Overview of Discrete and Continuous Variables

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Plan

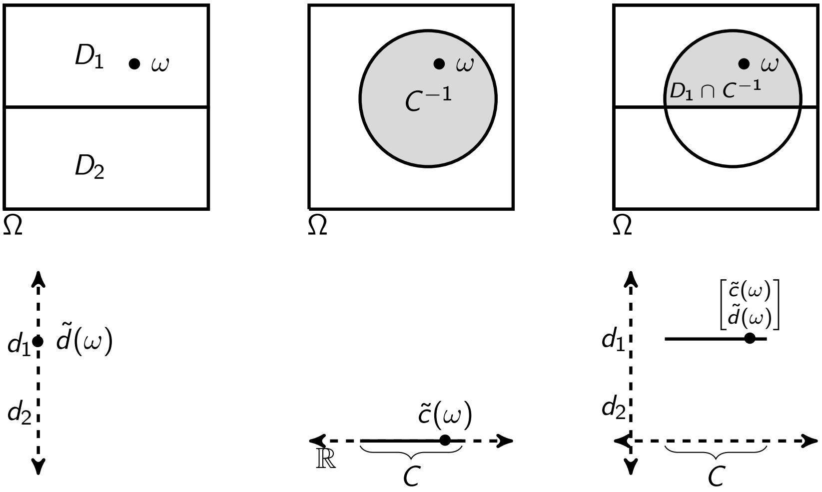
- Joint distribution of discrete and continuous variables
- Gaussian mixture models for classification and clustering
- Bayesian models

Discrete and continuous variables

How can we jointly model discrete and continuous quantities?

We represent them as random variables in the same probability space

Discrete and continuous variables



User interface

Joint pmf? ✗

Joint pdf? ✗

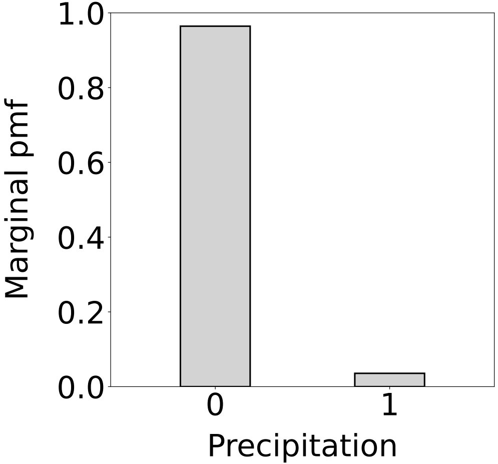
Joint cdf? ✓ but ☹

Alternatives? Marginal pmf and conditional pdf

Mauna Loa

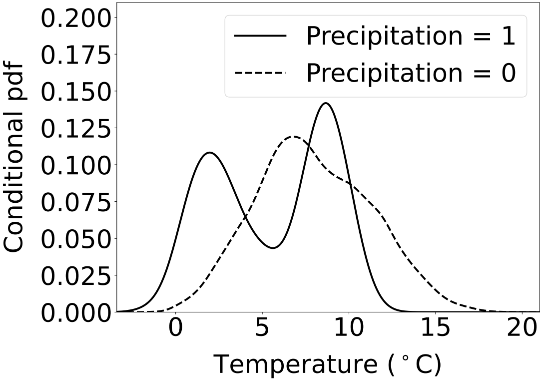
Temperature ($\tilde{c}$) and precipitation ($\tilde{d}$)

Marginal pmf of precipitation



Conditional pdf of temperature given precipitation

$$f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,d) := \lim_{\epsilon \to 0} \frac{\mathrm{P}(c - \epsilon < \tilde{c} \leq c\,|\,\tilde{d} = d)}{\epsilon}$$



Marginal distribution of $\tilde{c}$

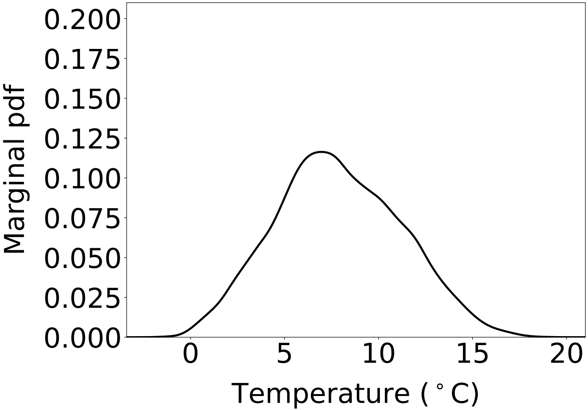
We know $p_{\tilde{d}}$ and $f_{\tilde{c}\,|\,\tilde{d}}(\cdot\,|\,d)$ for all d

Marginal distribution of $\tilde{c}$?

$$f_{\tilde{c}}(c) = \sum_{d \in D} p_{\tilde{d}}(d)\, f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,d)$$

Mauna Loa

$$f_{\tilde{c}}(c) = p_{\tilde{d}}(0)\, f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,0) + p_{\tilde{d}}(1)\, f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,1)$$



Conditional pmf

Conditional pmf of $\tilde{d}$ given $\tilde{c} = c$?

