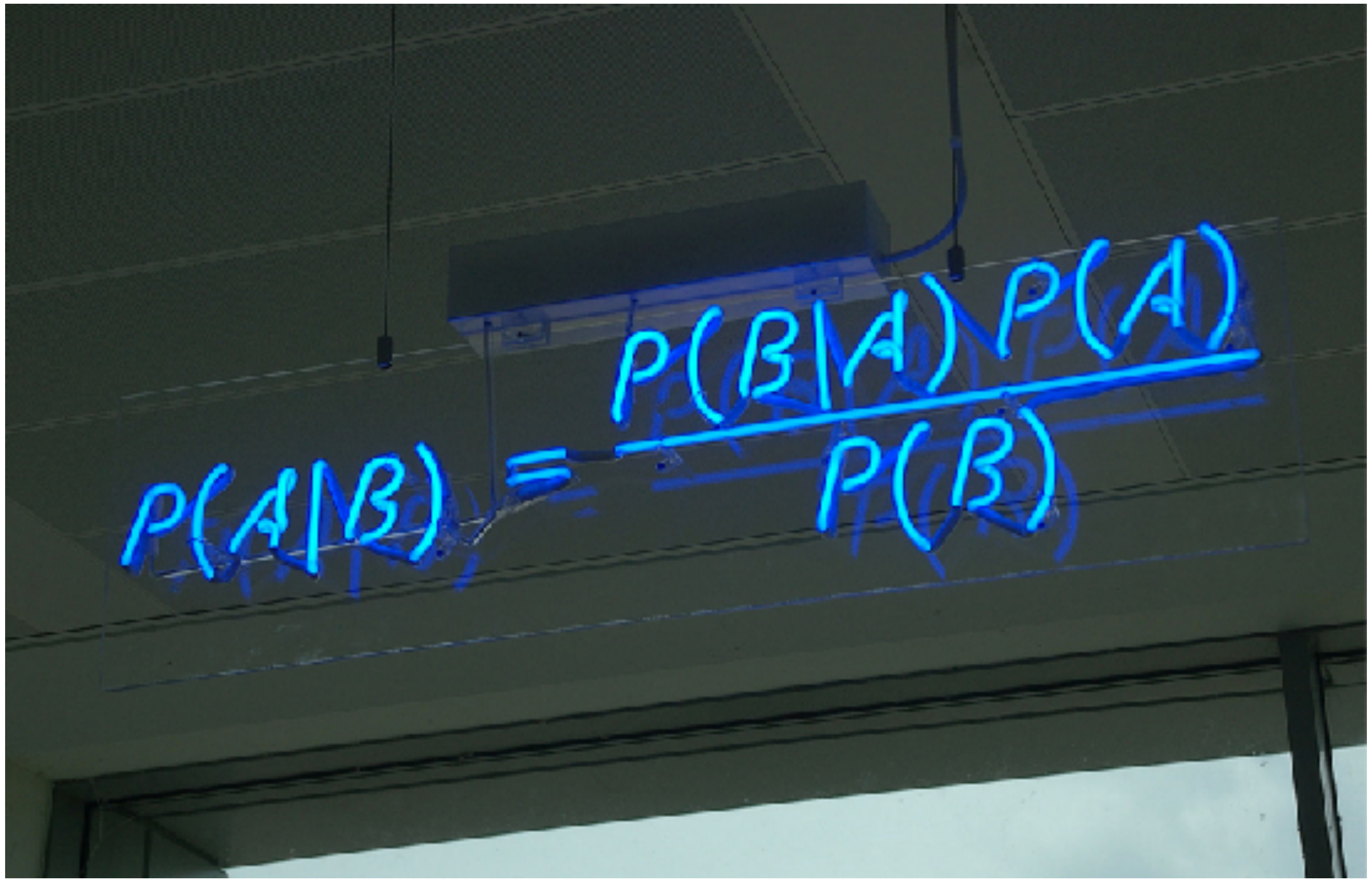
CO3091 - Computational Intelligence and Software Engineering

Lecture 19



Naïve Bayes — Part II

Leandro L. Minku

Overview

- Recap on Naïve Bayes for categorical attributes
- Naïve Bayes for numerical (continuous) input attributes
- Advantages and disadvantages
- Applications

Naïve Bayes Predictive Model

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2	0	2
!Like	1	2	3
Total:	3	2	5

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3	0	3
Short	0	2	2
Total:	3	2	5

Zero Frequency

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

P(Male | !Like,Long) = 0%

P(Long | Male) = 0%

Laplace Smoothing

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Use smoothing when computing $P(F_i|C)$.

