BAYESIAN INFERENCE

Antoine Deleforge

THANKS

Prof. Dr.-Ing. Walter Kellermann
Head of LMS chair
EARS project coordinator

Dr. Roland Maas
PhD at LMS chair
Research and development at Amazon

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

LMS

EARS
Embodied Audition for RobotS

SEVENTH FRAMEWORK PROGRAMME

OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- Bayesian inference by examples
  - Direct inference
  - The Expectation-Maximization algorithm
  - Variational Bayes methods
- Markov chain Monte Carlo

OUTLINE

- **What is Bayesian inference?**
  - **Overview**
  - **Classical vs. Bayesian approach**
  - **Bayes Theorem & Example**
  - **General Methodology**
- Bayesian inference by examples
  - Direct inference
  - The Expectation-Maximization algorithm
  - Variational Bayes methods
- Markov chain Monte Carlo

What is Bayesian Inference?

Inference

What is Bayesian Inference?

Inference

Ingredients Observations Model

What is Bayesian Inference?

Inference

Ingredients

Observations

Model

Goals

Estimation Quantitative deductions on causes or consequences of the observations, i.e., find underlying model parameters.

Example: *I observed a certain amount of rain drops forming on my window in the last minute. What is the current rainfall in milimeters?*

What is Bayesian Inference?

Inference

Ingredients

Observations

Model

Goals

Estimation Quantitative deductions on causes or consequences of the observations, i.e., find underlying model parameters.

Example: *I observed a certain amount of rain drops forming on my window in the last minute. What is the current rainfall in milimeters?*

Prediction From the infered model, predict what missing or future observations should be.

Example: *How many more raindrops will form on my window in the next hour?*

What is Bayesian Inference?

Inference

Ingredients Observations | Model

Goals

Estimation Quantitative deductions on causes or consequences of the observations, i.e., find underlying model parameters.

Example: *I observed a certain amount of rain drops forming on my window in the last minute. What is the current rainfall in milimeters?*

Prediction From the infered model, predict what missing or future observations should be.

Example: *How many more raindrops will form on my window in the next hour?*

Decision Take a decision out of a discrete set of choices

Example: *Is it safe to open my window 1 minute to get some fresh air?*

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

Classical/Frequentist Statistics	Bayesian Statistics

Definition

Features

Tools

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

	Classical/Frequentist Statistics	Bayesian Statistics
Definition	The model parameters which should be estimated are considered as **unknown constant**.	
Features		
Tools		

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

	Classical/Frequentist Statistics	Bayesian Statistics
Definition	The model parameters which should be estimated are considered as **unknown constant**.	The model parameters are considered as **hidden random variables**, following a hypothetical **probabilistic model**.
Features		
Tools		

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

	Classical/Frequentist Statistics	Bayesian Statistics
Definition	The model parameters which should be estimated are considered as **unknown constant**.	The model parameters are considered as **hidden random variables**, following a hypothetical **probabilistic model**.
Features	Inference entirely based on **observed data** and **frequentist arguments**. Useful when few prior knowledge exist on the underlying random process.	
Tools		

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

	Classical/Frequentist Statistics	Bayesian Statistics
Definition	The model parameters which should be estimated are considered as **unknown constant**.	The model parameters are considered as **hidden random variables**, following a hypothetical **probabilistic model**.
Features	Inference entirely based on **observed data** and **frequentist arguments**. Useful when few prior knowledge exist on the underlying random process.	Incorporate **prior knowledge on the hidden variables** in the form of a generative probabilistic model. Useful when some **reasonnable** probability density fundtions (PDFs) can be assumed.
Tools		

What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

	Classical/Frequentist Statistics	Bayesian Statistics
Definition	The model parameters which should be estimated are considered as **unknown constant**.	The model parameters are considered as **hidden random variables**, following a hypothetical **probabilistic model**.
Features	Inference entirely based on **observed data** and **frequentist arguments**. Useful when few prior knowledge exist on the underlying random process.	Incorporate **prior knowledge on the hidden variables** in the form of a generative probabilistic model. Useful when some **reasonnable** probability density fundtions (PDFs) can be assumed.
Tools	• Linear estimators • First and second order statistics • A lot of: $\mathbb{E}\{\dots\}$	

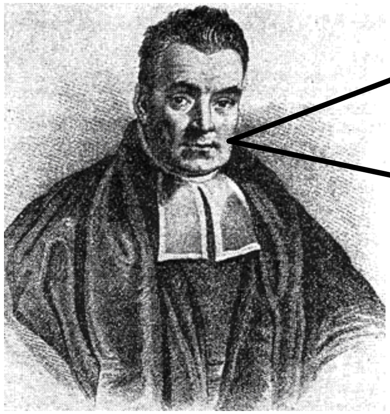
What is Bayesian Inference?

Statistics: Inference from the real world observations of a random phenomenon using probability theory

	Classical/Frequentist Statistics	Bayesian Statistics
Definition	The model parameters which should be estimated are considered as **unknown constant**.	The model parameters are considered as **hidden random variables**, following a hypothetical **probabilistic model**.
Features	Inference entirely based on **observed data** and **frequentist arguments**. Useful when few prior knowledge exist on the underlying random process.	Incorporate **prior knowledge on the hidden variables** in the form of a generative probabilistic model. Useful when some **reasonnable** probability density fundtions (PDFs) can be assumed.
Tools	• Linear estimators • First and second order statistics • A lot of: $\mathbb{E}\{\dots\}$	• **Bayes' Theorem** • Explicit PDFs • A lot of: $\mathcal{N}(\dots)$

What is Bayesian Inference?

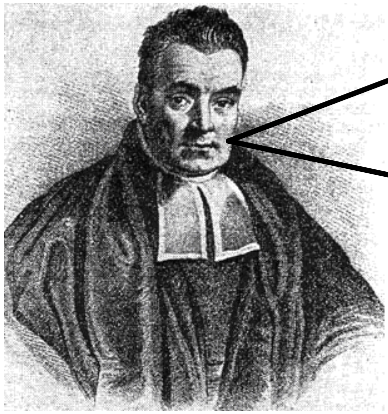
- Bayes' Theorem



$$p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x})=\frac{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})p(\boldsymbol{Z}=\boldsymbol{z})}{p(\boldsymbol{X}=\boldsymbol{x})}$$

What is Bayesian Inference?

- Bayes' Theorem

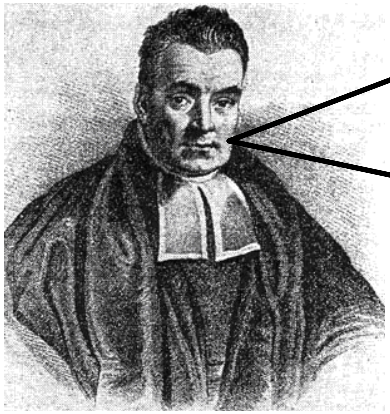


$\boldsymbol{X}$: Observed variables $\boldsymbol{Z}$: Hidden variables

$$p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x}) = \frac{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})p(\boldsymbol{Z}=\boldsymbol{z})}{p(\boldsymbol{X}=\boldsymbol{x})}$$

What is Bayesian Inference?

- Bayes' Theorem



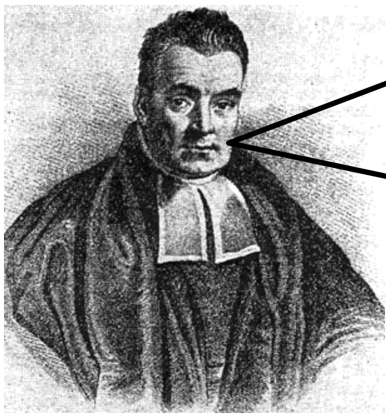
$\boldsymbol{X}$: Observed variables $\boldsymbol{Z}$: Hidden variables

Posterior

$$p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x}) = \frac{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})p(\boldsymbol{Z}=\boldsymbol{z})}{p(\boldsymbol{X}=\boldsymbol{x})}$$

What is Bayesian Inference?

- Bayes' Theorem



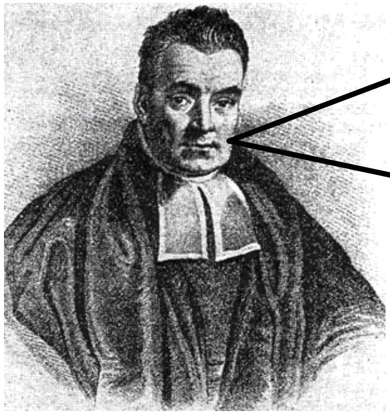
$\boldsymbol{X}$: Observed variables $\boldsymbol{Z}$: Hidden variables

Posterior Likelihood

$$p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x}) = \frac{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})p(\boldsymbol{Z}=\boldsymbol{z})}{p(\boldsymbol{X}=\boldsymbol{x})}$$

What is Bayesian Inference?

- Bayes' Theorem

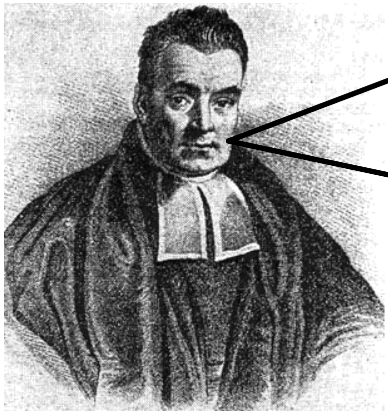


$\boldsymbol{X}$: Observed variables $\quad \boldsymbol{Z}$: Hidden variables

$$\underbrace{p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{Z}=\boldsymbol{z})}^{\text{Prior}}}{p(\boldsymbol{X}=\boldsymbol{x})}$$

What is Bayesian Inference?

- Bayes' Theorem



$\boldsymbol{X}$: Observed variables $\boldsymbol{Z}$: Hidden variables

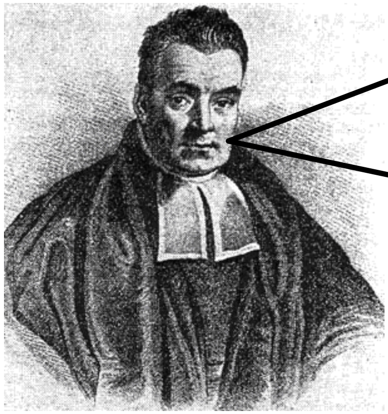
Posterior, Likelihood, Prior

$$\underbrace{p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{Z}=\boldsymbol{z})}^{\text{Prior}}}{p(\boldsymbol{X}=\boldsymbol{x})}$$

« Observed data » or « marginal » likelihood

What is Bayesian Inference?

- Bayes' Theorem



$\boldsymbol{X}$: Observed variables $\quad \boldsymbol{Z}$: Hidden variables

Posterior, Likelihood, Prior

$$\underbrace{p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{Z}=\boldsymbol{z})}^{\text{Prior}}}{p(\boldsymbol{X}=\boldsymbol{x})}$$

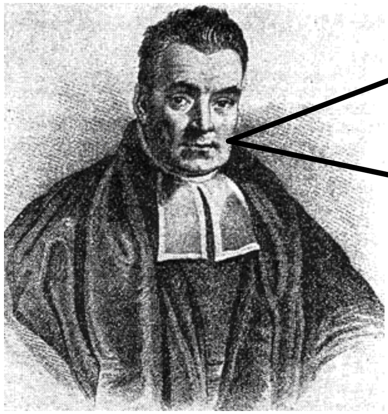
« Observed data » or
« marginal » likelihood

Remark 1: ***Bayes does not « forbid » model parameters !***

- No formal difference between a parameter and a hidden variable with constant prior
- Priors distributions often have parameters called « hyperparameters »

What is Bayesian Inference?

- Bayes’ Theorem



$\boldsymbol{X}$: Observed variables $\quad$ $\boldsymbol{Z}$: Hidden variables

$$\underbrace{p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z};\boldsymbol{\theta})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{Z}=\boldsymbol{z};\boldsymbol{\theta})}^{\text{Prior}}}{p(\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta})}$$

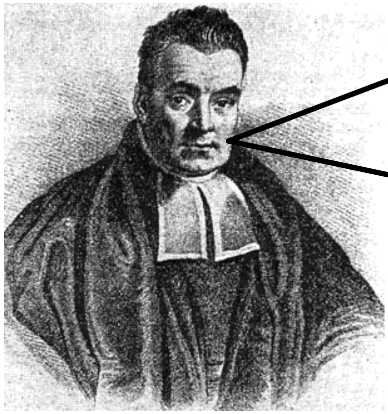
« Observed data » or
« marginal » likelihood

Remark 1: *Bayes does not « forbid » model parameters !*

- No formal difference between a parameter and a hidden variable with constant prior
- Priors distributions often have parameters called « hyperparameters »

What is Bayesian Inference?

- Bayes' Theorem



$\boldsymbol{X}$: Observed variables $\boldsymbol{Z}$: Hidden variables

$$\underbrace{p(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{Z}=\boldsymbol{z};\boldsymbol{\theta})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{Z}=\boldsymbol{z};\boldsymbol{\theta})}^{\text{Prior}}}{p(\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta})}$$

« Observed data » or
« marginal » likelihood

Remark 1: ***Bayes does not « forbid » model parameters !***
- No formal difference between a parameter and a hidden variable with constant prior
- Priors distributions often have parameters called « hyperparameters »

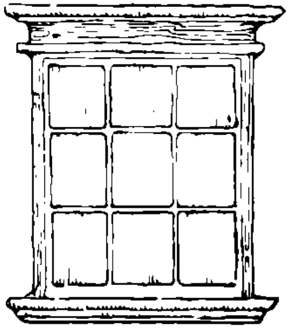
Remark 2: ***Why hidden variables?***
- Formally not needed: $p(\boldsymbol{X})$ can be obtained by marginalizing out hidden variables
- A convenient and powerful **view point** which makes inference possible in complex scenarios through a variety of methods

What is Bayesian Inference?

An Example:

Observations

X_{D}	X_{C}
I see drops on my window	I see a cat or a dog at my window

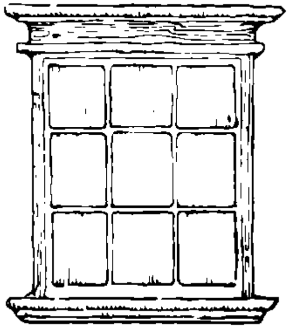


What is Bayesian Inference?

An Example:

Observations

X_{D}	X_{C}
I see drops on my window	I see a cat or a dog at my window



Hypothesis

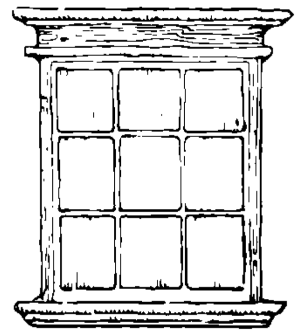
$Z = 0$	$Z = 1$	$Z = 2$
It's not raining	It's raining rain	It's raining cats and dogs

What is Bayesian Inference?

An Example:

Observations

X_{D}	X_{C}
I see drops on my window	I see a cat or a dog at my window



Hypothesis

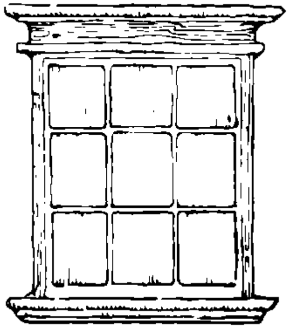
$Z = 0$	$Z = 1$	$Z = 2$
It's not raining	It's raining rain	It's raining cats and dogs

$p(X_i | Z = j)$

	X_{D}	X_{C}
$Z = 0$	0.1	0.1
$Z = 1$	0.99	0.1
$Z = 2$	0.1	0.99

What is Bayesian Inference?

An Example:



Observations

X_{D}	X_{C}
I see drops on my window	I see a cat or a dog at my window

Hypothesis

$Z=0$	$Z=1$	$Z=2$
It’s not raining	It’s raining rain	It’s raining cats and dogs

$p(X_i|Z=j)$

	X_{D}	X_{C}
$Z=0$	0.1	0.1
$Z=1$	0.99	0.1
$Z=2$	0.1	0.99

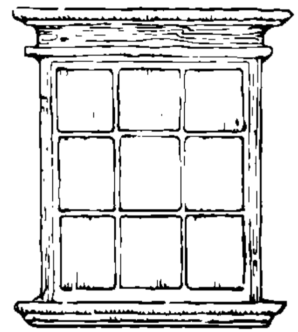
$p(X_{\mathrm{D}},X_{\mathrm{C}}|Z=1)=p(X_{\mathrm{D}},X_{\mathrm{C}}|Z=2)=0.099$: equal likelihood!

What is Bayesian Inference?

An Example:

Observations

X_{D}	X_{C}
I see drops on my window	I see a cat or a dog at my window



Hypothesis

$Z=0$	$Z=1$	$Z=2$
It's not raining	It's raining rain	It's raining cats and dogs

$p(X_i|Z=j)$

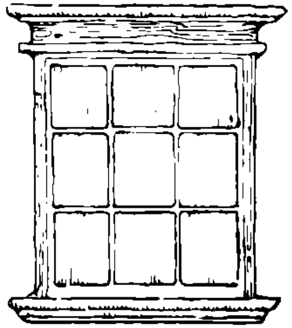
	X_{D}	X_{C}
$Z=0$	0.1	0.1
$Z=1$	0.99	0.1
$Z=2$	0.1	0.99

$p(X_{\mathrm{D}},X_{\mathrm{C}}|Z=1)=p(X_{\mathrm{D}},X_{\mathrm{C}}|Z=2)=0.099$: equal likelihood!

- **Add priors**: $p(Z=0)=49.995\%$ $p(Z=1)=49.995\%$ $p(Z=2)=0.01\%$

What is Bayesian Inference?

An Example:



Observations

X_{D}	X_{C}
I see drops on my window	I see a cat or a dog at my window

Hypothesis

$Z=0$	$Z=1$	$Z=2$
It’s not raining	It’s raining rain	It’s raining cats and dogs

$p(X_i|Z=j)$

	X_{D}	X_{C}
$Z=0$	0.1	0.1
$Z=1$	0.99	0.1
$Z=2$	0.1	0.99

$p(X_{\mathrm{D}},X_{\mathrm{C}}|Z=1)=p(X_{\mathrm{D}},X_{\mathrm{C}}|Z=2)=0.099$: equal likelihood!

- **Add priors**: $p(Z=0)=49.995\%$ $p(Z=1)=49.995\%$ $p(Z=2)=0.01\%$
- **Bayes’ theorem:**

$$p(Z=1|X_{\mathrm{D}},X_{\mathrm{C}})\approx 99.98\% \quad p(Z=2|X_{\mathrm{D}},X_{\mathrm{C}})\approx 0.02\%$$

What is Bayesian Inference?

General Methodology

What is Bayesian Inference?

General Methodology

Modeling

What is Bayesian Inference?

General Methodology

Modeling

- What is hidden?
- What is observed?
- Dependencies?

General Methodology

Modeling

- What is hidden?
- What is observed?
- Dependencies?

Graphical model

What is Bayesian Inference?

General Methodology

Modeling

- What is hidden?
- What is observed?
- Dependencies?

Graphical model + Choice of prior and conditional PDFs

General Methodology

Modeling

What is hidden?

What is observed?

Dependencies?

Graphical model + Choice of prior and conditional PDFs

=

Joint PDF

What is Bayesian Inference?

General Methodology

Modeling

- What is hidden?
- What is observed?
- Dependencies?

Graphical model + Choice of prior and conditional PDFs = **Joint PDF**

Inference

What is Bayesian Inference?

General Methodology

Modeling

- What is hidden?
- What is observed?
- Dependencies?

Graphical model + Choice of prior and conditional PDFs = **Joint PDF**

Inference

Apply Baye's Theorem

What is Bayesian Inference?

General Methodology

Modeling

- What is hidden?
- What is observed?
- Dependencies?

Graphical model + Choice of prior and conditional PDFs = **Joint PDF**

Inference

Apply Baye's Theorem →

Choice of method:
- Exact / Approximate
- Direct / Iterative

What is Bayesian Inference?

General Methodology

Modeling

What is hidden?

What is observed?

Dependencies?

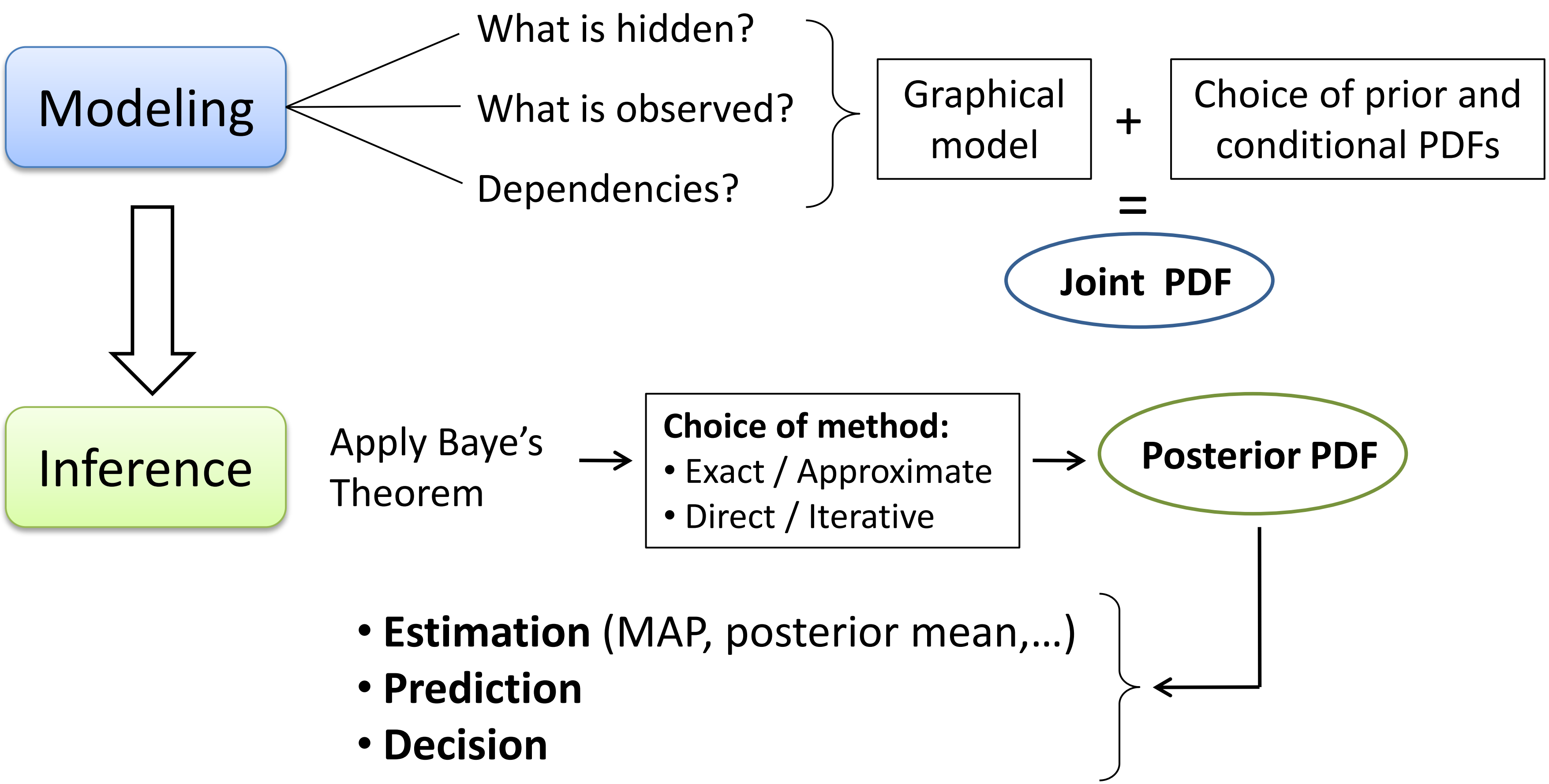
Graphical model + Choice of prior and conditional PDFs = **Joint PDF**

Inference

Apply Baye's Theorem → **Choice of method:**
- Exact / Approximate
- Direct / Iterative

→ **Posterior PDF**

General Methodology



OUTLINE

- **What is Bayesian inference?**
  - **Overview**
  - **Classical vs. Bayesian approach**
  - **Bayes Theorem & Example**
  - **General Methodology**
- Bayesian inference by examples
  - Direct inference
  - The Expectation-Maximization algorithm
  - Variational Bayes methods
- Markov chain Monte Carlo

OUTLINE

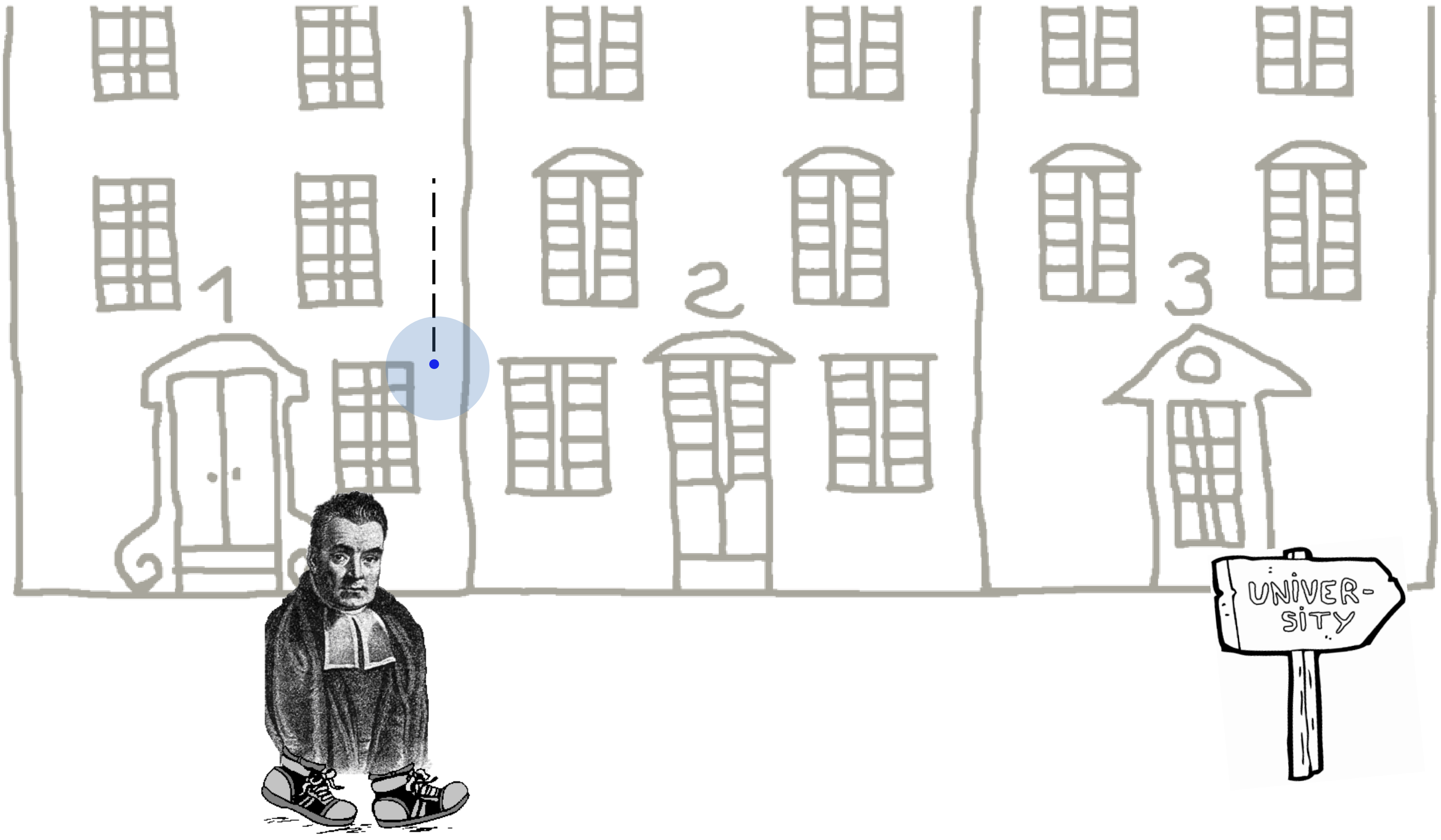
- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