Problem: $\mathrm{P}(\tilde{c} = c) = 0$

As usual, we resort to limits

$$p_{\tilde{d}\,|\,\tilde{c}}(d\,|\,c) := \lim_{\epsilon \to 0} \mathrm{P}\left(\tilde{d} = d\,|\,c - \epsilon < \tilde{c} \leq c\right)$$

Marginal distribution of $\tilde{d}$

We know $f_{\tilde{c}}$ and $p_{\tilde{d}\,|\,\tilde{c}}(\cdot\,|\,c)$ for all c

Marginal distribution of $\tilde{d}$?

$$p_{\tilde{d}}(d) = \int_{c=-\infty}^{\infty} f_{\tilde{c}}(c)\, p_{\tilde{d}\,|\,\tilde{c}}(d\,|\,c)\, \mathrm{d}c$$

Chain rule

For discrete $\tilde{a}$ and $\tilde{b}$

$$\begin{aligned} p_{\tilde{a},\tilde{b}}(a,b) &= p_{\tilde{a}}(a)\, p_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) \\ &= p_{\tilde{b}}(b)\, p_{\tilde{a}\,|\,\tilde{b}}(a\,|\,b) \end{aligned}$$

For continuous $\tilde{a}$ and $\tilde{b}$

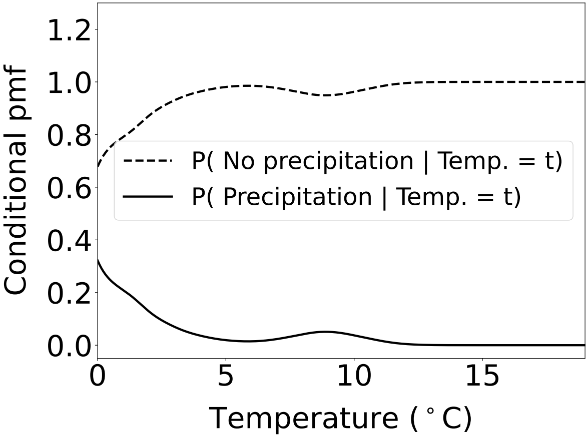
$$\begin{aligned} f_{\tilde{a},\tilde{b}}(a,b) &= f_{\tilde{a}}(a)\, f_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) \\ &= f_{\tilde{b}}(b)\, f_{\tilde{a}\,|\,\tilde{b}}(a\,|\,b) \end{aligned}$$

For discrete $\tilde{d}$ and continuous $\tilde{c}$

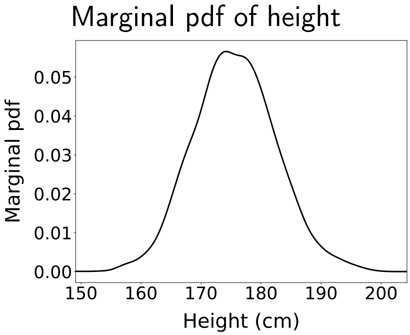
$$p_{\tilde{d}}(d)\, f_{\tilde{c}\,|\,\tilde{d}}(c\,|\,d) = f_{\tilde{c}}(c)\, p_{\tilde{d}\,|\,\tilde{c}}(d\,|\,c)$$

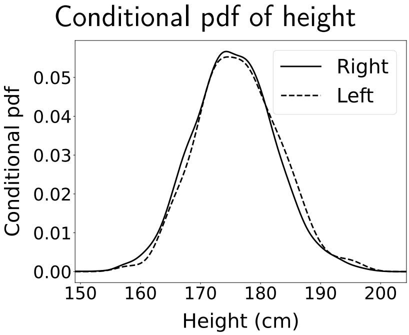
Mauna Loa

$$p_{\tilde{d} \mid \tilde{c}}(d \mid c) = \frac{p_{\tilde{d}}(d)\, f_{\tilde{c} \mid \tilde{d}}(c \mid d)}{f_{\tilde{c}}(c)}$$

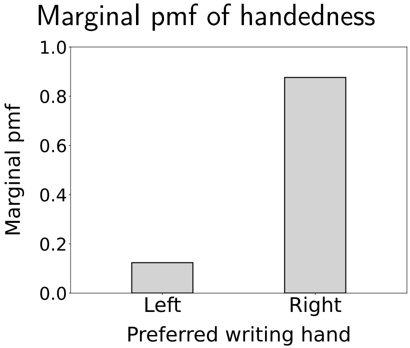


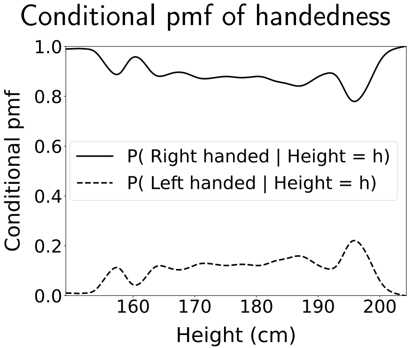
Height and handedness





Height and handedness





Independence

A pair of continuous and discrete random variables $\tilde{c}$ and $\tilde{d}$ are independent if and only if