Still use the original frequency tables for computing P(C).

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	0+1	4
!Like	1+1	2+1	5
Total:	5	4	9

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3+1	0+1	5
Short	0+1	2+1	4
Total:	5	4	9

Laplace Smoothing

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Gender = Female	Gender = Male	Total:
3	2	5

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	0+1	4
!Like	1+1	2+1	5
Total:	5	4	9

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3+1	0+1	5
Short	0+1	2+1	4
Total:	5	4	9

Naïve Bayes

Assuming that each input attribute is conditionally independent of all other input attributes given the output and making some simplifications:

$$P(C|F_1,\ldots,F_n) = P(C)\,P(F_1|C)\,P(F_2|C)\,\ldots\,P(F_n|C)$$

$$P(C|F_1,\ldots,F_n) = P(C)\prod_{i=1}^{n} P(F_i|C)$$

Naïve bayes predicts the class with the maximum $P(C|F_1,\ldots,F_n)$.

Making Predictions

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	0+1	4
!Like	1+1	2+1	5
Total:	5	4	9

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3+1	0+1	5
Short	0+1	2+1	4
Total:	5	4	9

Gender = Female	Gender = Male	Total:
3	2	5

(!Like, Long, Gender = ?)

P(Female | !Like,Long) =

= P(Female) P(!Like|Female) P(Long|Female)

= 3/5 * 2/5 * 4/5 = 19.2%

Predicted class: Gender = Female

P(Male | !Like,Long) =

= P(Male) P(!Like|Male) P(Long|Male)

= 2/5 * 3/4 * 1/4 = 7.5%

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

Categorical vs Numeric Attributes

Categorical attributes: the frequency tables are created by counting occurrences of each possible input and output attribute value among the examples in the training set (and summing one).

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	0+1	4
!Like	1+1	2+1	5
Total:	5	4	9

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3+1	0+1	5
Short	0+1	2+1	4
Total:	5	4	9

Gender = Female	Gender = Male	Total:
3	2	5

Categorical vs Numeric Attributes

Numerical attributes: e.g.: input attribute height.

- There are potentially infinite possible attribute values.

Frequency Table for Height	Gender = Female	Gender = Male	Total:
1.00m			
1.01m			
1.02m			
1.03m			
…			
Total:			

Dealing with Numeric Input Attributes — Discretising Values

- Transform numerical input attributes into categories (i.e., discrete values).
- E.g.: height —> short, medium, tall.
- However, one person may consider tall as >1.70m. Another may consider it as >1.90m.
- Moreover, if a tall person is a person with > 1.90m, should a person with 1.89m be considered medium height? Not tall?
- This level of subjectiveness and crispiness may lead to loss of information.