OUTLINE

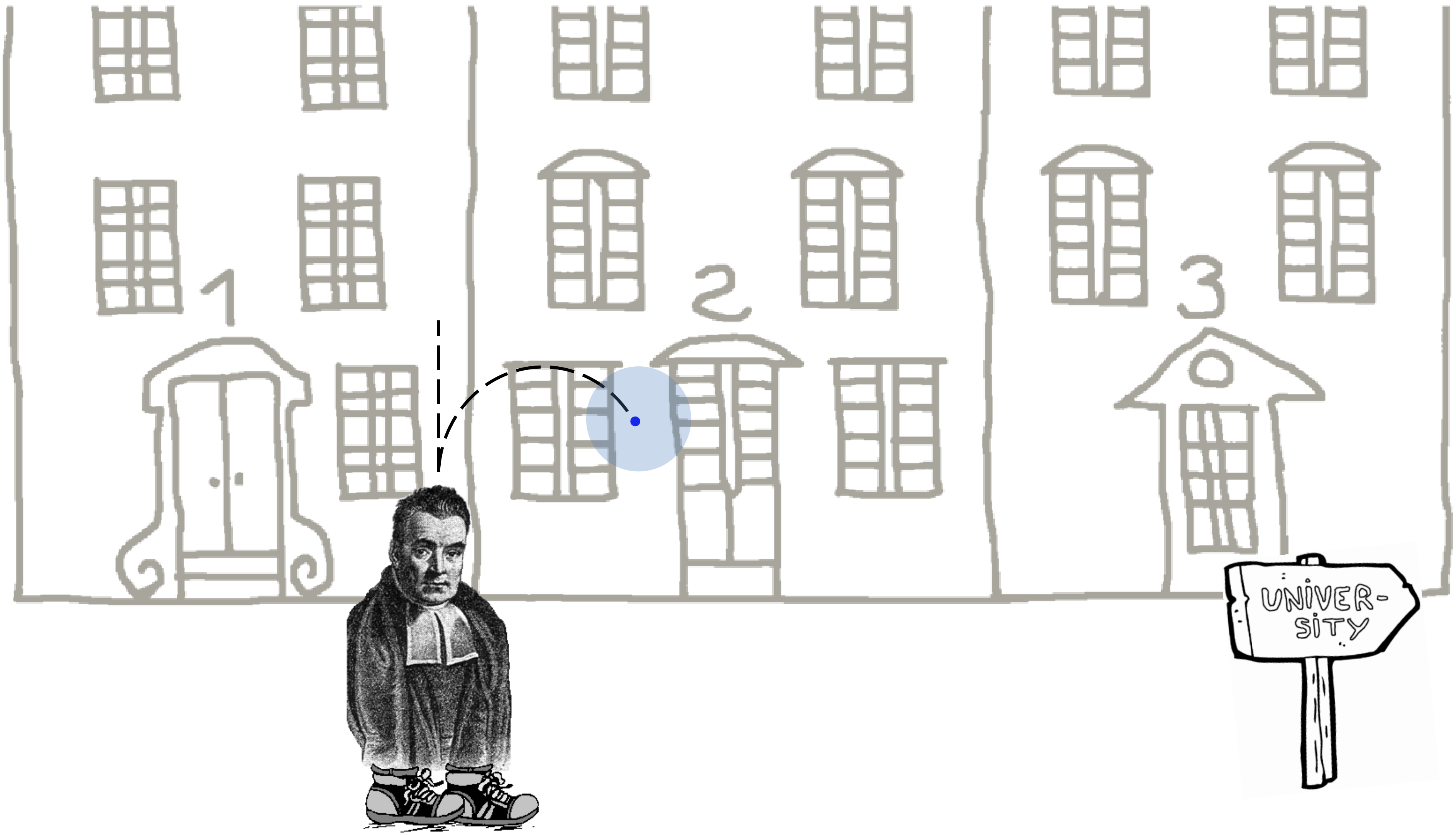
- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

Bayesian Inference: Examples

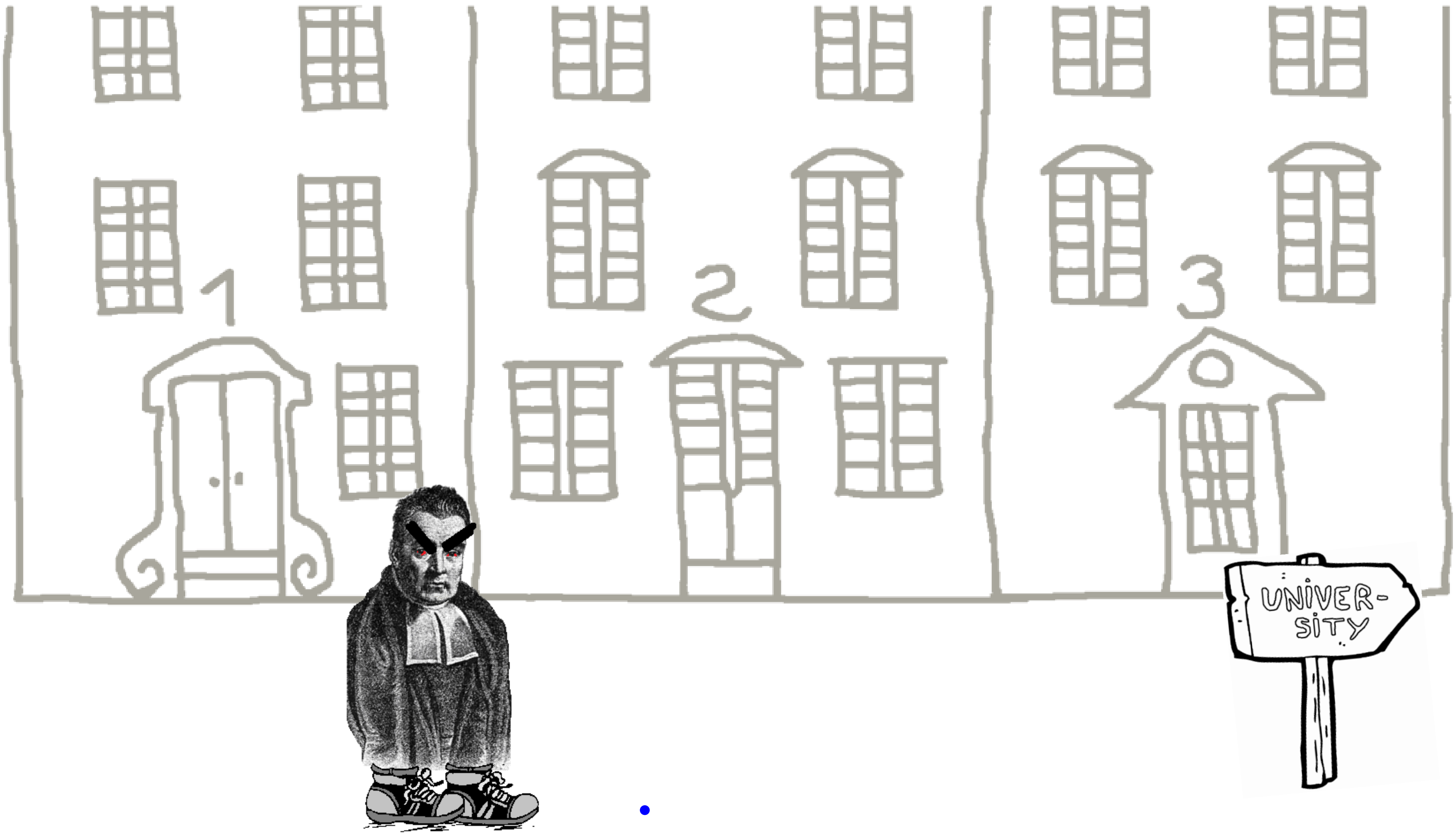
Bayesian Inference: Examples



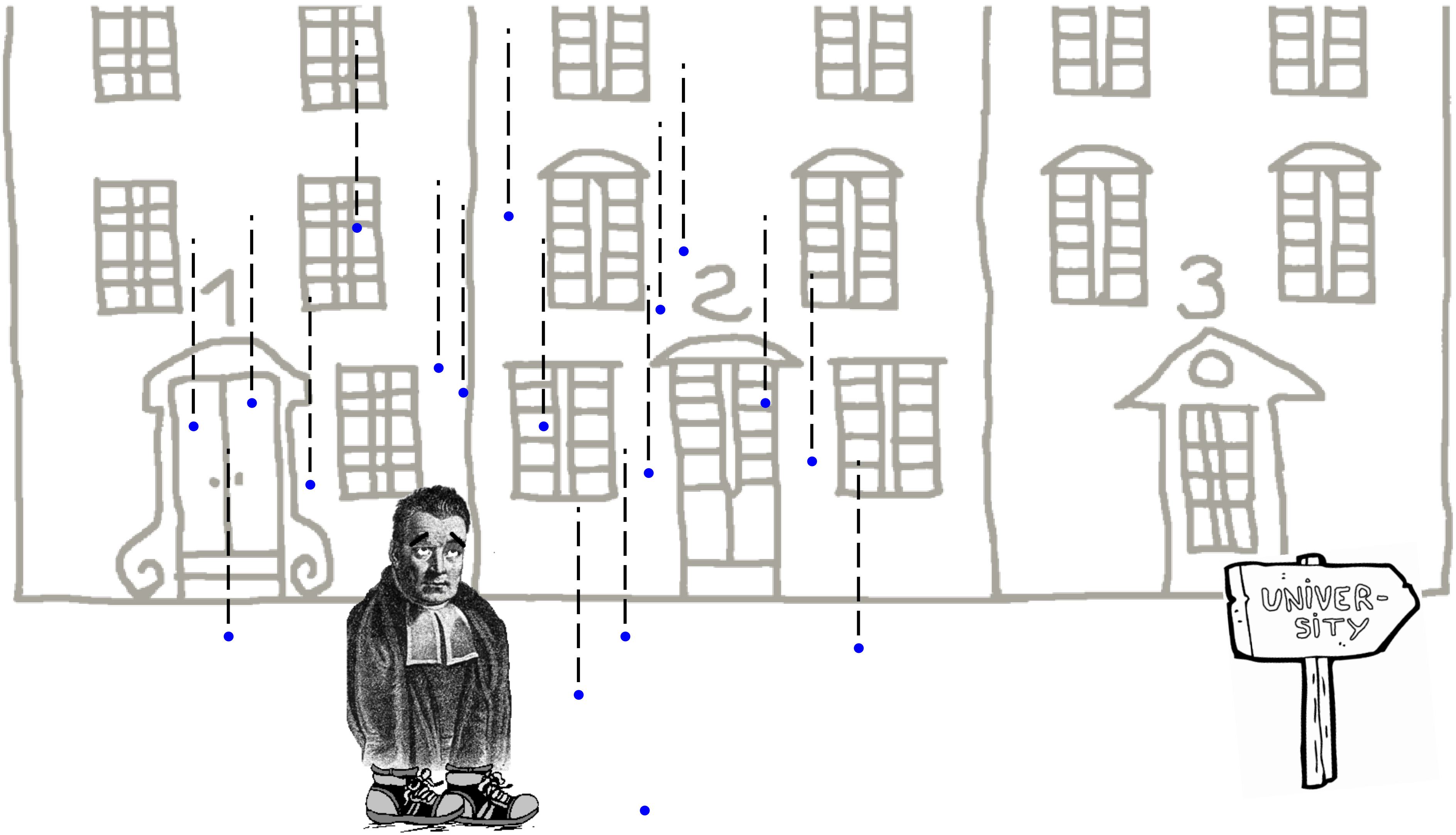
Bayesian Inference: Examples



Bayesian Inference: Examples

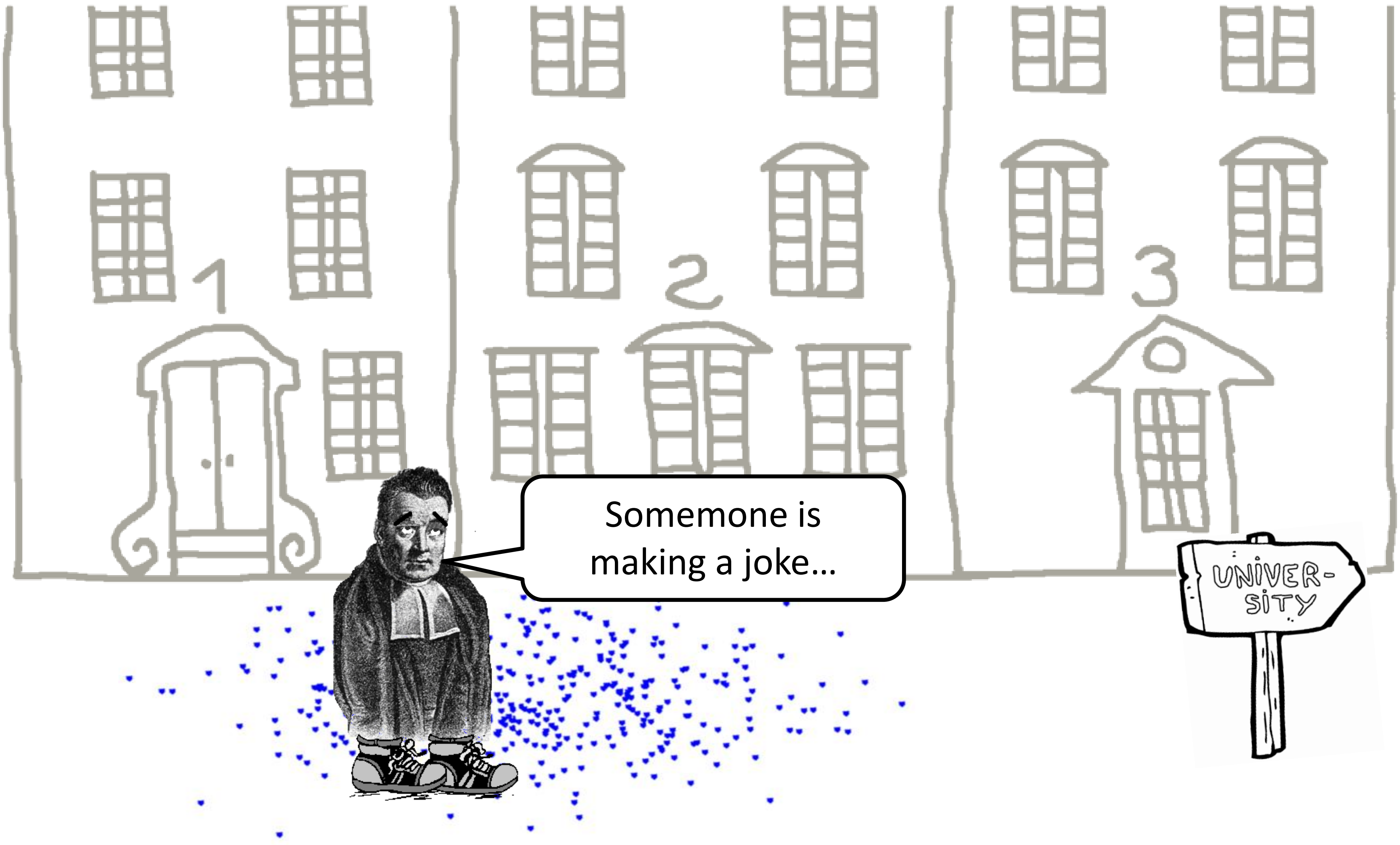


Bayesian Inference: Examples



Bayesian Inference: Examples

Bayesian Inference: Examples



Bayesian Inference: Examples

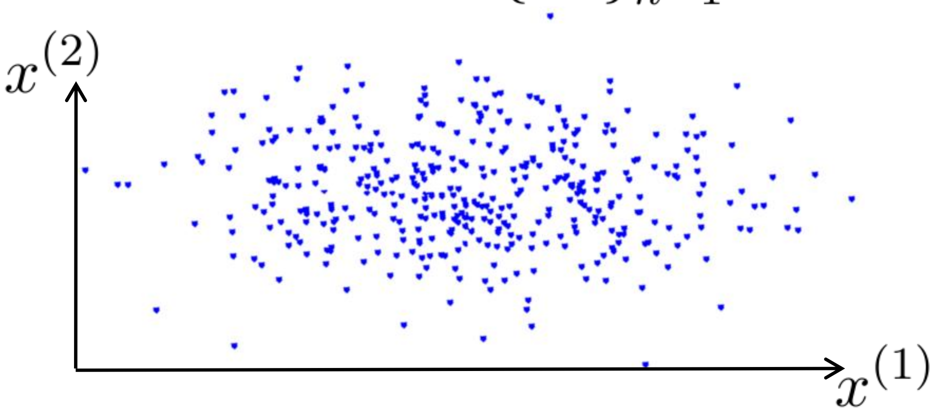
Bayesian Inference: Examples

Modeling

Bayesian Inference: Examples

Modeling

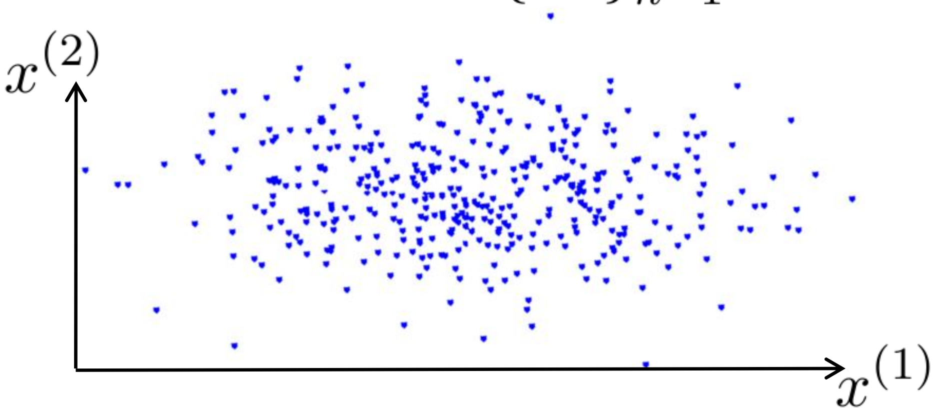
Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$



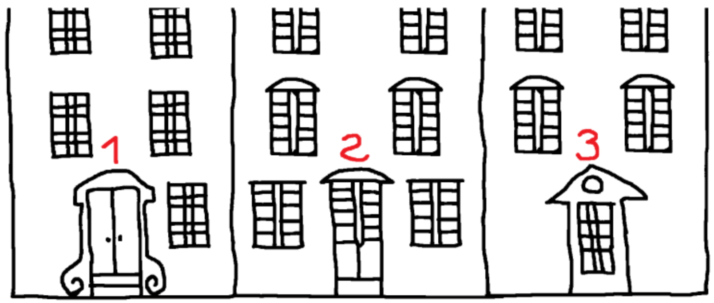
Bayesian Inference: Examples

Modeling

Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$



Hidden variable: $Z \in \{1, 2, 3\}$

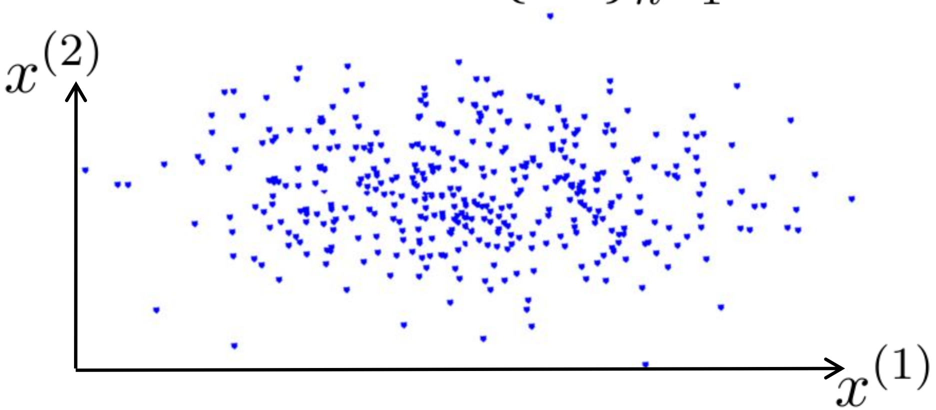


Guilty house number?

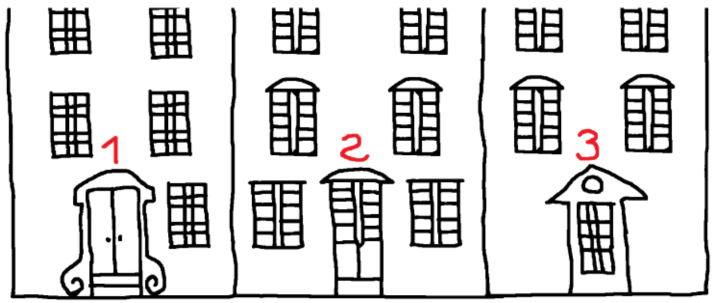
Bayesian Inference: Examples

Modeling

Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^N \subset \mathbb{R}^2$

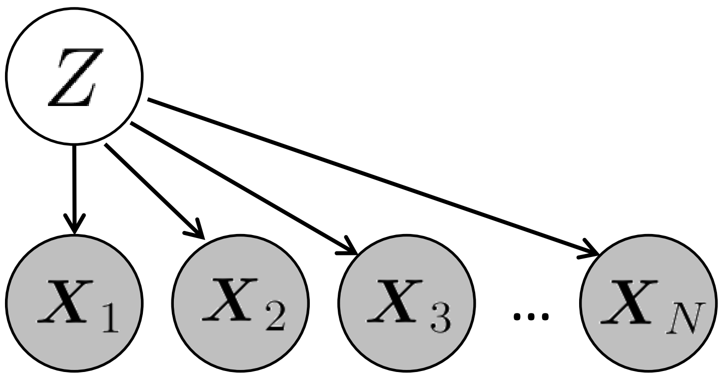


Hidden variable: $Z \in \{1, 2, 3\}$



Guilty house number?

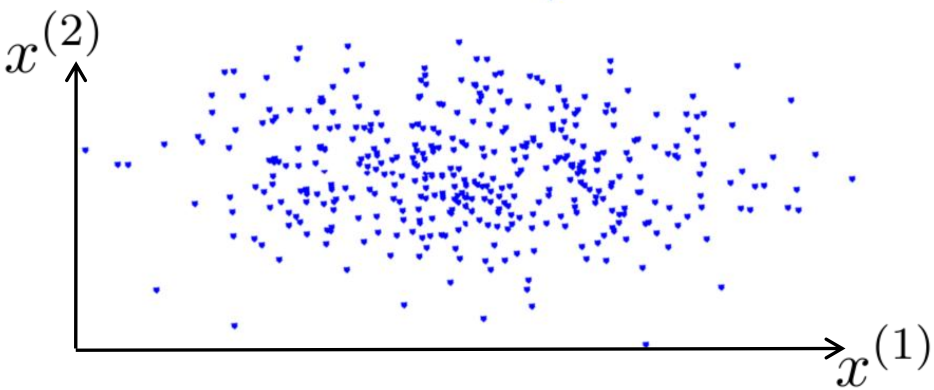
Graphical Model:



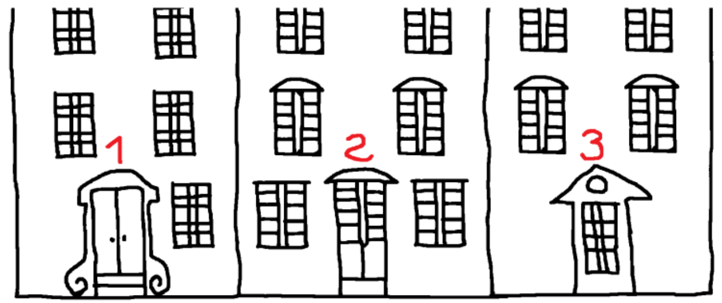
Bayesian Inference: Examples

Modeling

Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^N \subset \mathbb{R}^2$

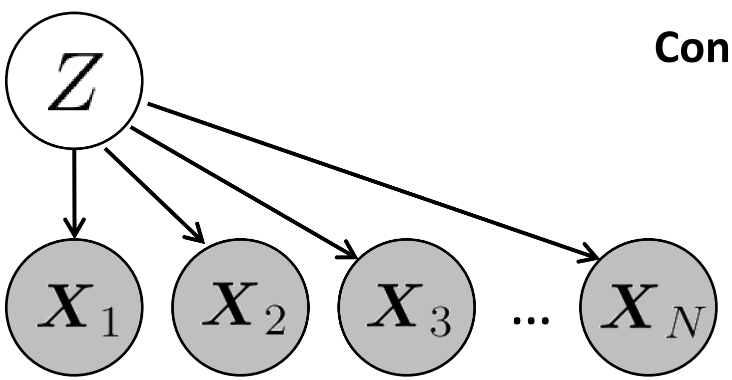


Hidden variable: $Z \in \{1, 2, 3\}$



Guilty house number?

Graphical Model:



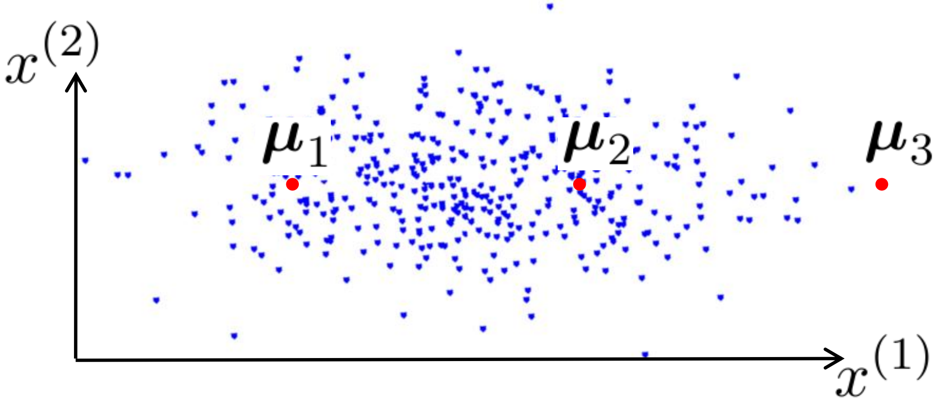
Conditionals

$$\begin{cases} p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 1) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_1, \mathbf{I}) \\ p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 2) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_2, \mathbf{I}) \\ p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 3) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_3, \mathbf{I}) \end{cases}$$

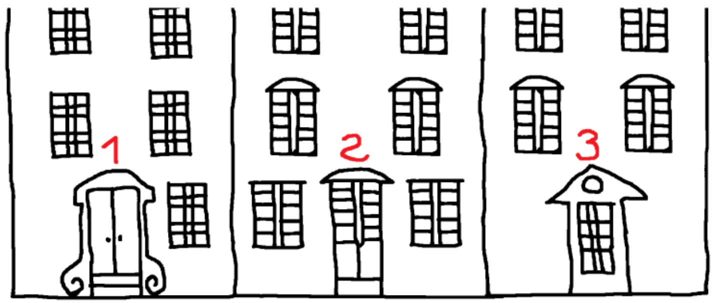
Bayesian Inference: Examples

Modeling

Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$

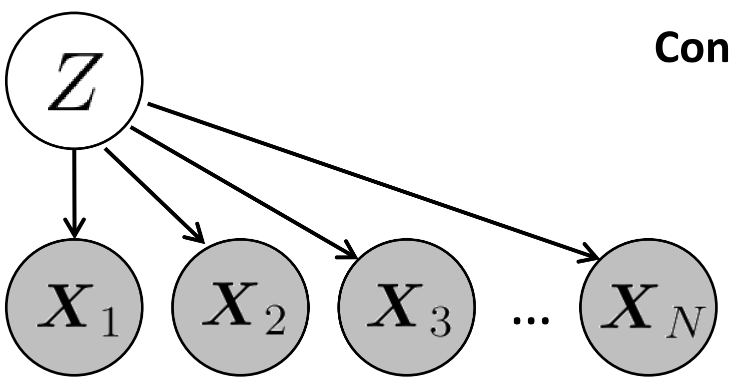


Hidden variable: $Z \in \{1, 2, 3\}$



Guilty house number?

Graphical Model:



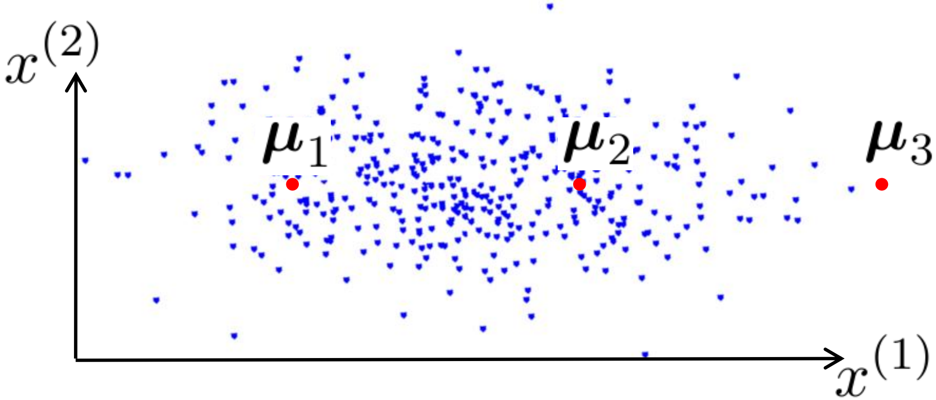
Conditionals

$$\begin{cases} p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 1) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_1, \mathbf{I}) \\ p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 2) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_2, \mathbf{I}) \\ p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 3) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_3, \mathbf{I}) \end{cases}$$

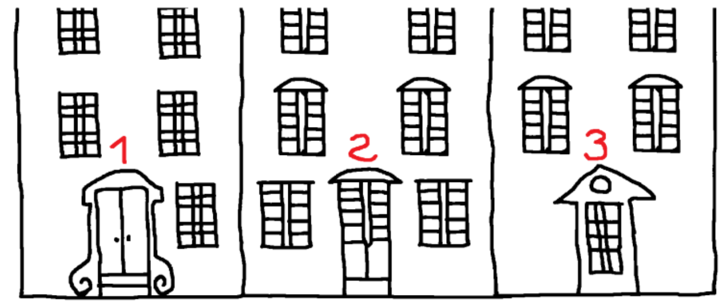
Bayesian Inference: Examples

Modeling

Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$

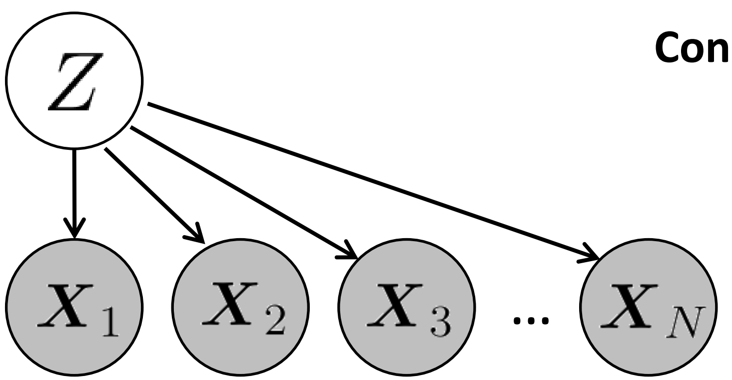


Hidden variable: $Z \in \{1, 2, 3\}$



Guilty house number?