$$p_{\tilde{d} \mid \tilde{c}}(d \mid c) = p_{\tilde{d}}(d)$$

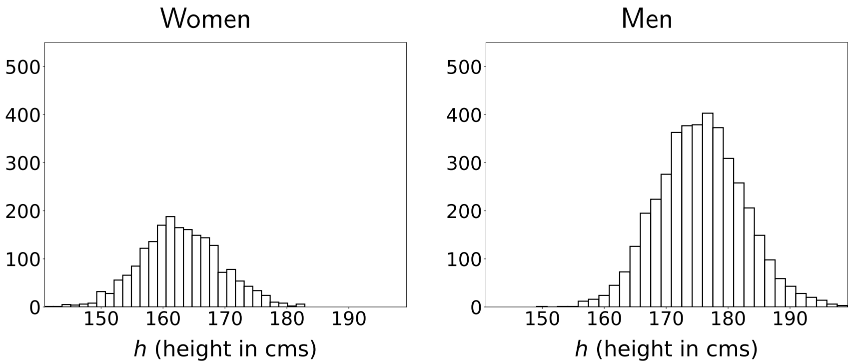
$$f_{\tilde{c} \mid \tilde{d}}(c \mid d) = f_{\tilde{c}}(c) \quad \text{for all } c, d$$

Conditional independence

A pair of continuous and discrete random variables $\tilde{c}$ and $\tilde{d}$ are conditionally independent given $\tilde{a}$ if and only if

$$\begin{aligned} p_{\tilde{d}\,|\,\tilde{c},\tilde{a}}(d\,|\,c,a) &= p_{\tilde{d}\,|\,\tilde{a}}(d\,|\,a) \\ f_{\tilde{c}\,|\,\tilde{d},\tilde{a}}(c\,|\,d,a) &= f_{\tilde{c}\,|\,\tilde{a}}(c\,|\,a) \quad \text{for all } a, c, d \end{aligned}$$

Mixture models



Gaussian mixture model

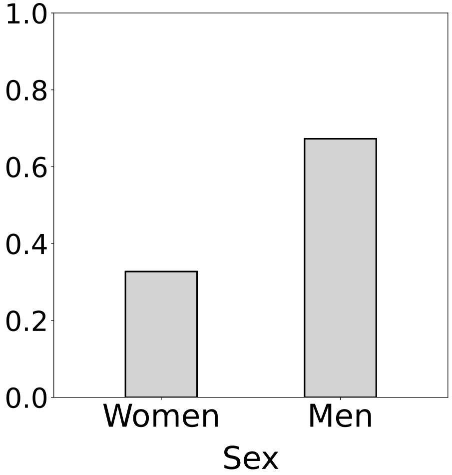
Height: Continuous random variable $\tilde{h}$

Sex: Discrete random variable $\tilde{s}$

Conditional distribution of $\tilde{h}$ given $\tilde{s}$ is Gaussian

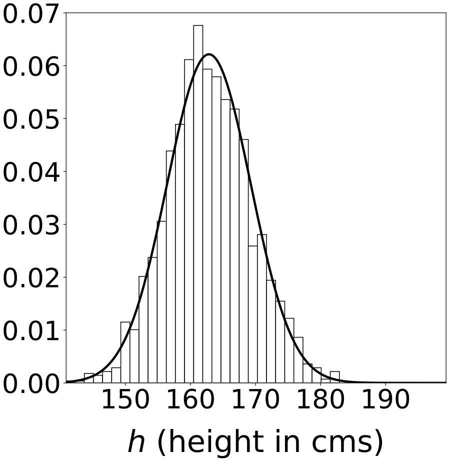
Distribution of $\tilde{s}$?

1,986 women and 4,082 men



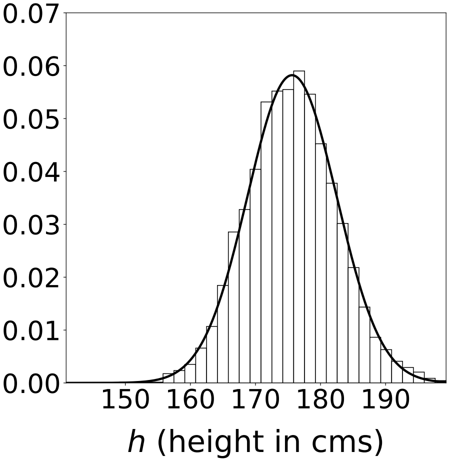
Conditional distribution of $\tilde{h}$ given $\tilde{s}$ = woman?

Gaussian with $\mu_{\text{women}} = 163$ cm and $\sigma_{\text{women}} = 6.4$ cm



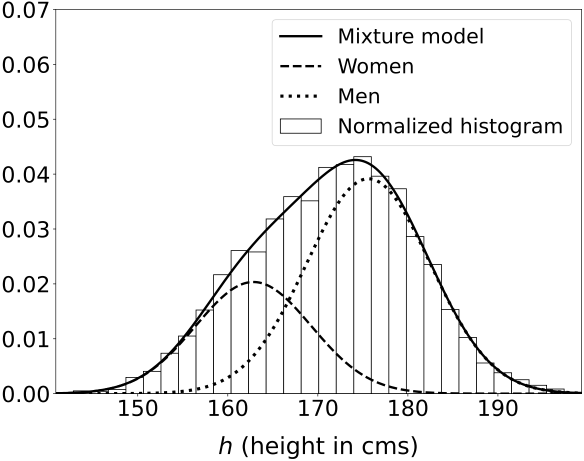
Conditional distribution of $\tilde{h}$ given $\tilde{s}$ = man?