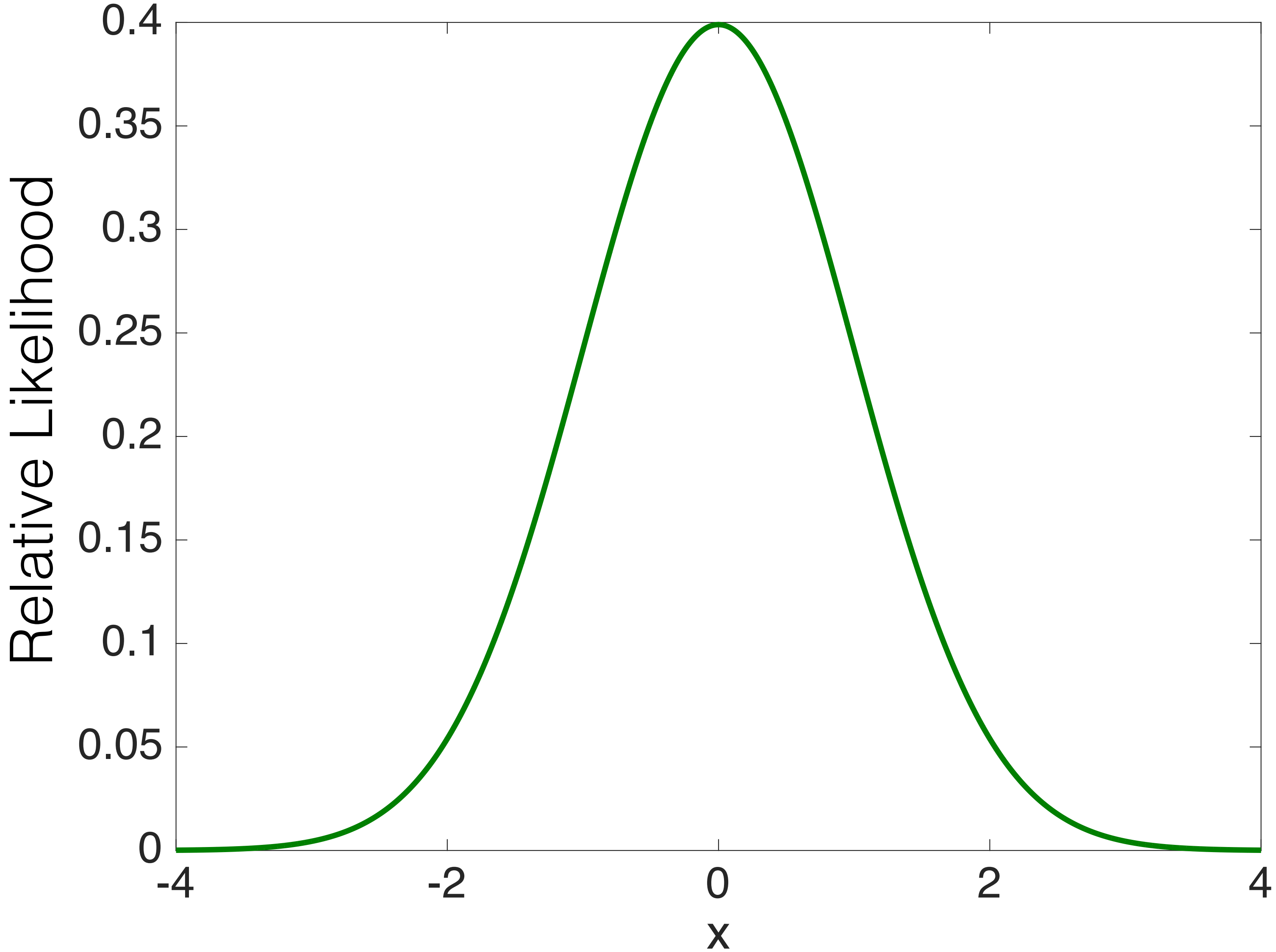
Dealing with Numeric Input Attributes — Probability Density Functions

We assume that examples are drawn from probability distributions.

To determine $P(F|C)$ for numerical input attributes, we could learn parameters for a probability density function instead of counting frequencies.