Graphical Model:



Conditionals

$$\begin{cases} p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 1) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_1, \mathbf{I}) \\ p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 2) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_2, \mathbf{I}) \\ p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z = 3) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_3, \mathbf{I}) \end{cases}$$

Priors

$$\begin{cases} p(Z = 1) = 0.1 & \text{(Grandma Jane)} \\ p(Z = 2) = 0.3 & \text{(Student house)} \\ p(Z = 3) = 0.6 & \text{(Family with kids)} \end{cases}$$

Bayesian Inference: Examples

Inference

Bayes' Theorem:

$$p(Z = i | \boldsymbol{x}_1, \dots, \boldsymbol{x}_N) = \frac{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N | Z = i) p(Z = i)}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)}$$

Bayesian Inference: Examples

Inference

Bayes' Theorem:

$$p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N) = \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N|Z=i)p(Z=i)}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)}$$

$$= \frac{\prod_{n=1}^{N} p(\boldsymbol{x}_n|Z=i)p(Z=i)}{\sum_{k=1}^{3}\prod_{n=1}^{N} p(\boldsymbol{x}_n|Z=k)p(Z=k)}$$

Bayesian Inference: Examples

Inference

Bayes' Theorem:

$$p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N) = \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N|Z=i)p(Z=i)}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)}$$

$$= \frac{\prod_{n=1}^N p(\boldsymbol{x}_n|Z=i)p(Z=i)}{\sum_{k=1}^3 \prod_{n=1}^N p(\boldsymbol{x}_n|Z=k)p(Z=k)}$$

⇨ Direct computation

Bayesian Inference: Examples

Inference

Bayes' Theorem:

$$p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N) = \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N|Z=i)p(Z=i)}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)}$$

$$= \frac{\prod_{n=1}^N p(\boldsymbol{x}_n|Z=i)p(Z=i)}{\sum_{k=1}^3 \prod_{n=1}^N p(\boldsymbol{x}_n|Z=k)p(Z=k)}$$

$\Rightarrow$ Direct computation

Estimation: Maximum a Posteriori (MAP)

$$\hat{z} = \underset{i}{\operatorname{argmax}}\Big[p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)\Big]$$

Bayesian Inference: Examples

Inference

Bayes' Theorem:

$$p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N) = \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N|Z=i)p(Z=i)}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)}$$

$$= \frac{\prod_{n=1}^{N} p(\boldsymbol{x}_n|Z=i)p(Z=i)}{\sum_{k=1}^{3}\prod_{n=1}^{N} p(\boldsymbol{x}_n|Z=k)p(Z=k)} \Rightarrow$$ Direct computation

Estimation: Maximum a Posteriori (MAP)

$$\hat{z} = \underset{i}{\operatorname{argmax}}\left[p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)\right] \Rightarrow \hat{z} = 2$$, the student house

Bayesian Inference: Examples

Inference

Bayes' Theorem:

$$p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)=\frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N|Z=i)p(Z=i)}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)}$$

$$=\frac{\prod_{n=1}^N p(\boldsymbol{x}_n|Z=i)p(Z=i)}{\sum_{k=1}^3\prod_{n=1}^N p(\boldsymbol{x}_n|Z=k)p(Z=k)}$$

$\Rightarrow$ Direct computation

Estimation: Maximum a Posteriori (MAP)

$$\hat{z}=\underset{i}{\operatorname{argmax}}\Big[p(Z=i|\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)\Big]$$

$\Rightarrow$ $\hat{z}=2$, the student house

Decision:

These pranksters will hear from me at the Uni!

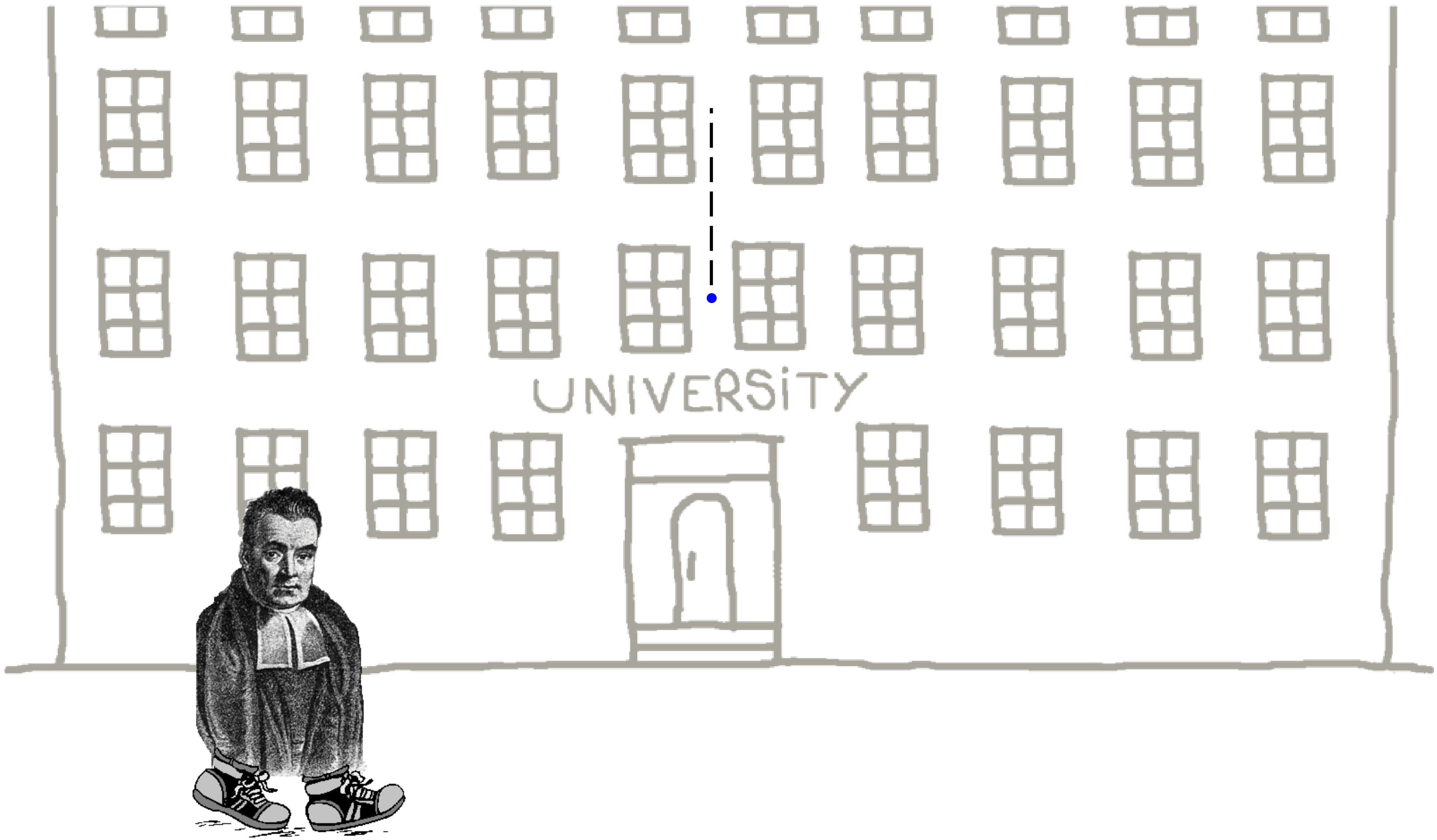
OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

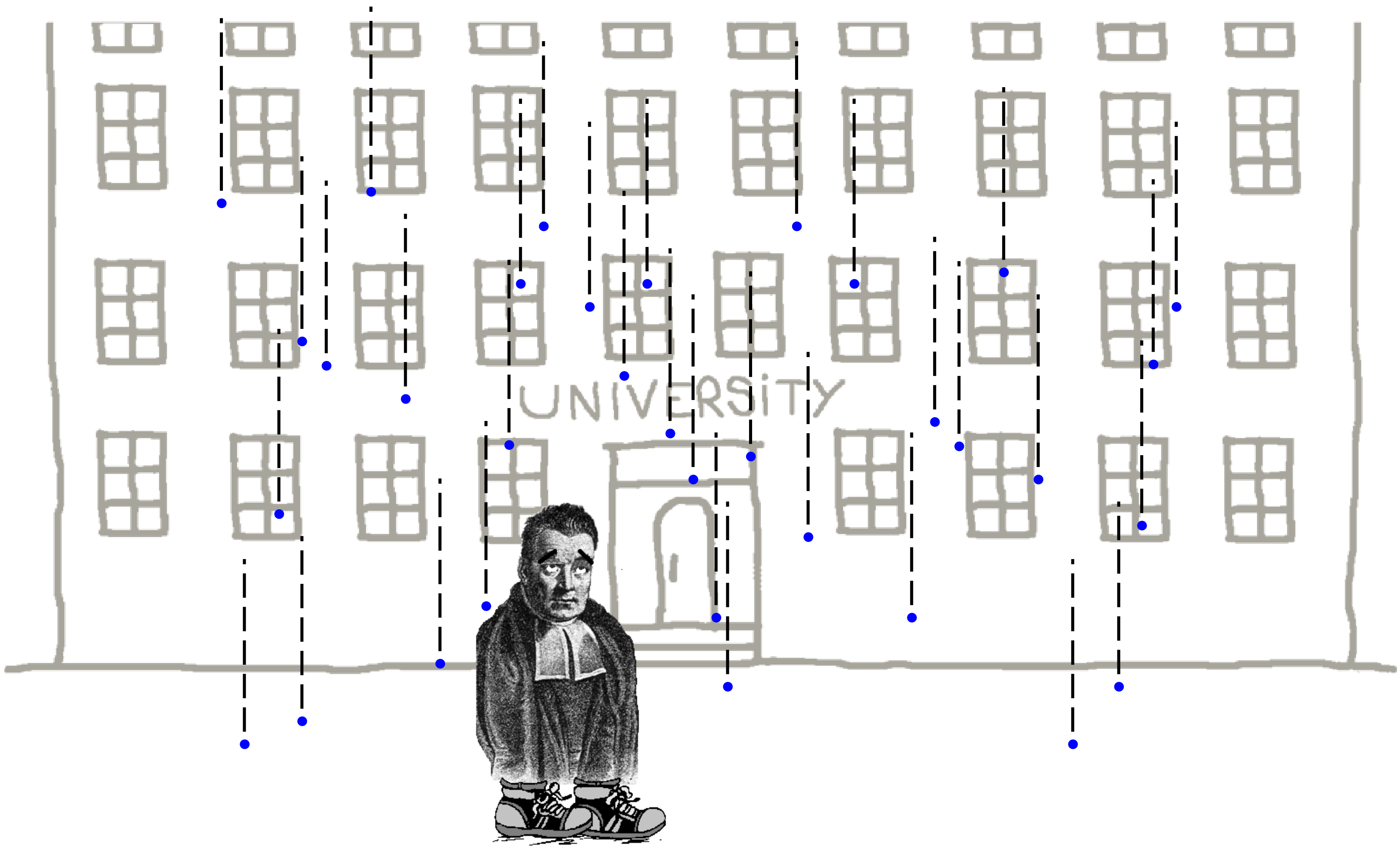
OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

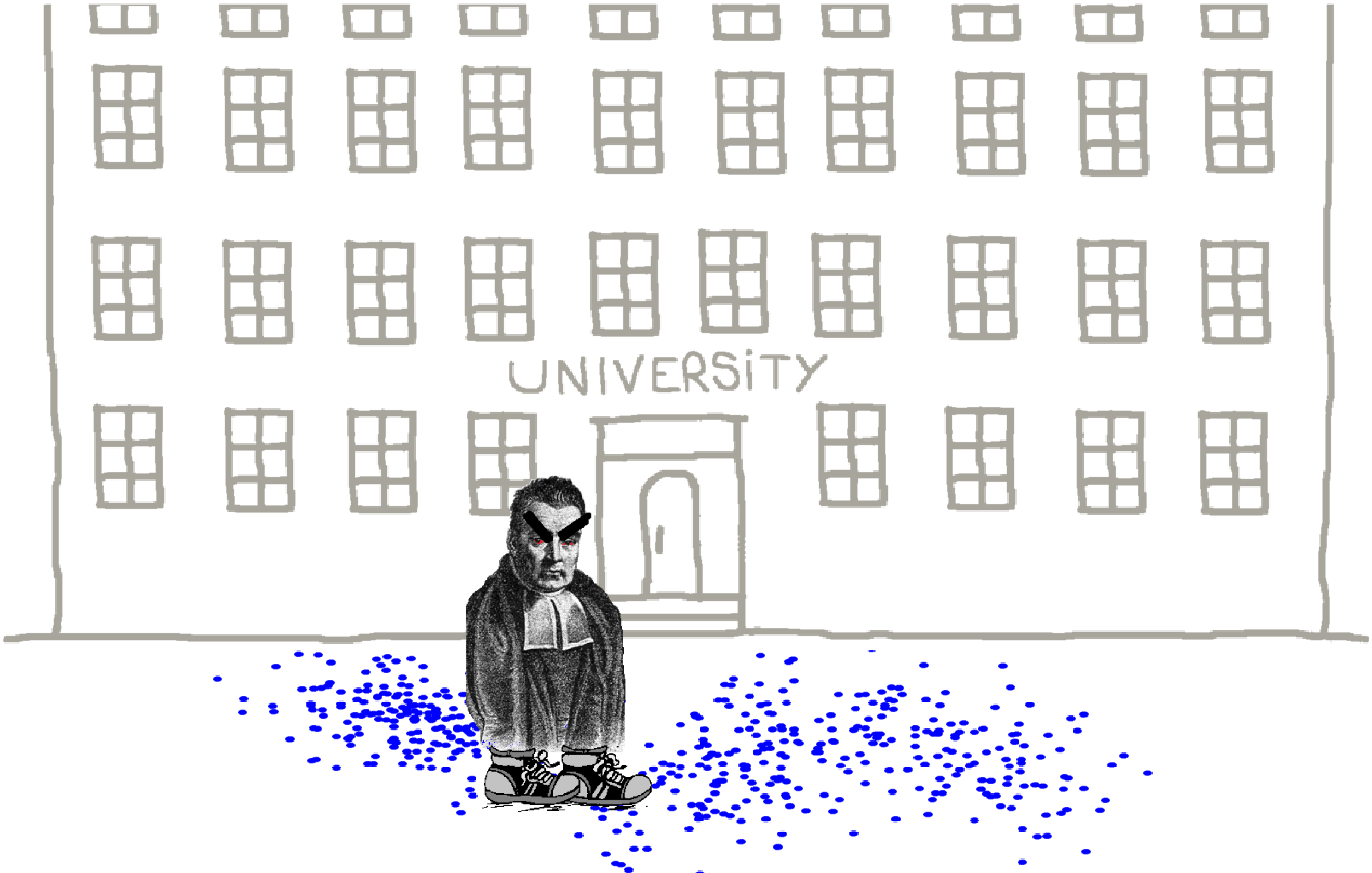
Bayesian Inference: Examples



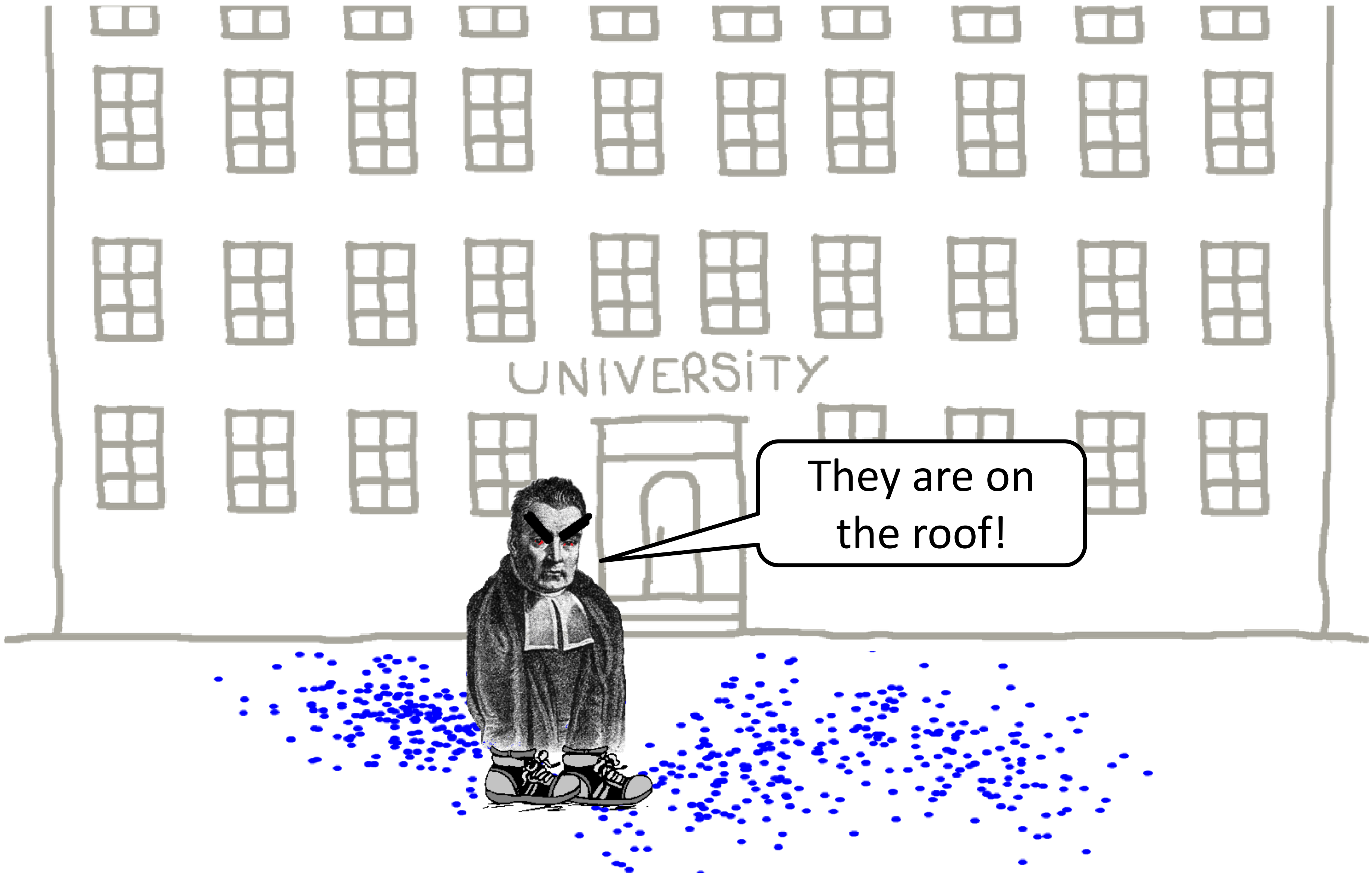
Bayesian Inference: Examples



Bayesian Inference: Examples



Bayesian Inference: Examples



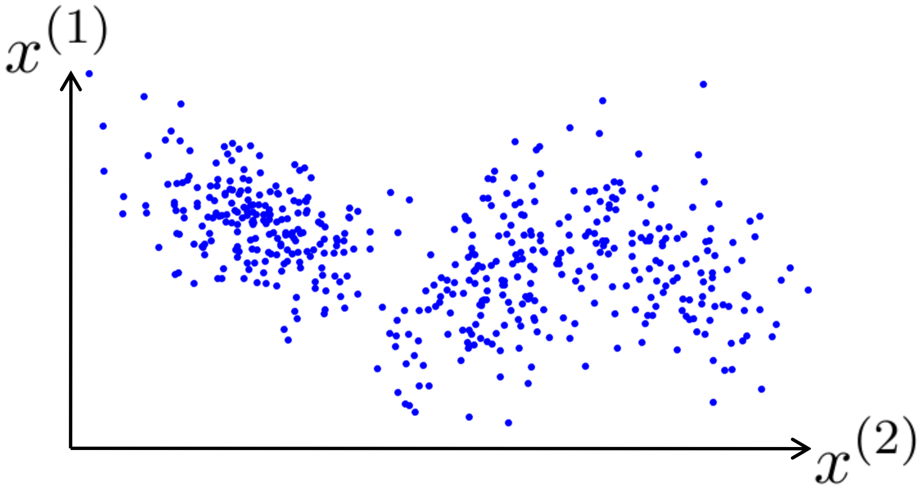
Bayesian Inference: Examples

Modeling

Bayesian Inference: Examples

Modeling

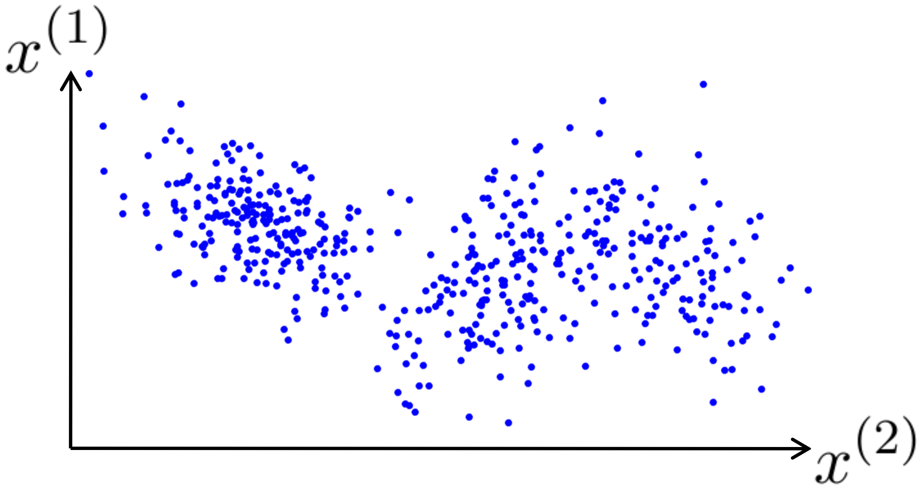
Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$



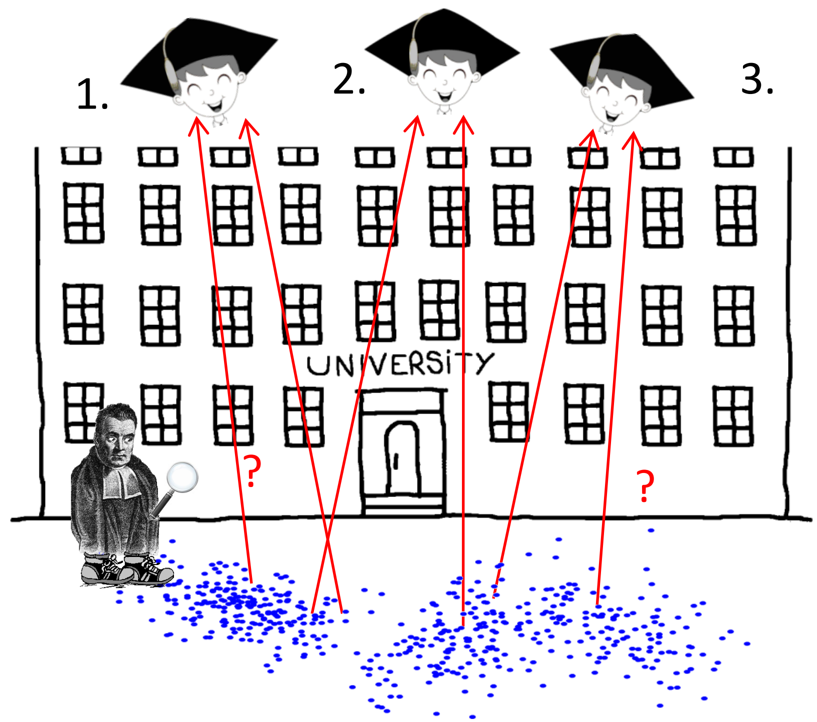
Bayesian Inference: Examples

Modeling

Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^N \subset \mathbb{R}^2$



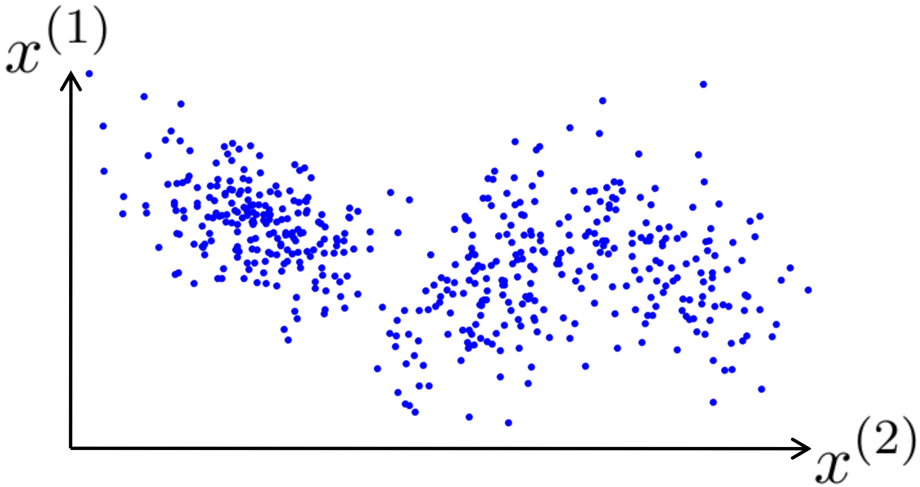
Hidden Variables: $\{Z_n\}_{n=1}^N \in \{1,2,3\}^N$



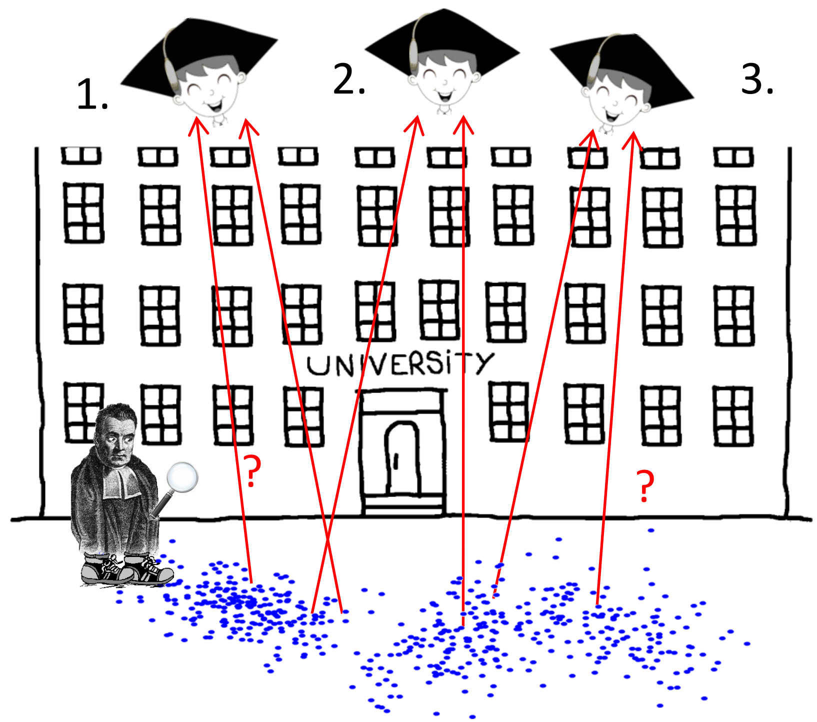
Bayesian Inference: Examples

Modeling

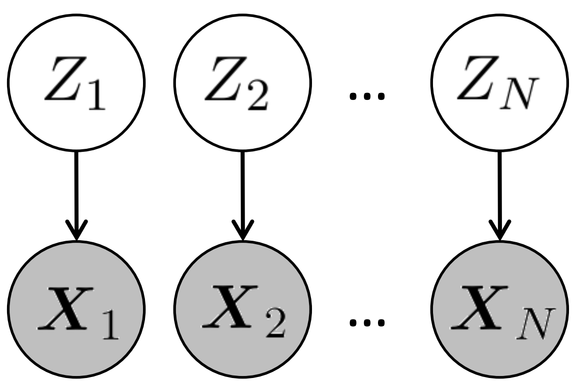
Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$