Gaussian with $\mu_{\text{men}} = 176$ cm and $\sigma_{\text{men}} = 6.9$ cm



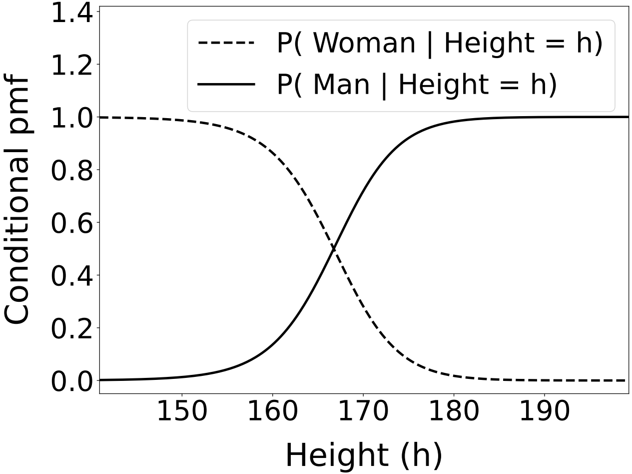
Gaussian mixture model

$$f_{\tilde{h}}(h) = p_{\tilde{s}}\,(\text{woman})\, f_{\tilde{h}\,|\,\tilde{s}}\,(h\,|\,\text{woman}) + p_{\tilde{s}}\,(\text{man})\, f_{\tilde{h}\,|\,\tilde{s}}\,(h\,|\,\text{man})$$



Conditional distribution of $\tilde{s}$ given $\tilde{h}$

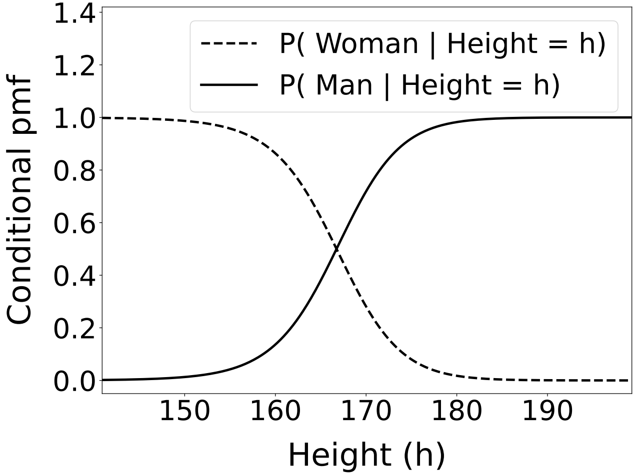
$$p_{\tilde{s}\,|\,\tilde{h}}(0\,|\,h) = \frac{p_{\tilde{s}}(0)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,0)}{p_{\tilde{s}}(0)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,0) + p_{\tilde{s}}(1)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,1)}$$



Gaussian discriminant analysis

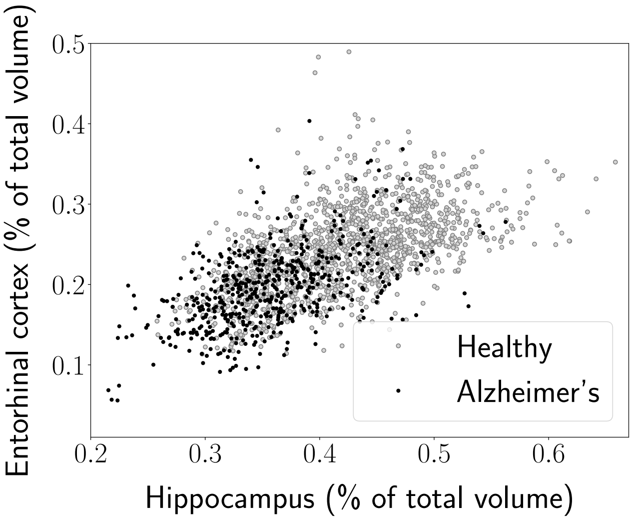
Idea: Use Gaussian mixture model for classification

$$p_{\tilde{s}\,|\,\tilde{h}}(0\,|\,h) = \frac{p_{\tilde{s}}(0)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,0)}{p_{\tilde{s}}(0)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,0) + p_{\tilde{s}}(1)\, f_{\tilde{h}\,|\,\tilde{s}}(h\,|\,1)}$$

