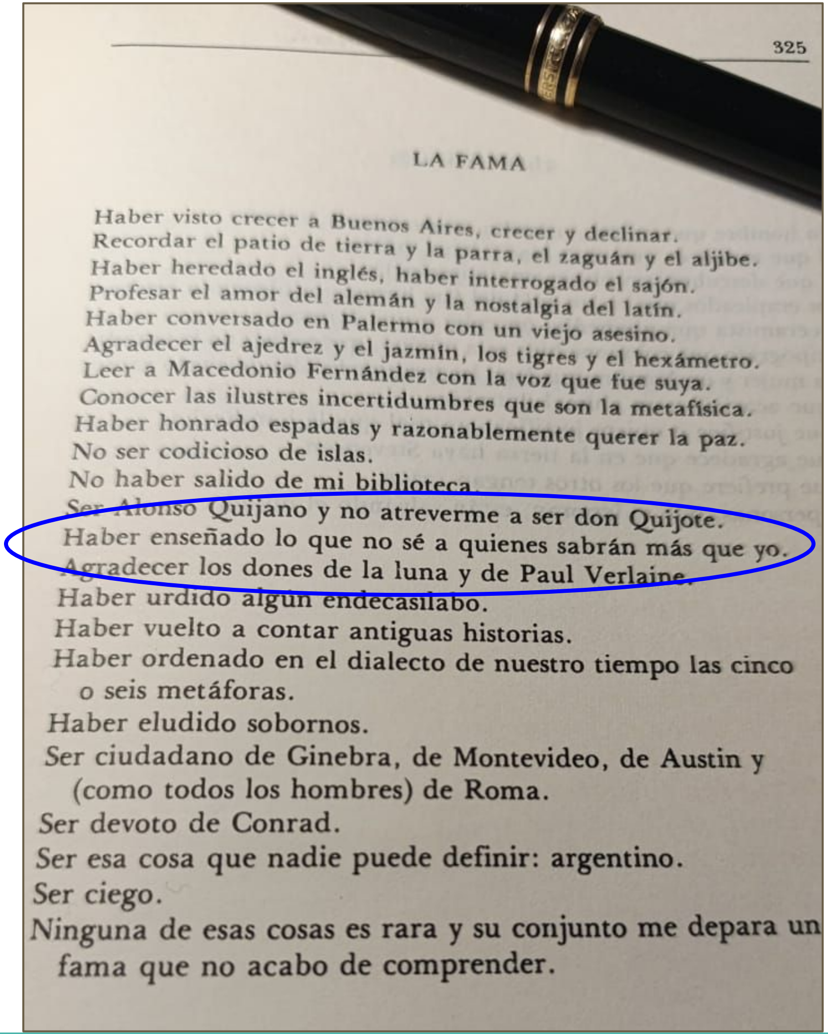
Disclaimer

I'm not a statistician...I'm learning!

Disclaimer

I'm not a statistician...I'm learning!

J.L.Borges: *La fama* (1981)



325

LA FAMA

Haber visto crecer a Buenos Aires, crecer y declinar.
Recordar el patio de tierra y la parra, el zaguán y el aljibe.
Haber heredado el inglés, haber interrogado el sajón.
Profesar el amor del alemán y la nostalgia del latín.
Haber conversado en Palermo con un viejo asesino.
Agradecer el ajedrez y el jazmín, los tigres y el hexámetro.
Leer a Macedonio Fernández con la voz que fue suya.
Conocer las ilustres incertidumbres que son la metafísica.
Haber honrado espadas y razonablemente querer la paz.
No ser codicioso de islas.
No haber salido de mi biblioteca.
Ser Alonso Quijano y no atreverme a ser don Quijote.
Haber enseñado lo que no sé a quienes sabrán más que yo.
Agradecer los dones de la luna y de Paul Verlaine.
Haber urdido algún endecasílabo.
Haber vuelto a contar antiguas historias.
Haber ordenado en el dialecto de nuestro tiempo las cinco
o seis metáforas.
Haber eludido sobornos.
Ser ciudadano de Ginebra, de Montevideo, de Austin y
(como todos los hombres) de Roma.
Ser devoto de Conrad.
Ser esa cosa que nadie puede definir: argentino.
Ser ciego.
Ninguna de esas cosas es rara y su conjunto me depara una
fama que no acabo de comprender.

What is Machine Learning ?

What is Machine Learning ?

The capacity of a machine to provide an answer in the cases for which it has not been programmed

Popular definition

What is Machine Learning ?

The capacity of a machine to provide an answer in the cases for which it has not been programmed

Popular definition

To learn the pattern in the data and provide answers according to that pattern

More technical

What is Machine Learning ?

The capacity of a machine to provide an answer in the cases for which it has not been programmed

Popular definition

To learn the pattern in the data and provide answers according to that pattern

More technical

To learn the PDF of the data and then being able to assess, predict , generate, etc.

More scientific

What is Machine Learning ?

Bayesian: assume data being sampled from a PDF, infer its parameters and learn the internal structure of the data

The capacity of a machine to provide an answer in the cases for which it has not been programmed

Popular definition

To learn the pattern in the data and provide answers according to that pattern

More technical

To learn the PDF of the data and then being able to assess, predict , generate, etc.

More scientific

What is Machine Learning ?

The capacity to provide the cases fo not been p

Popular

attern in the ide answers hat pattern

hnical

Bayesian: assume data being sampled from a PDF, infer its parameters and learn the internal structure of the data

To learn the PDF of the data and then being able to assess, predict , generate, etc.

More scientific

Data & Information

Data & Information

- **Data:** Measure N times the length of the table

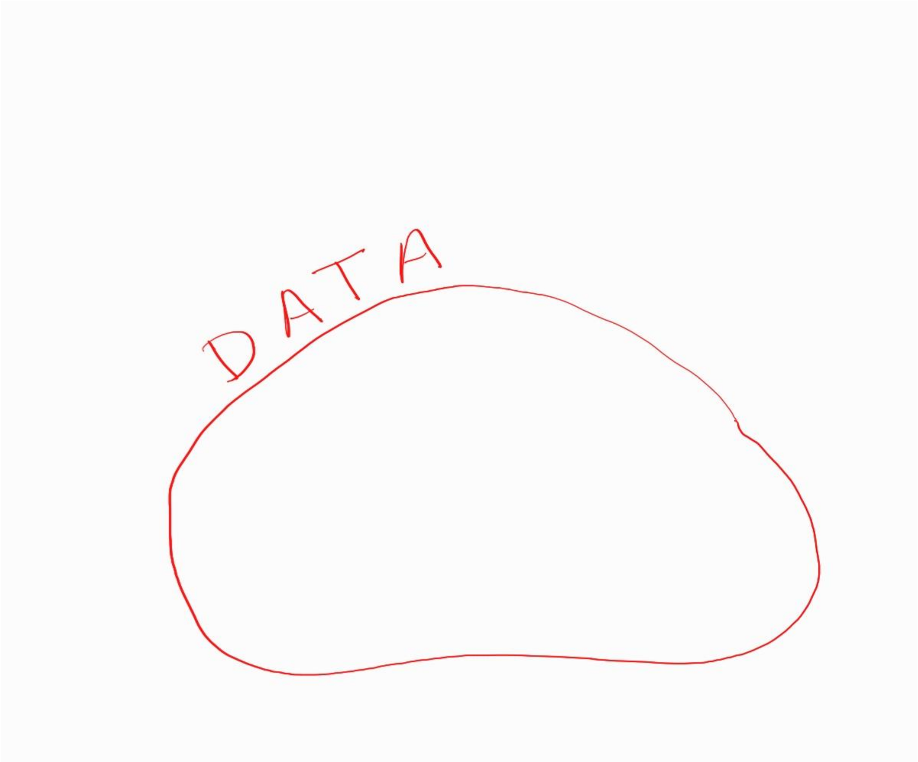
- **Information:** An estimation on the length of the table

Data & Information

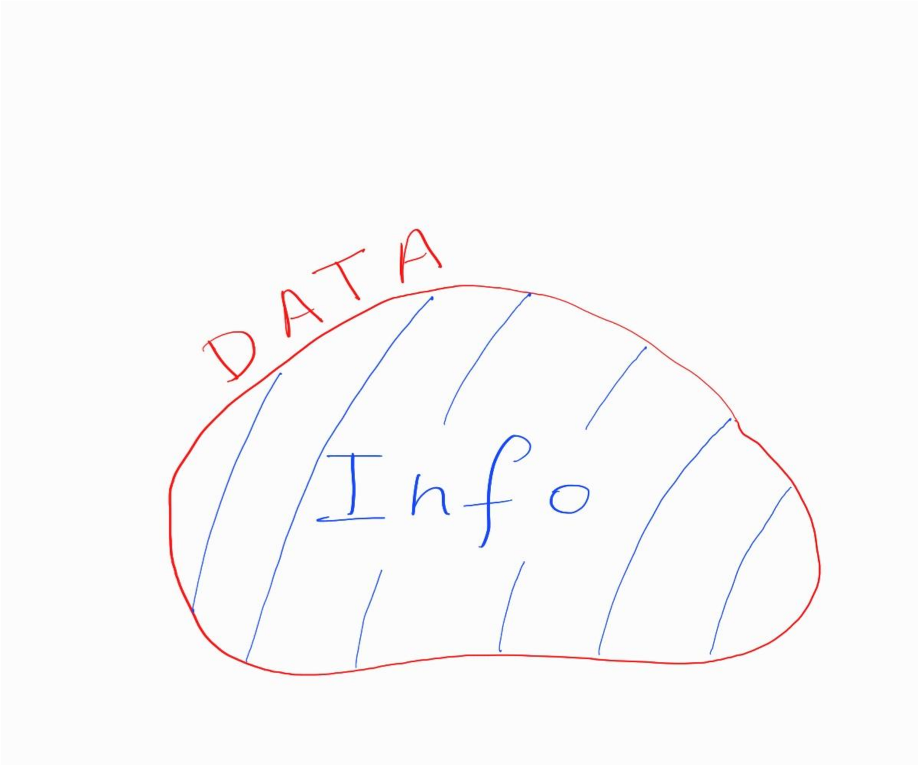
- **Data:** Measure N times the length of the table
- **Information:** An estimation on the length of the table