Hidden Variables: $\{Z_n\}_{n=1}^{N} \in \{1, 2, 3\}^N$



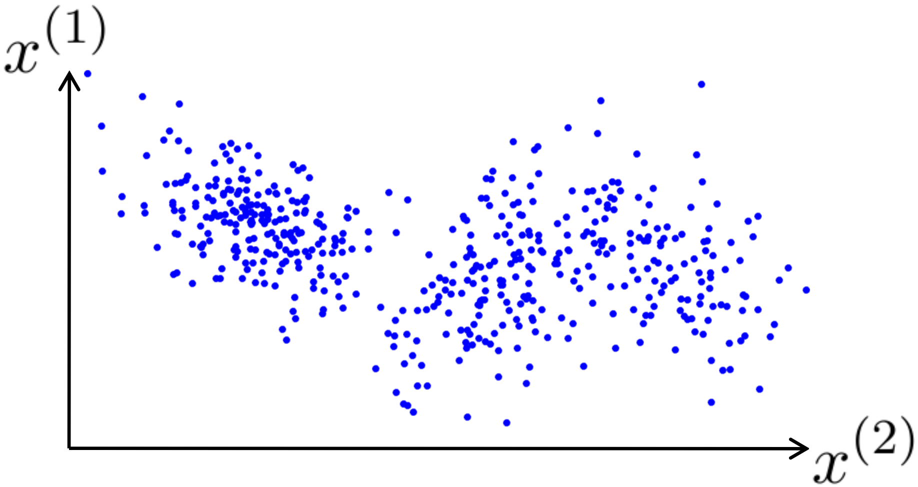
Graphical Model:



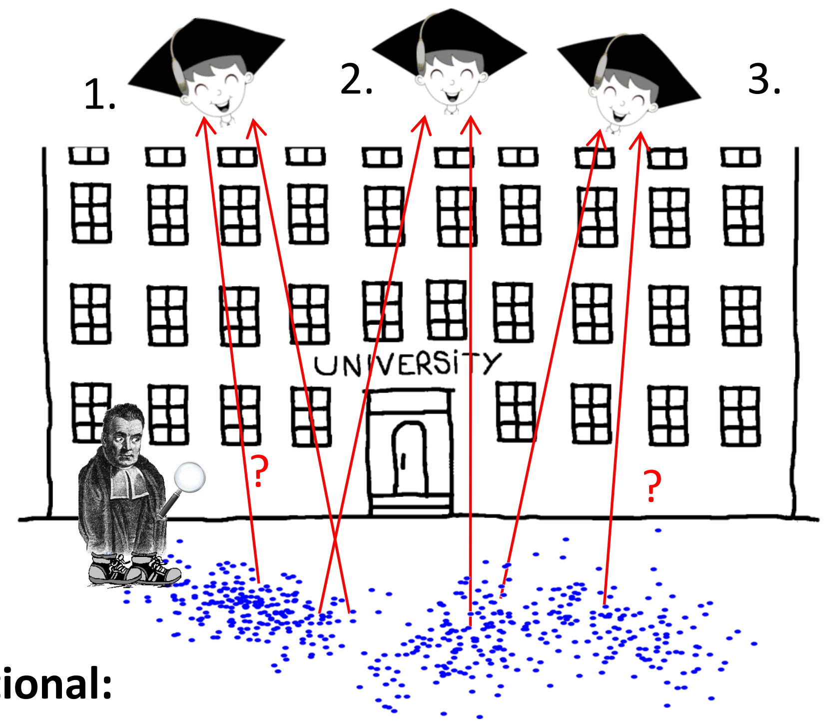
Bayesian Inference: Examples

Modeling

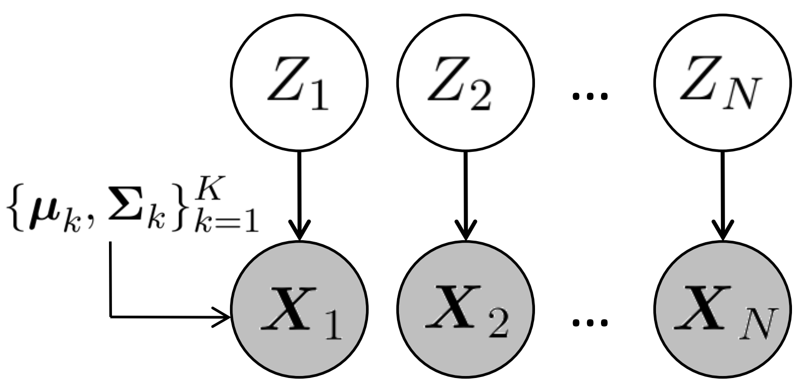
Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$



Hidden Variables: $\{Z_n\}_{n=1}^{N} \in \{1,2,3\}^N$



Graphical Model:



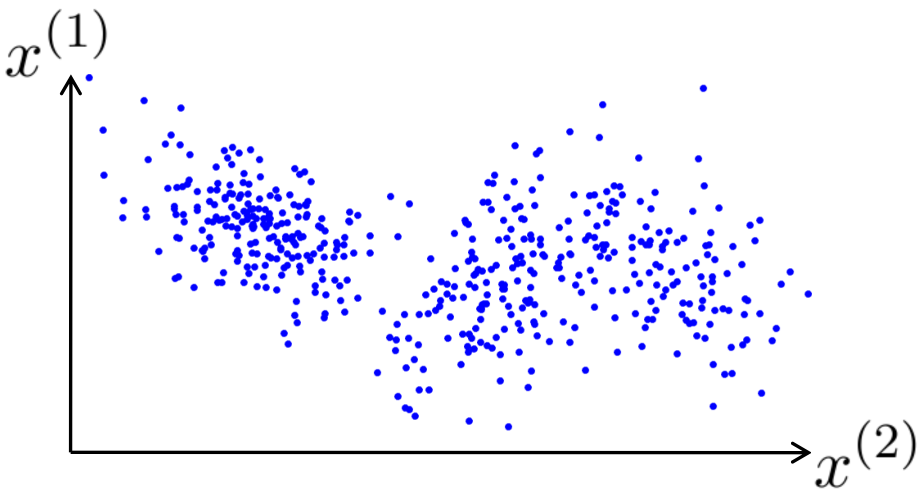
Conditional:

$$p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

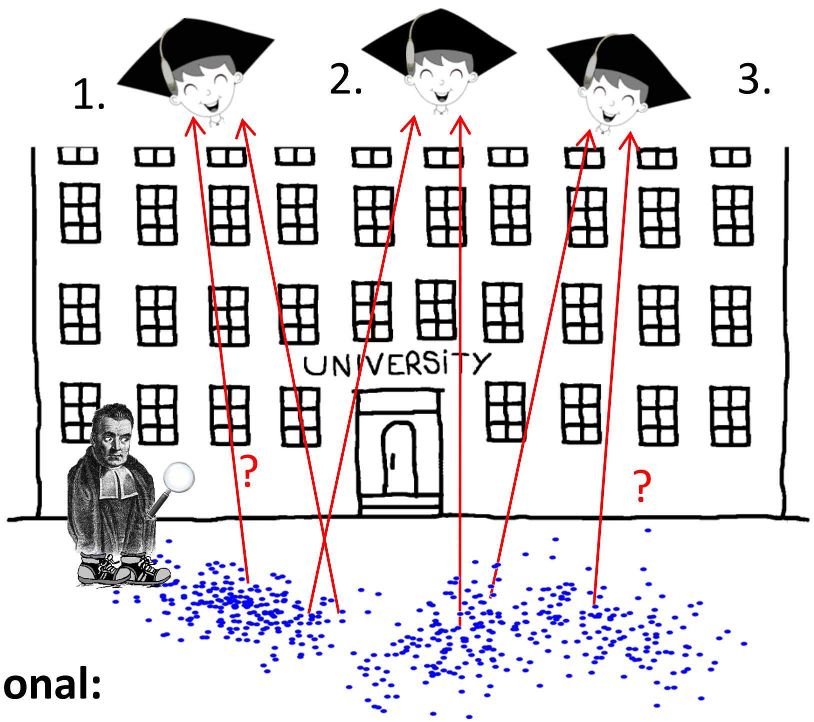
Bayesian Inference: Examples

Modeling

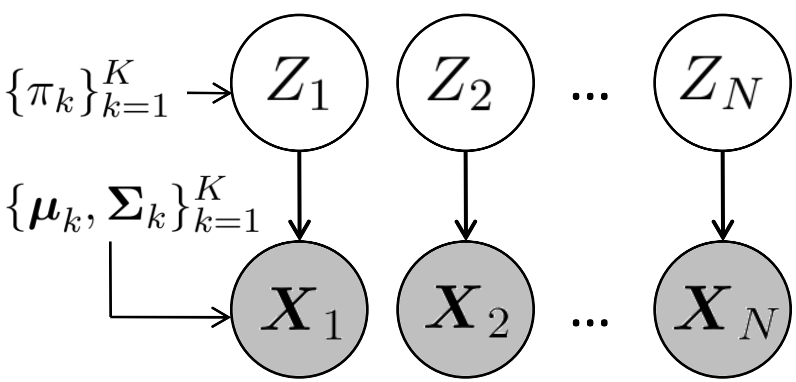
Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^N \subset \mathbb{R}^2$



Hidden Variables: $\{Z_n\}_{n=1}^N \in \{1,2,3\}^N$



Graphical Model:



Conditional:

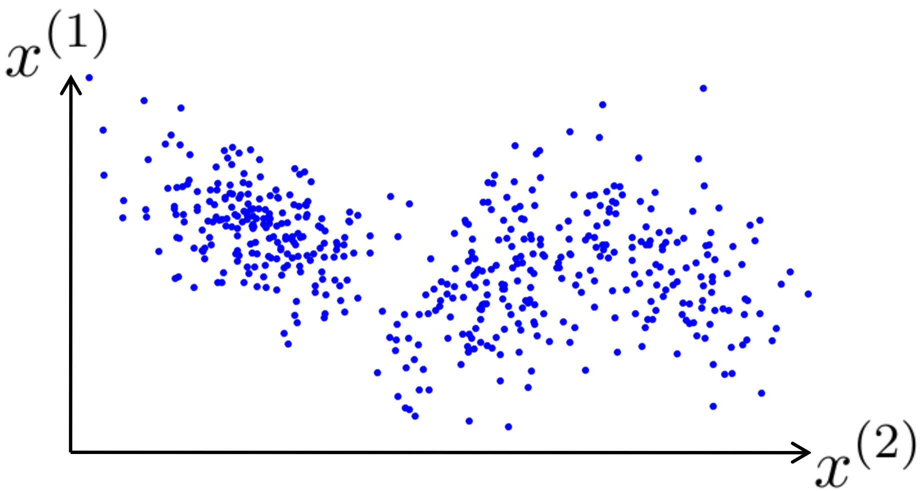
$$p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Priors: $p(Z_n = k; \boldsymbol{\theta}) = \pi_k, \qquad \sum_{k=1}^K \pi_k = 1$

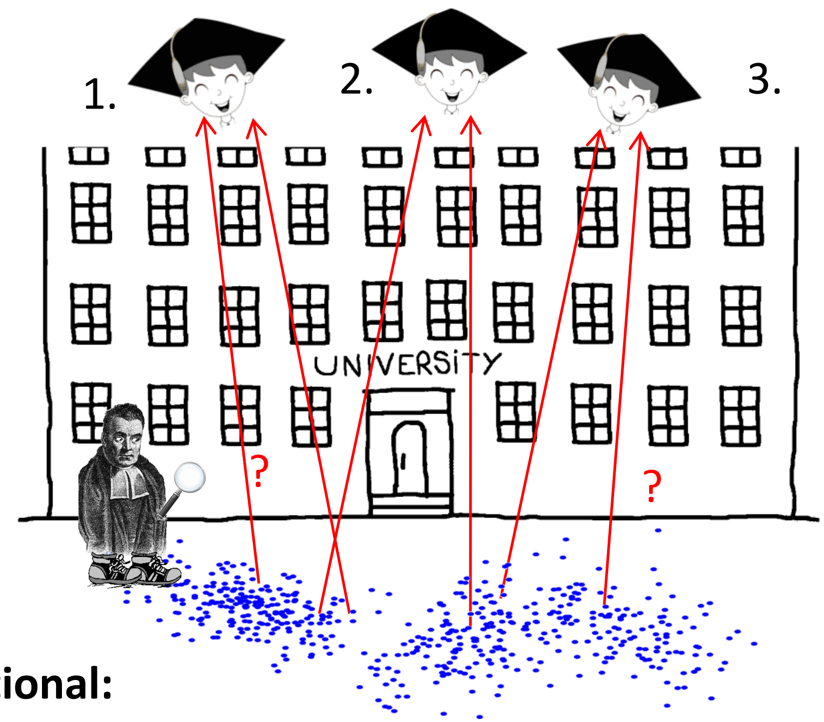
Bayesian Inference: Examples

Modeling

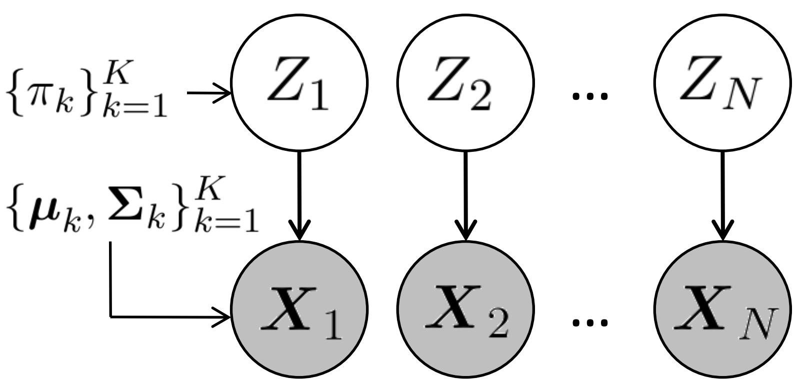
Observed variables: $\{\boldsymbol{x}_n\}_{n=1}^{N} \subset \mathbb{R}^2$



Hidden Variables: $\{Z_n\}_{n=1}^{N} \in \{1,2,3\}^N$



Graphical Model:



Conditional:

$$p(\boldsymbol{X}_n = \boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Priors: $p(Z_n = k; \boldsymbol{\theta}) = \pi_k, \qquad \sum_{k=1}^{K} \pi_k = 1$

Parameters: $\boldsymbol{\theta} = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k\}_{k=1}^{K}$

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \ldots, Z_N = z_N | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where $$p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \frac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta})}$$
$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \ldots, Z_N = z_N | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where $p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \dfrac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta})}$

$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Simple form, but $\boldsymbol{\theta}$ is unknown

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \dots, Z_N = z_N | \boldsymbol{x}_1, \dots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where
$$p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \frac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N; \boldsymbol{\theta})}$$
$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Simple form, but $\boldsymbol{\theta}$ is unknown => Maximum likelihood? $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, \mathcal{L}(\boldsymbol{\theta}; \mathbf{X})$

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \ldots, Z_N = z_N | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where $p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \dfrac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta})}$

$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Simple form, but $\boldsymbol{\theta}$ is unknown => Maximum likelihood? $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{X})$

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{X}) = p(\boldsymbol{X}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{X}_N = \boldsymbol{x}_N; \boldsymbol{\theta})$$

$$= \log \left(\sum_{z_1, \ldots, z_N = 1}^{K} \prod_{n=1}^{N} \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$$

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \ldots, Z_N = z_N | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where $p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \dfrac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{\theta})}$

$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Simple form, but $\boldsymbol{\theta}$ is unknown => Maximum likelihood? $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, \mathcal{L}(\boldsymbol{\theta}; \mathbf{X})$

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{X}) = p(\boldsymbol{X}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{X}_N = \boldsymbol{x}_N; \boldsymbol{\theta})$$

$$= \log \left(\sum_{z_1, \ldots, z_N = 1}^{K} \prod_{n=1}^{N} \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$$

- Non-convex
- Combinatorial
- Intractable

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \dots, Z_N = z_N | \boldsymbol{x}_1, \dots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where $$p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \frac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N; \boldsymbol{\theta})}$$
$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Simple form, but $\boldsymbol{\theta}$ is unknown => Maximum likelihood? $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{X})$

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{X}) = p(\boldsymbol{X}_1 = \boldsymbol{x}_1, \dots, \boldsymbol{X}_N = \boldsymbol{x}_N; \boldsymbol{\theta})$$
$$= \log \left(\sum_{z_1, \dots, z_N = 1}^{K} \prod_{n=1}^{N} \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$$

- Non-convex
- Combinatorial
- Intractable

- The joint probability $p(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})$ has a much simpler form than the marginal $p(\mathbf{X}; \boldsymbol{\theta})$
- $\mathbf{Z}$ is a hidden variable, and cannot be estimated without knowing $\boldsymbol{\theta}$

Bayesian Inference: Examples

Inference

Bayes' Theorem: $p(Z_1 = z_1, \dots, Z_N = z_N | \boldsymbol{x}_1, \dots, \boldsymbol{x}_N; \boldsymbol{\theta}) = \prod_{n=1}^{N} p(Z_n = z_n | \boldsymbol{x}_n; \boldsymbol{\theta})$

where $p(Z_n = k | \boldsymbol{x}_n; \boldsymbol{\theta}) = \dfrac{p(\boldsymbol{x}_n | Z_n = k; \boldsymbol{\theta}) p(Z_n = k; \boldsymbol{\theta})}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N; \boldsymbol{\theta})}$

$$\propto \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Simple form, but $\boldsymbol{\theta}$ is unknown => Maximum likelihood? $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{X})$

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{X}) = p(\boldsymbol{X}_1 = \boldsymbol{x}_1, \dots, \boldsymbol{X}_N = \boldsymbol{x}_N; \boldsymbol{\theta})$$

$$= \log \left(\sum_{z_1, \dots, z_N = 1}^{K} \prod_{n=1}^{N} \pi_k \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$$

- Non-convex
- Combinatorial
- Intractable

- The joint probability $p(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})$ has a much simpler form than the marginal $p(\mathbf{X}; \boldsymbol{\theta})$
- $\mathbf{Z}$ is a hidden variable, and cannot be estimated without knowing $\boldsymbol{\theta}$

$\Longrightarrow$ **Expectation-Maximization** (EM) **algorithm**

Bayesian Inference: Examples

Posterior expectation

EM algorithm

- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Bayesian Inference: Examples

Posterior expectation

EM algorithm

- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}) = \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye's theorem)

Bayesian Inference: Examples

Posterior expectation

EM algorithm

- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye's theorem)

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

Bayesian Inference: Examples

Posterior expectation

EM algorithm

- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye's theorem)

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

Bayesian Inference: Examples

Posterior expectation

EM algorithm
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye’s theorem)

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$$

where $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ is the conditional cross entropy of $\mathbf{Z}$ given$\mathbf{X},\boldsymbol{\theta}$ for the distribution $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)})$.

Bayesian Inference: Examples

Posterior expectation

EM algorithm
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye's theorem)

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$$

where $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ is the conditional cross entropy of $\mathbf{Z}$ given $\mathbf{X},\boldsymbol{\theta}$ for the distribution $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)})$.

In particular: $\mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Bayesian Inference: Examples

Posterior expectation

EM algorithm
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye's theorem)

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$$

where $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ is the conditional cross entropy of $\mathbf{Z}$ given $\mathbf{X},\boldsymbol{\theta}$ for the distribution $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)})$.

In particular: $\mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Therefore: $\mathcal{L}(\boldsymbol{\theta};\mathbf{X}) - \mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Bayesian Inference: Examples

Posterior expectation

EM algorithm
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})} \quad \text{(Baye's theorem)}$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$$

where $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ is the conditional cross entropy of $\mathbf{Z}$ given $\mathbf{X},\boldsymbol{\theta}$ for the distribution $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)})$.

In particular: $\mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Therefore: $\mathcal{L}(\boldsymbol{\theta};\mathbf{X}) - \mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

According to Gibb's inequality, we have $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) \geq H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Bayesian Inference: Examples

Posterior expectation

EM algorithm
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})}$$ (Baye's theorem)

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$$

where $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ is the conditional cross entropy of $\mathbf{Z}$ given $\mathbf{X},\boldsymbol{\theta}$ for the distribution $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)})$.

In particular: $\mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Therefore: $\mathcal{L}(\boldsymbol{\theta};\mathbf{X}) - \mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

According to Gibb's inequality, we have $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) \geq H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

$$\mathcal{L}(\boldsymbol{\theta};\mathbf{X}) - \mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) \geq Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$$

Bayesian Inference: Examples

Posterior expectation

EM algorithm
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$

Complete-data log-likelihood

Proof of correctness:

$$p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) = \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{p(\mathbf{X}|\boldsymbol{\theta})} \quad \text{(Baye's theorem)}$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log p(\mathbf{X}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}) \log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$$

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$$

where $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ is the conditional cross entropy of $\mathbf{Z}$ given $\mathbf{X},\boldsymbol{\theta}$ for the distribution $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)})$.

In particular: $\mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

Therefore: $\mathcal{L}(\boldsymbol{\theta};\mathbf{X}) - \mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

According to Gibb's inequality, we have $H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) \geq H(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})$.

$$\boxed{\mathcal{L}(\boldsymbol{\theta};\mathbf{X}) - \mathcal{L}(\boldsymbol{\theta}^{(i)};\mathbf{X}) \geq Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) - Q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(i)})}$$

The likelihood can only increase at each step!

Bayesian Inference: Examples

Derivations for the Gaussian mixture model

Bayesian Inference: Examples

Derivations for the Gaussian mixture model

- **E-step:** computing the current posterior probabilities

$$r_{n,k}^{(i)} = p(Z_n = k|\boldsymbol{x}_n;\boldsymbol{\theta}^{(i)}) = \frac{p(\boldsymbol{x}_n|Z_n = k;\boldsymbol{\theta}^{(i)})p(Z_n = k;\boldsymbol{\theta}^{(i)})}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N;\boldsymbol{\theta}^{(i)})} \propto \boxed{\pi_k^{(i)}\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k^{(i)},\boldsymbol{\Sigma}_k^{(i)})}$$

Bayesian Inference: Examples

Derivations for the Gaussian mixture model

- **E-step:** computing the current posterior probabilities

$$r_{n,k}^{(i)} = p(Z_n = k|\boldsymbol{x}_n;\boldsymbol{\theta}^{(i)}) = \frac{p(\boldsymbol{x}_n|Z_n = k;\boldsymbol{\theta}^{(i)})p(Z_n = k;\boldsymbol{\theta}^{(i)})}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N;\boldsymbol{\theta}^{(i)})} \propto \boxed{\pi_k^{(i)}\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k^{(i)},\boldsymbol{\Sigma}_k^{(i)})}$$

We deduce $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$:

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\{\log\mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})\} \\
&= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\left\{\log\prod_{n=1}^{N}\sum_{k=1}^{K}\mathbb{I}\{Z_n = k\}\pi_k\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\right\} \\
&= \sum_{k,n=1}^{K,N}\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\{\mathbb{I}\{Z_n = k\}\}\log\big(\pi_k\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\big) \\
&= \sum_{k,n=1}^{K,N} r_{n,k}^{(i)}\big(\log\pi_k - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)^\top\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)\big) + \text{const.}
\end{aligned}$$

Bayesian Inference: Examples

Derivations for the Gaussian mixture model

- **E-step:** computing the current posterior probabilities

$$r_{n,k}^{(i)} = p(Z_n = k|\boldsymbol{x}_n;\boldsymbol{\theta}^{(i)}) = \frac{p(\boldsymbol{x}_n|Z_n = k;\boldsymbol{\theta}^{(i)})p(Z_n = k;\boldsymbol{\theta}^{(i)})}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N;\boldsymbol{\theta}^{(i)})} \propto \boxed{\pi_k^{(i)}\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k^{(i)},\boldsymbol{\Sigma}_k^{(i)})}$$

We deduce $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$:

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\{\log\mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})\} \\
&= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\left\{\log\prod_{n=1}^{N}\sum_{k=1}^{K}\mathbb{I}\{Z_n = k\}\pi_k\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\right\} \\
&= \sum_{k,n=1}^{K,N}\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\{\mathbb{I}\{Z_n = k\}\}\log\big(\pi_k\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\big) \\
&= \sum_{k,n=1}^{K,N} r_{n,k}^{(i)}\Big(\log\pi_k - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)\Big) + \text{const.}
\end{aligned}$$

- **M-step:** maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ by finding the zeros of the derivative

Bayesian Inference: Examples

Derivations for the Gaussian mixture model

- **E-step:** computing the current posterior probabilities

$$r_{n,k}^{(i)} = p(Z_n = k|\boldsymbol{x}_n;\boldsymbol{\theta}^{(i)}) = \frac{p(\boldsymbol{x}_n|Z_n = k;\boldsymbol{\theta}^{(i)})p(Z_n = k;\boldsymbol{\theta}^{(i)})}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N;\boldsymbol{\theta}^{(i)})} \propto \boxed{\pi_k^{(i)}\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k^{(i)},\boldsymbol{\Sigma}_k^{(i)})}$$

We deduce $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$:

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\{\log\mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})\} \\
&= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\left\{\log\prod_{n=1}^{N}\sum_{k=1}^{K}\mathbb{I}\{Z_n = k\}\pi_k\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\right\} \\
&= \sum_{k,n=1}^{K,N}\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}\{\mathbb{I}\{Z_n = k\}\}\log\big(\pi_k\mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\big) \\
&= \sum_{k,n=1}^{K,N} r_{n,k}^{(i)}\big(\log\pi_k - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)^\top\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)\big) + \text{const.}
\end{aligned}$$

- **M-step:** maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$ by finding the zeros of the derivative

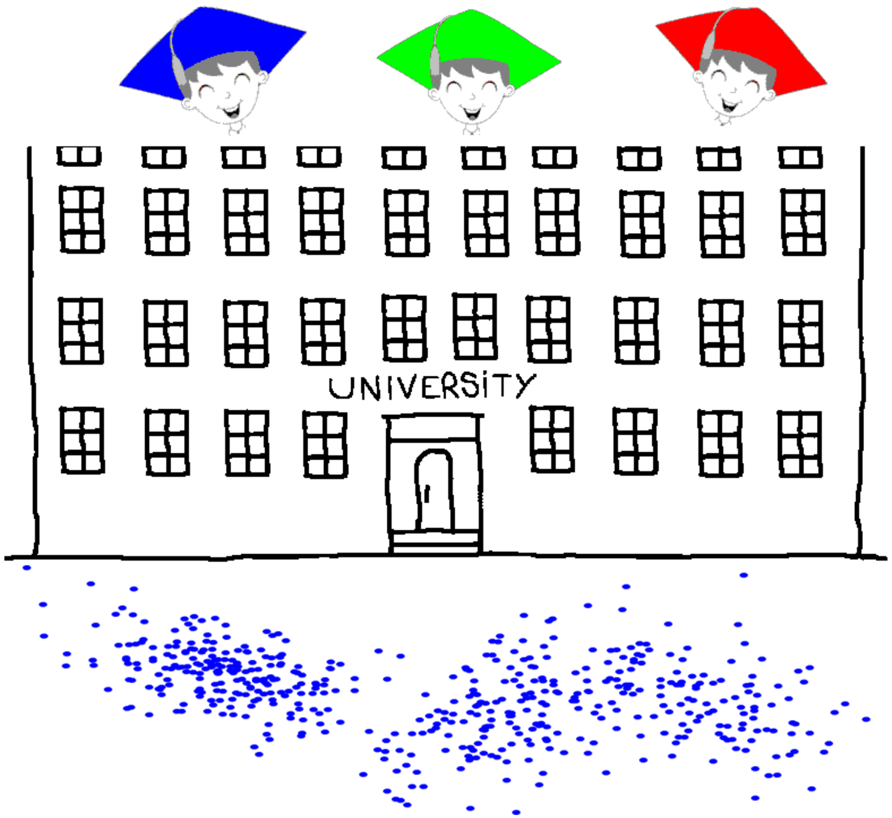
$$\boxed{\pi_k^{(i+1)} = \frac{1}{N}\sum_{n=1}^{N} r_{n,k}^{(i)}\,,} \quad \boxed{\boldsymbol{\mu}_k^{(i+1)} = \frac{1}{N}\sum_{n=1}^{N} r_{n,k}^{(i)}\boldsymbol{x}_n\,,} \quad \boxed{\boldsymbol{\Sigma}_k^{(i+1)} = \frac{1}{N}\sum_{n=1}^{N} r_{n,k}^{(i)}(\boldsymbol{x}_n-\boldsymbol{\mu}_k^{(i+1)})(\boldsymbol{x}_n-\boldsymbol{\mu}_k^{(i+1)})^\top}$$

Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

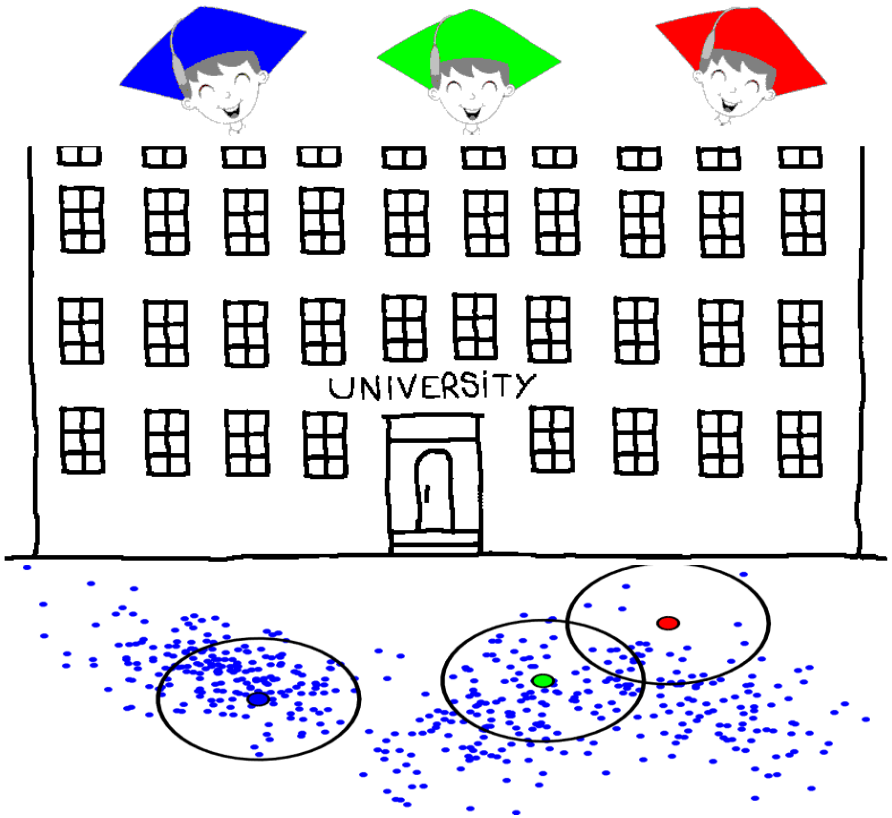


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

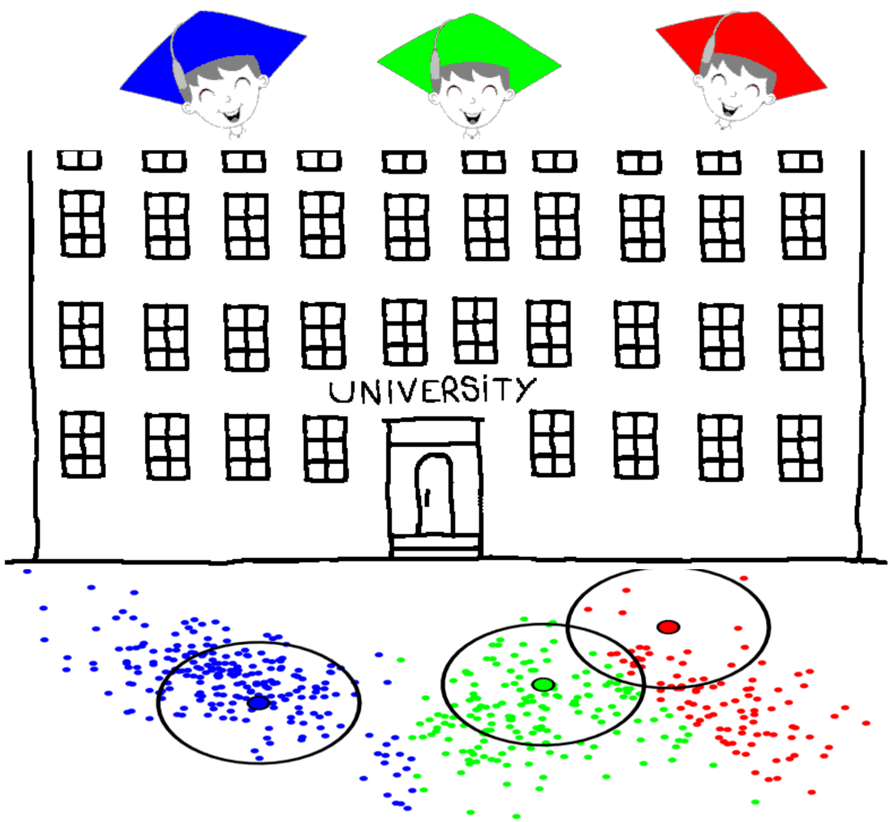


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

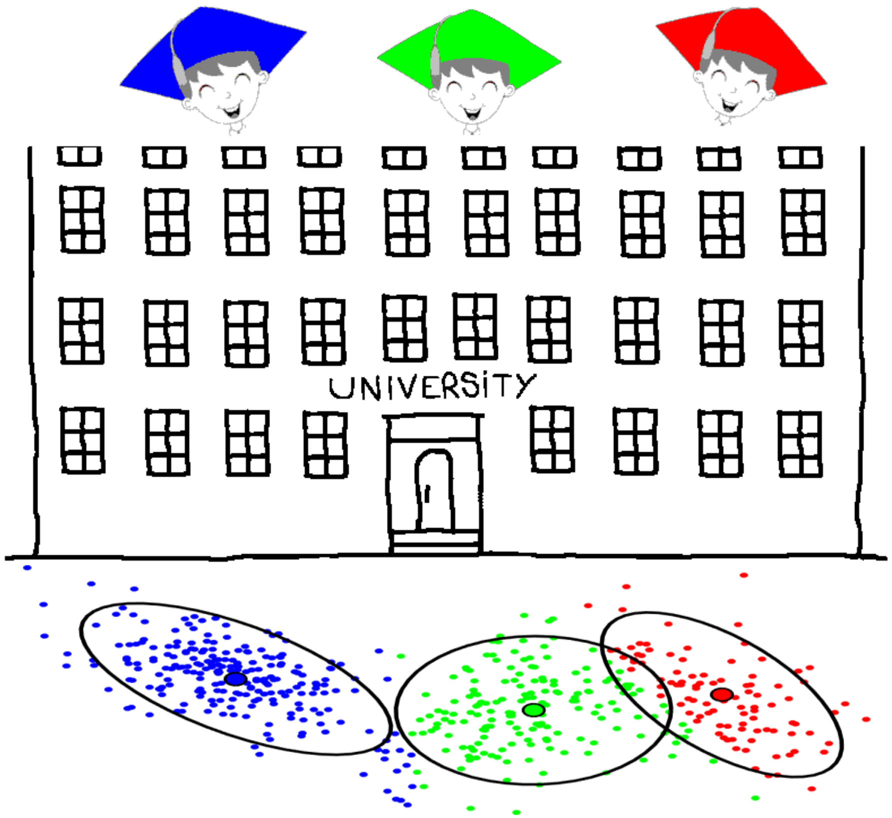


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

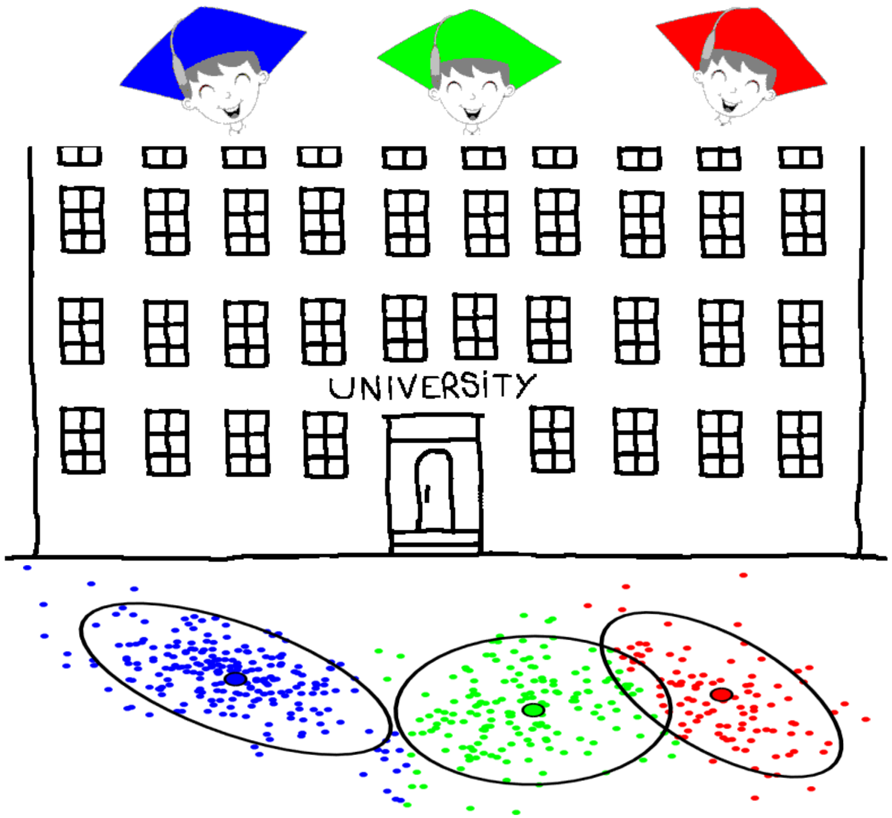


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

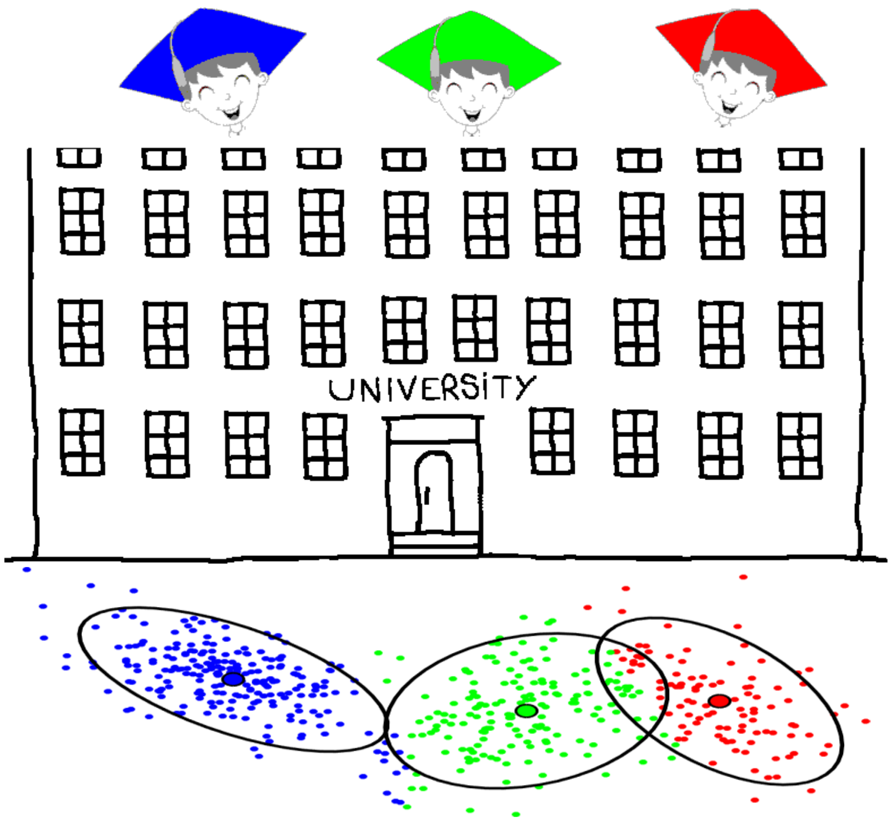


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

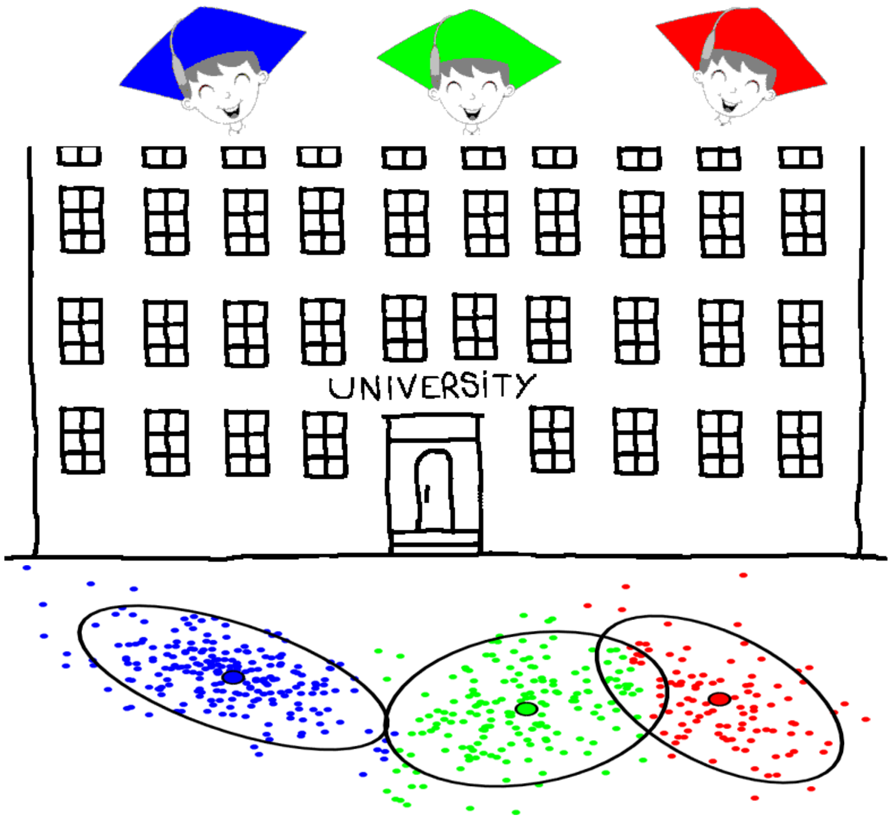


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

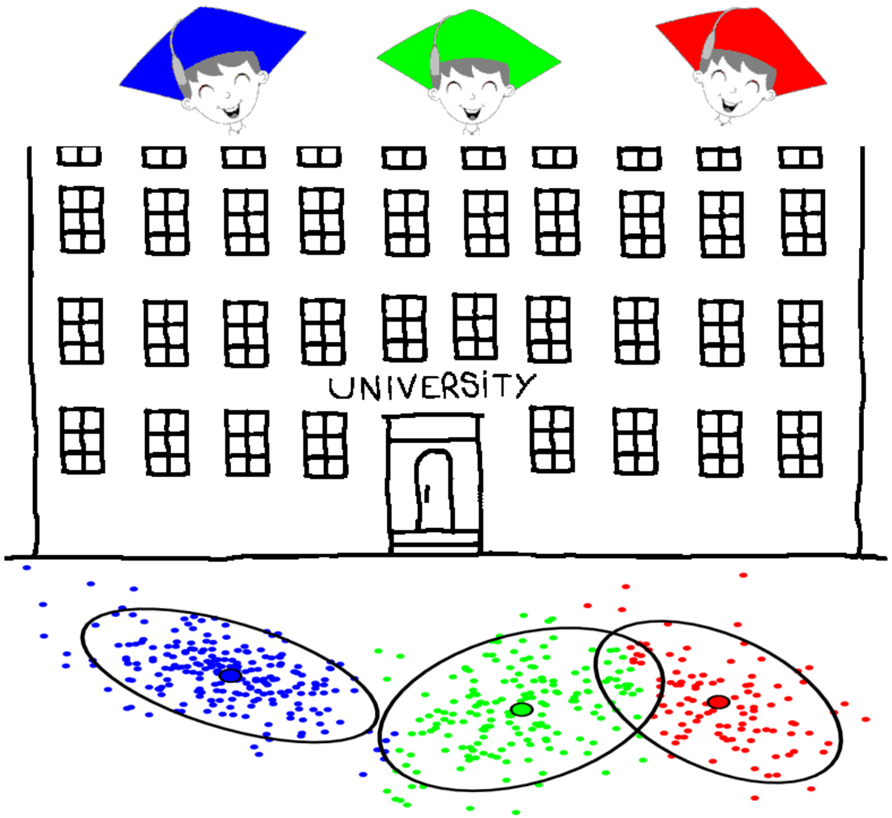


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

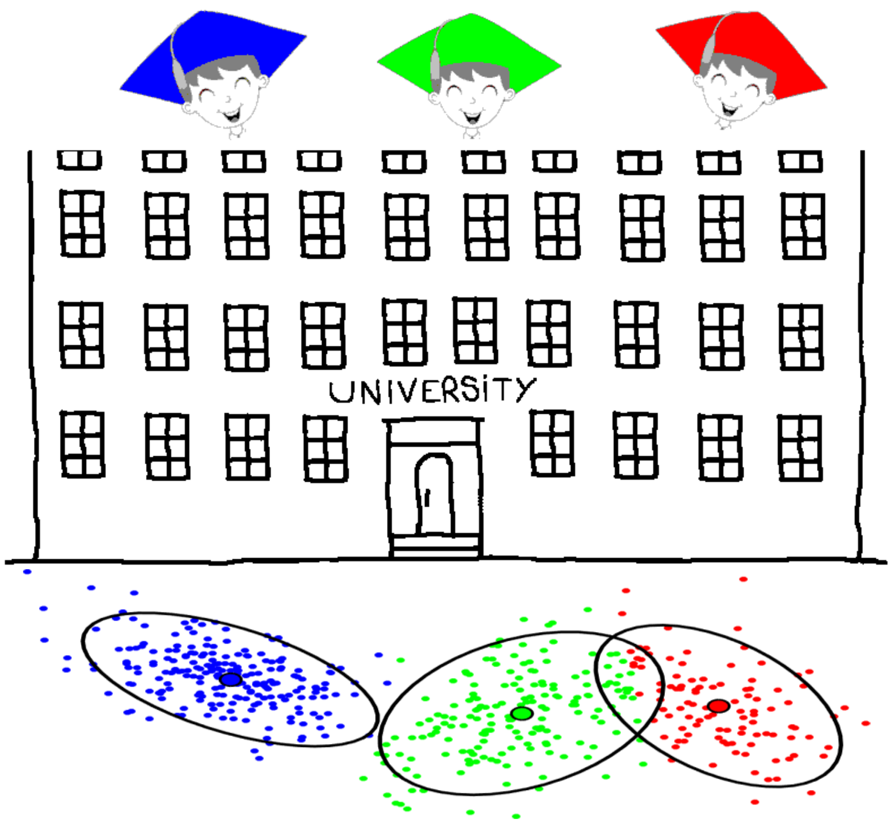


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

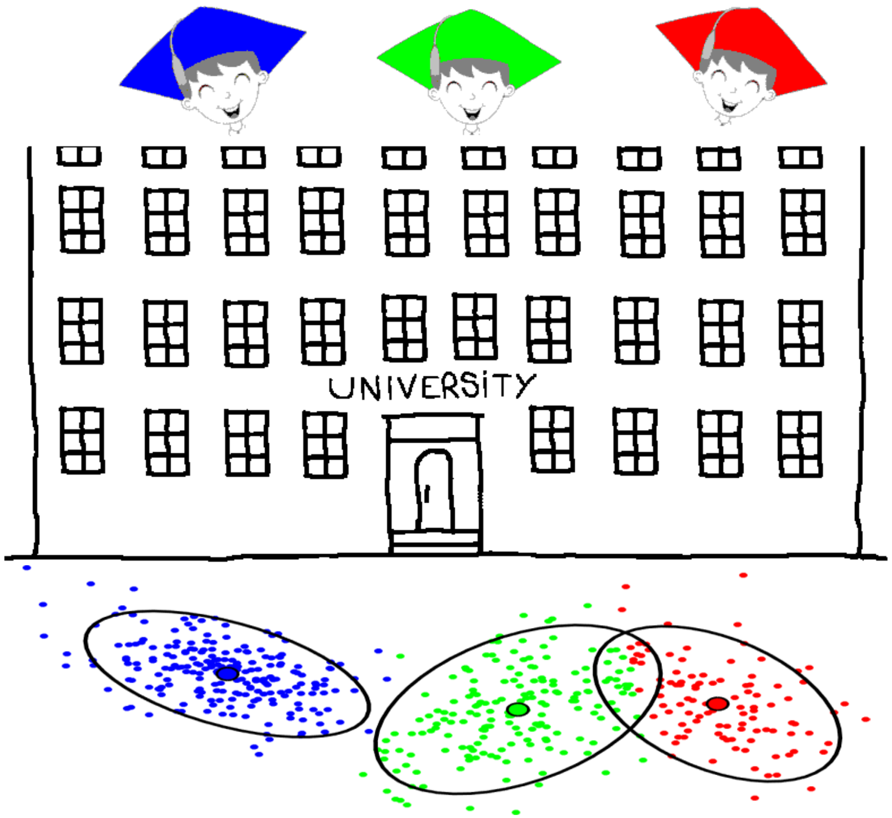


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

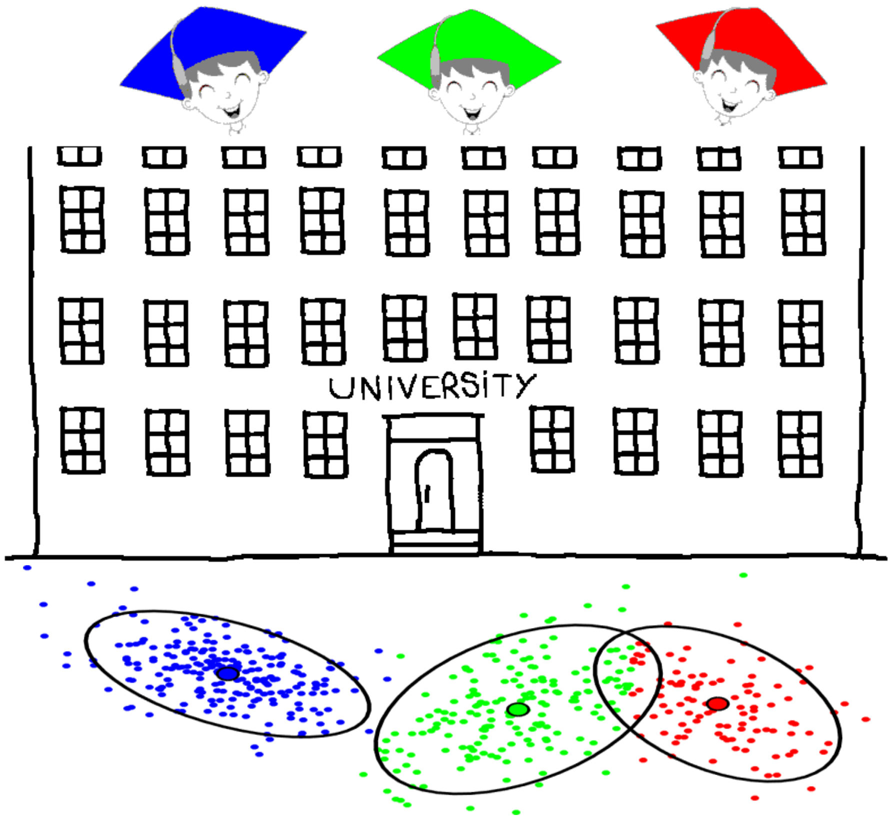


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

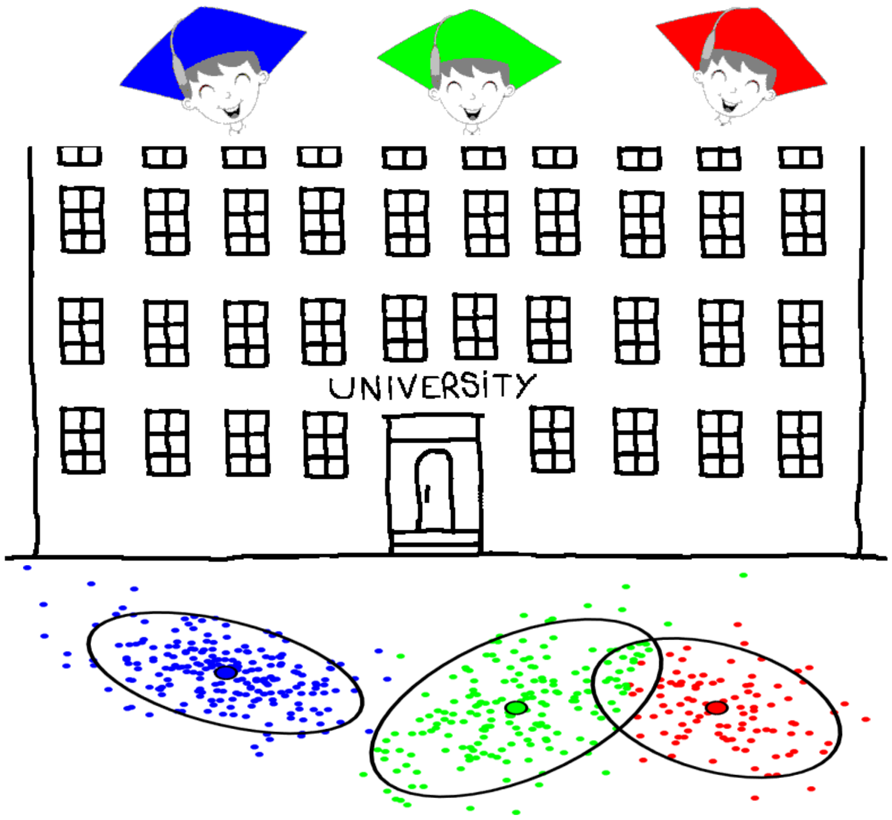


Bayesian Inference: Examples

Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**

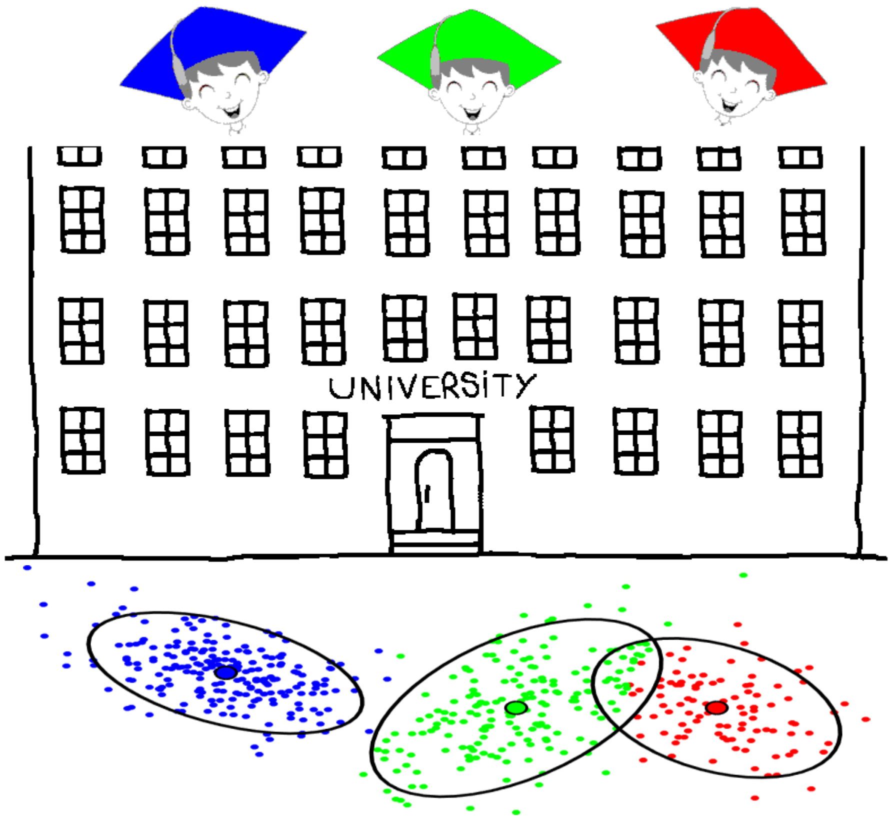


Bayesian Inference: Examples

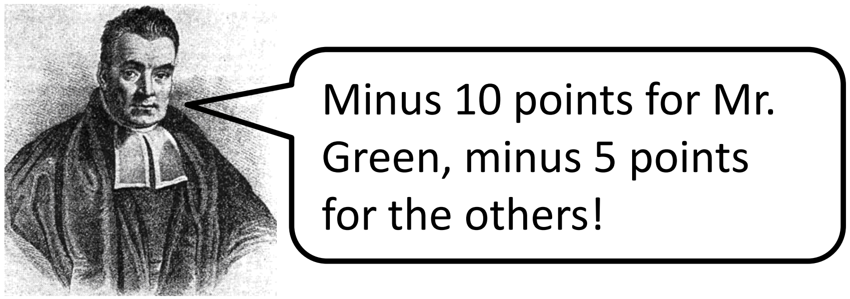
Inference

EM algorithm

- **Initialization:** Random « guess » for $\boldsymbol{\theta}$
- **E-step:** $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(i)}}[\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})]$
- **M-step:** $\boldsymbol{\theta}^{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$
- **Convergence**



Decision:



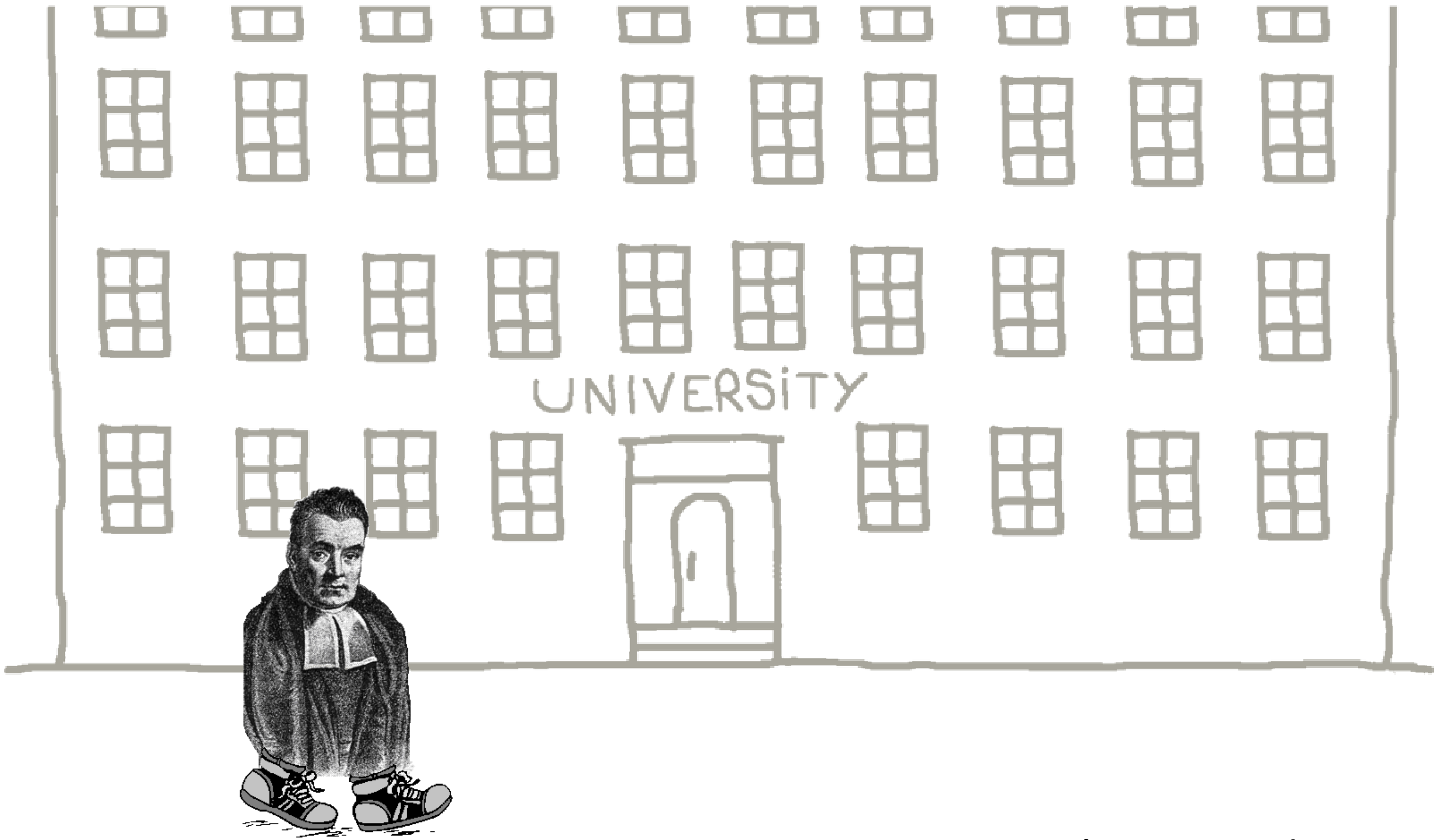
OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

Bayesian Inference: Examples



The next day…

Bayesian Inference: Examples

The next day…

Bayesian Inference: Examples

The next day…

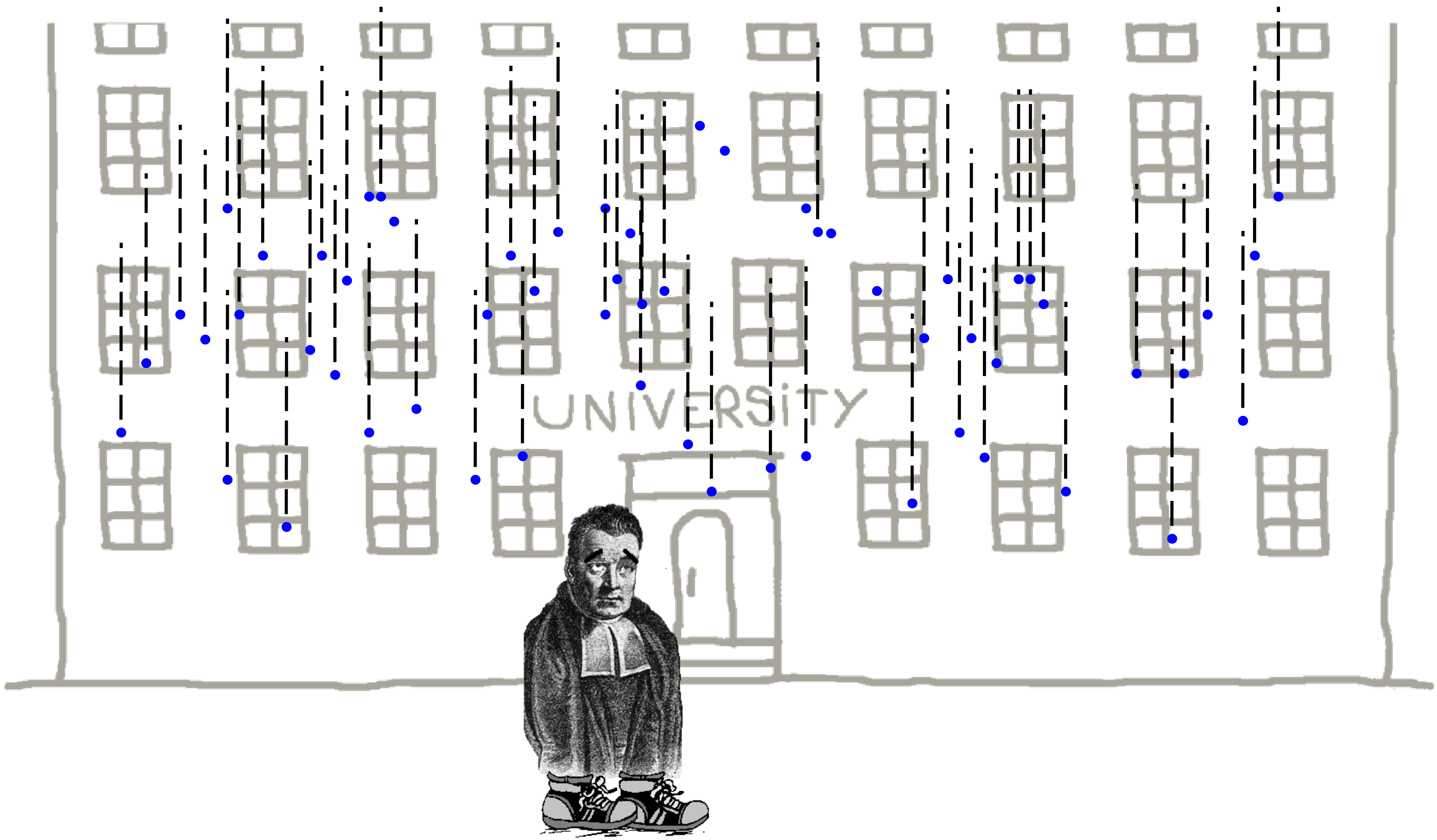
Bayesian Inference: Examples

The next day...

Bayesian Inference: Examples

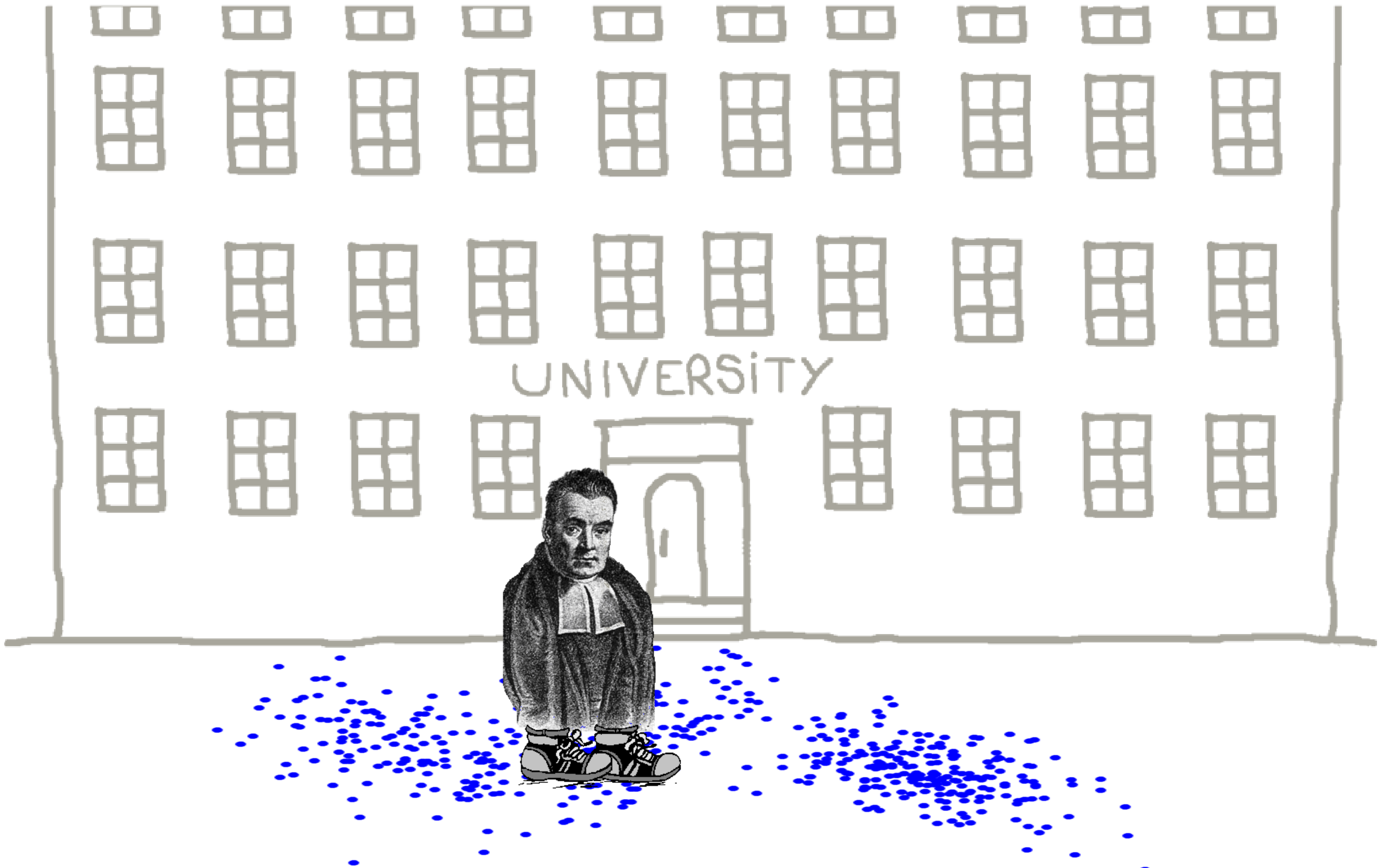
The next day…

Bayesian Inference: Examples

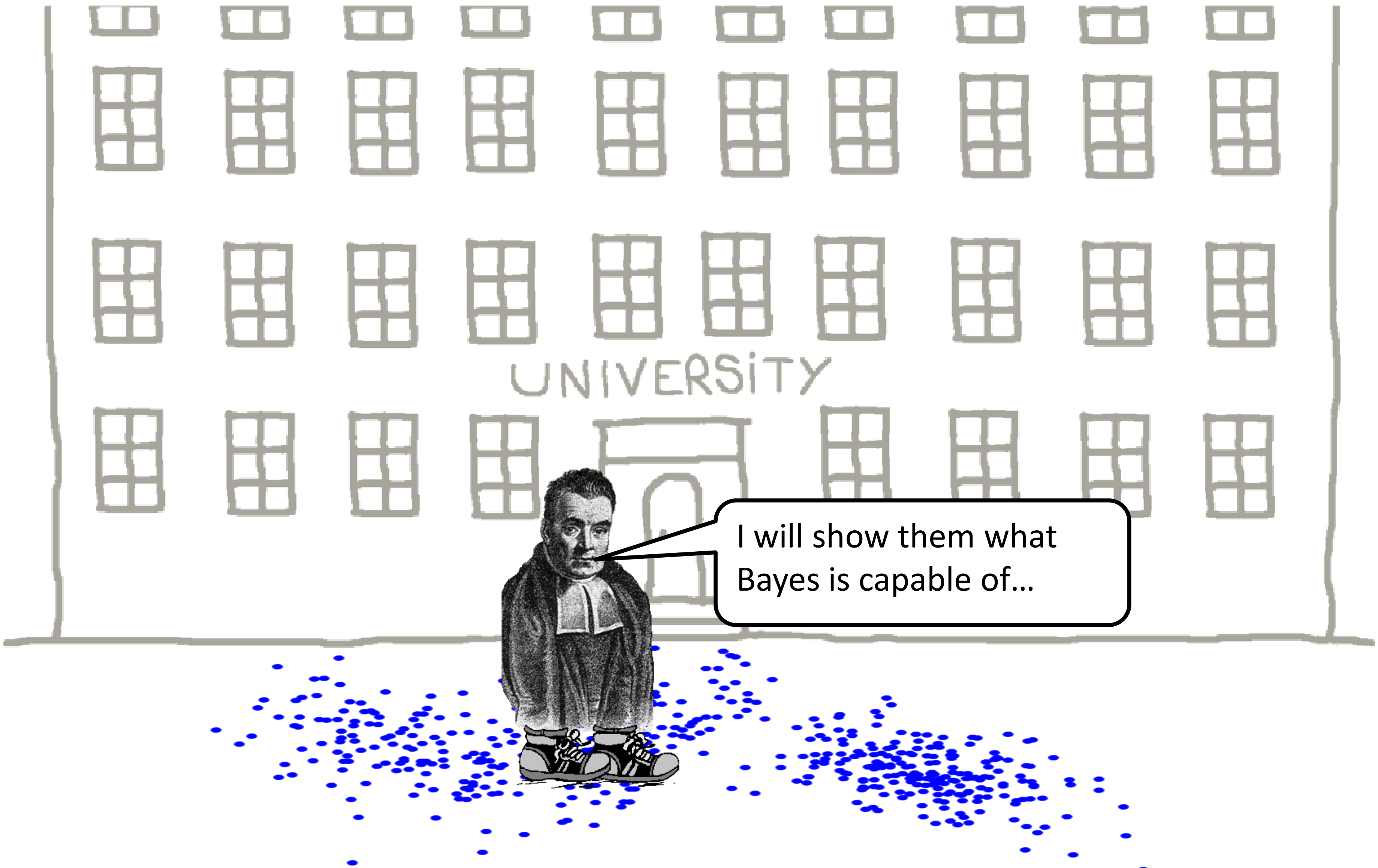


The next day...

Bayesian Inference: Examples



Bayesian Inference: Examples



Bayesian Inference: Examples

Modeling

A « fully Bayesian » model

$$p(\boldsymbol{x}_n | Z_n = k, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$$
$$p(Z_n = k | \boldsymbol{\pi}) = \pi_k$$

GMM (same as before)

Bayesian Inference: Examples

Modeling

A « fully Bayesian » model

$$p(\boldsymbol{x}_n|Z_n = k, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$$
$$p(Z_n = k|\boldsymbol{\pi}) = \pi_k$$

GMM (same as before)

$$p(\boldsymbol{\Lambda}_k) = \mathcal{W}(\boldsymbol{\Lambda}_k; \mathbf{W}_0, \nu_0)$$
$$p(\boldsymbol{\mu}_k|\boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{\mu}_k; \boldsymbol{m}_0, \boldsymbol{\Lambda}_k^{-1})$$
$$p(\boldsymbol{\pi}) = \mathrm{SymDir}(\boldsymbol{\pi}|\alpha_0)$$

Priors on all parameters

Note: These are the *conjugate priors* for the normal and the multinomial distributions, i.e., they are such that $p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k|\boldsymbol{x}_n, Z_n = k) \cong p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ and $p(\boldsymbol{\pi}|z_n) \cong p(\boldsymbol{\pi})$

Bayesian Inference: Examples

Modeling A « fully Bayesian » model

$$p(\boldsymbol{x}_n | Z_n = k, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$$
$$p(Z_n = k | \boldsymbol{\pi}) = \pi_k$$

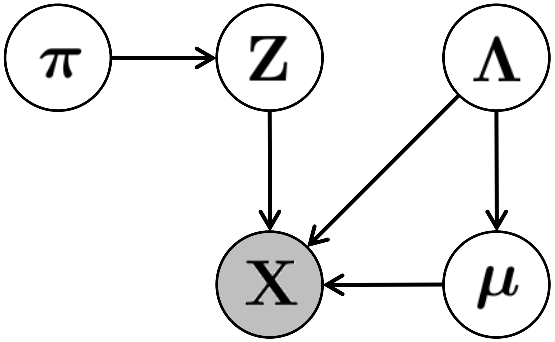
GMM (same as before)

$$p(\boldsymbol{\Lambda}_k) = \mathcal{W}(\boldsymbol{\Lambda}_k; \mathbf{W}_0, \nu_0)$$
$$p(\boldsymbol{\mu}_k | \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{\mu}_k; \boldsymbol{m}_0, \boldsymbol{\Lambda}_k^{-1})$$
$$p(\boldsymbol{\pi}) = \mathrm{SymDir}(\boldsymbol{\pi} | \alpha_0)$$

Priors on all parameters

Note: These are the *conjugate priors* for the normal and the multinomial distributions, i.e., they are such that $p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k | \boldsymbol{x}_n, Z_n = k) \cong p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ and $p(\boldsymbol{\pi} | z_n) \cong p(\boldsymbol{\pi})$

Graphical model:



Bayesian Inference: Examples

Modeling A « fully Bayesian » model

$$p(\boldsymbol{x}_n|Z_n = k, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$$
$$p(Z_n = k|\boldsymbol{\pi}) = \pi_k$$

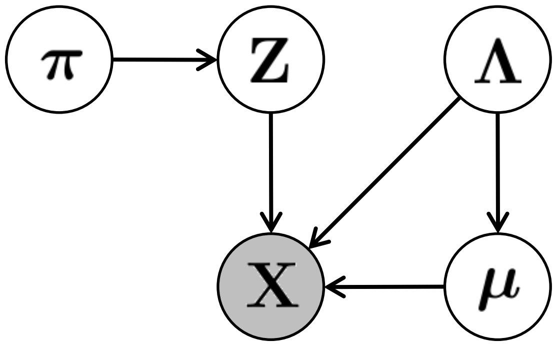
GMM (same as before)

$$p(\boldsymbol{\Lambda}_k) = \mathcal{W}(\boldsymbol{\Lambda}_k; \mathbf{W}_0, \nu_0)$$
$$p(\boldsymbol{\mu}_k|\boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{\mu}_k; \boldsymbol{m}_0, \boldsymbol{\Lambda}_k^{-1})$$
$$p(\boldsymbol{\pi}) = \mathrm{SymDir}(\boldsymbol{\pi}|\alpha_0)$$

Priors on all parameters

Note: These are the *conjugate priors* for the normal and the multinomial distributions, i.e., they are such that $p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k|\boldsymbol{x}_n, Z_n = k) \cong p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ and $p(\boldsymbol{\pi}|z_n) \cong p(\boldsymbol{\pi})$

Graphical model:



Choice of hyperparameters:

$$\mathbf{W}_0 = \mathbf{I}$$
$$\nu_0 = D = 2$$
$$\boldsymbol{m}_0 = \mathrm{mean}(\mathbf{X})$$

$\alpha_0 > 0$: low values will allow Gaussian weights to be close to 0

Bayesian Inference: Examples

Inference

- The posterior distribution $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X})$ is intractable

Bayesian Inference: Examples

Inference

- The posterior distribution $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X})$ is intractable

- **Technique**: use a *variational approximation* $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}) \approx q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$, where $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is restricted to a family of distributions having a simpler form than the true $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X})$

Bayesian Inference: Examples

Inference

- The posterior distribution $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X})$ is intractable

- **Technique**: use a *variational approximation* $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}) \approx q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$, where $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is restricted to a family of distributions having a simpler form than the true $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X})$

- The variational distribution $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is typically assumed to **factorize** over some partition of the latent variables.

Bayesian Inference: Examples

Inference

- The posterior distribution $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X})$ is intractable

- **Technique**: use a *variational approximation* $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X}) \approx q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$, where $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is restricted to a family of distributions having a simpler form than the true $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X})$

- The variational distribution $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is typically assumed to **factorize** over some partition of the latent variables.

- Here we use: $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) = q_Z(\mathbf{Z}) q_{\Lambda\mu\pi}(\mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$. Remarkably, this is the **only** assumption needed to obtain a tractable *EM-like* inference procedure.

Bayesian Inference: Examples

Inference

- The posterior distribution $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X})$ is intractable

- **Technique**: use a *variational approximation* $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X}) \approx q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$, where $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is restricted to a family of distributions having a simpler form than the true $p(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{X})$

- The variational distribution $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$ is typically assumed to **factorize** over some partition of the latent variables.

- Here we use: $q(\mathbf{Z}, \mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) = q_Z(\mathbf{Z}) q_{\Lambda\mu\pi}(\mathbf{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})$. Remarkably, this is the **only** assumption needed to obtain a tractable *EM-like* inference procedure.

- Such procedures are referred to as ***Variational Bayesian EM algorithms.***

Bayesian Inference: Examples

Inference

- The posterior distribution $p(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X})$ is intractable
- **Technique**: use a *variational approximation* $p(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X}) \approx q(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})$, where $q(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})$ is restricted to a family of distributions having a simpler form than the true $p(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X})$
- The variational distribution $q(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})$ is typically assumed to **factorize** over some partition of the latent variables.
- Here we use: $q(\mathbf{Z},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = q_Z(\mathbf{Z})q_{\Lambda\mu\pi}(\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})$. Remarkably, this is the **only** assumption needed to obtain a tractable *EM-like* inference procedure.
- Such procedures are referred to as ***Variational Bayesian EM algorithms.***
  For $p(\mathbf{Z},\mathbf{W}|\mathbf{X}) \approx q_Z(\mathbf{Z})q_W(\mathbf{W})$:

VB-EM
- **E-Z step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_W^{(i-1)}(\mathbf{W})}\{\log p(\mathbf{Z}|\mathbf{X},\mathbf{W})\}\right)$
- **E-W step:** $q_W^{(i)}(\mathbf{W}) \propto \exp\left(\mathbb{E}_{q_Z^{(i)}(\mathbf{Z})}\{\log p(\mathbf{W}|\mathbf{X},\mathbf{Z})\}\right)$

Bayesian Inference: Examples

Inference

Proof of correctness

• Using a similar reasonning as for EM, we can show that the VB-EM iteratively minimizes the Kulback-Leibler divergence between the true posterior $p(\mathbf{Z},\mathbf{W}|\mathbf{X})$ and its variational approximation $q(\mathbf{Z},\mathbf{W}) = q_Z(\mathbf{Z})q_W(\mathbf{W})$:

$$(q_Z^{(\infty)}, q_W^{(\infty)}) = \operatorname*{argmin}_{q_Z, q_W} \mathrm{KL}(q||p)$$

VB-EM

- **E-Z step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_W^{(i-1)}(\mathbf{W})}\{\log p(\mathbf{Z}|\mathbf{X},\mathbf{W})\}\right)$
- **E-W step:** $q_W^{(i)}(\mathbf{W}) \propto \exp\left(\mathbb{E}_{q_Z^{(i)}(\mathbf{Z})}\{\log p(\mathbf{W}|\mathbf{X},\mathbf{Z})\}\right)$

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

• **E-Λμπ-Step:**

$$\log q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = \log p(\boldsymbol{\pi}) + \sum_{k=1}^{K} \log p(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k) + \sum_{n,k=1}^{N,K} q^{(i-1)}_{Z_n}(k) \log \pi_k \mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1}) + \mathrm{const.}$$

Using the decomposition $q^{(i-1)}_{Z}(\mathbf{Z}) = \prod_{n=1}^{N} q^{(i-1)}_{Z_n}(k)$ (see E-Z-step).

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

•**E-Λμπ-Step:**

$$\log q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = \log p(\boldsymbol{\pi}) + \sum_{k=1}^{K} \log p(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k) + \sum_{n,k=1}^{N,K} q^{(i-1)}_{Z_n}(k) \log \pi_k \mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1}) + \text{const.}$$

Using the decomposition $q^{(i-1)}_{Z}(\mathbf{Z}) = \prod_{n=1}^{N} q^{(i-1)}_{Z_n}(k)$ (see E-Z-step).

This leads to the factorization: $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = q^{(i)}_{\pi}(\boldsymbol{\pi}) \prod_{k=1}^{K} q^{(i)}_{\Lambda_k\mu_k}(\boldsymbol{\Lambda}_k,\boldsymbol{\mu}_k)$

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

- **E-Λμπ-Step:**

$$\log q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = \log p(\boldsymbol{\pi}) + \sum_{k=1}^{K} \log p(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k) + \sum_{n,k=1}^{N,K} q^{(i-1)}_{Z_n}(k) \log \pi_k \mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1}) + \text{const.}$$

Using the decomposition $q^{(i-1)}_{Z}(\mathbf{Z}) = \prod_{n=1}^{N} q^{(i-1)}_{Z_n}(k)$ (see E-Z-step).

This leads to the factorization: $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = q^{(i)}_{\pi}(\boldsymbol{\pi}) \prod_{k=1}^{K} q^{(i)}_{\Lambda_k\mu_k}(\boldsymbol{\Lambda}_k,\boldsymbol{\mu}_k)$

with $\begin{cases} q^{(i)}_{\pi}(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi};\alpha_0 + N^{(i)}_1,\ldots,\alpha_0 + N^{(i)}_K) \\ q^{(i)}_{\Lambda_k\mu_k}(\boldsymbol{\Lambda}_k,\boldsymbol{\mu}_k) = \mathcal{N}\left(\boldsymbol{\mu}_k;\boldsymbol{m}^{(i)}_k,\dfrac{\boldsymbol{\Lambda}_k^{-1}}{1+N^{(i)}_k}\right)\mathcal{W}\left(\boldsymbol{\Lambda}_k;\mathbf{W}^{(i)}_k,\nu^{(i)}_k\right), \end{cases}$

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

•E-Λμπ-Step:

$$\log q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = \log p(\boldsymbol{\pi}) + \sum_{k=1}^{K}\log p(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k) + \sum_{n,k=1}^{N,K} q^{(i-1)}_{Z_n}(k)\log \pi_k \mathcal{N}(\boldsymbol{x}_n;\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1}) + \text{const.}$$

Using the decomposition $q^{(i-1)}_{Z}(\mathbf{Z}) = \prod_{n=1}^{N} q^{(i-1)}_{Z_n}(k)$ (see E-Z-step).