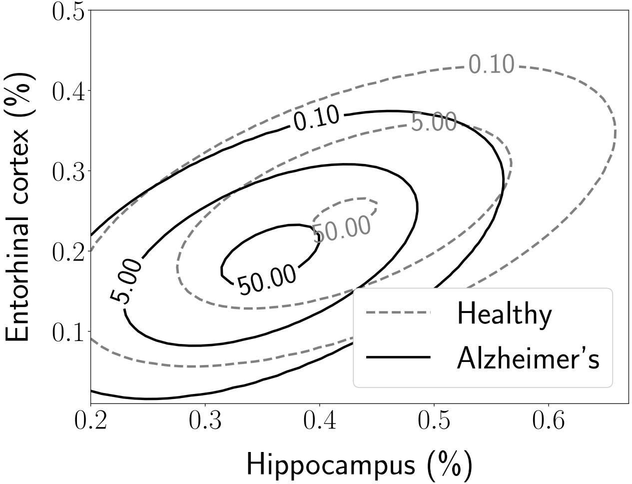


Training data

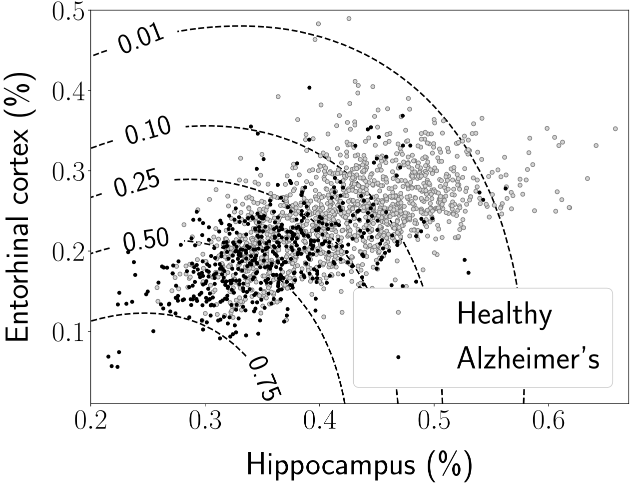
Alzheimer's Disease Neuroimaging Initiative



Conditional density of features given class

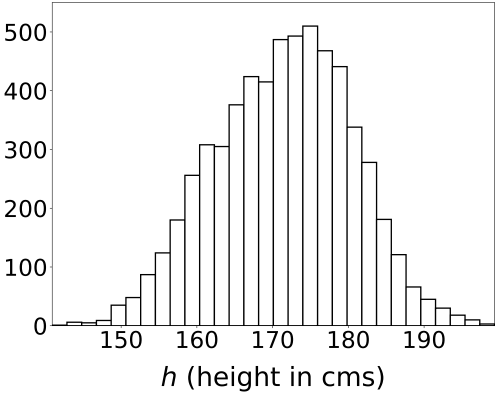


Conditional probability of class given features



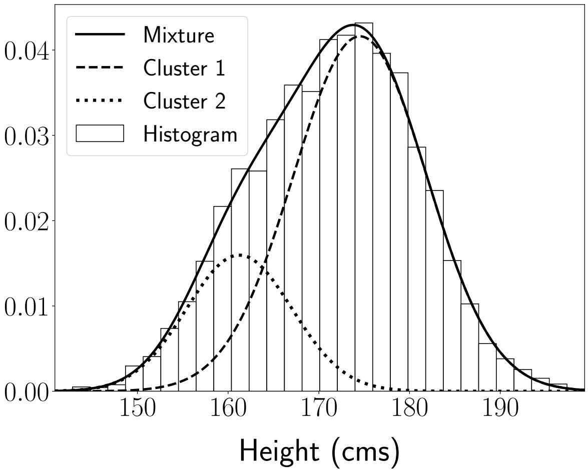
Clustering

Unsupervised learning: No training labels

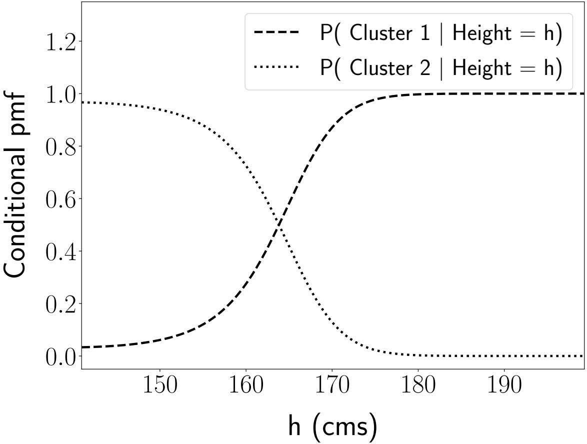


Strategy: Fit mixture model to cluster the data

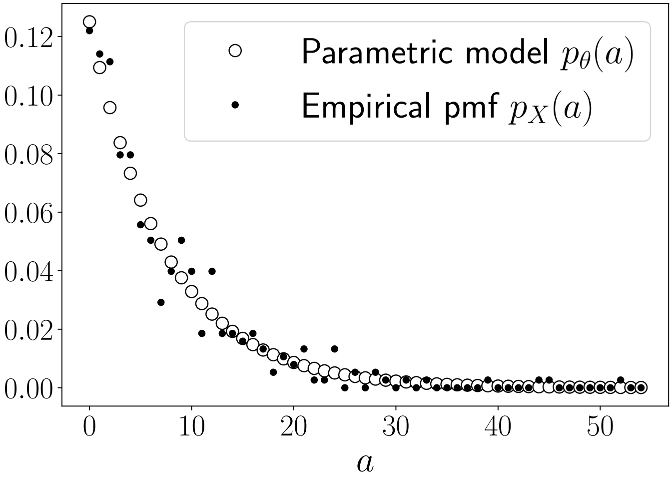
Gaussian mixture model for clustering



Gaussian mixture model for clustering



Parametric modeling



Bayesian parametric modeling

Key idea: Interpret parameters as random variables

Building a Bayesian model

Parameters: $\tilde{\theta}$

Data: $\tilde{x}$

1. Prior distribution of parameters: $f_{\tilde{\theta}}$
2. Conditional distribution or likelihood of the data given the parameters $p_{\tilde{x} \mid \tilde{\theta}}$ or $f_{\tilde{x} \mid \tilde{\theta}}$