There is not such a thing as intrinsic information in the Data

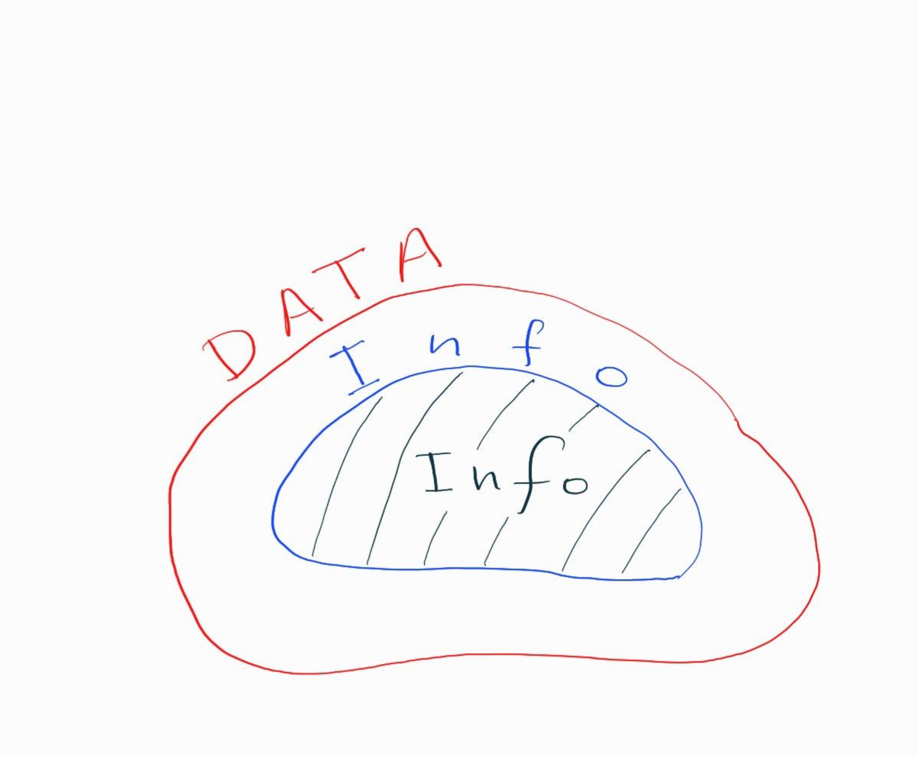
Data & Information



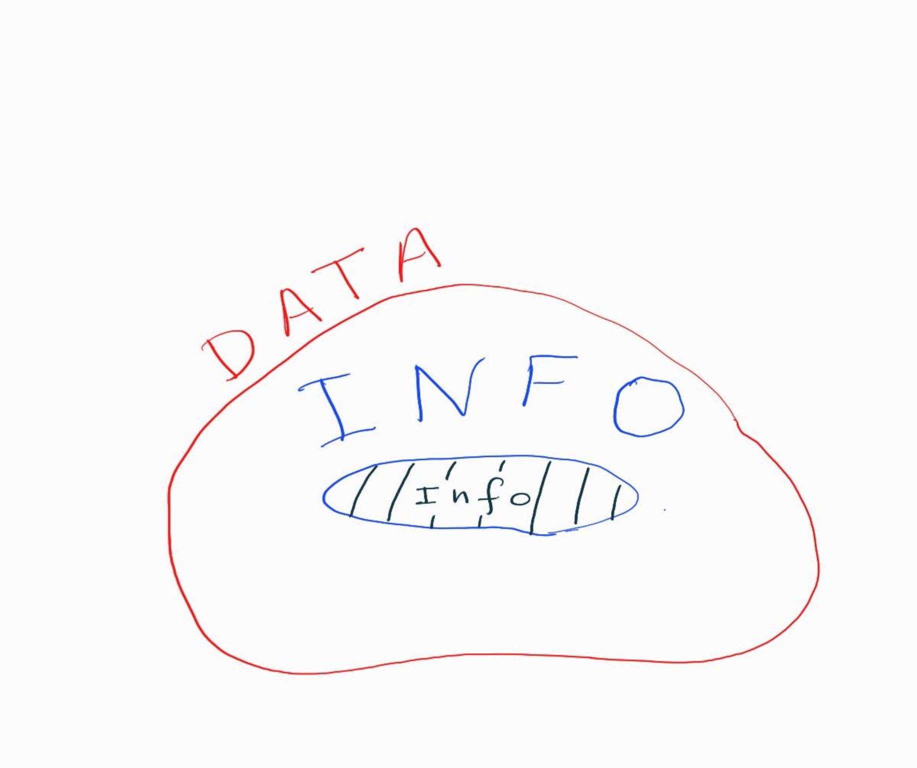
Data & Information



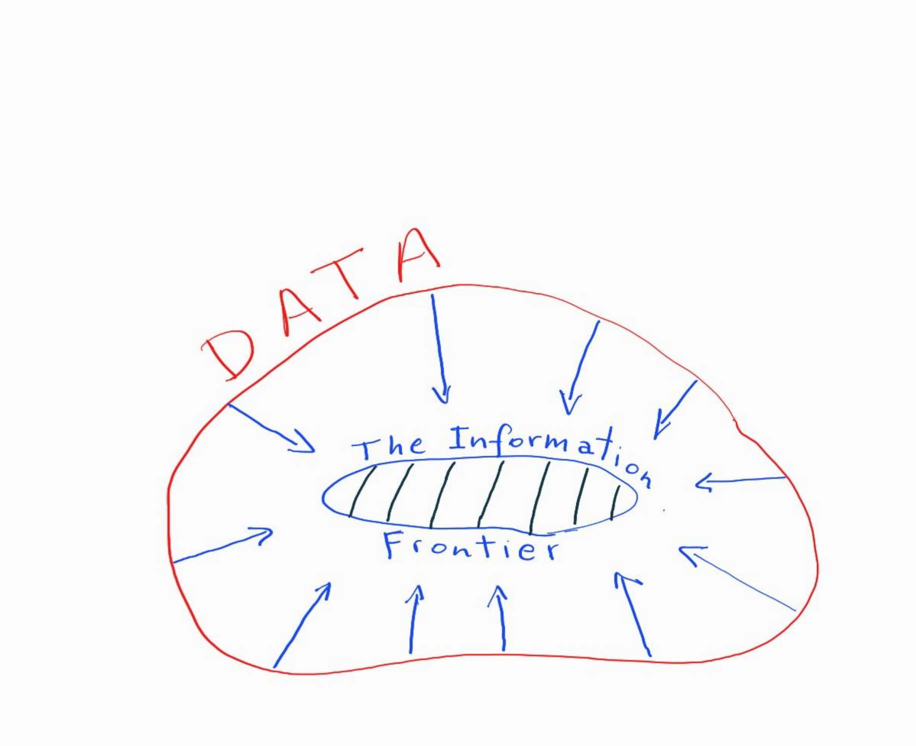
Data & Information



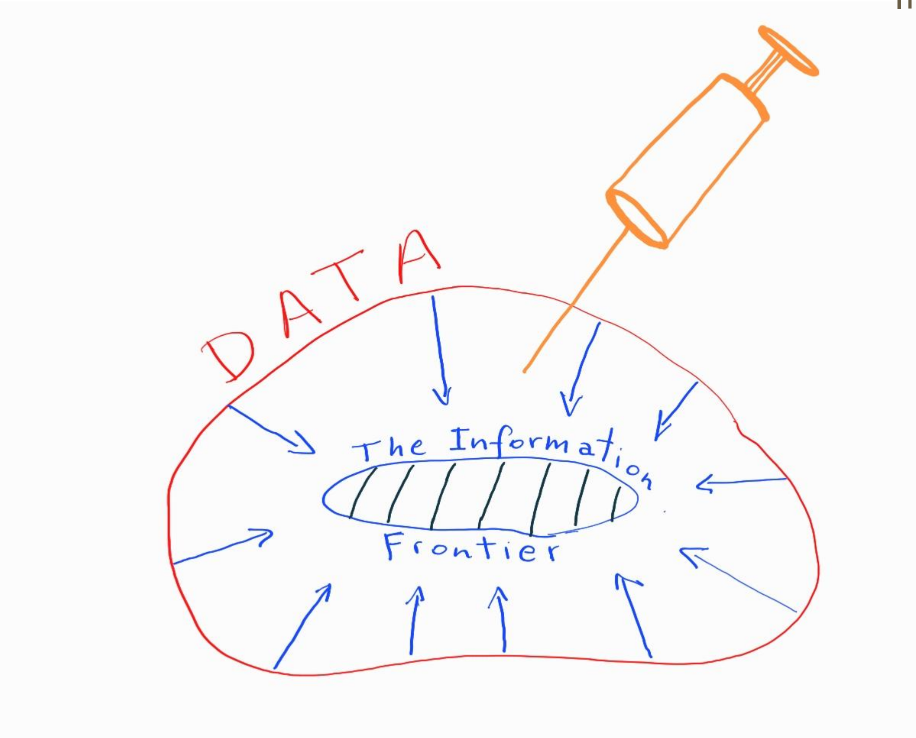
The Information Frontier



The Information Frontier



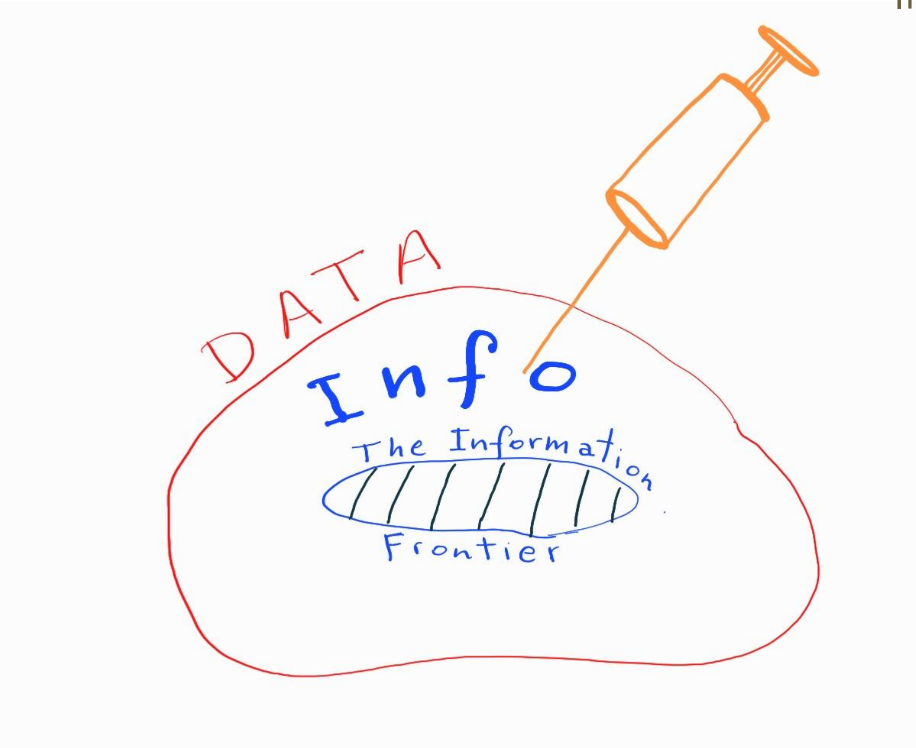
The Information Frontier



Inject *catalysts*:

- Modeling
- Tools & techniques
- Prior info

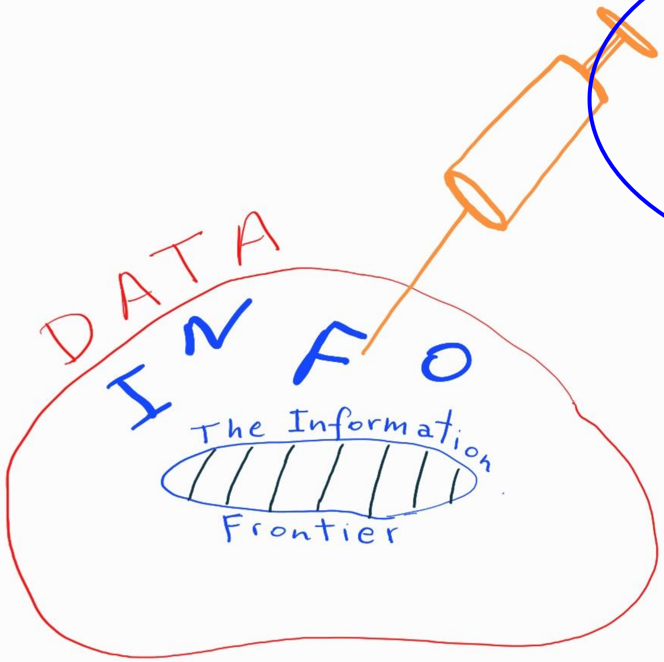
The Information Frontier



Inject *catalysts*:

- Modeling
- Tools & techniques
- Prior info

The Information Frontier

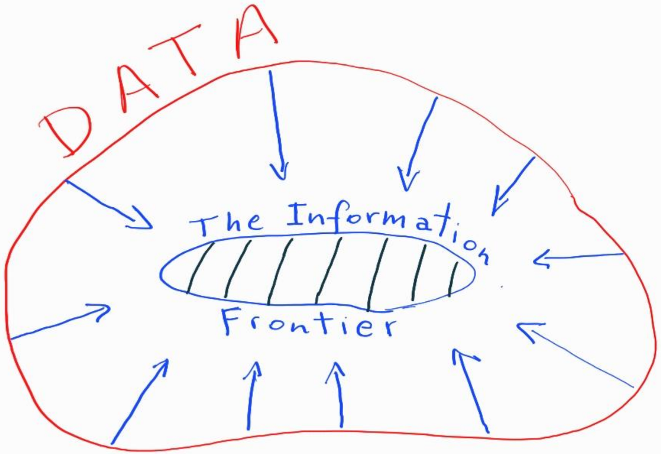


Inject *catalysts*:

- Modeling
- Tools & techniques
- Prior info

The Information Frontier

Science data is very complex and sophisticated



Different tools can explore differently this frontier

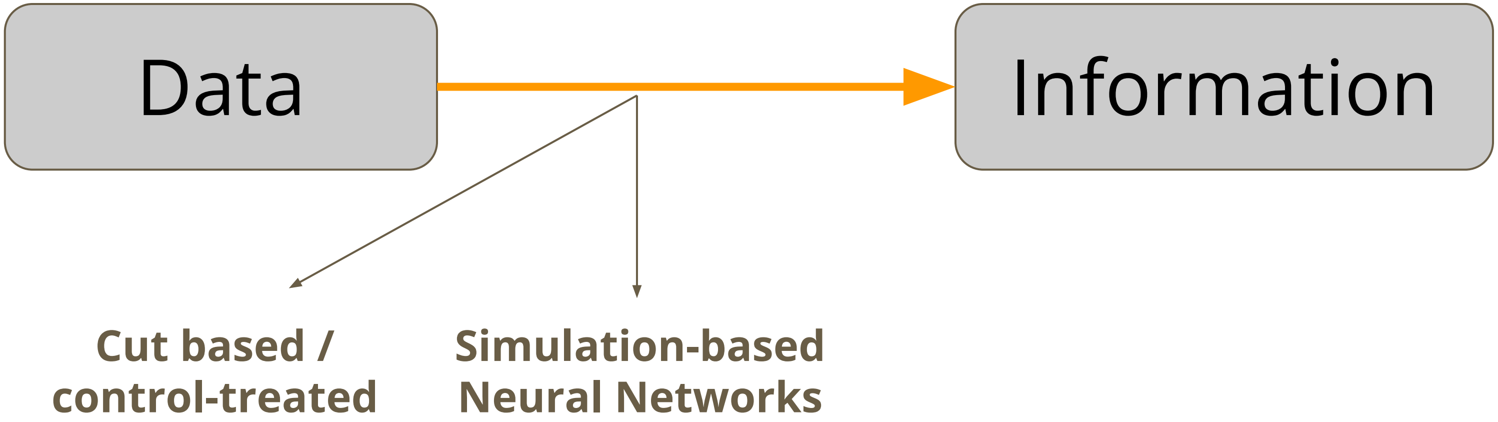
The Information Frontier

Scien
comp
soph

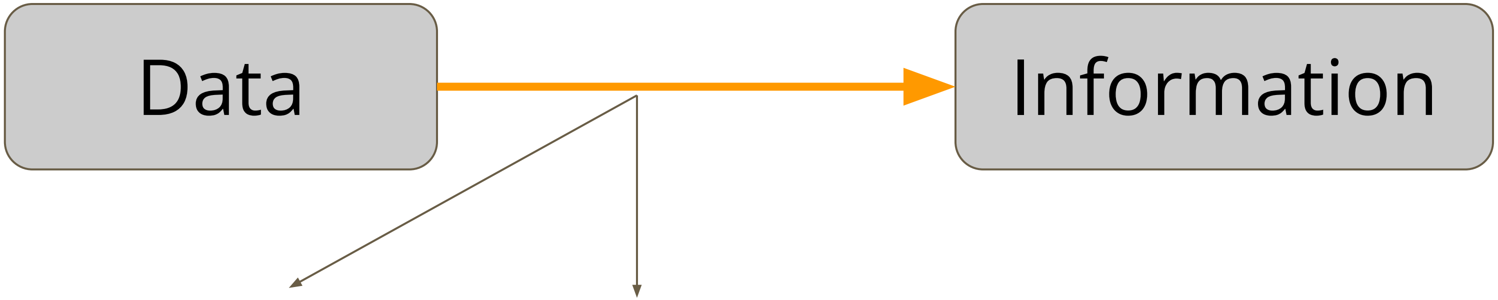
Different tools can explore differently this frontier

Typical problem in science

Typical problem in science



Typical problem in science

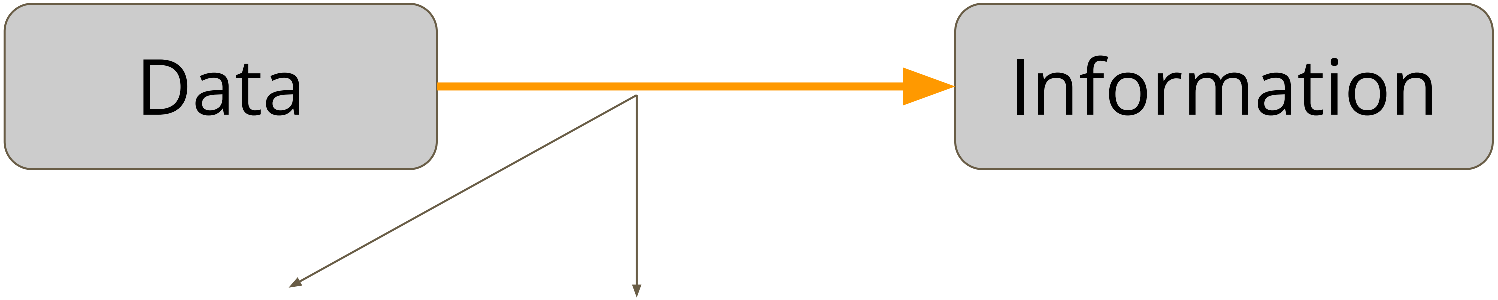


Cut based / control-treated

Simulation-based Neural Networks

Misses to take benefit of correlations

Typical problem in science



Cut based / control-treated

Misses to take benefit of correlations

Simulation-based Neural Networks

Extracts correlations, but learns "too much" from simulations

Typical problem in science

Data → Information

Cut based / control-treated

Misses to take benefit of correlations

Simulation-based Neural Networks

Extracts correlations, but learns "too much" from simulations

Bayesian tools

This minicourse! Reduce simulation's impact at the price of increasing modelling impact

The danger of NN in natural sciences

arXiv > stat > arXiv:2405.18095

Search...

Hel

Statistics > Machine Learning

[Submitted on 28 May 2024 (v1), last revised 31 May 2024 (this version, v2)]

Is machine learning good or bad for the natural sciences?

David W. Hogg (NYU, MPIA, Flatiron), Soledad Villar (JHU, Flatiron)

Machine learning (ML) methods are having a huge impact across all of the sciences. However, ML has a strong ontology - in which only the data exist - and a strong epistemology - in which a model is considered good if it performs well on held-out training data. These philosophies are in strong conflict with both standard practices and key philosophies in the natural sciences. Here we identify some locations for ML in the natural sciences at which the ontology and epistemology are valuable. For example, when an expressive machine learning model is used in a causal inference to represent the effects of confounders, such as foregrounds, backgrounds, or instrument calibration parameters, the model capacity and loose philosophy of ML can make the results more trustworthy. We also show that there are contexts in which the introduction of ML introduces strong, unwanted statistical biases. For one, when ML models are used to emulate physical (or first-principles) simulations, they amplify confirmation biases. For another, when expressive regressions are used to label datasets, those labels cannot be used in downstream joint or ensemble analyses without taking on uncontrolled biases. The question in the title is being asked of all of the natural sciences; that is, we are calling on the scientific communities to take a step back and consider the role and value of ML in their fields; the (partial) answers we give here come from the particular perspective of physics.