This leads to the factorization: $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) = q^{(i)}_{\pi}(\boldsymbol{\pi})\prod_{k=1}^{K} q^{(i)}_{\Lambda_k\mu_k}(\boldsymbol{\Lambda}_k,\boldsymbol{\mu}_k)$

with $\begin{cases} q^{(i)}_{\pi}(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi};\alpha_0+N_1^{(i)},\dots,\alpha_0+N_K^{(i)}) \\ q^{(i)}_{\Lambda_k\mu_k}(\boldsymbol{\Lambda}_k,\boldsymbol{\mu}_k) = \mathcal{N}\left(\boldsymbol{\mu}_k;\boldsymbol{m}_k^{(i)},\dfrac{\boldsymbol{\Lambda}_k^{-1}}{1+N_k^{(i)}}\right)\mathcal{W}\left(\boldsymbol{\Lambda}_k;\mathbf{W}_k^{(i)},\nu_k^{(i)}\right), \end{cases}$

where: $\overline{\boldsymbol{x}}_k^{(i)} = \dfrac{1}{N_k^{(i)}}\sum_{n=1}^{N} q^{(i-1)}_{Z_n}(k)\boldsymbol{x}_n,\qquad \mathbf{S}_k^{(i)} = \dfrac{1}{N_k^{(i)}}\sum_{n=1}^{N} q^{(i-1)}_{Z_n}(k)(\boldsymbol{x}_n-\overline{\boldsymbol{x}}_k^{(i)})(\boldsymbol{x}_n-\overline{\boldsymbol{x}}_k^{(i)})^\top,$

$$N_k^{(i)} = \sum_{n=1}^{N} q^{(i-1)}_{Z_n}(k),\qquad \boldsymbol{m}_k^{(i)} = \frac{\boldsymbol{m}_0 + N_k^{(i)}\overline{\boldsymbol{x}}_k^{(i)}}{1+N_k^{(i)}},$$

$$\mathbf{W}_k^{(i)-1} = \mathbf{W}_0^{-1} + N_k^{(i)}\mathbf{S}_k^{(i)} + \frac{N_k^{(i)}}{N_k^{(i)}+1}(\overline{\boldsymbol{x}}_k^{(i)}-\boldsymbol{m}_0)(\overline{\boldsymbol{x}}_k^{(i)}-\boldsymbol{m}_0)^\top,\qquad \nu_k^{(i)} = \nu_0 + N_k^{(i)}$$

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

- **E-Z-Step:**

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

•**E-Z-Step:**

$$\log q_Z^{(i)}(\mathbf{Z}) = \mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\} + \text{const.}$$

$$= \sum_{n,k=1}^{N,K} \mathbb{I}\{Z_n = k\} \log \rho^{(i)}_{n,k} + \text{const.}$$

where $\rho^{(i)}_{n,k} = \mathbb{E}_{q^{(i)}_{\pi}}\{\log \pi_k\} + \frac{1}{2}\mathbb{E}_{q^{(i)}_{\Lambda_k}}\{\log|\mathbf{\Lambda}_k|\} - \frac{D}{2}\log 2\pi - \frac{1}{2}\mathbb{E}_{q^{(i)}_{\Lambda_k\mu_k}}\{(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \mathbf{\Lambda}_k(\boldsymbol{x}_n - \boldsymbol{\mu}_k)\}$

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

•**E-Z-Step:**

$$\log q_Z^{(i)}(\mathbf{Z}) = \mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\} + \text{const.}$$
$$= \sum_{n,k=1}^{N,K} \mathbb{I}\{Z_n = k\} \log \rho_{n,k}^{(i)} + \text{const.}$$

where $\rho_{n,k}^{(i)} = \mathbb{E}_{q_\pi^{(i)}}\{\log \pi_k\} + \frac{1}{2}\mathbb{E}_{q^{(i)}_{\Lambda_k}}\{\log|\boldsymbol{\Lambda}_k|\} - \frac{D}{2}\log 2\pi - \frac{1}{2}\mathbb{E}_{q^{(i)}_{\Lambda_k \mu_k}}\{(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Lambda}_k(\boldsymbol{x}_n - \boldsymbol{\mu}_k)\}$

It follows that $q_Z^{(i)}(\mathbf{Z}) = \prod_{n=1}^{N} q_{Z_n}^{(i)}(k)$ where $q_{Z_n}^{(i)}(k) = \frac{\rho_{n,k}^{(i)}}{\sum_{j=1}^{K}\rho_{n,j}^{(i)}} = r_{n,k}^{(i)}$

Bayesian Inference: Examples

Inference

Derivations for the Bayesian mixture of Gaussian

•**E-Z-Step:**

$$\log q_Z^{(i)}(\mathbf{Z}) = \mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\} + \text{const.}$$
$$= \sum_{n,k=1}^{N,K} \mathbb{I}\{Z_n = k\} \log \rho_{n,k}^{(i)} + \text{const.}$$

where $\rho_{n,k}^{(i)} = \mathbb{E}_{q_\pi^{(i)}}\{\log \pi_k\} + \frac{1}{2}\mathbb{E}_{q^{(i)}_{\Lambda_k}}\{\log|\mathbf{\Lambda}_k|\} - \frac{D}{2}\log 2\pi - \frac{1}{2}\mathbb{E}_{q^{(i)}_{\Lambda_k\mu_k}}\{(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \mathbf{\Lambda}_k(\boldsymbol{x}_n - \boldsymbol{\mu}_k)\}$

It follows that $q_Z^{(i)}(\mathbf{Z}) = \prod_{n=1}^{N} q_{Z_n}^{(i)}(k)$ where $q_{Z_n}^{(i)}(k) = \frac{\rho_{n,k}^{(i)}}{\sum_{j=1}^K \rho_{n,j}^{(i)}} = r_{n,k}^{(i)}$

Finally, we can express $r_{n,k}^{(i)}$ as a function of the parameters calculated in previous step:

$$r_{n,k}^{(i)} \propto |\mathbf{W}_k^{(i)}|^{1/2} \exp\left(\psi(\alpha_0 + N_k^{(i)}) + \sum_{i=1}^{D}\psi\left(\frac{\nu_k^{(i)} + 1 - i}{2}\right) - \frac{D}{2N_k^{(i)} + 2} - \frac{\nu_k^{(i)}}{2}(\boldsymbol{x}_n - \boldsymbol{m}_k^{(i)})^\top \mathbf{W}_k^{(i)}(\boldsymbol{x}_n - \boldsymbol{m}_k^{(i)})\right)$$

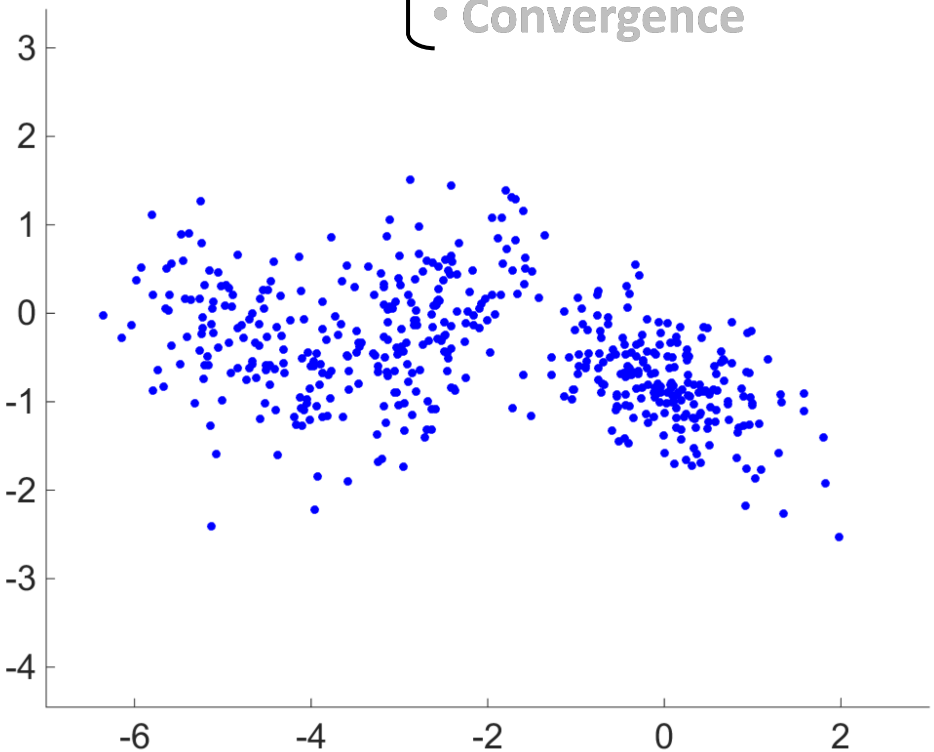
where $\psi(.)$ denotes the digamma function.

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\}\right)$
- **Convergence**



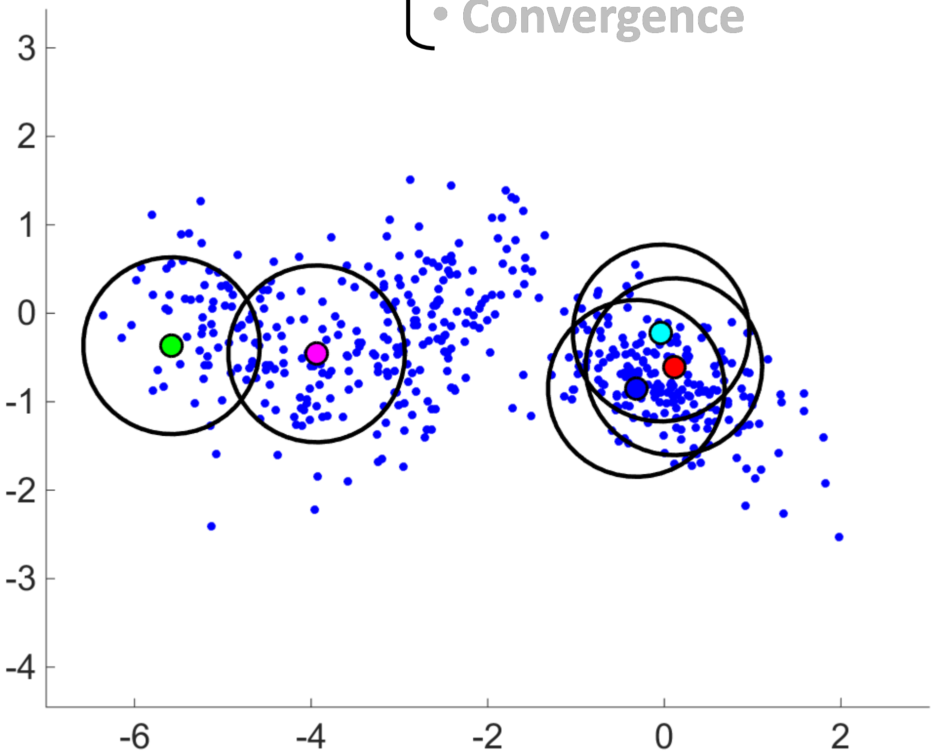
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



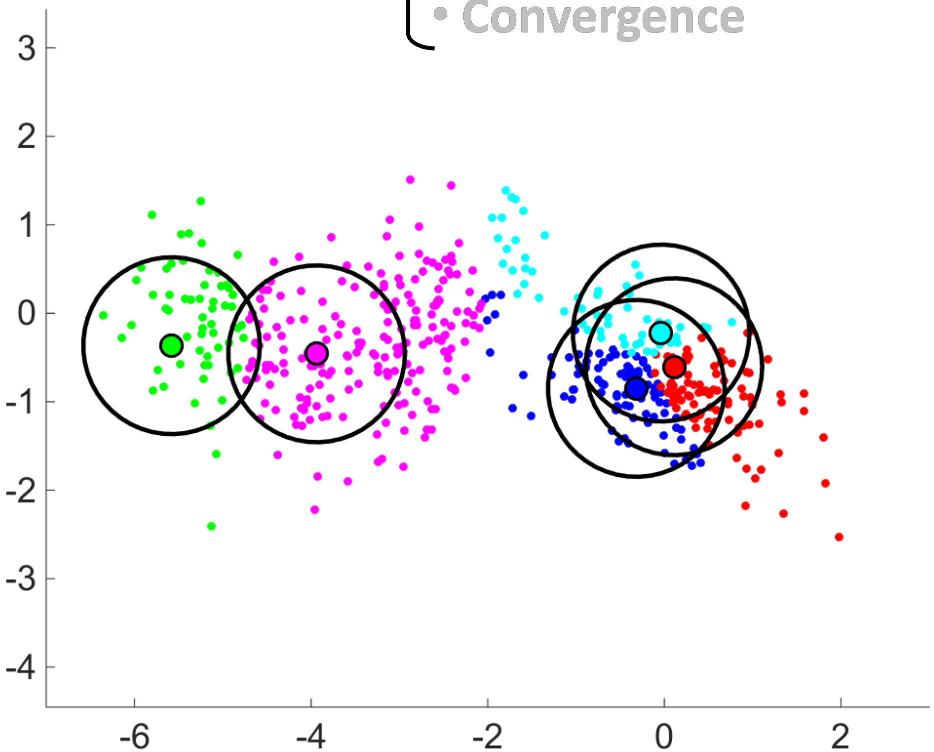
$K = 5, \ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



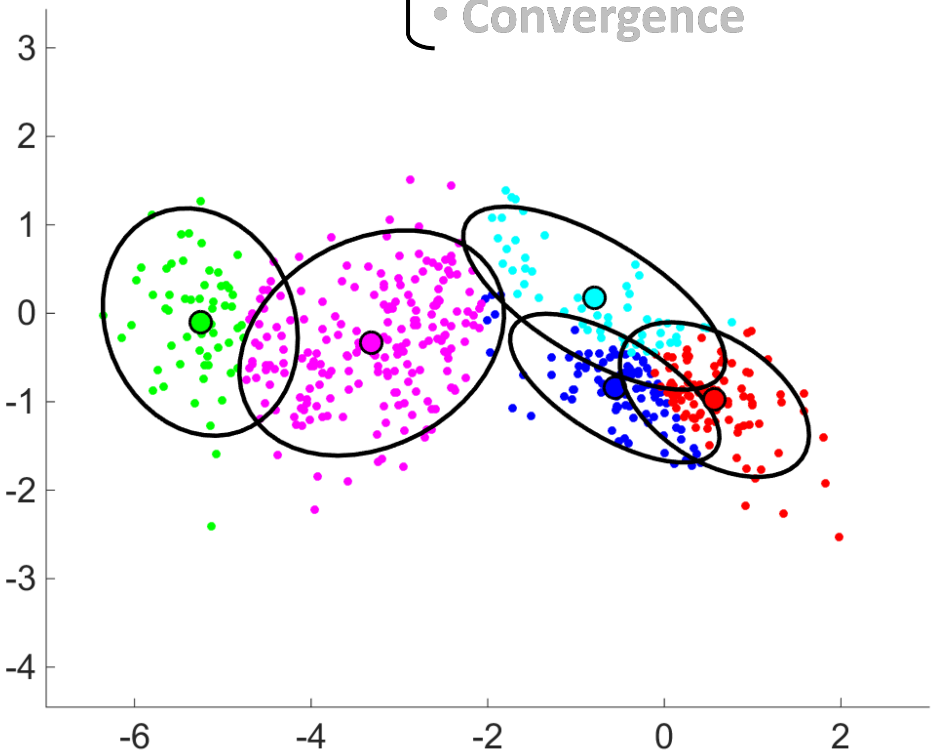
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



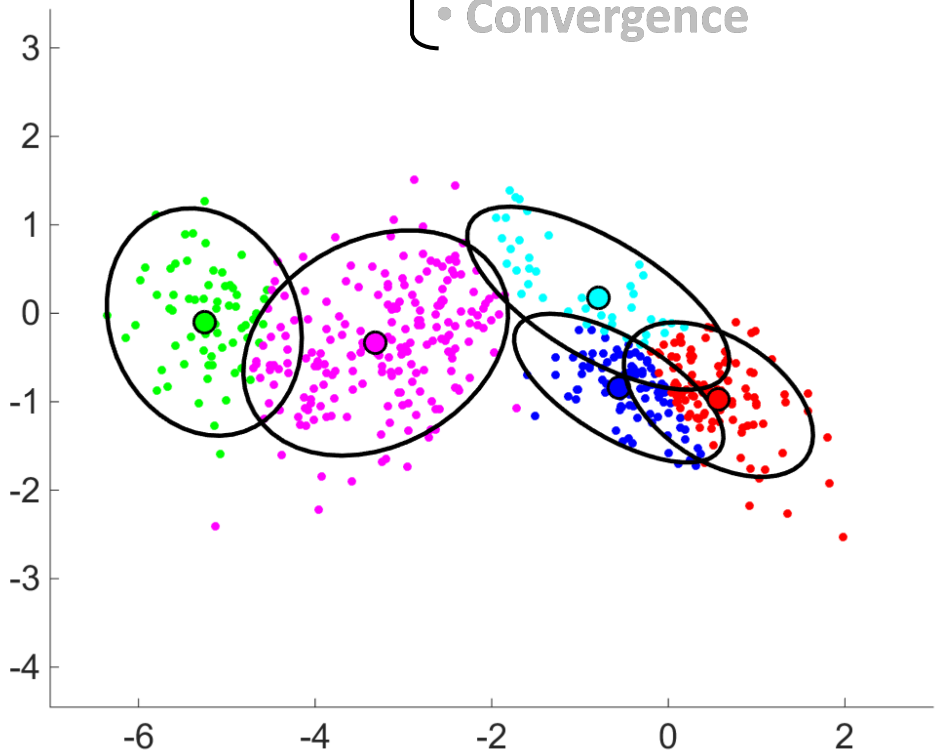
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



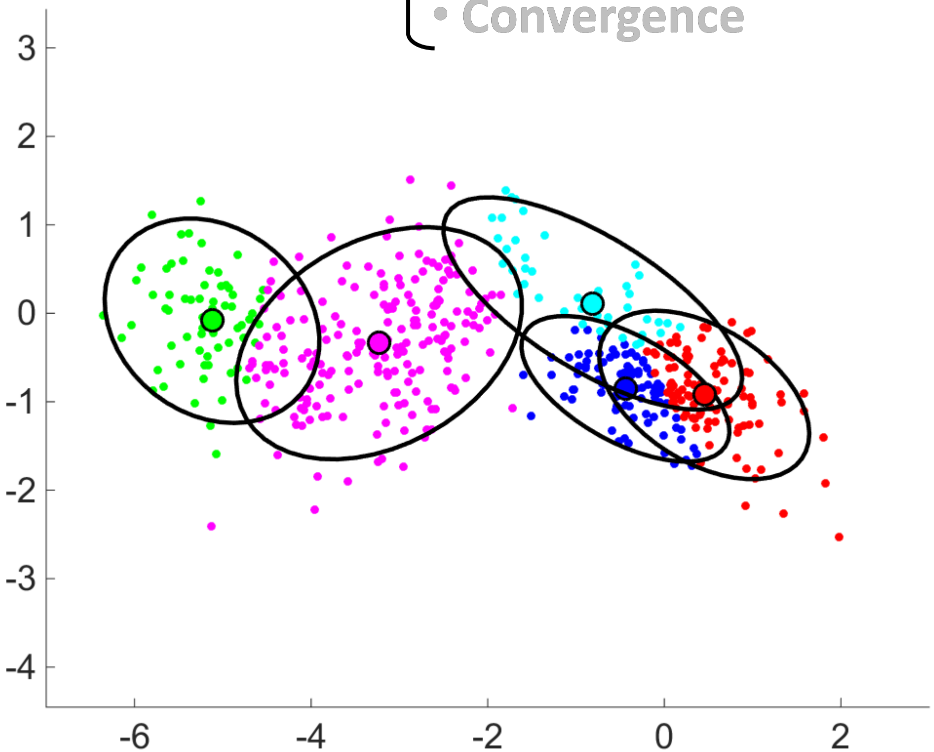
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\}\right)$
- **Convergence**



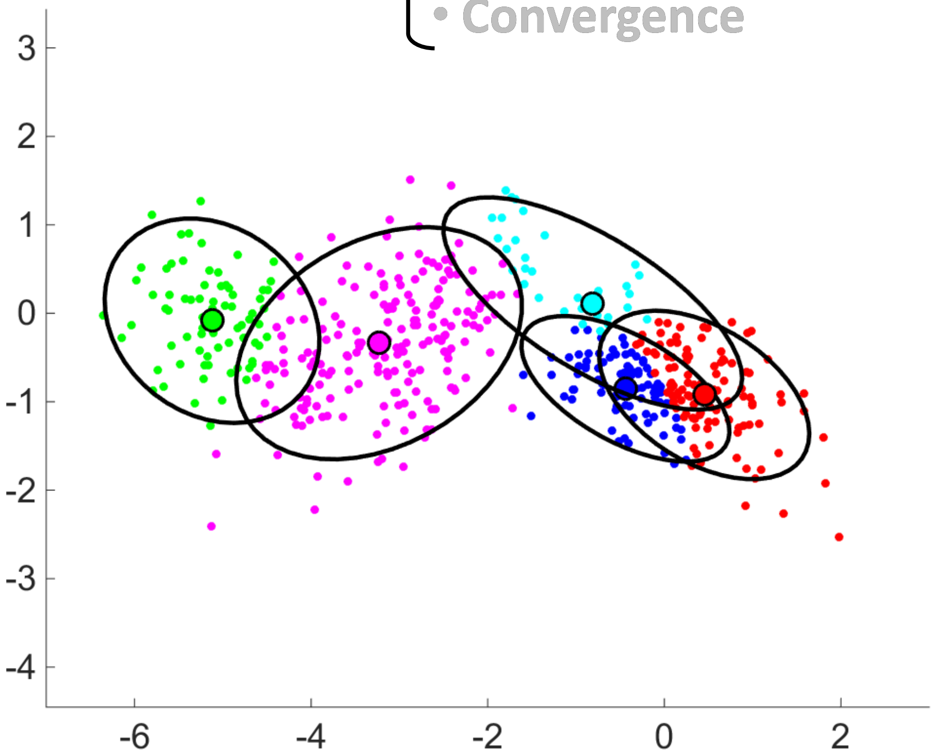
$K = 5, \ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



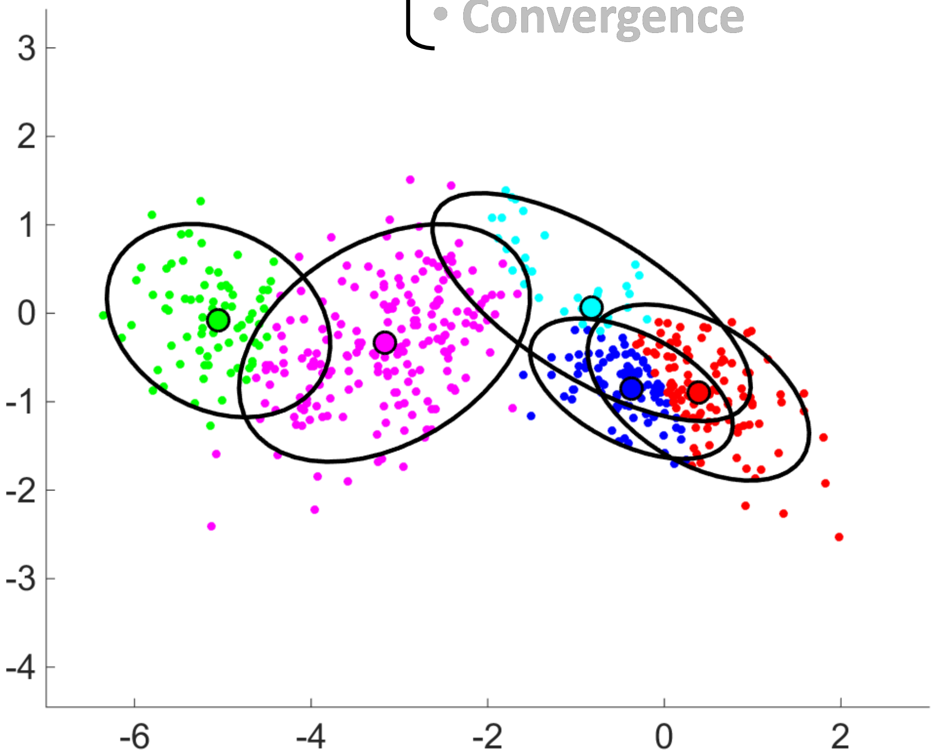
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



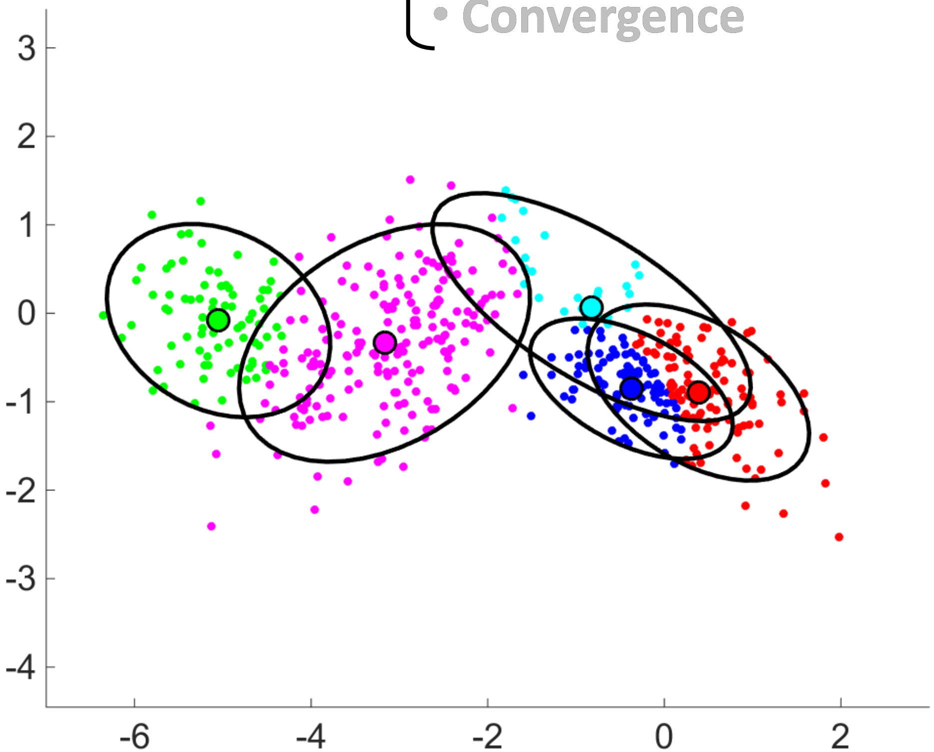
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\mathbf{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



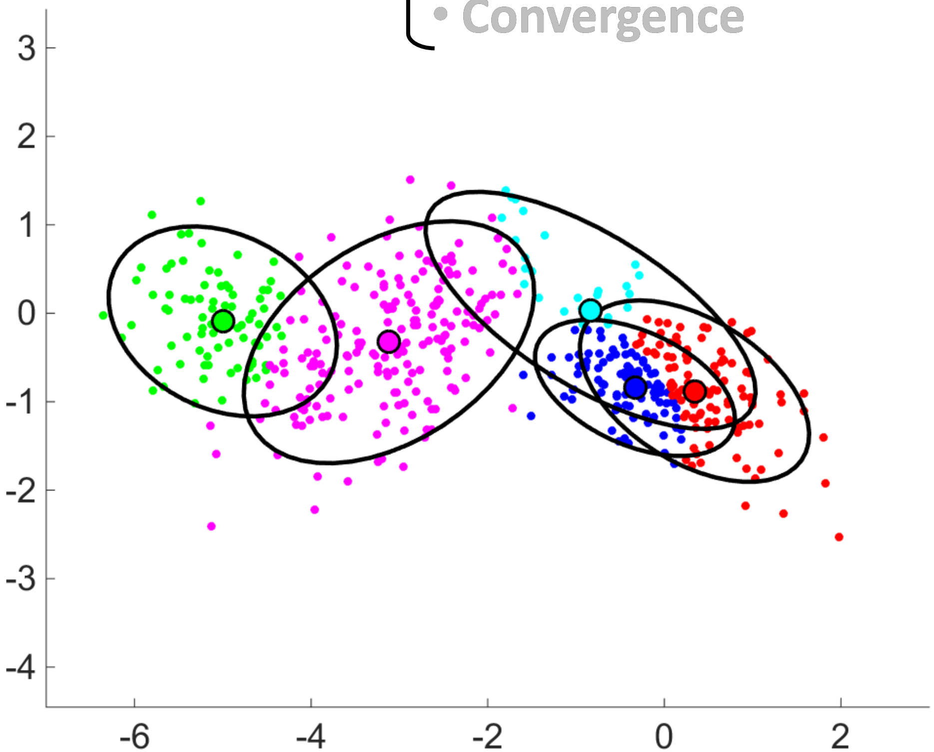
$K=5,\ \alpha_0=0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\}\right)$
- **Convergence**



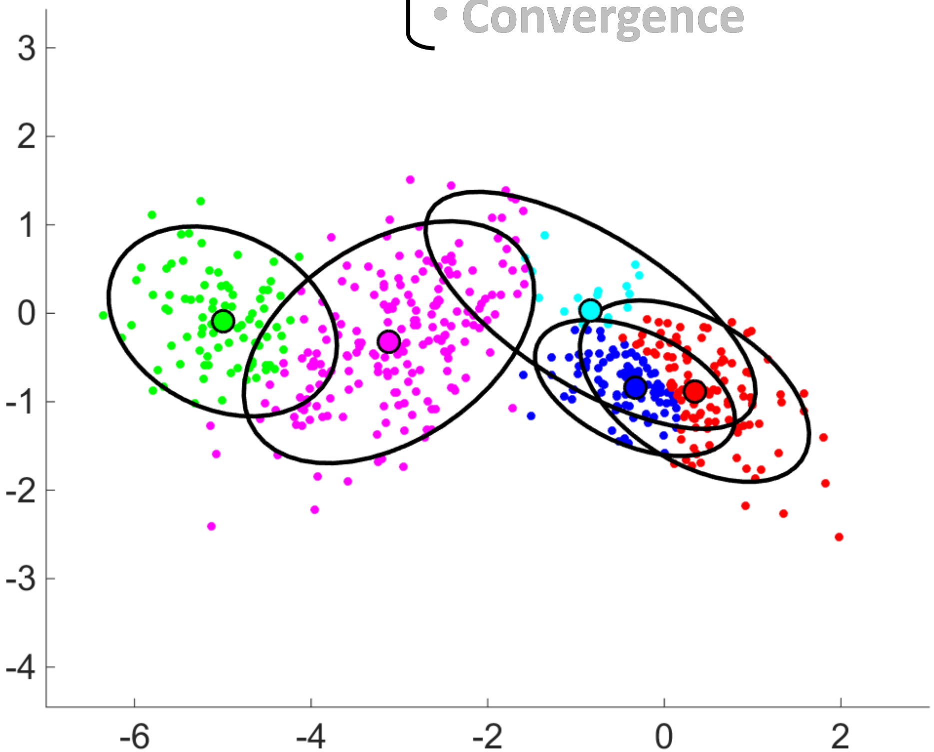
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