Goal: Compute posterior distribution of parameters given data

Single coin flip

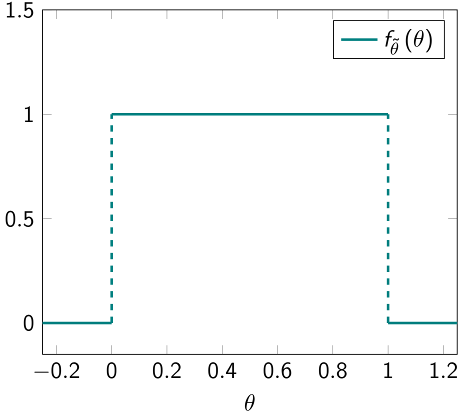
Parameter: Probability of heads $\tilde{\theta}$

Prior: $f_{\tilde{\theta}}$

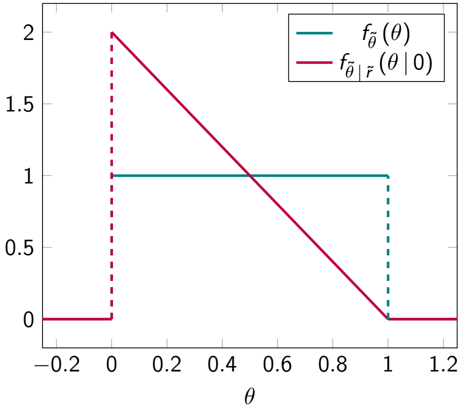
Likelihood

$$p_{\tilde{r} \mid \tilde{\theta}}(r \mid \theta) = \begin{cases} \theta & \text{if } r = 1 \\ 1 - \theta & \text{if } r = 0 \end{cases}$$

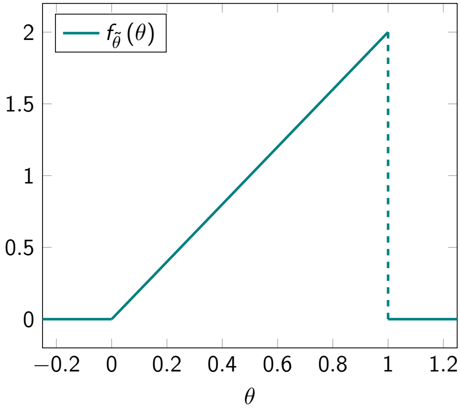
Uniform prior



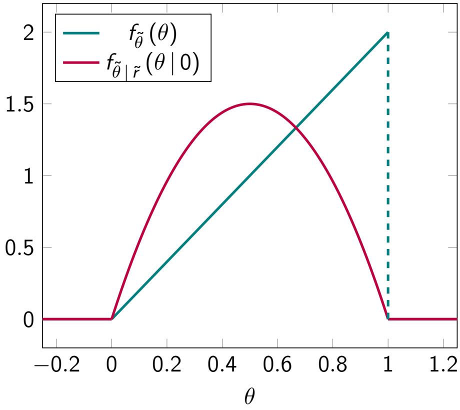
Posterior pdf after coin lands on tails



Triangular prior



Posterior if coin lands on tails



Conditional independence

What if we have more data?

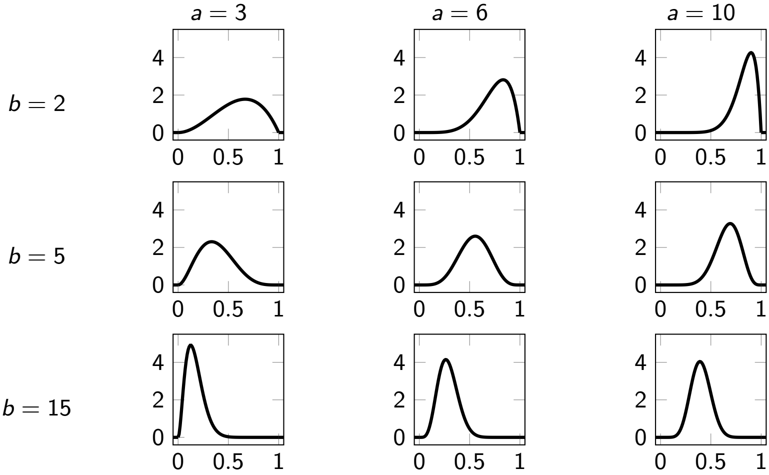
Common assumption: Data are conditionally independent given parameters

Same effect as iid assumption: likelihood factorizes

$$p_{\tilde{x} \,|\, \tilde{\theta}}(x \,|\, \theta) = \prod_{i=1}^{n} p_{\tilde{x}[i] \,|\, \tilde{\theta}}(x[i] \,|\, \theta)$$

$$f_{\tilde{x} \,|\, \tilde{\theta}}(x \,|\, \theta) = \prod_{i=1}^{n} f_{\tilde{x}[i] \,|\, \tilde{\theta}}(x[i] \,|\, \theta)$$