Bayesian ML for Scientific Research

Summary

Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*

Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*
- Lecture 2: Simple Bayesian examples

Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*
- Lecture 2: Simple Bayesian examples
- Lecture 3: Mixture models

Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*
- Lecture 2: Simple Bayesian examples
- Lecture 3: Mixture models
- Lecture 4: Consistency between data, model and results

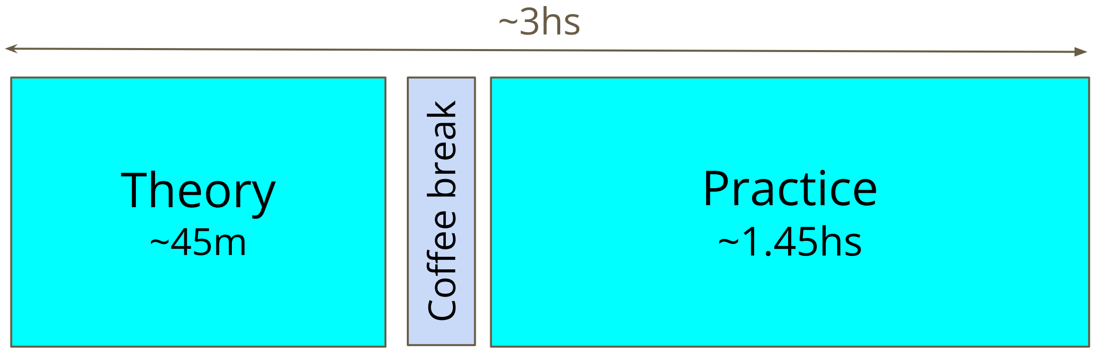
Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*
- Lecture 2: Simple Bayesian examples
- Lecture 3: Mixture models
- Lecture 4: Consistency between data, model and results
- Lecture 5: Mixture models for non-parametric distributions

Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*
- Lecture 2: Simple Bayesian examples
- Lecture 3: Mixture models
- Lecture 4: Consistency between data, model and results
- Lecture 5: Mixture models for non-parametric distributions

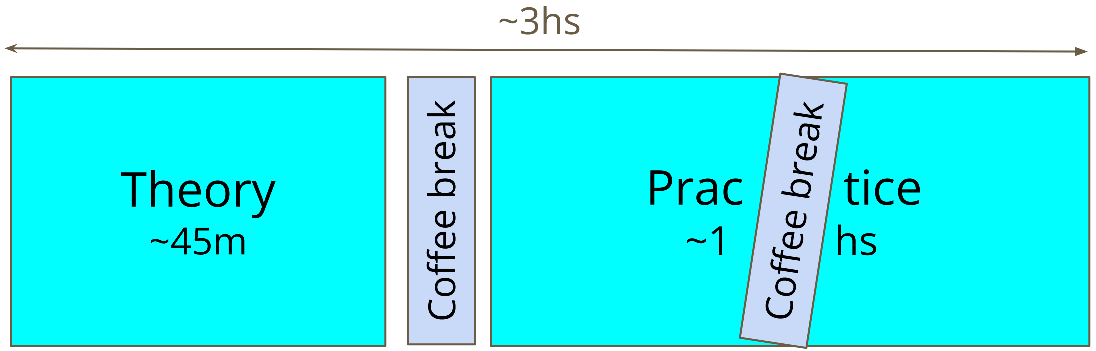
Lectures setup



Lectures summary

- Lecture 1: Introduction to Bayesian *techniques*
- Lecture 2: Simple Bayesian examples
- Lecture 3: Mixture models
- Lecture 4: Consistency between data, model and results
- Lecture 5: Mixture models for non-parametric distributions

Lectures setup



Introduction to Bayesian *techniques*

Bayesian Inference

Bayes Theorem:

$$p(\theta | X) = \frac{p(X | \theta) * p(\theta)}{p(x)}$$

Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

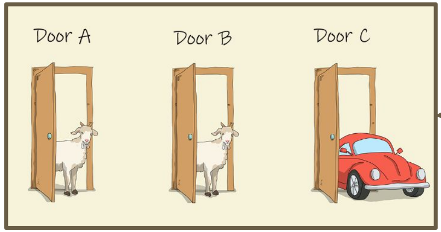
Used in very different contexts

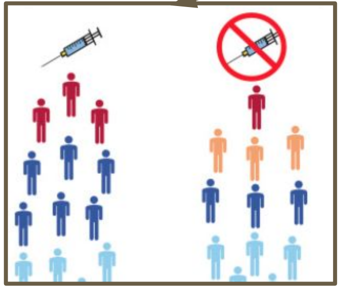
Bayesian Inference

Bayes Theorem:

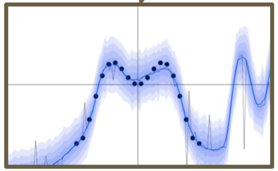
$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Used in very different contexts





... ...

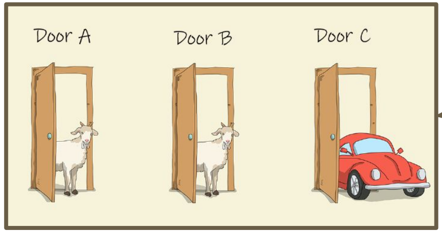


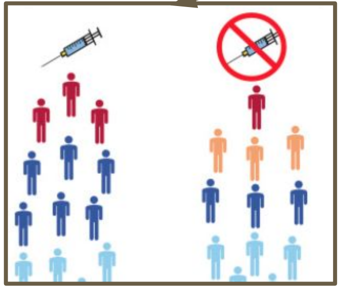
Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

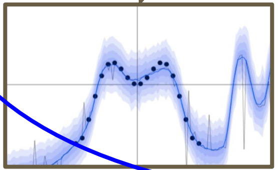
Used in very different contexts

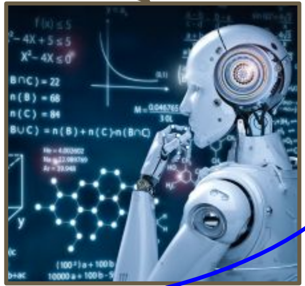




… …

These lectures





Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Our utility: X = data, θ=parameters

Model data as being sampled from a clever PDF with parameters θ

Bayesian Inference

Bayes Theorem:

$$p(\theta | X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Our utility: X = data, θ=parameters

Model data as being sampled from a clever PDF with parameters θ

→

Infer θ once you see the data X

Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Our utility: X = data, θ=parameters

Model data as being sampled from a clever PDF with parameters θ → Infer θ once you see the data X → Connect θ to physical parameters of interest

Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Our utility: X = data, θ=parameters

Data comes from real physical process not from a PDF!

Model data as being sampled from a clever PDF with parameters θ → Infer θ once you see the data X → Connect θ to physical parameters of interest

Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Our utility: X = data, θ=parameters

Data comes from real physical process not from a PDF!

Model data as being sampled from a clever PDF with parameters θ

Infer θ once you see the data X

Connect θ to physical parameters of interest

Inject prior knowledge exactly where it goes!

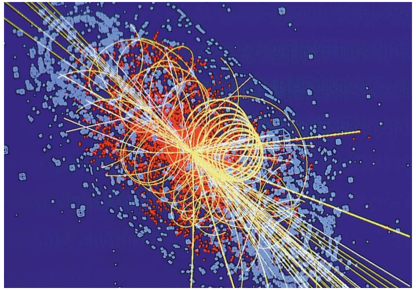
The Art of Modelling

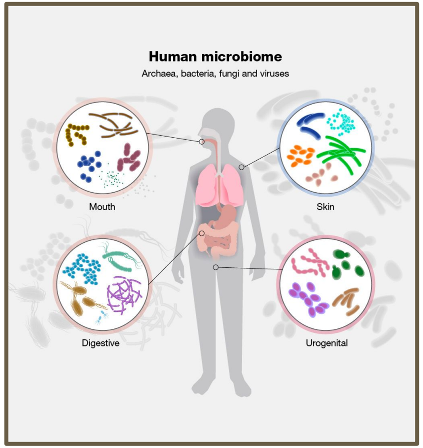
Graphical Models

Graphical Models

- Build sophisticated PDFs
- Visualize inner structure
- Maximize prior knowledge exploitation

Science: Mixture models

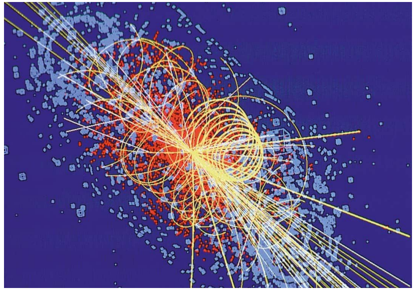


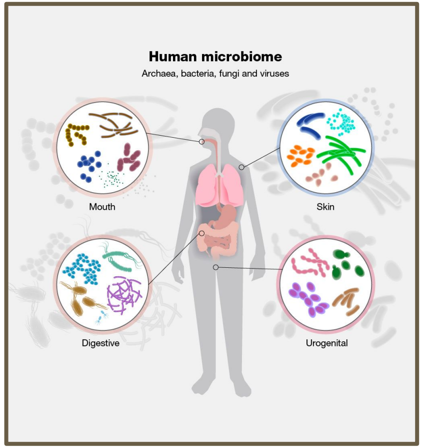


Dataset X:

- Signal
- Few backgrounds

Science: Mixture models



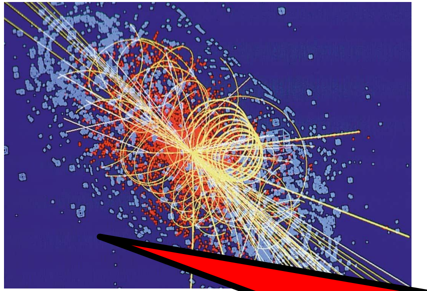


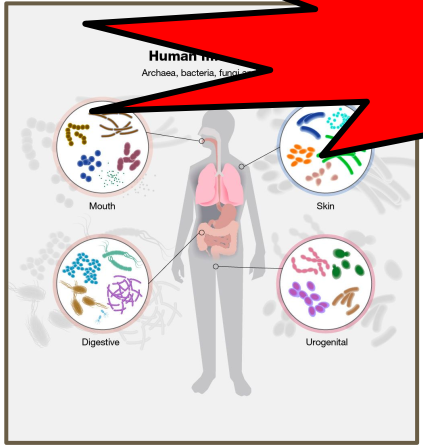
Dataset X:

- Signal
- Few backgrounds
- Each event is either

  signal or one of the backgrounds

Science: Mixture models





How to create such a PDF !?

Dataset X:

- Signal
- backgrounds
- event is either

signal or one of the backgrounds

Bayesian Inference: Graphical models

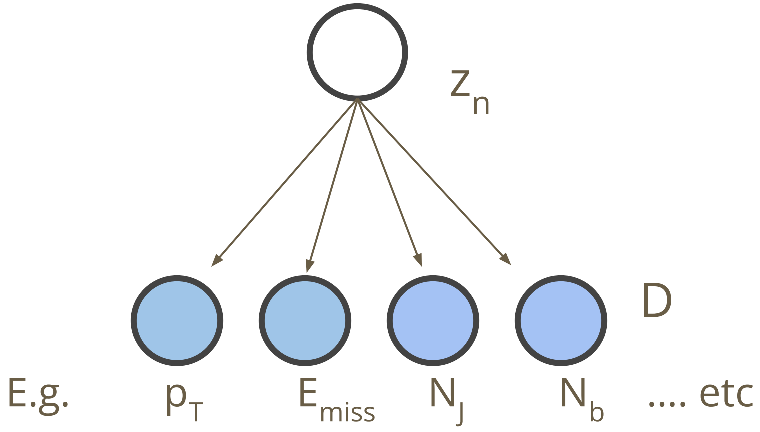
Graphical representation of a PDF to easily visualize the internal structure of the random variables

Bayesian Inference: Graphical models

At each event, sample a multinomial random variable that decides whether is signal or some of the backgrounds

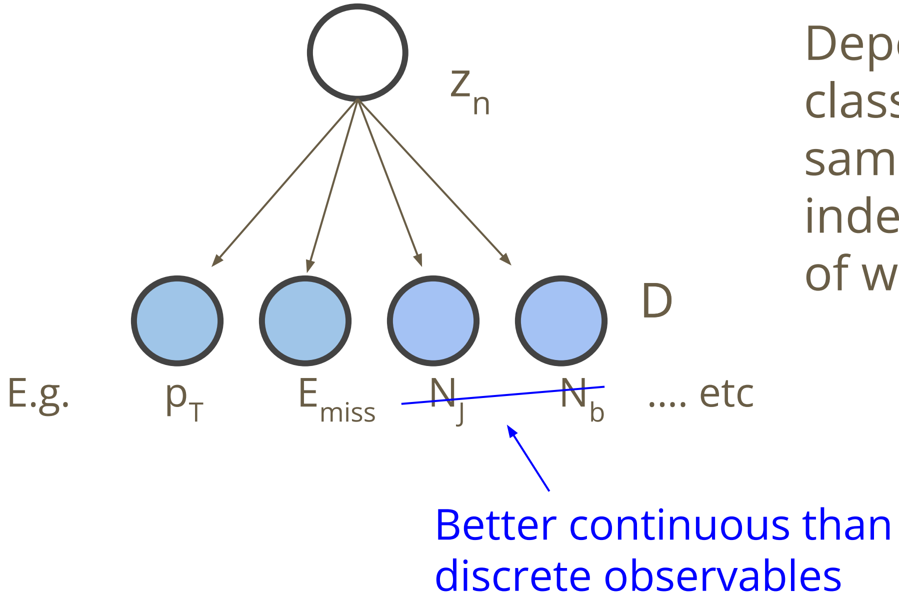
(K classes)

Bayesian Inference: Graphical models



Depending on the class of the event, we sample D random independent variables of what *we measure*

Bayesian Inference: Graphical models



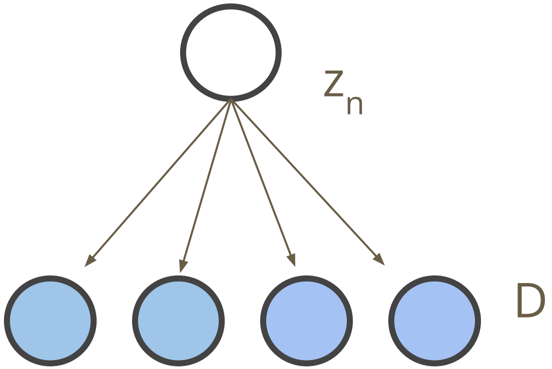
Depending on the class of the event, we sample D random independent variables of what *we measure*

Bayesian Inference: Graphical models

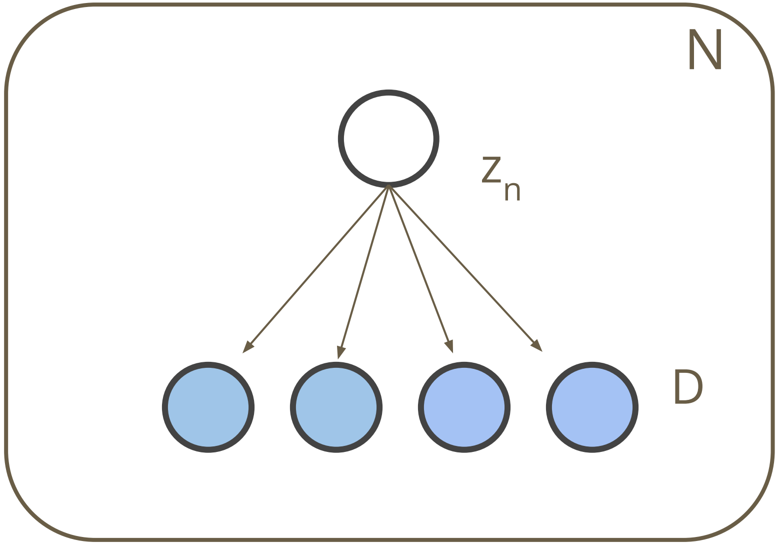
Convention:

Empty circles:
Sampled and unobserved RV

Filled circles:
Sampled and observed RV

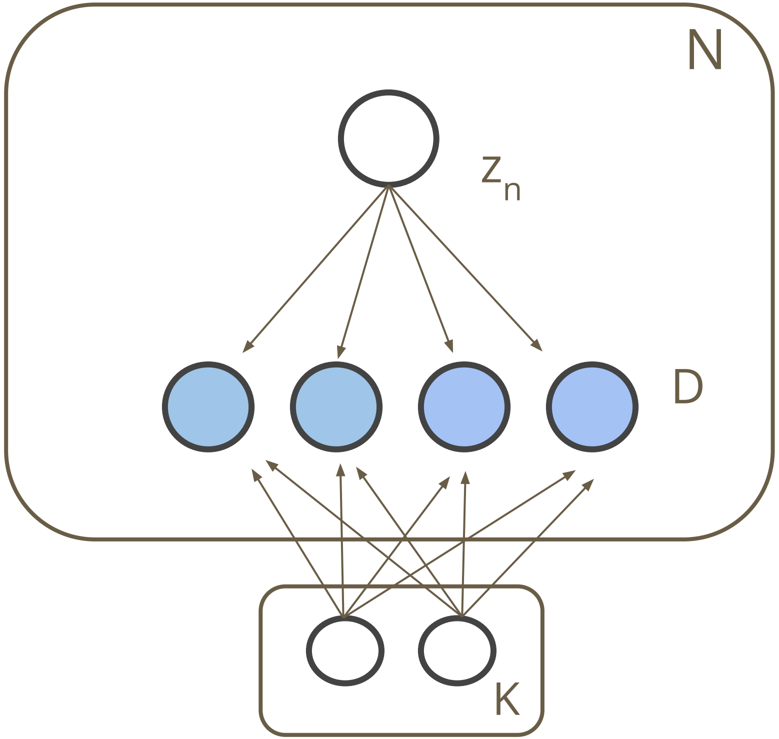


Bayesian Inference: Graphical models



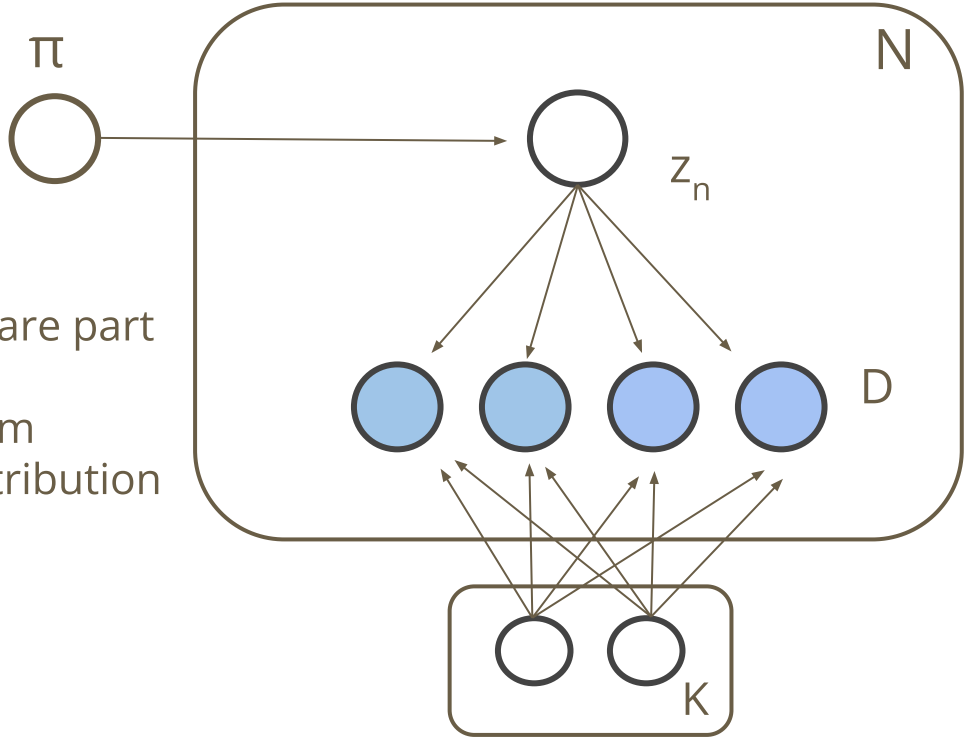
Procedure that is repeated N times

Bayesian Inference: Graphical models



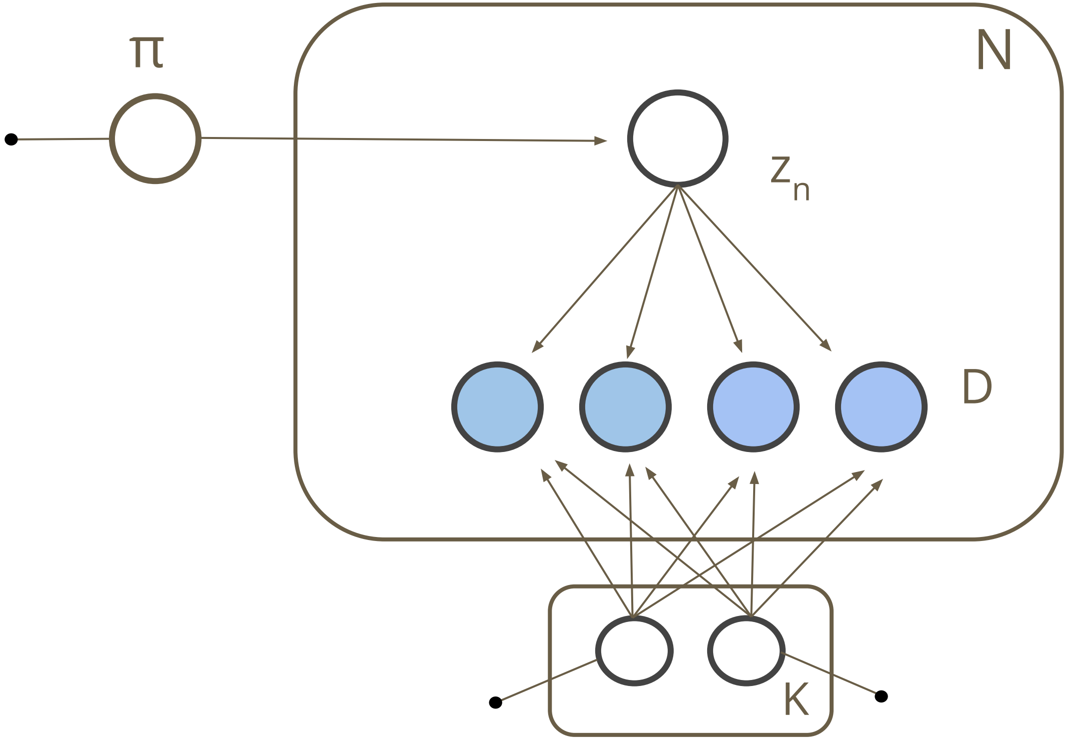
Each one of the K classes has an *expected distribution* over the measured quantities

Bayesian Inference: Graphical models



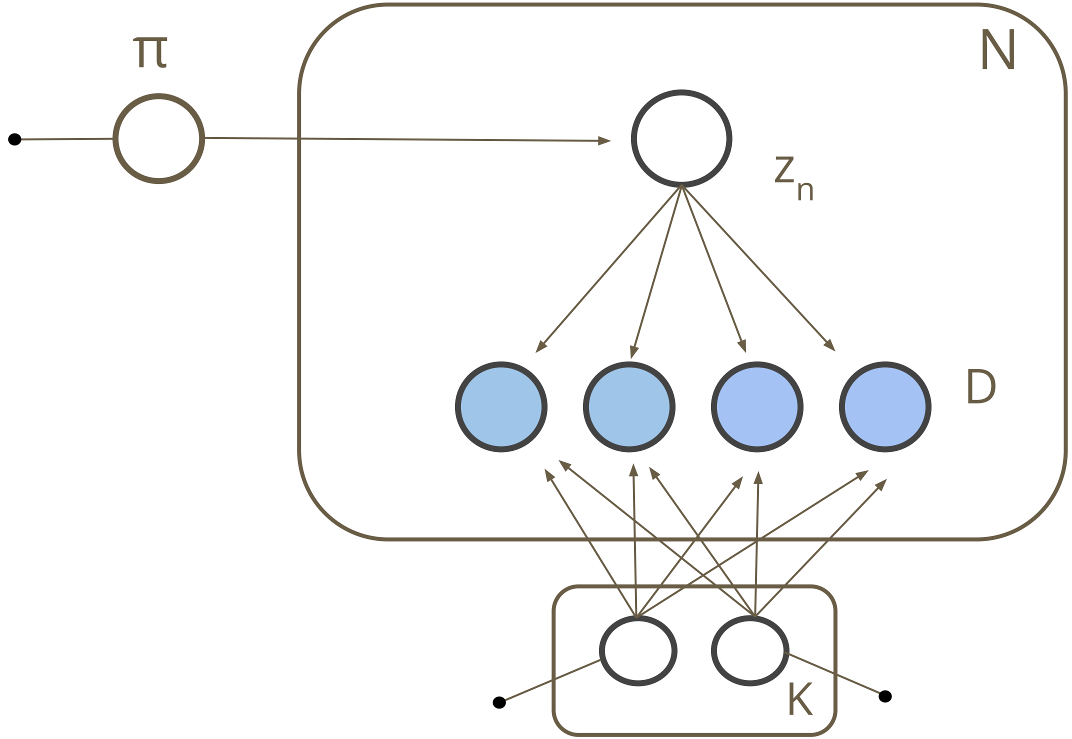
Multinomial parameters are part of the PDF. Sampled from Dirichlet distribution

Bayesian Inference: Graphical models



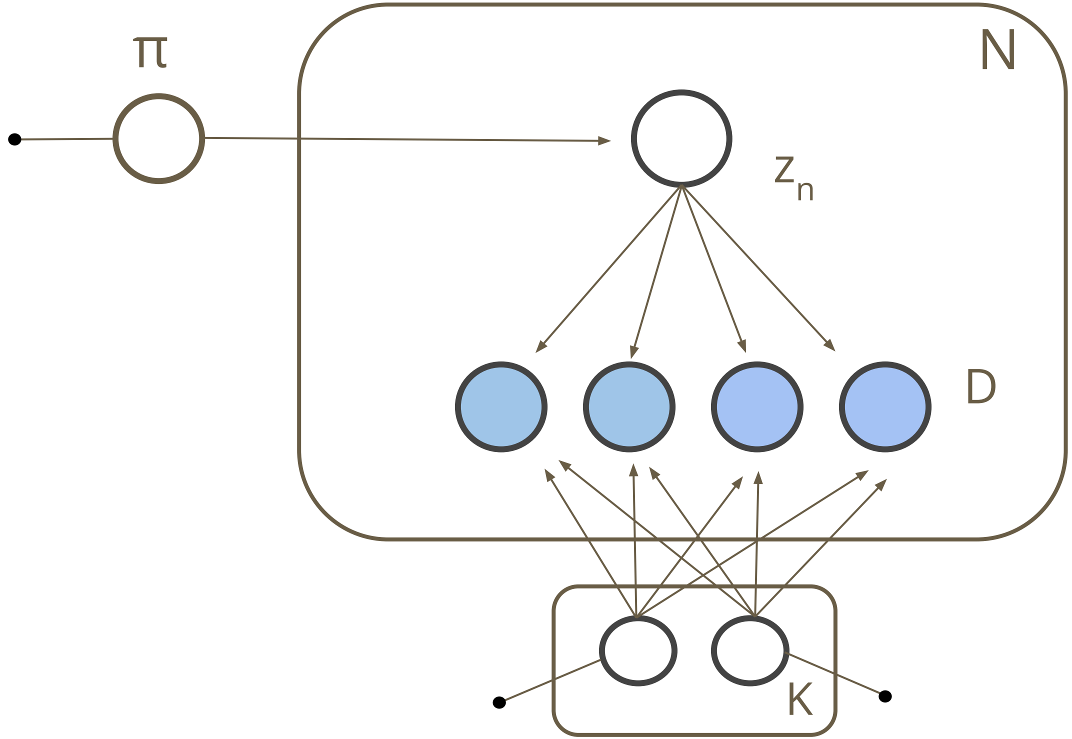
Hyperparameters to define how to sample parameters

Bayesian Inference: Graphical models



Mixture Model

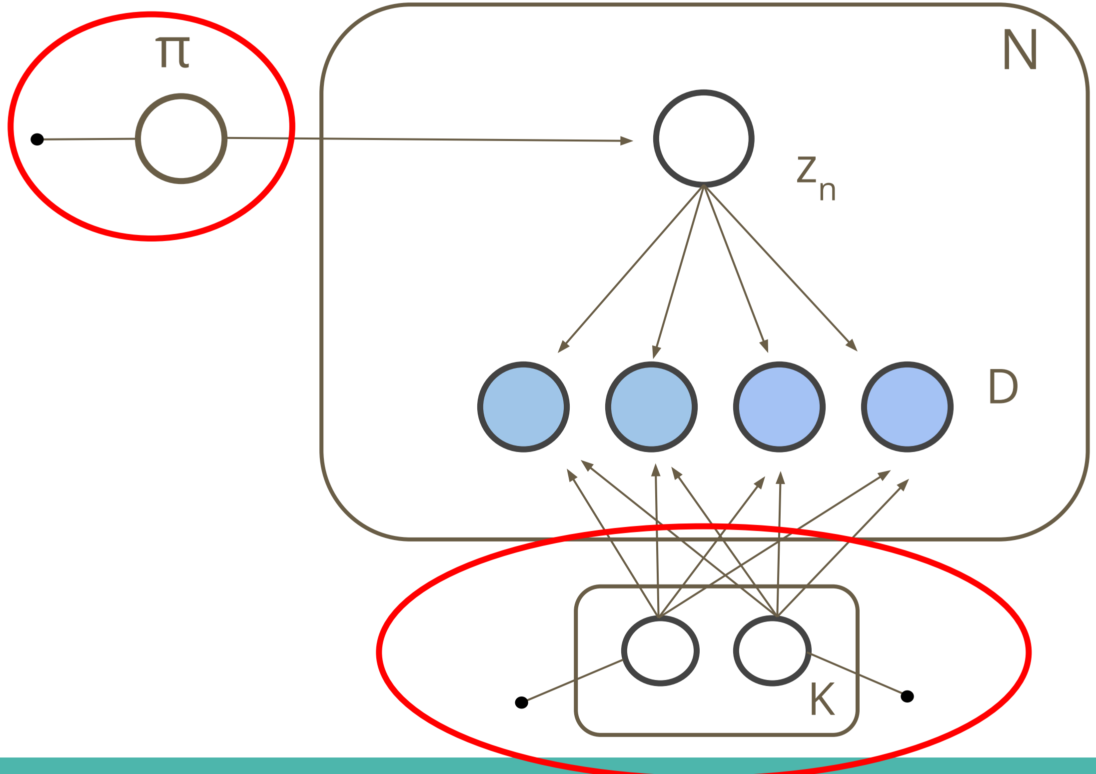
Bayesian Inference: Graphical models



Mixture Model

- Model data as being sampled from a PDF

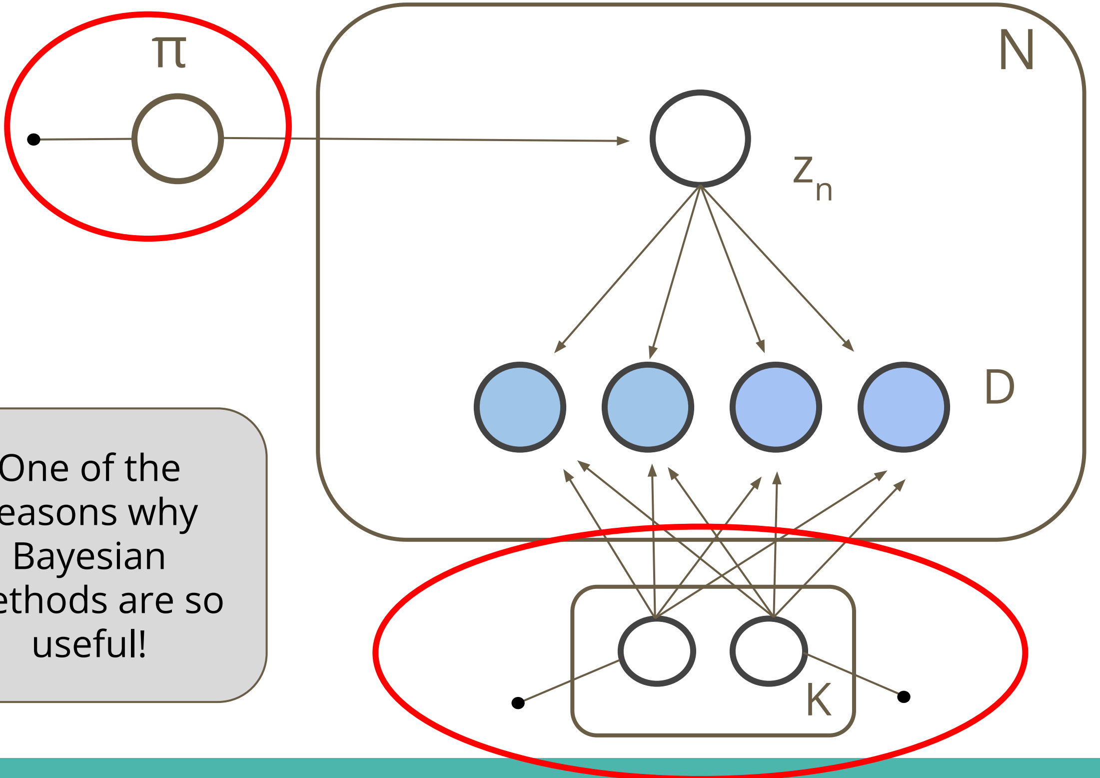
Bayesian Inference: Graphical models



Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge

Bayesian Inference: Graphical models

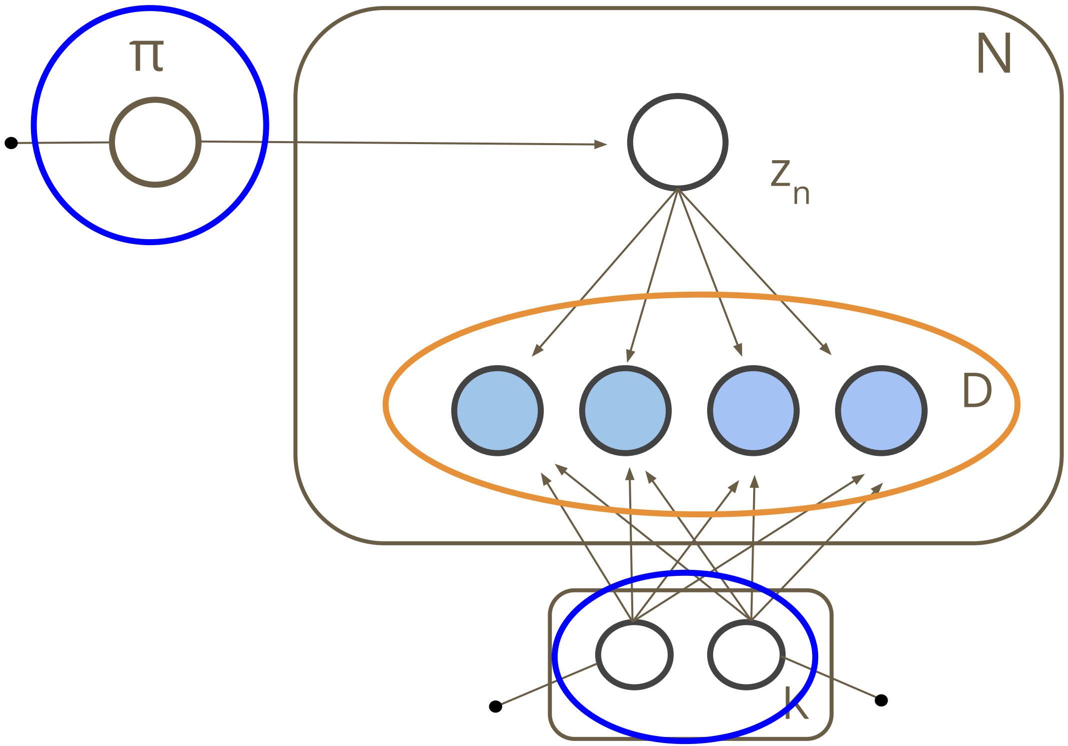


One of the reasons why Bayesian methods are so useful!

Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge

Bayesian Inference: Graphical models

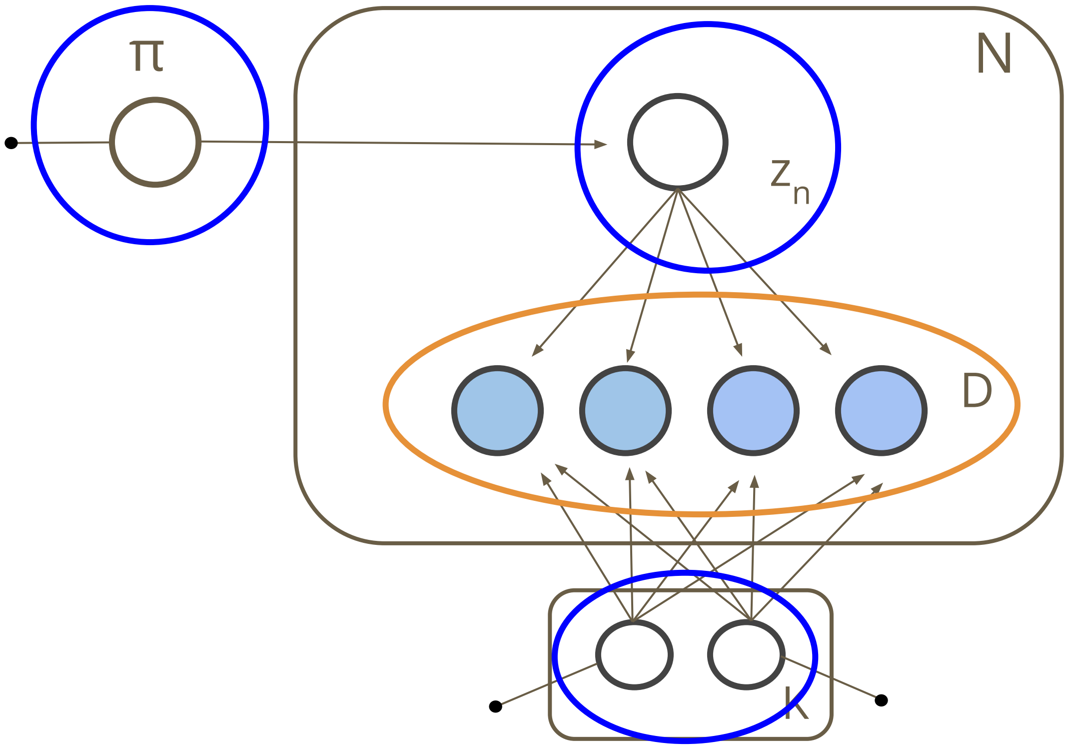


Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge
- Infer the parameters conditioned in the observed data

$$p(\theta|X) = p(X|\theta)\ p(\theta)\ /\ p(x)$$

Bayesian Inference: Graphical models



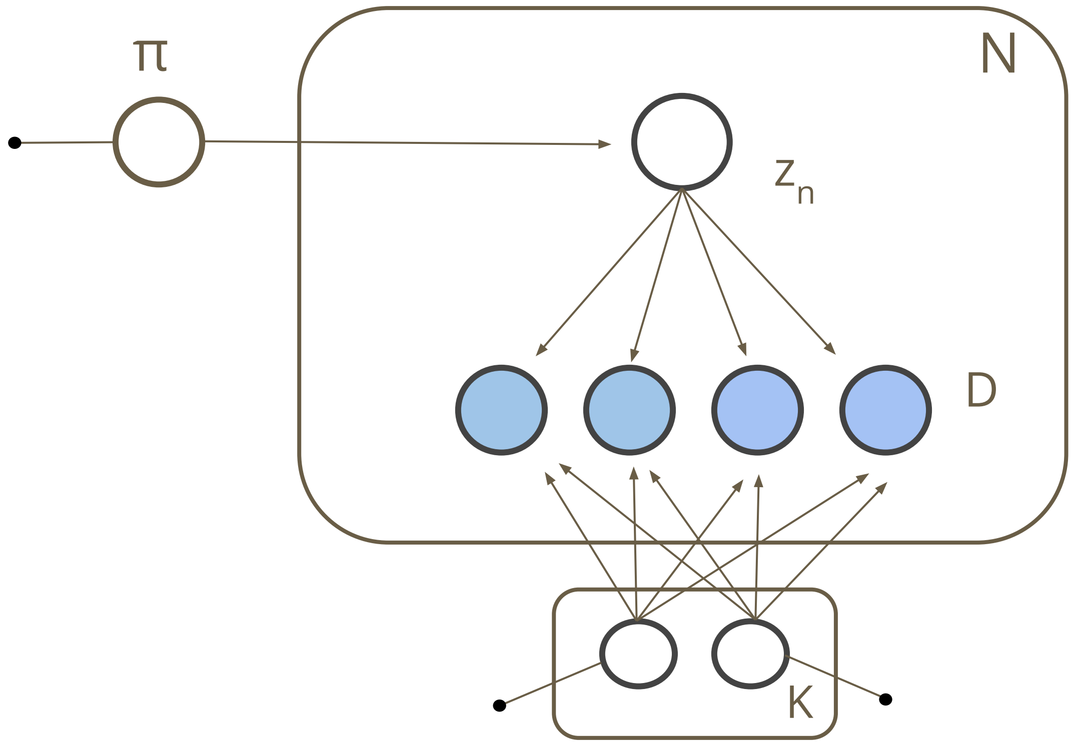
Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge
- Infer the parameters conditioned in the observed data

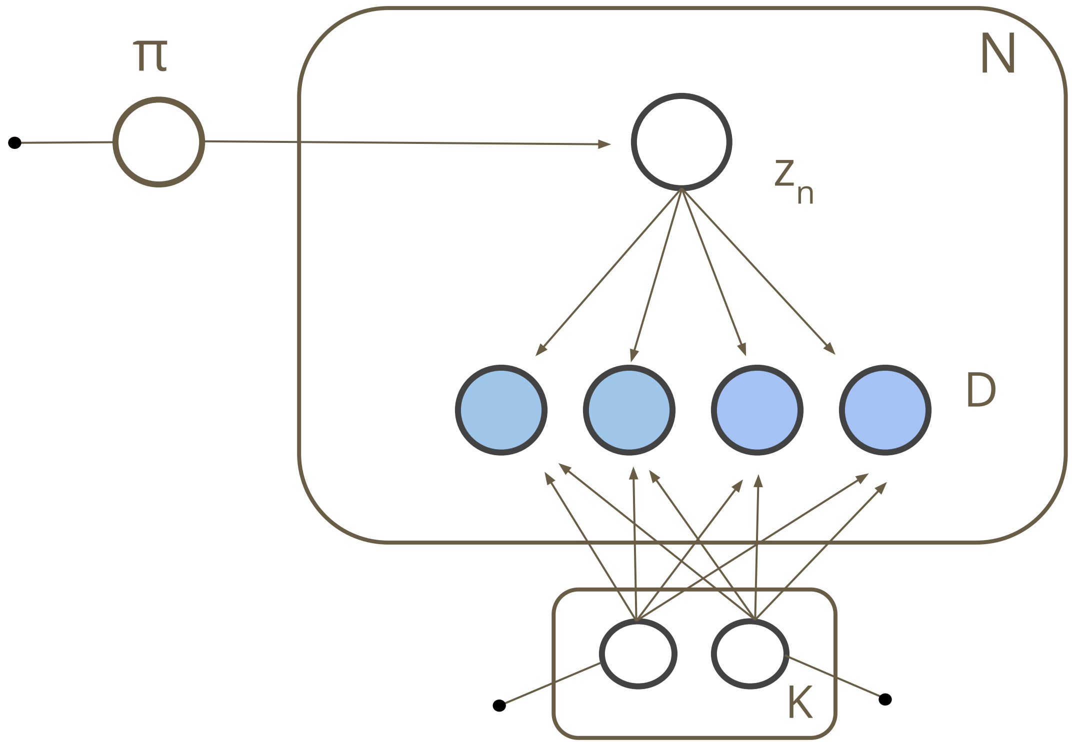
$$p(\theta | X) = p(X | \theta)\, p(\theta) / p(x)$$