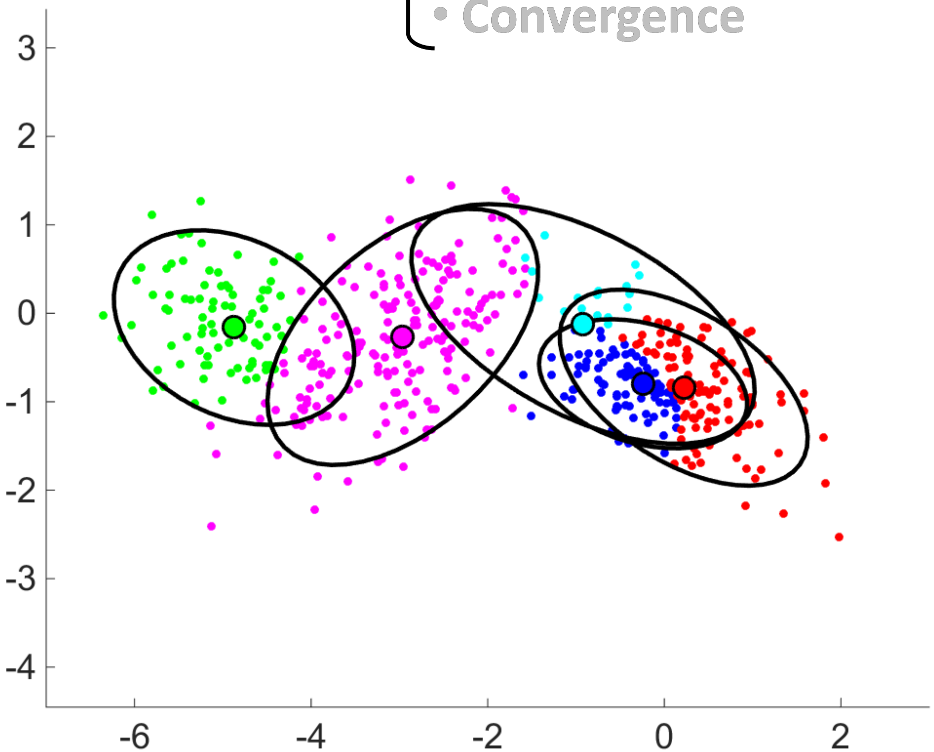
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



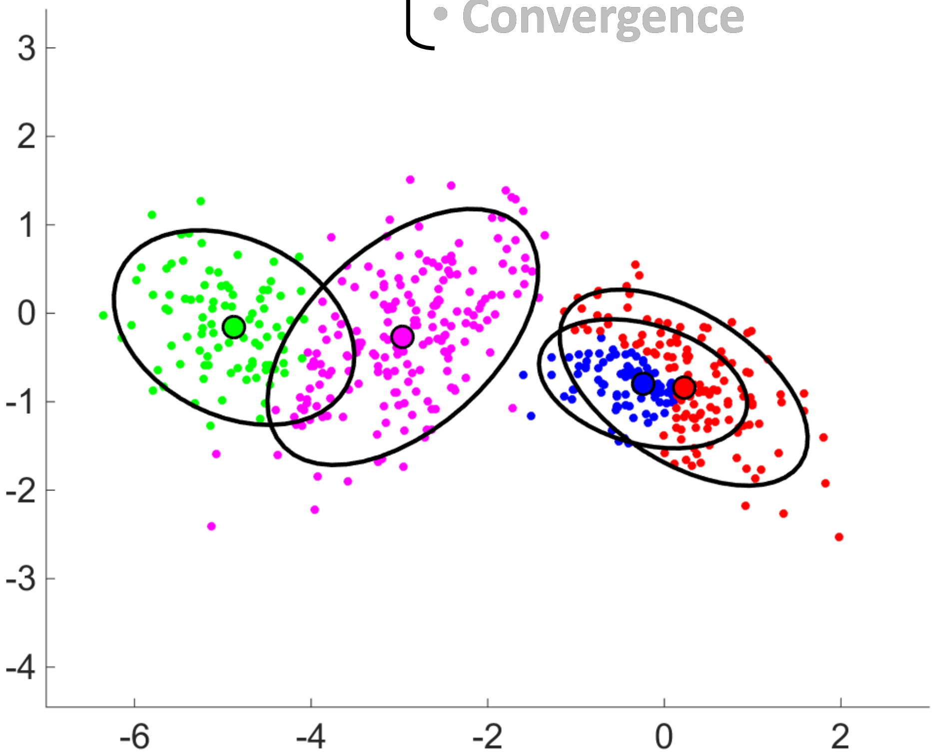
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



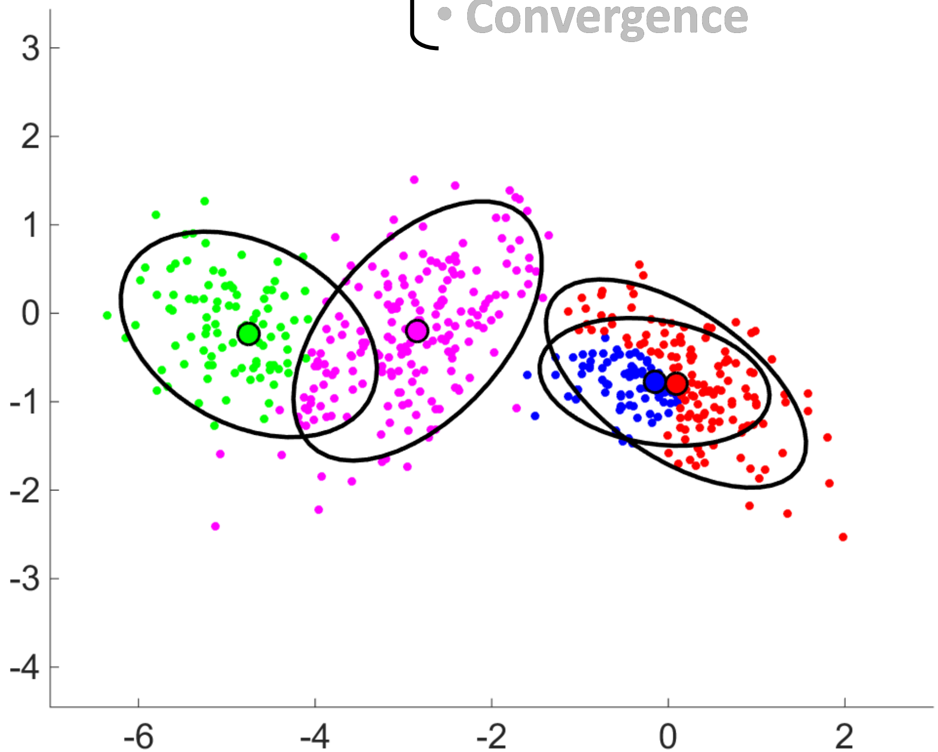
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



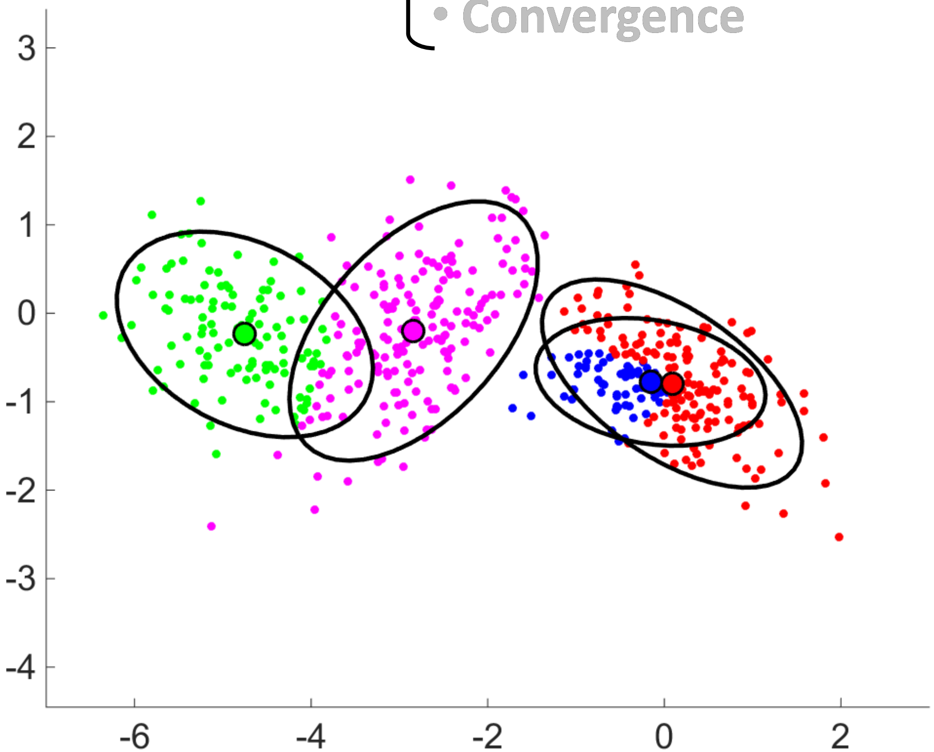
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



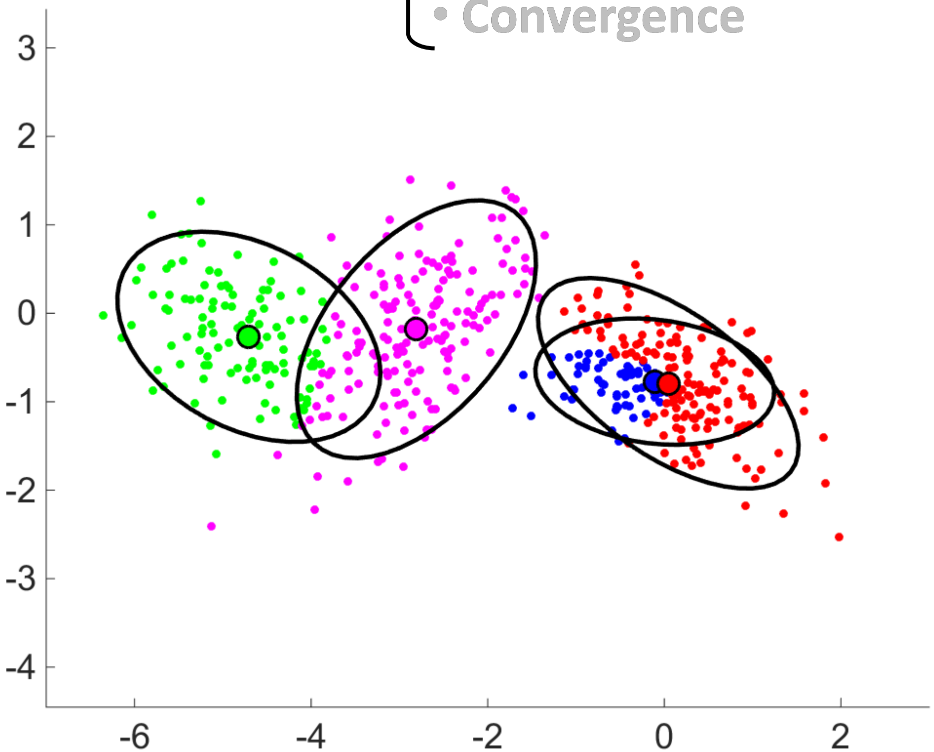
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\}\right)$
- **Convergence**



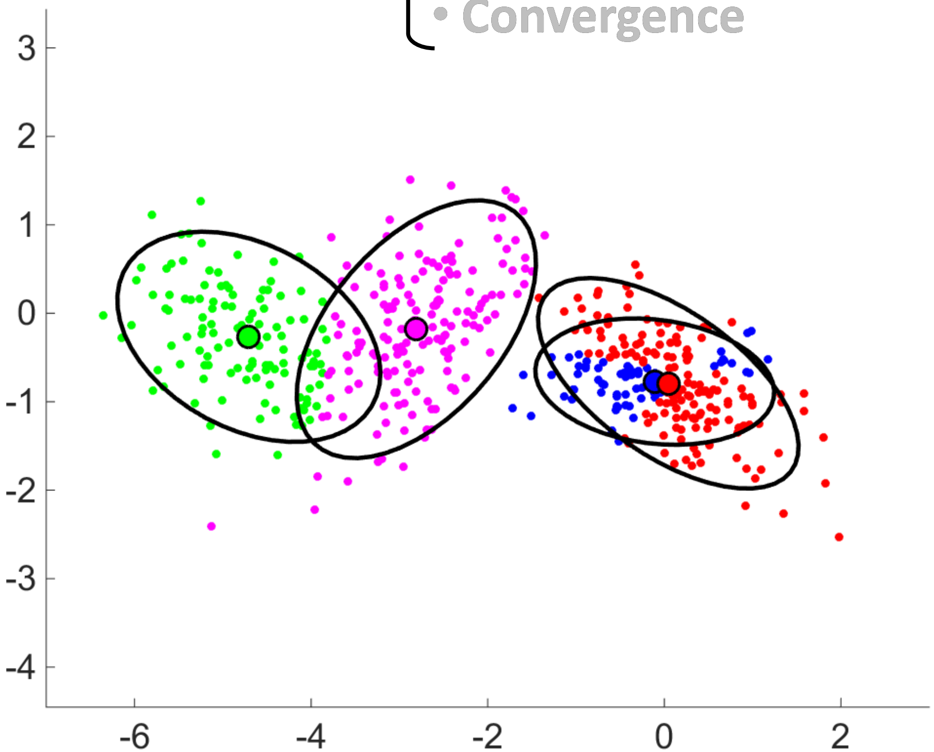
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



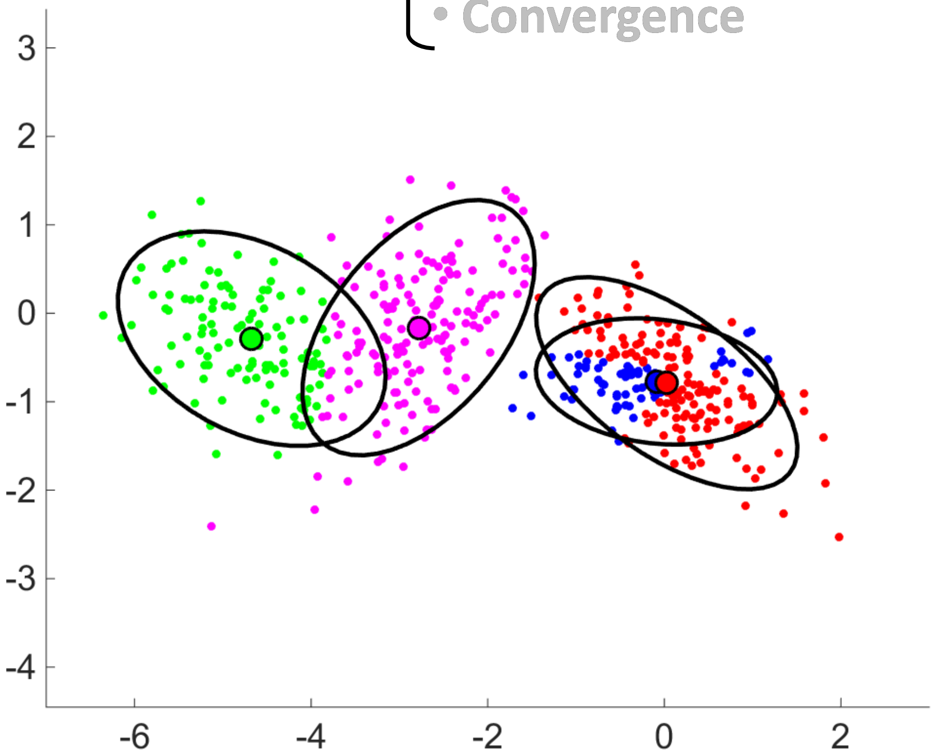
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



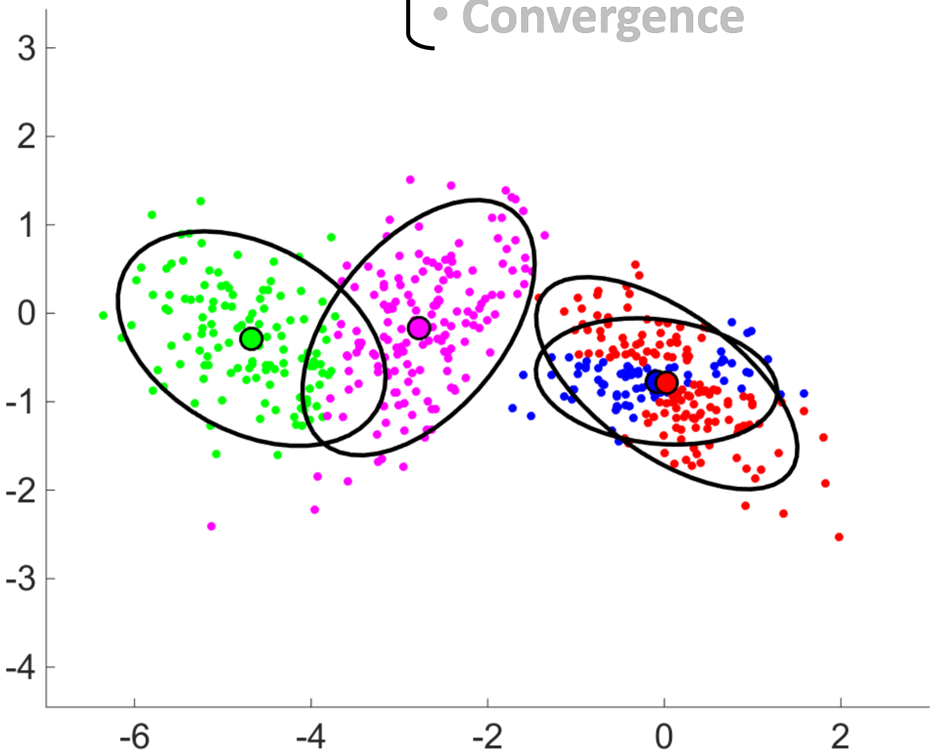
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\}\right)$
- **Convergence**



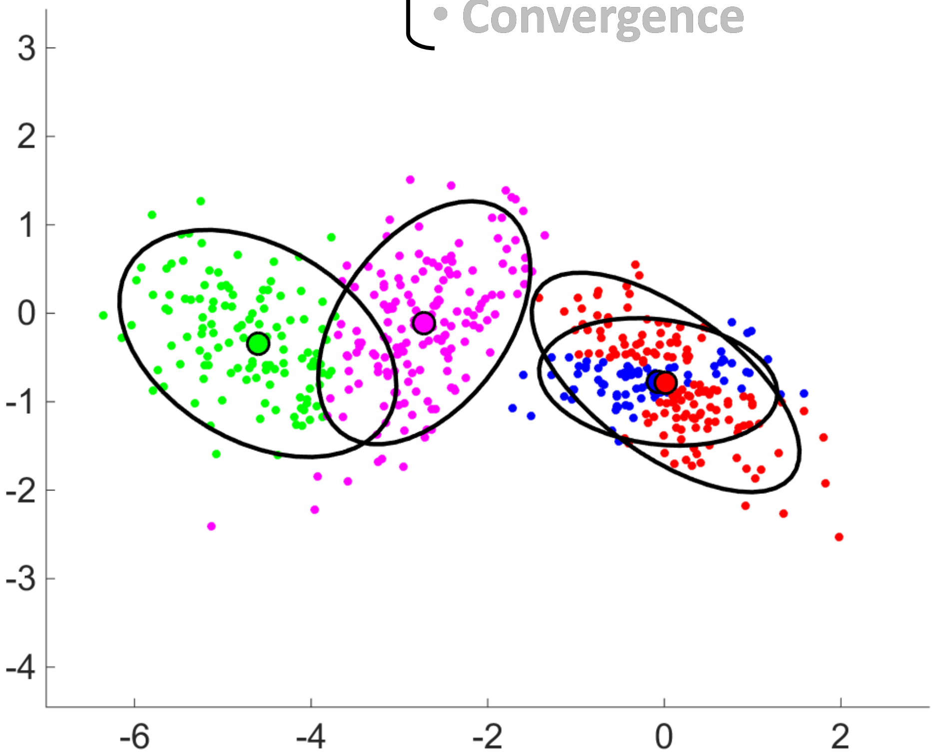
$K = 5, \ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



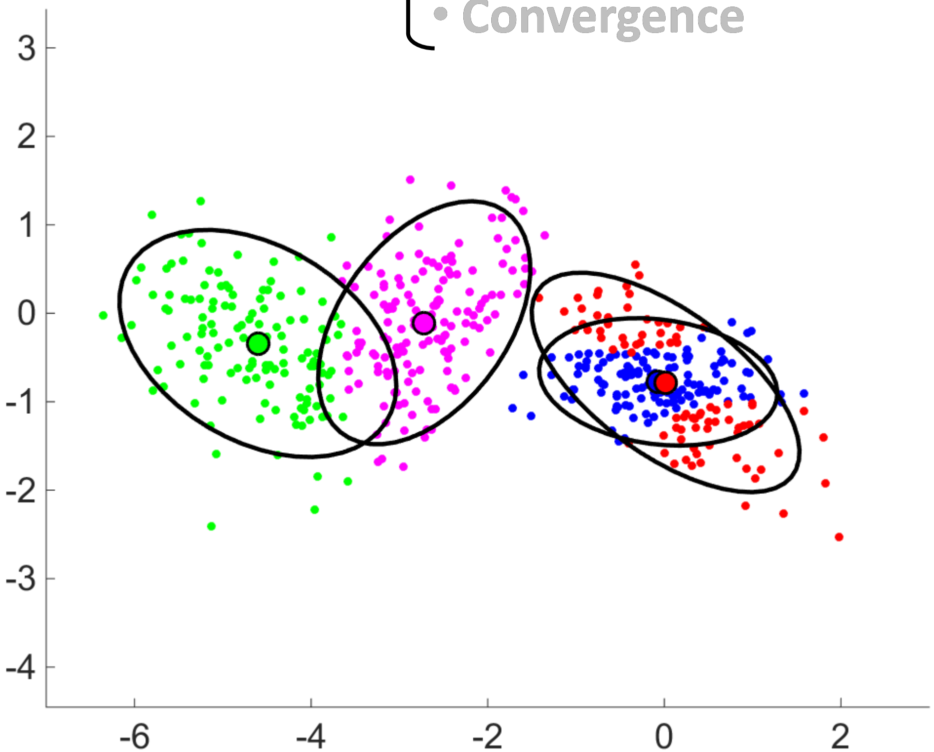
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q^{(i)}_{\Lambda\mu\pi}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q^{(i)}_{\Lambda\mu\pi}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



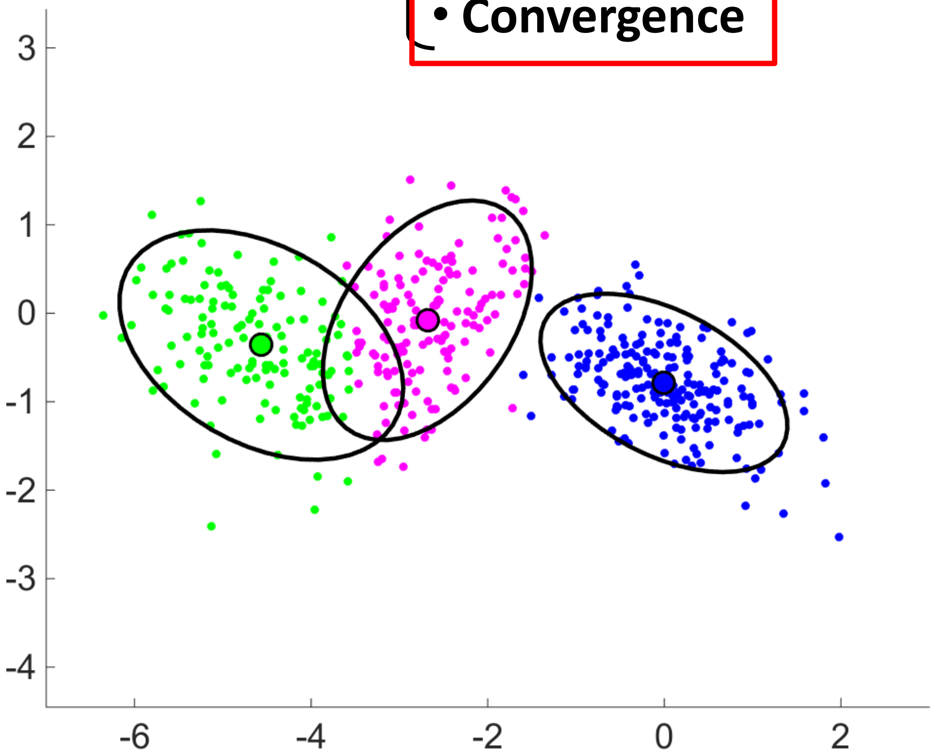
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi}|\mathbf{X}, \mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{\pi})\}\right)$
- **Convergence**



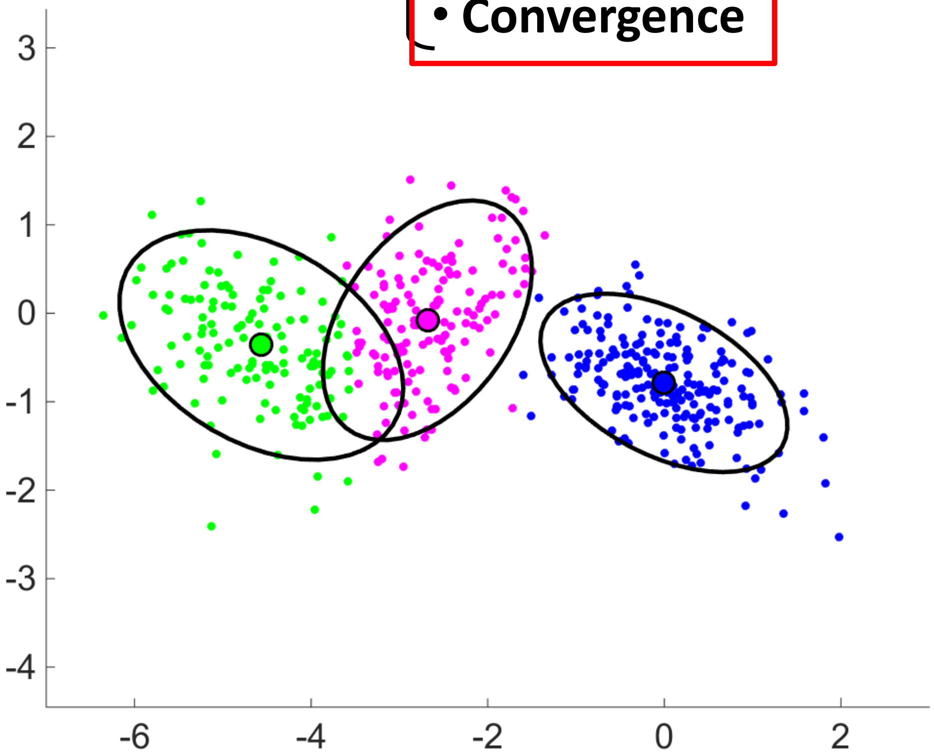
$K = 5,\ \alpha_0 = 0.01$

Bayesian Inference: Examples

Inference

VB-EM in action

- **Initialization:** Random means + GMM E-step for $q_Z^{(0)}(\mathbf{Z})$
- **E-Λμπ-Step :** $q_{\Lambda\mu\pi}^{(i)}(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}) \propto \exp\left(\mathbb{E}_{q_Z^{(i-1)}}\{\log p(\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi}|\mathbf{X},\mathbf{Z})\}\right)$
- **E-Z-Step:** $q_Z^{(i)}(\mathbf{Z}) \propto \exp\left(\mathbb{E}_{q_{\Lambda\mu\pi}^{(i)}}\{\log p(\mathbf{Z}|\mathbf{X},\boldsymbol{\Lambda},\boldsymbol{\mu},\boldsymbol{\pi})\}\right)$
- **Convergence**



$K = 5,\ \alpha_0 = 0.01$

Conclusions on GMM VB-EM

- Similar computational time as GMM-EM (though slightly more iterations)
- Priors on Gaussian weights handle automatically degenerate or unused clusters
- Determination of K
- Works even for very small data samples

OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- **Bayesian inference by examples**
  - **Direct inference**
  - **The Expectation-Maximization algorithm**
  - **Variational Bayes methods**
- Markov chain Monte Carlo

OUTLINE

- What is Bayesian inference?
  - Overview
  - Classical vs. Bayesian approach
  - Bayes Theorem & Example
  - General Methodology
- Bayesian inference by examples
  - Direct inference
  - The Expectation-Maximization algorithm
  - Variational Bayes methods
- **Markov chain Monte Carlo**