Beta distribution



Real poll (Pennsylvania)

Data: 281 people intend to vote for Trump, 300 for Biden

Parameter: Fraction of Trump voters $\tilde{\theta}$

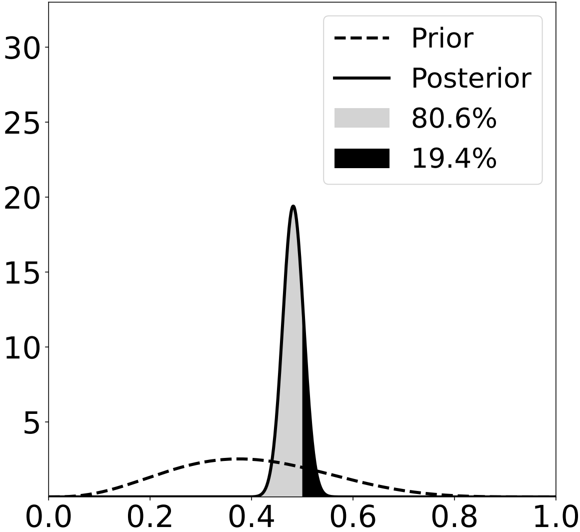
Prior: $f_{\tilde{\theta}}$ Beta with parameters a and b

Likelihood: $p_{\tilde{x}\,|\,\tilde{\theta}}$ Binomial with parameters n and θ

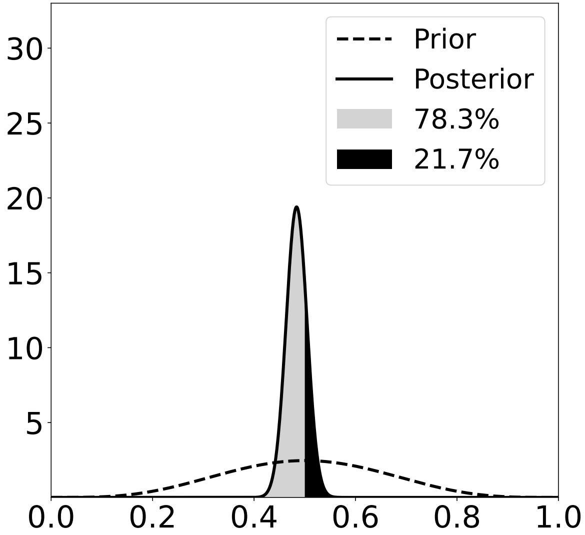
Posterior: $f_{\tilde{\theta}\,|\,\tilde{r}}$ Beta with parameters $a + 281$ and $b + 300$

Probability that Trump wins in Pennsylvania? $P(\tilde{\theta} > 0.5 \,|\, \tilde{x} = x)$