- Infer the latent variables

Bayesian Inference

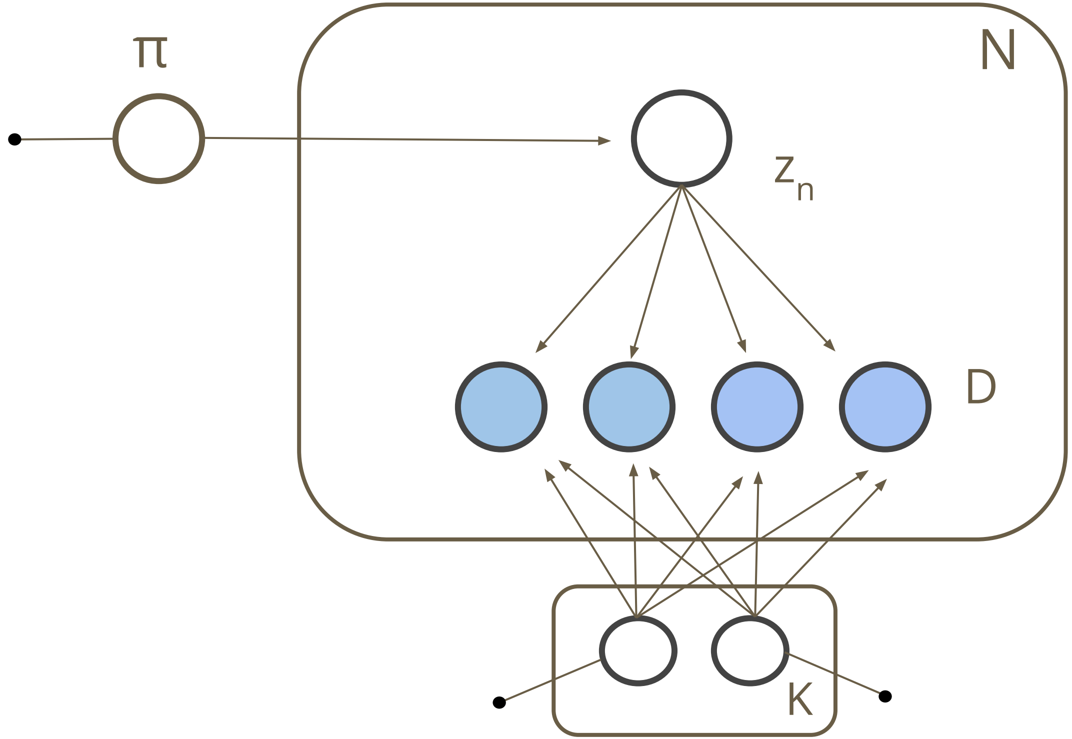


Bayesian Inference



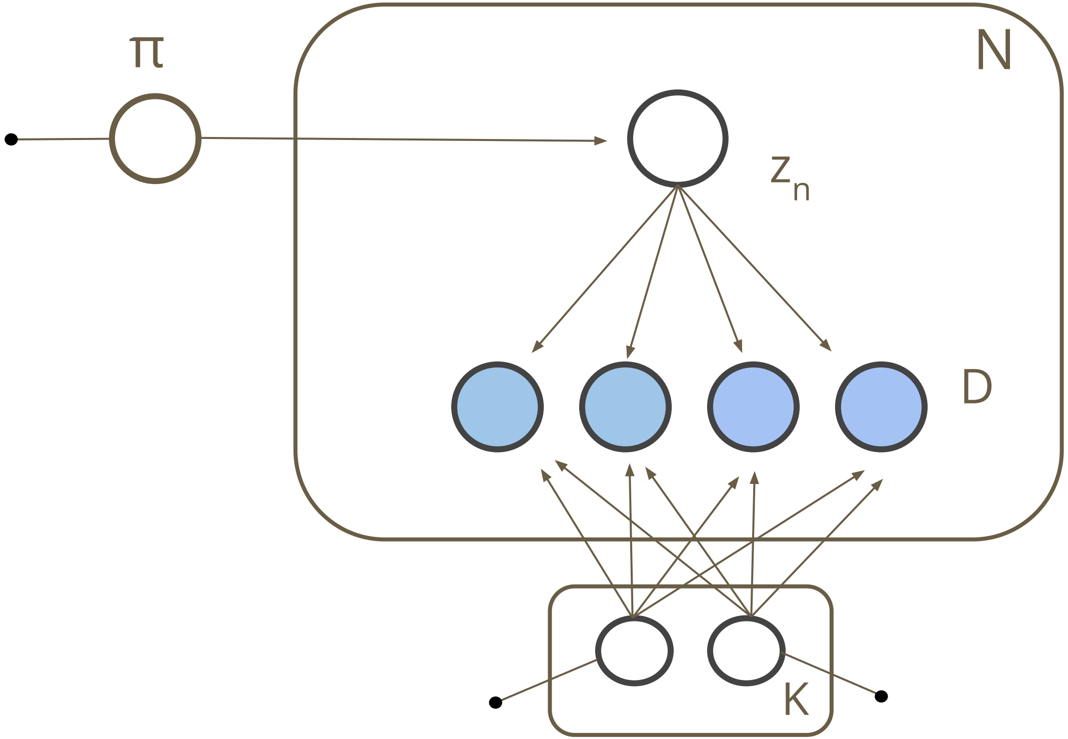
- No hard cuts

Bayesian Inference



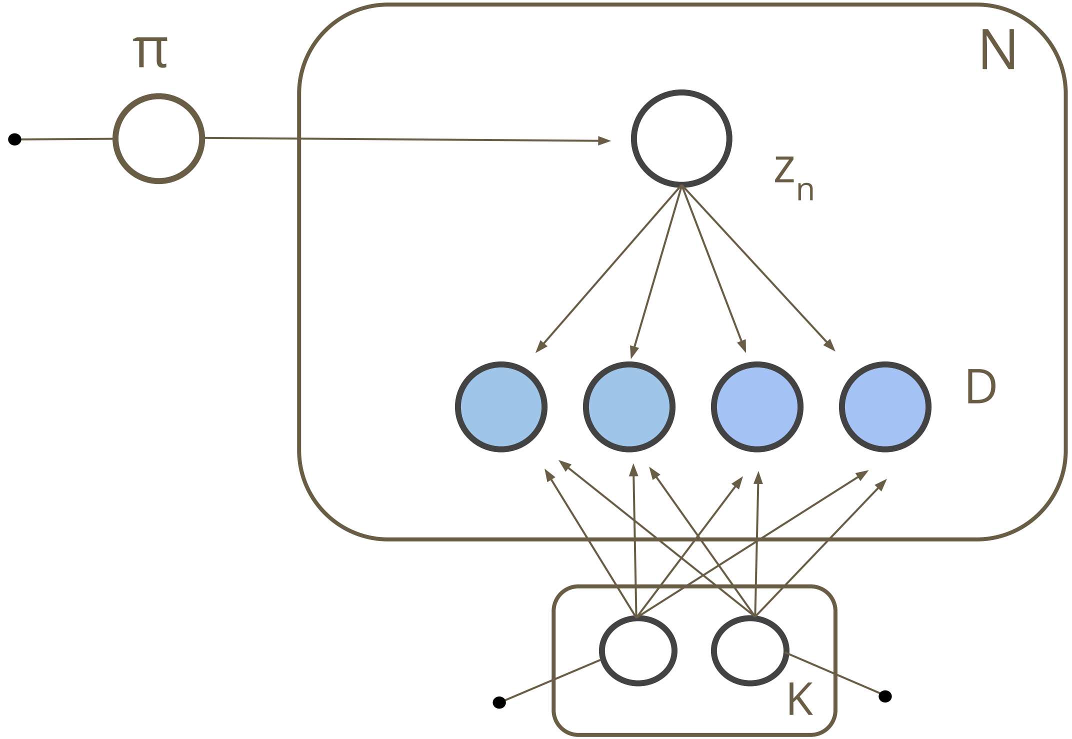
- No hard cuts
- Soft assignments

Bayesian Inference



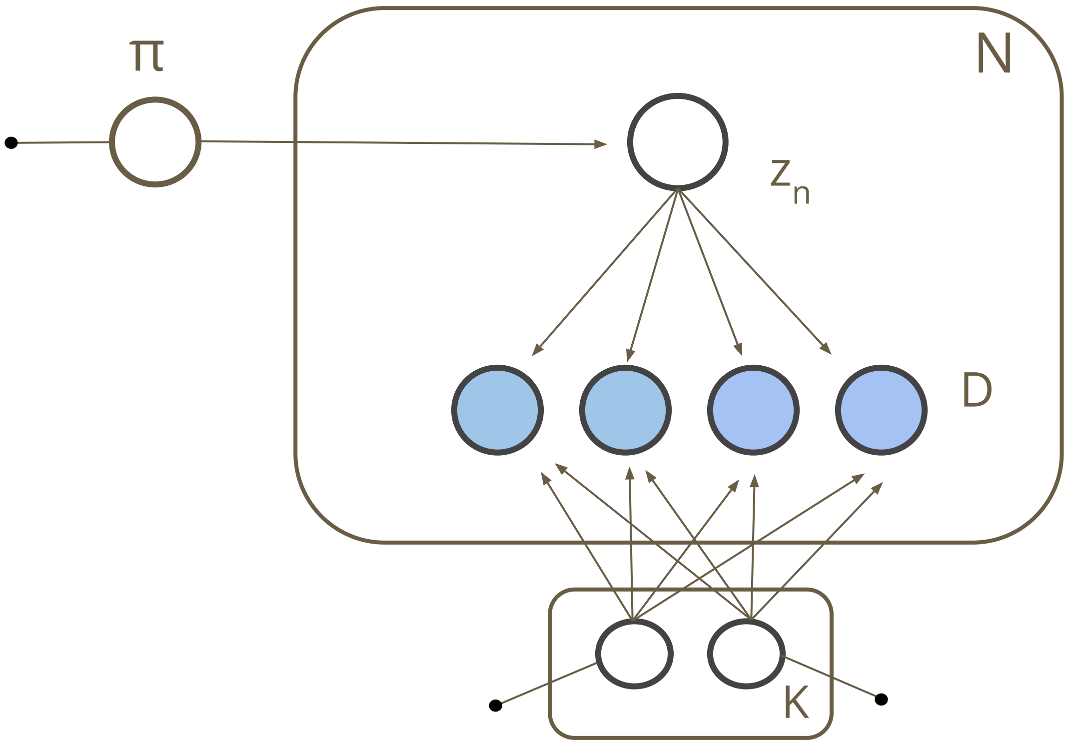
- No hard cuts
- Soft assignments
- No signal/control regions

Bayesian Inference



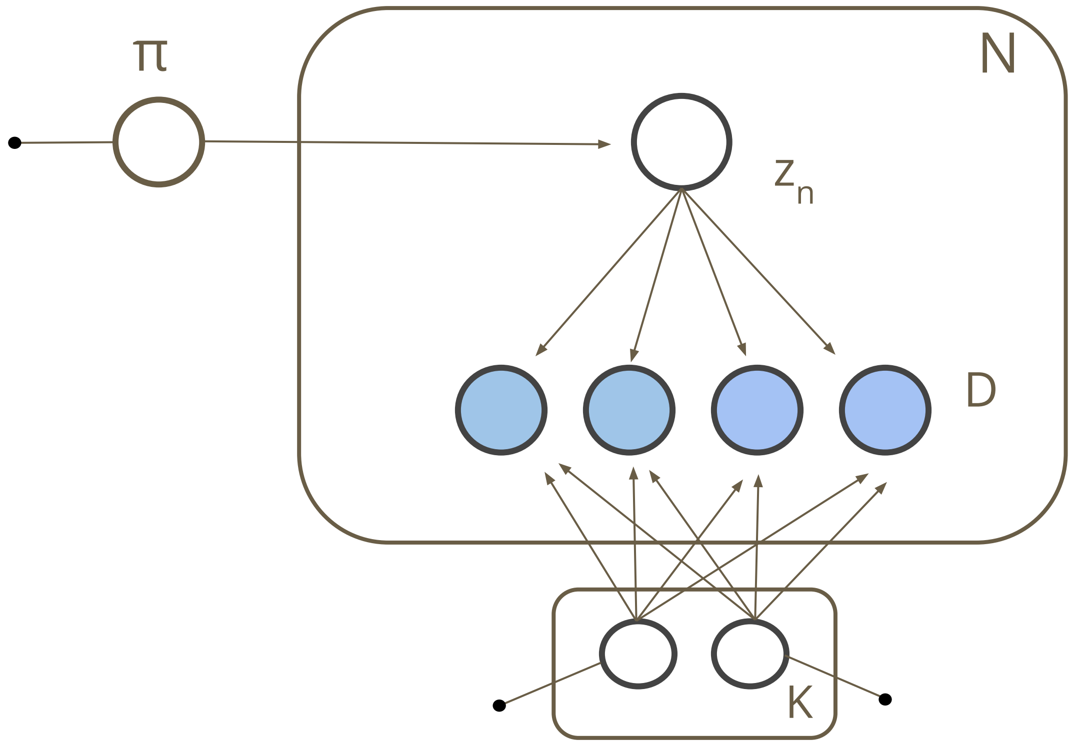
- No hard cuts
- Soft assignments
- No signal/control regions
- K classes & D observables

Bayesian Inference



- No hard cuts
- Soft assignments
- No signal/control regions
- K classes & D observables
- Unfolding of data internal structure

Bayesian Inference



- No hard cuts
- Soft assignments
- No signal/control regions
- K classes & D observables
- Unfolding of data internal structure
- Controlled injection of prior knowledge

1D Gaussian Mixture

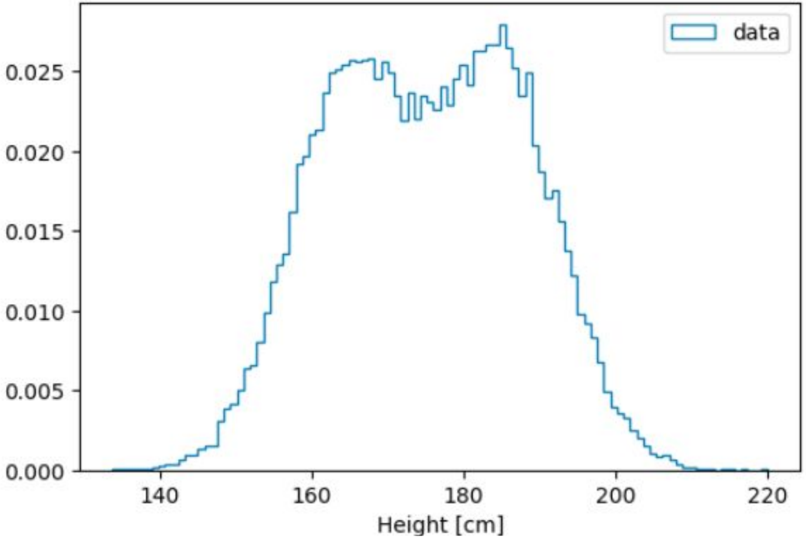
Simple example to settle down the ideas

1D Gaussian Mixture

Measure the height
of many people

1D Gaussian Mixture

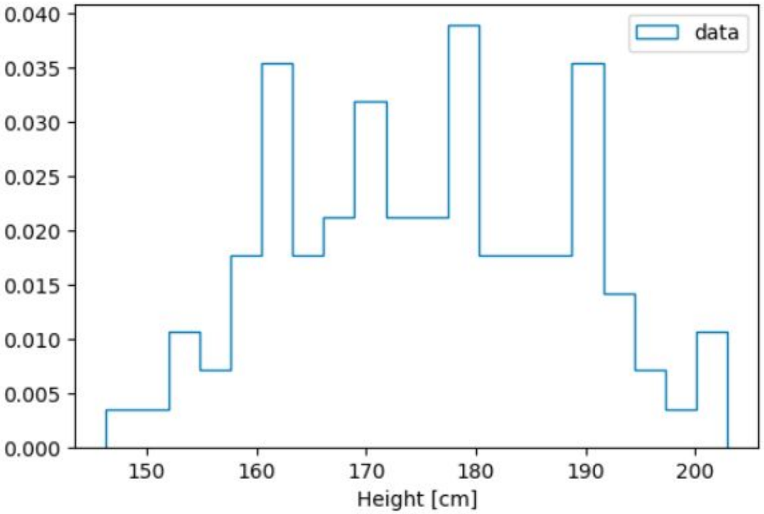
Measure the height
of many people



Is not hard to see with enough events
(500k)

1D Gaussian Mixture

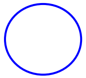
Measure the height
of many people



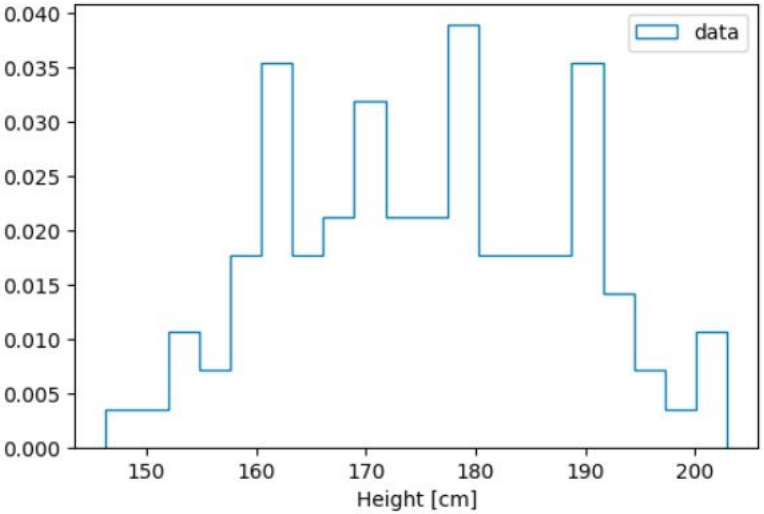
Gets more involved with less events!
(100)

1D Gaussian Mixture

z_n ~Bernoulli

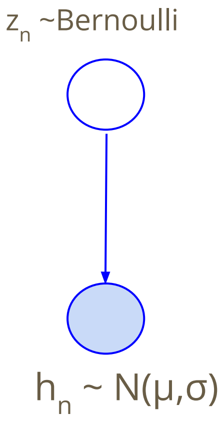


Graphical Model

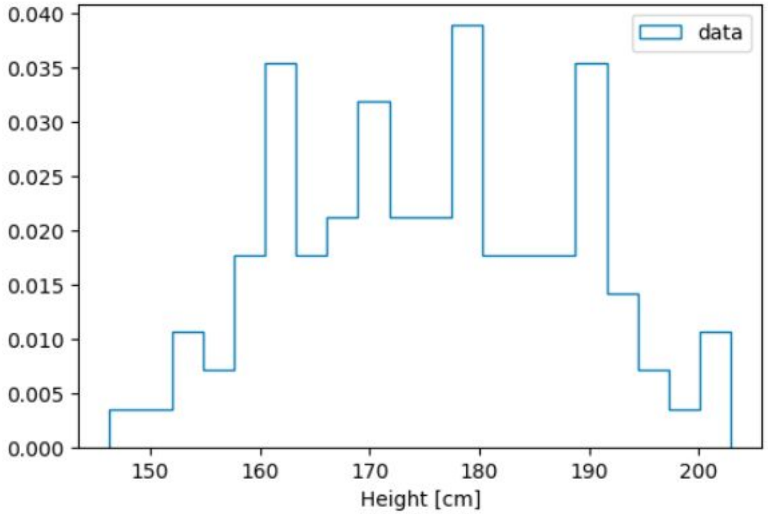


Gets more involved with less events!
(100)

1D Gaussian Mixture

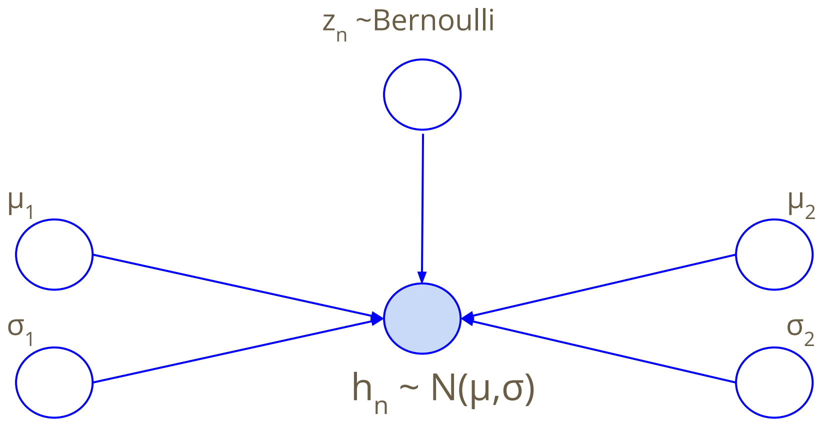


Graphical Model

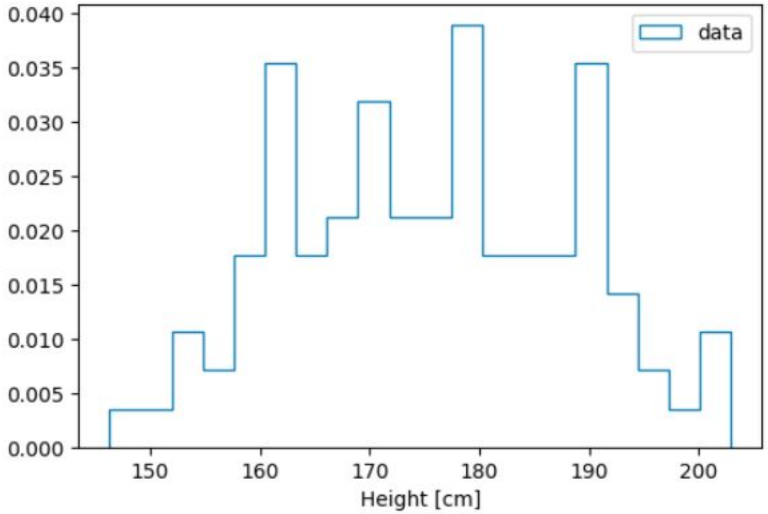


Gets more involved with less events!
(100)

1D Gaussian Mixture

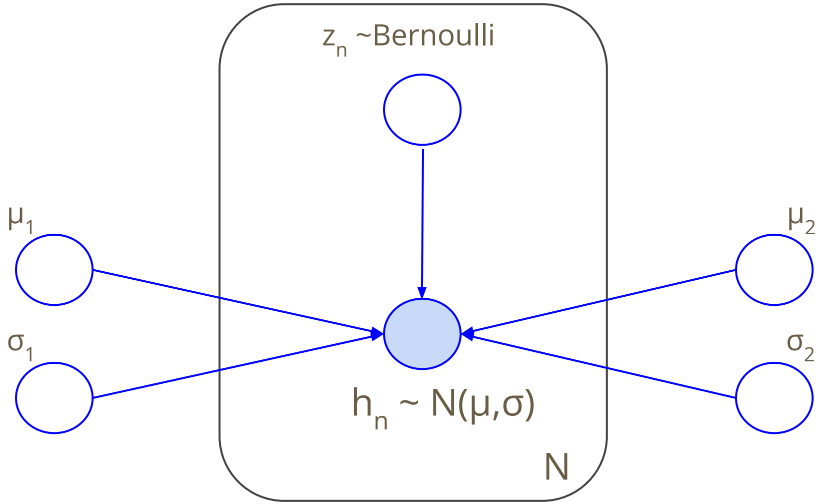


Graphical Model

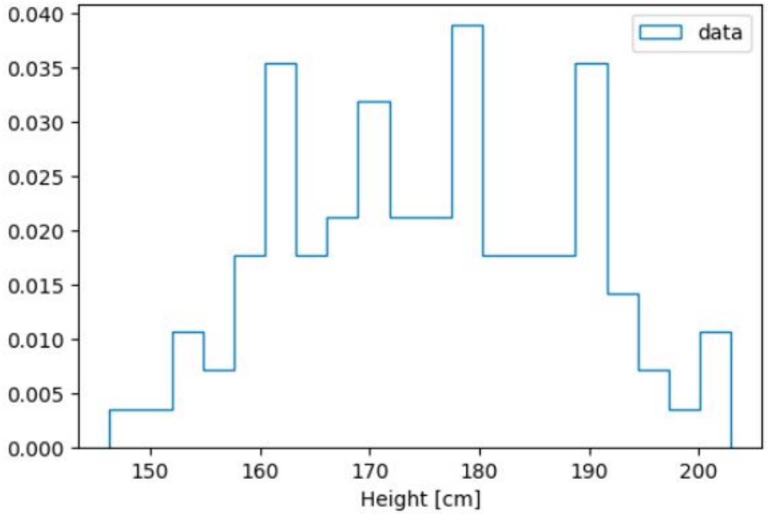


Gets more involved with less events!
(100)