Dealing with Numeric Input Attributes — Probability Density Functions

Normal Distribution (μ, σ^2)

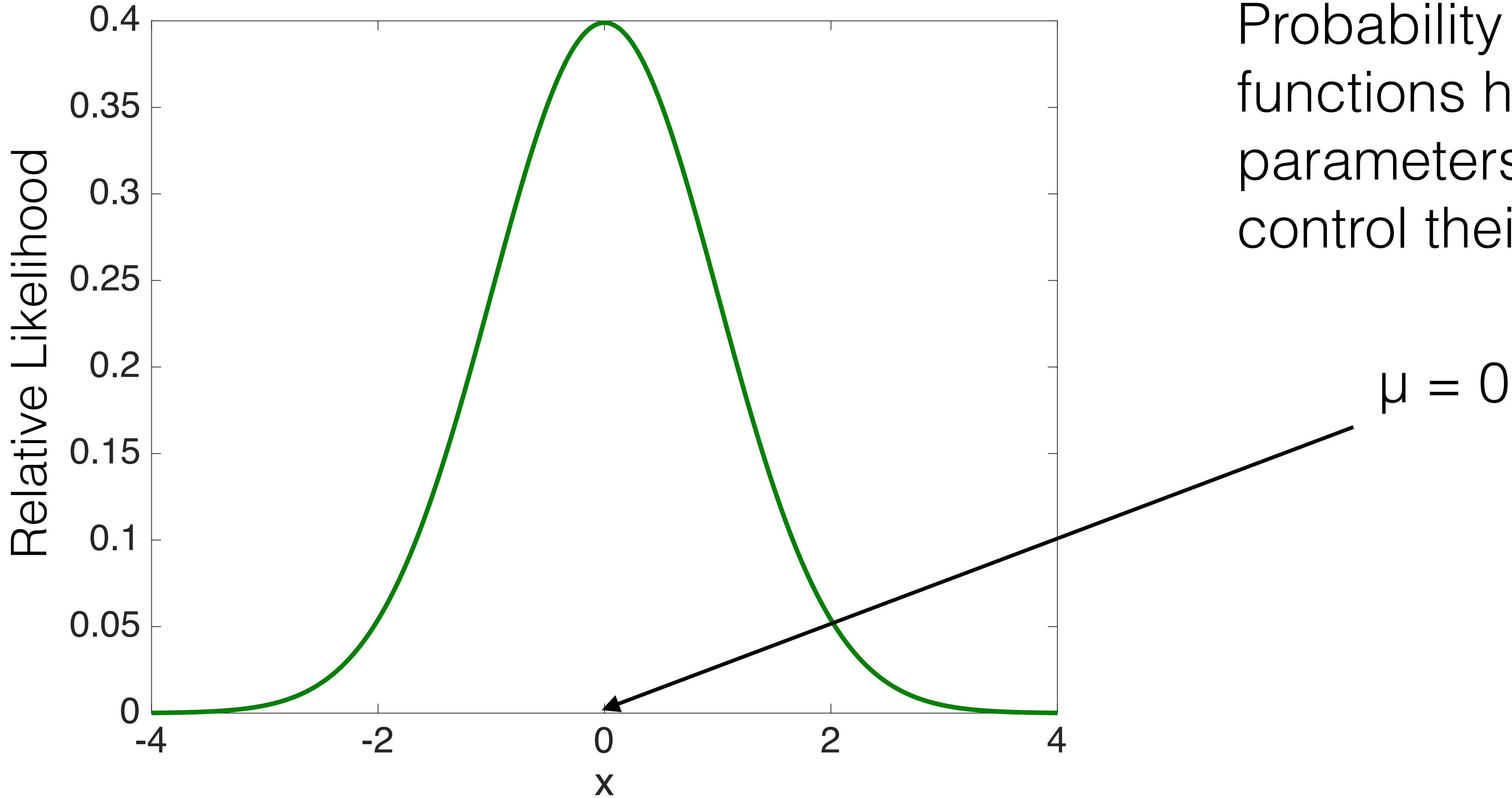


x axis represents values of the input attribute.

y axis represents the [relative] likelihood of observing such values given a certain class.