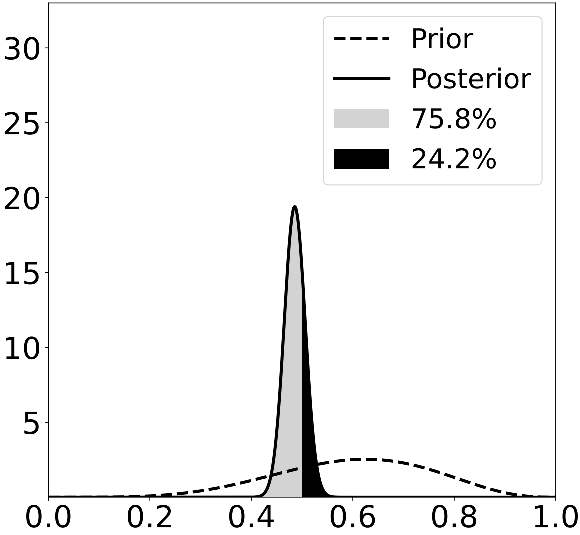
Poll in Pennsylvania



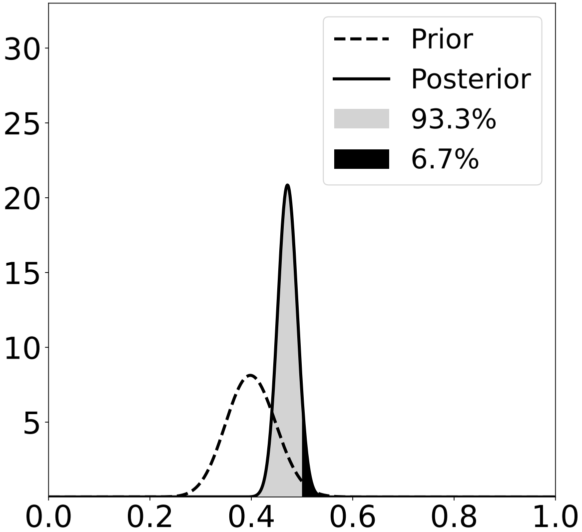
Poll in Pennsylvania



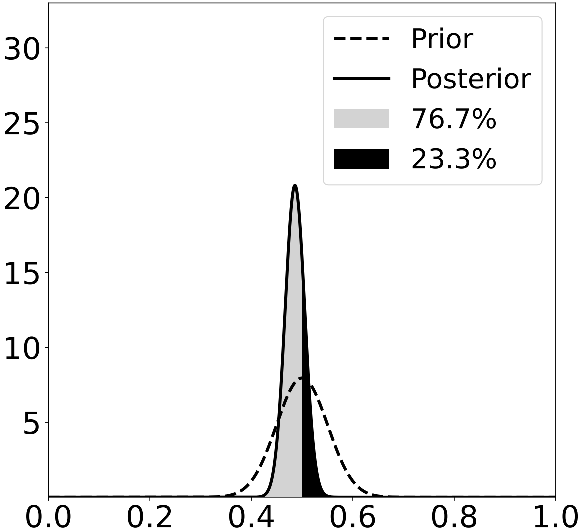
Poll in Pennsylvania



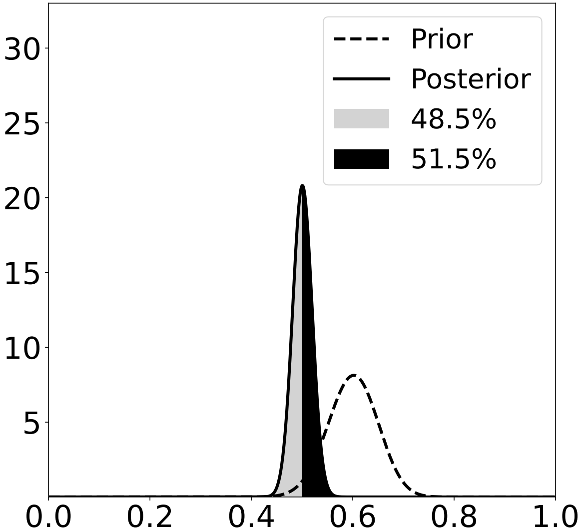
Poll in Pennsylvania



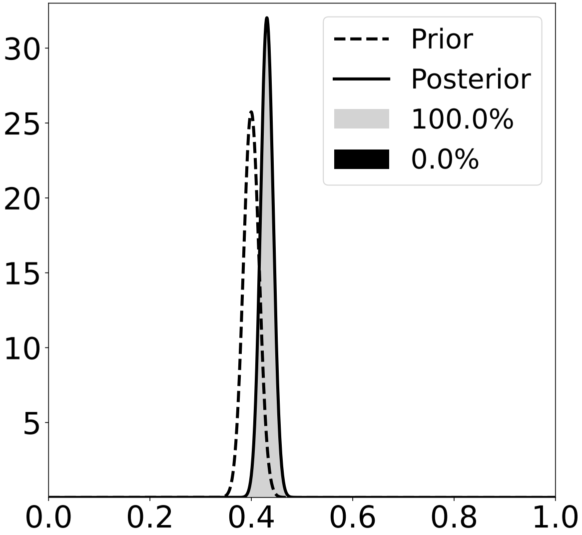
Poll in Pennsylvania



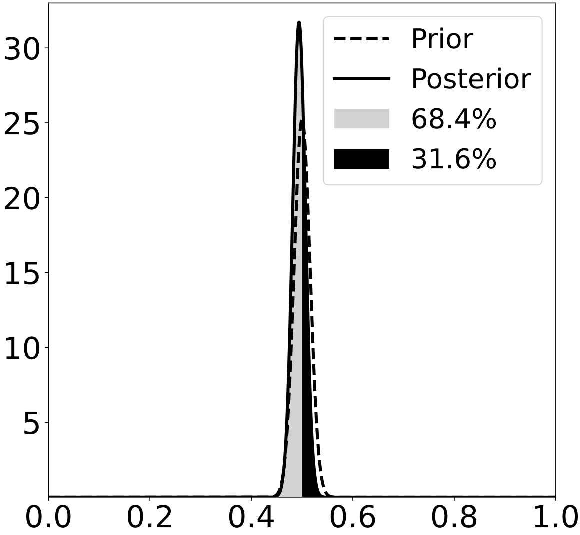
Poll in Pennsylvania



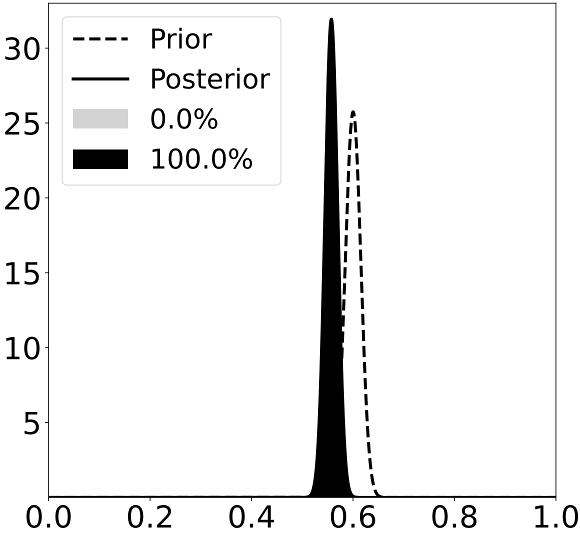
Poll in Pennsylvania



Poll in Pennsylvania



Poll in Pennsylvania



What have we learned?

- Joint distribution of discrete and continuous variables
- Gaussian mixture models for classification and clustering
- Bayesian models