1D Gaussian Mixture

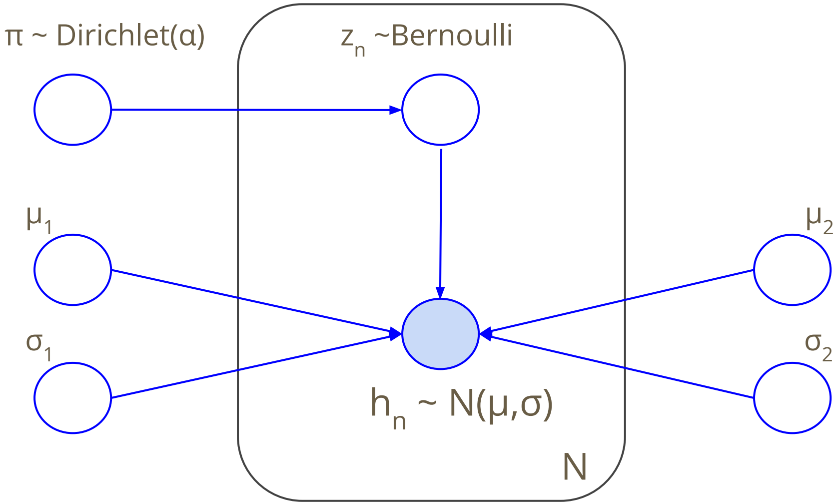


Graphical Model

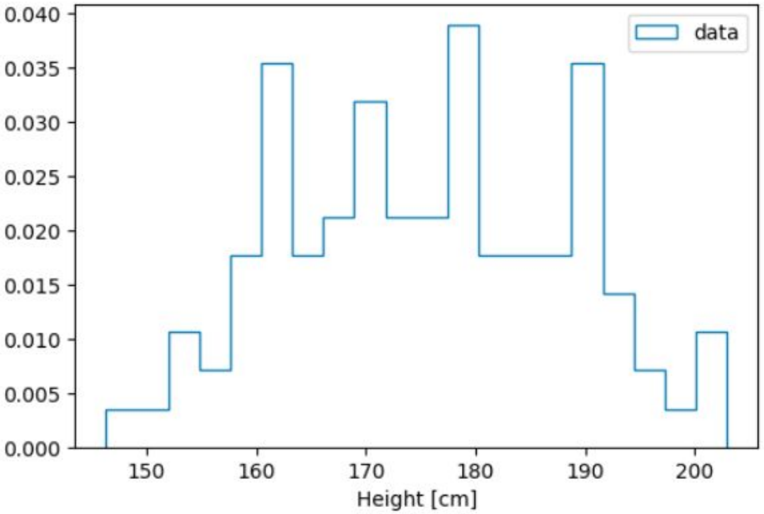


Gets more involved with less events!
(100)

1D Gaussian Mixture

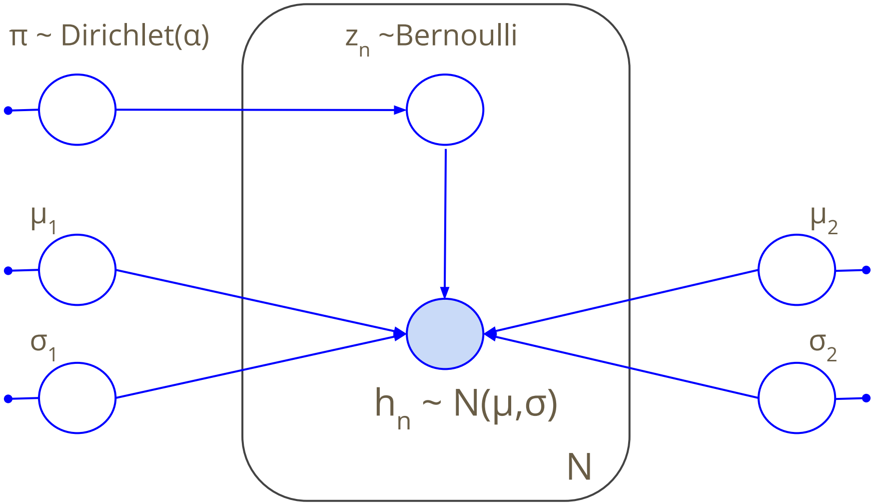


Graphical Model

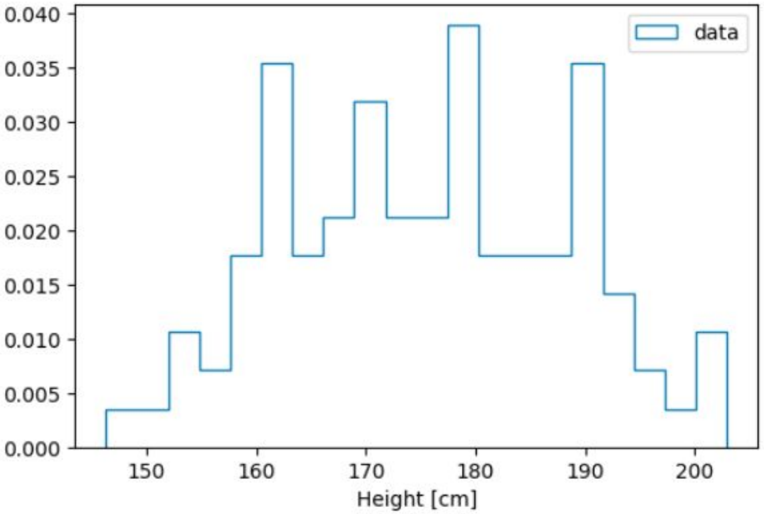


Gets more involved with less events!
(100)

1D Gaussian Mixture

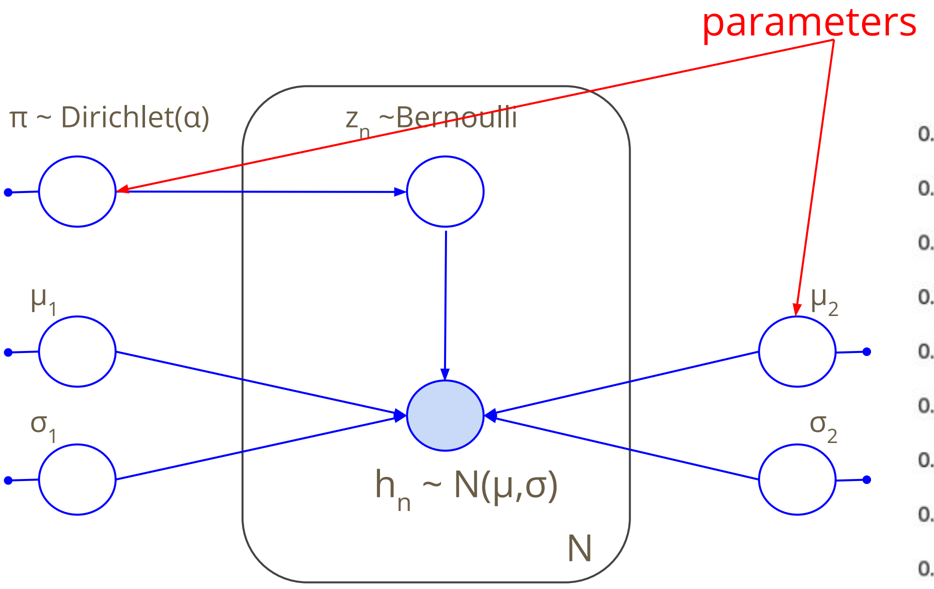


Graphical Model

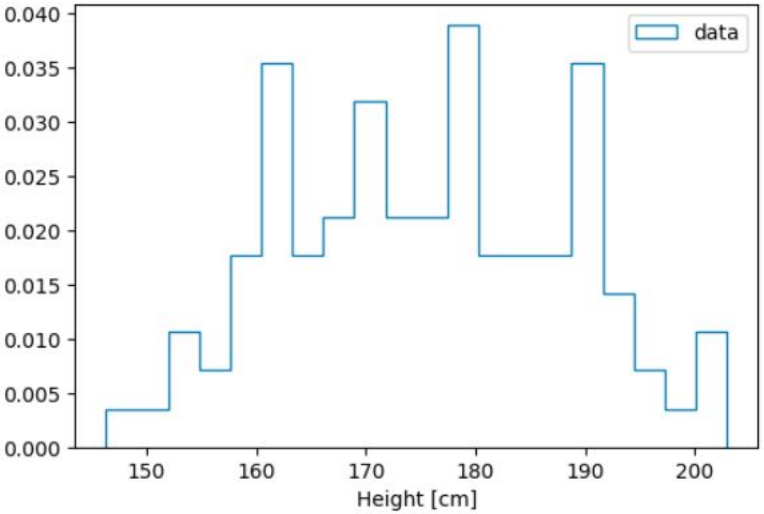


Gets more involved with less events!
(100)

1D Gaussian Mixture

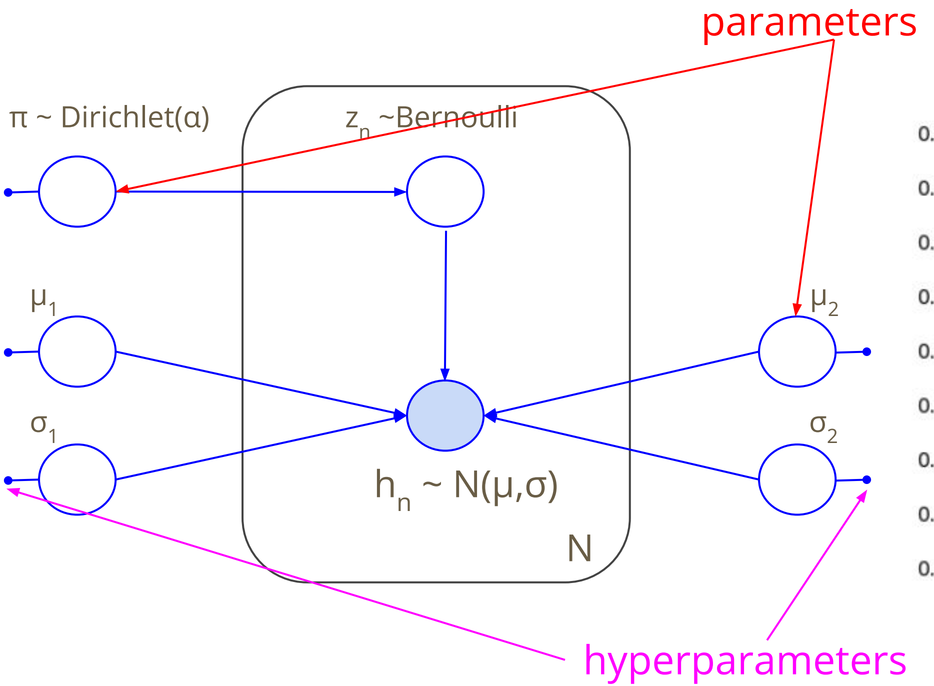


Graphical Model

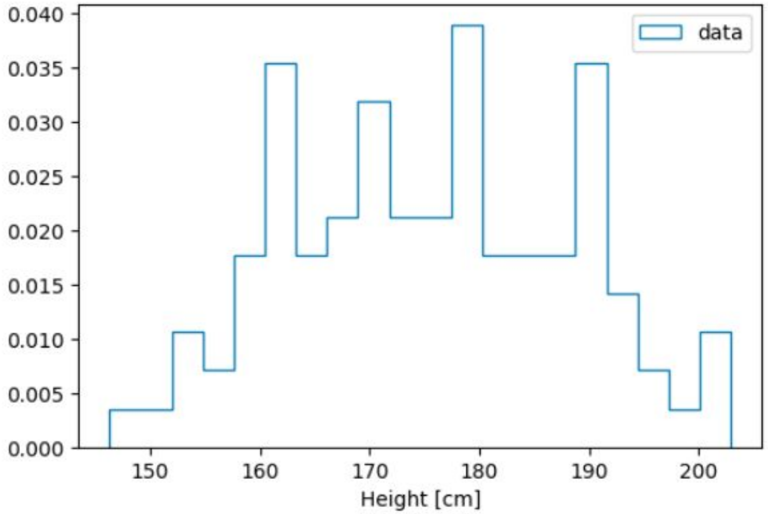


Gets more involved with less events!
(100)

1D Gaussian Mixture

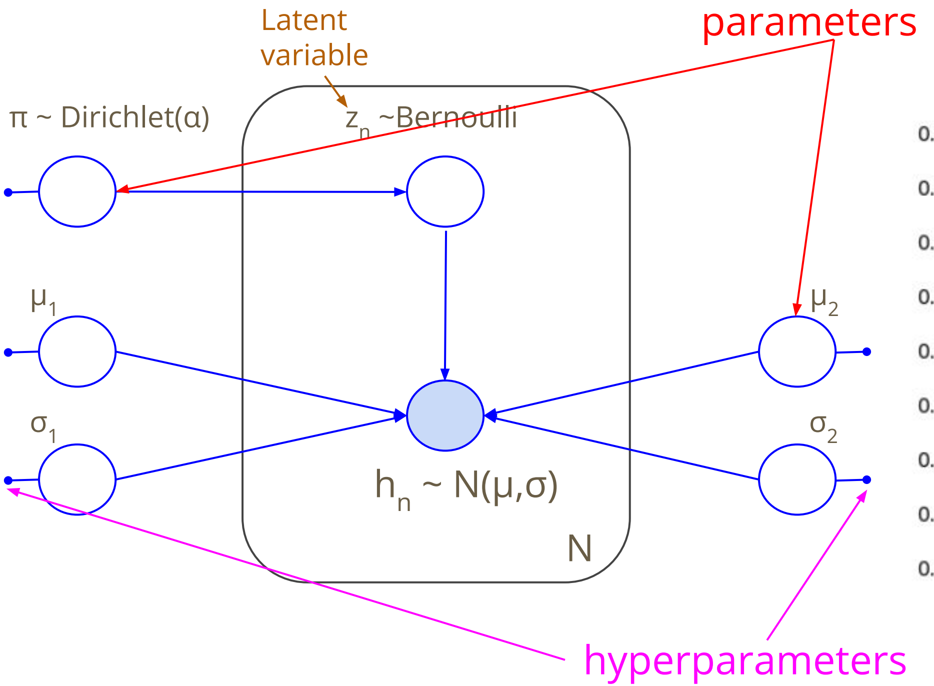


Graphical Model

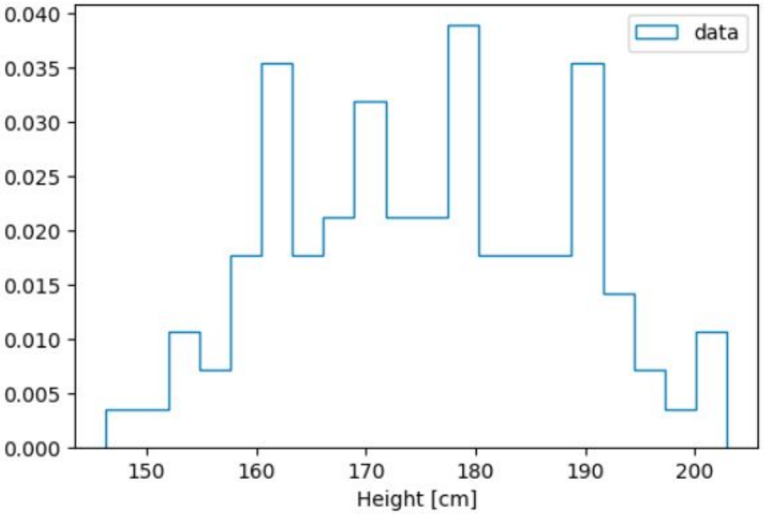


Gets more involved with less events!
(100)

1D Gaussian Mixture

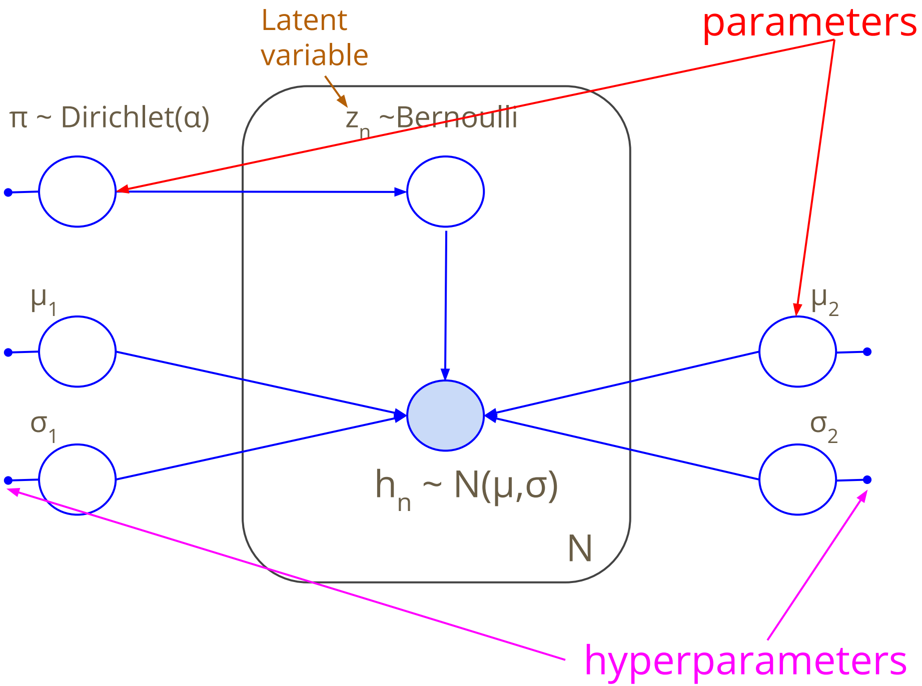


Graphical Model



Gets more involved with less events!
(100)

1D Gaussian Mixture

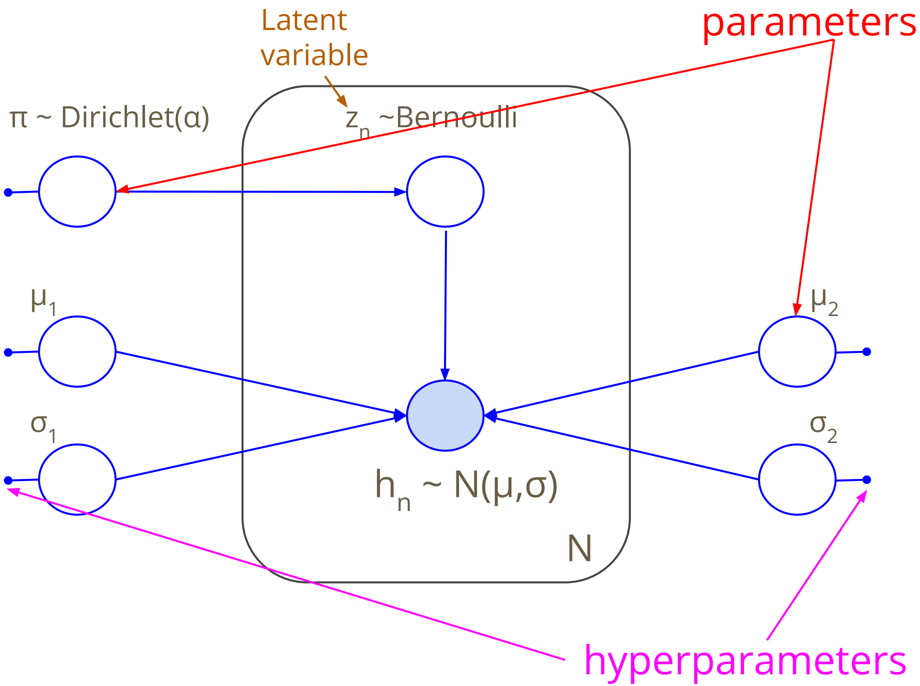


Graphical Model

$X : \{ h_n \}$

$\theta : \mu_{1,2}, \sigma_{1,2}, \pi, \{ z_n \}$

1D Gaussian Mixture



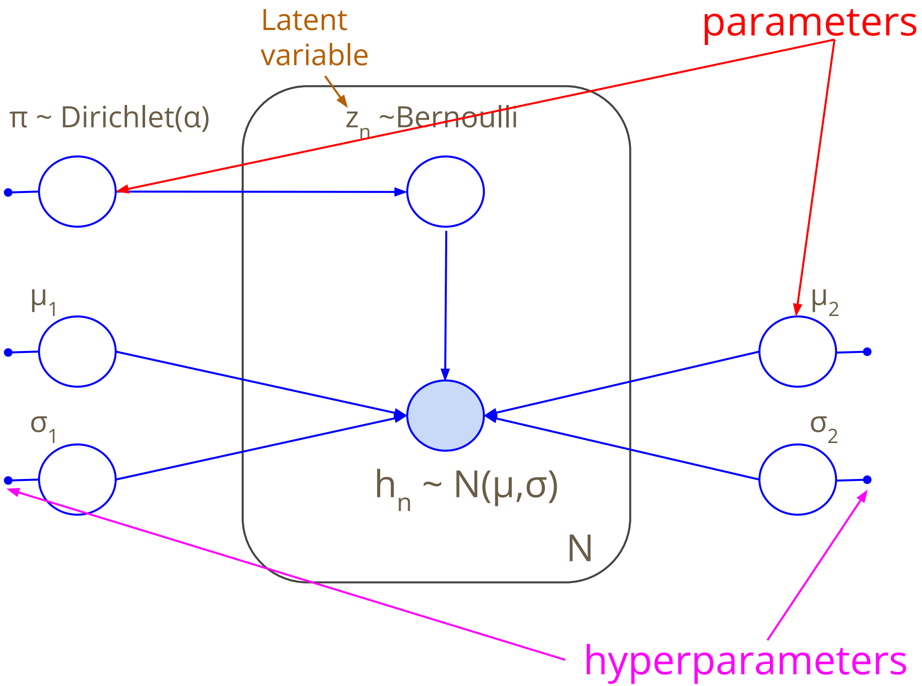
Graphical Model

$X : \{ h_n \}$

$\theta : \mu_{1,2}, \sigma_{1,2}, \pi, \{ z_n \}$

marginalize over z_n

1D Gaussian Mixture



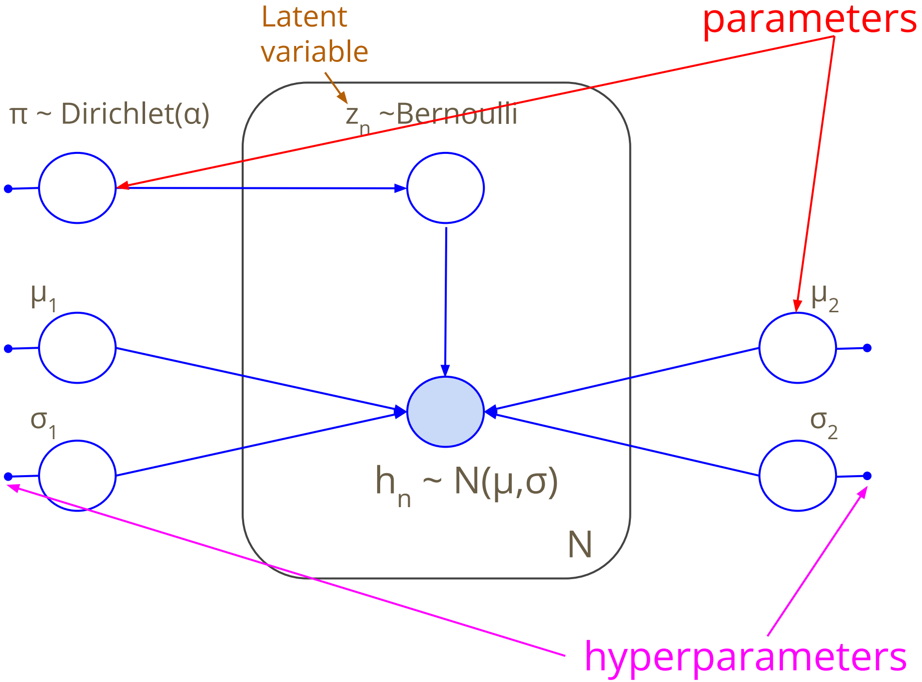
Graphical Model

$X : \{ h_n \}$

$\theta : \ \mu_{1,2} , \sigma_{1,2} , \pi , \{ z_n \}$

$P(x_n | \theta) = \pi \ N(h_n;\mu_1,\sigma_1) + (1\text{-}\pi) \ N(h_n;\mu_2, \sigma_2)$

1D Gaussian Mixture



Graphical Model

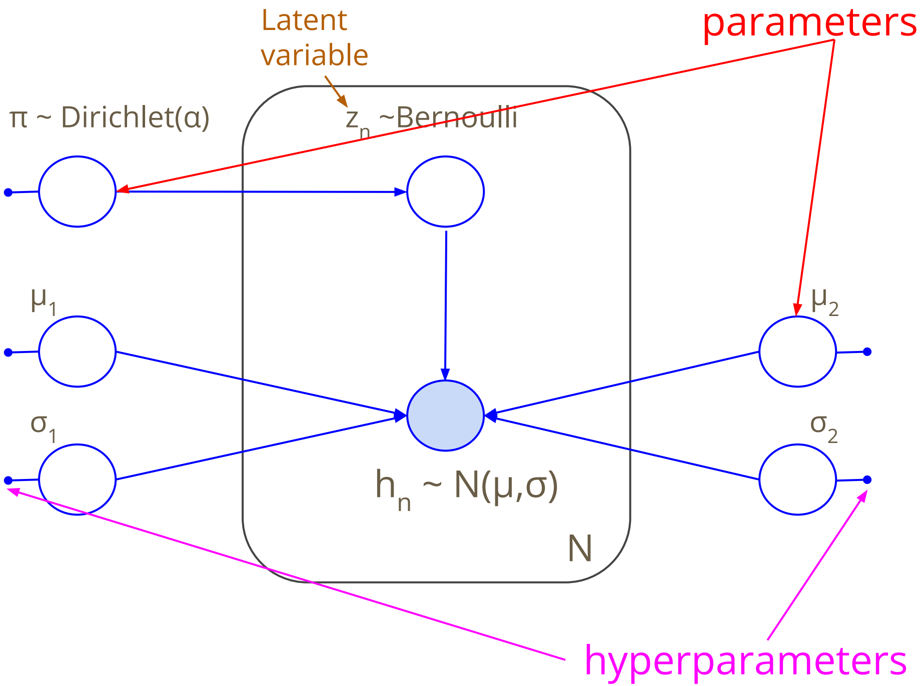
$X : \{ h_n \}$

$\theta : \mu_{1,2}, \sigma_{1,2}, \pi, \{ z_n \}$

$P(x_n | \theta) = \pi \, N(h_n; \mu_1, \sigma_1) + (1-\pi) \, N(h_n; \mu_2, \sigma_2)$

$P(X | \theta) = \text{prod}(P(x_n | \theta))$

1D Gaussian Mixture



Graphical Model

$X : \{ h_n \}$

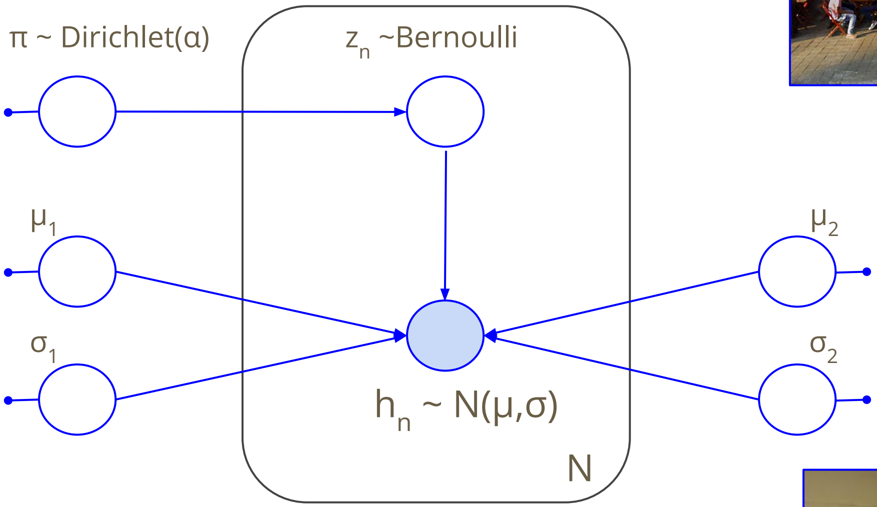
$\theta : \mu_{1,2} , \sigma_{1,2} , \pi , \{ z_n \}$

Likelihood

$P(x_n | \theta) = \pi \, N(h_n;\mu_1,\sigma_1) + (1-\pi) \, N(h_n;\mu_2, \sigma_2)$

$P(X | \theta) = \text{prod}(P(x_n | \theta))$

1D Gaussian Mixture



Gaussian Mixture
@lunch time!

Final remarks @ Lecture 1

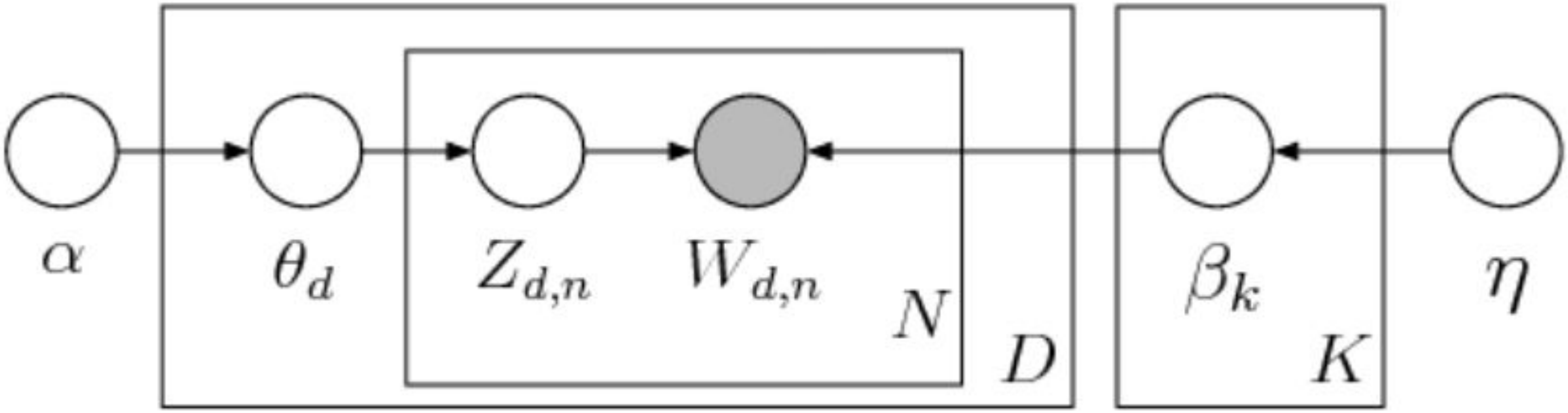
Final remarks @ Lecture 1

- Bayes equation very insightful !
- Model data as being sampled from a PDF
- Graphical Models to create PDFs
- Bayesian: learn PDF from data as it comes out
- Many tools and art-et-metier to learn !

Final remarks @ Lecture 1

iArxiv.org model
(LDA Model)

Simple game



- Propose model
- Infer parameters' posterior from observed data

Some open questions @ Lecture 1

- How do we actually *infer* ?
- How do we check consistency ?
- How much can priors affect my results ?
- etc…

Some open questions @ Lecture 1

- How do we actually *infer* ?
- How do we check consistency ?
- How much can priors affect my results ?
- etc…

For next lectures!

Some open questions @ Lecture 1

- How do we actually *infer* ?
- How do we check consistency ?
- How much can priors affect my results ?
- etc...

For next lectures!

OBRIGADO!!