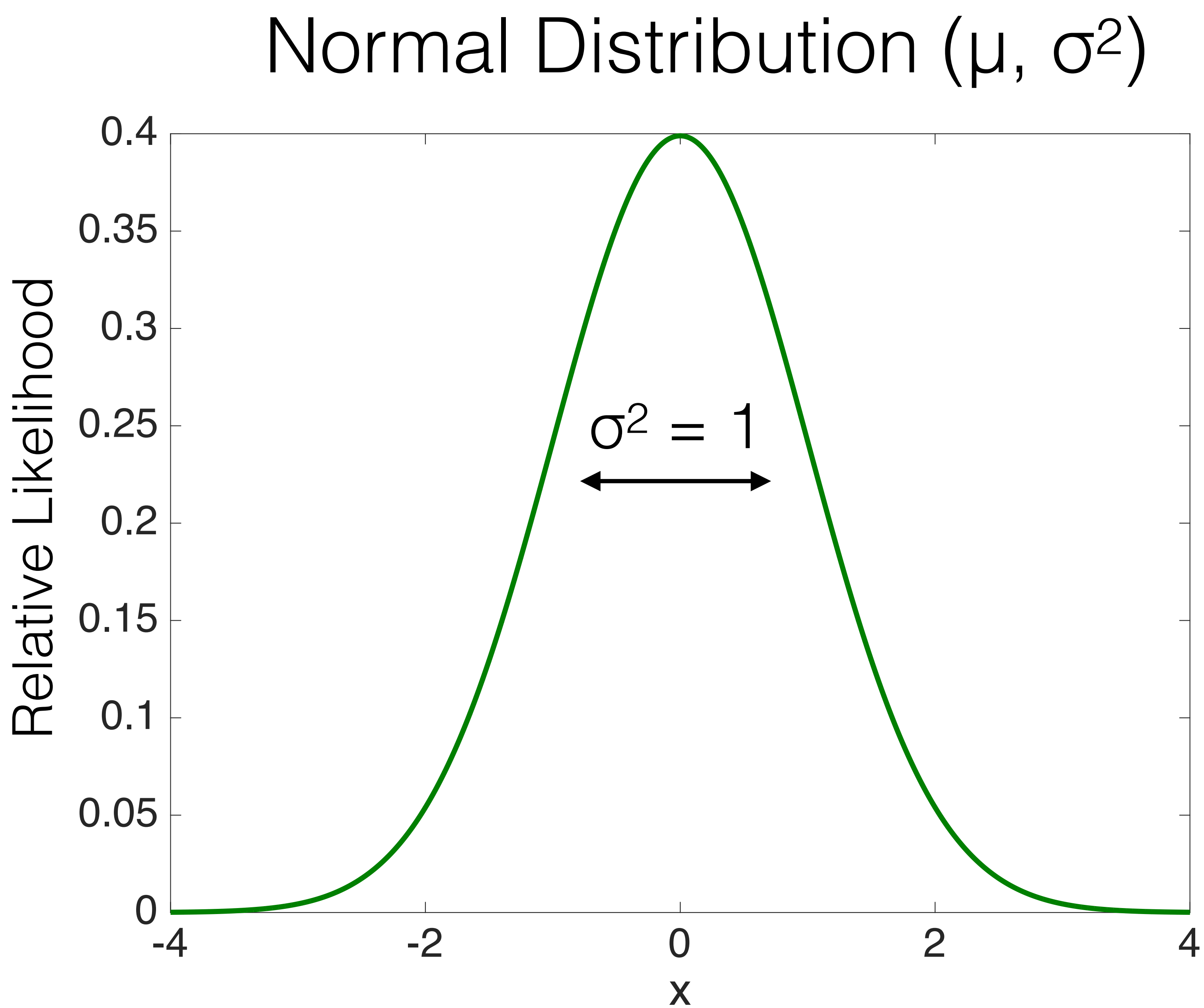
Parameters of Probability Density Functions

Normal Distribution (μ, σ^2)



Probability density functions have parameters that control their shape.

Parameters of Probability Density Functions



Probability density functions have parameters that control their shape.

In order to learn P(F|C), we can learn such parameters.

General Idea

1. Decide on a type of probability density function for each numeric input attribute.

2. For each different output attribute value (class) and numeric input attribute, choose parameters for these functions.

3. Use these functions whenever we need to get a probability P(F|C) for the corresponding numeric input attributes.

General Idea

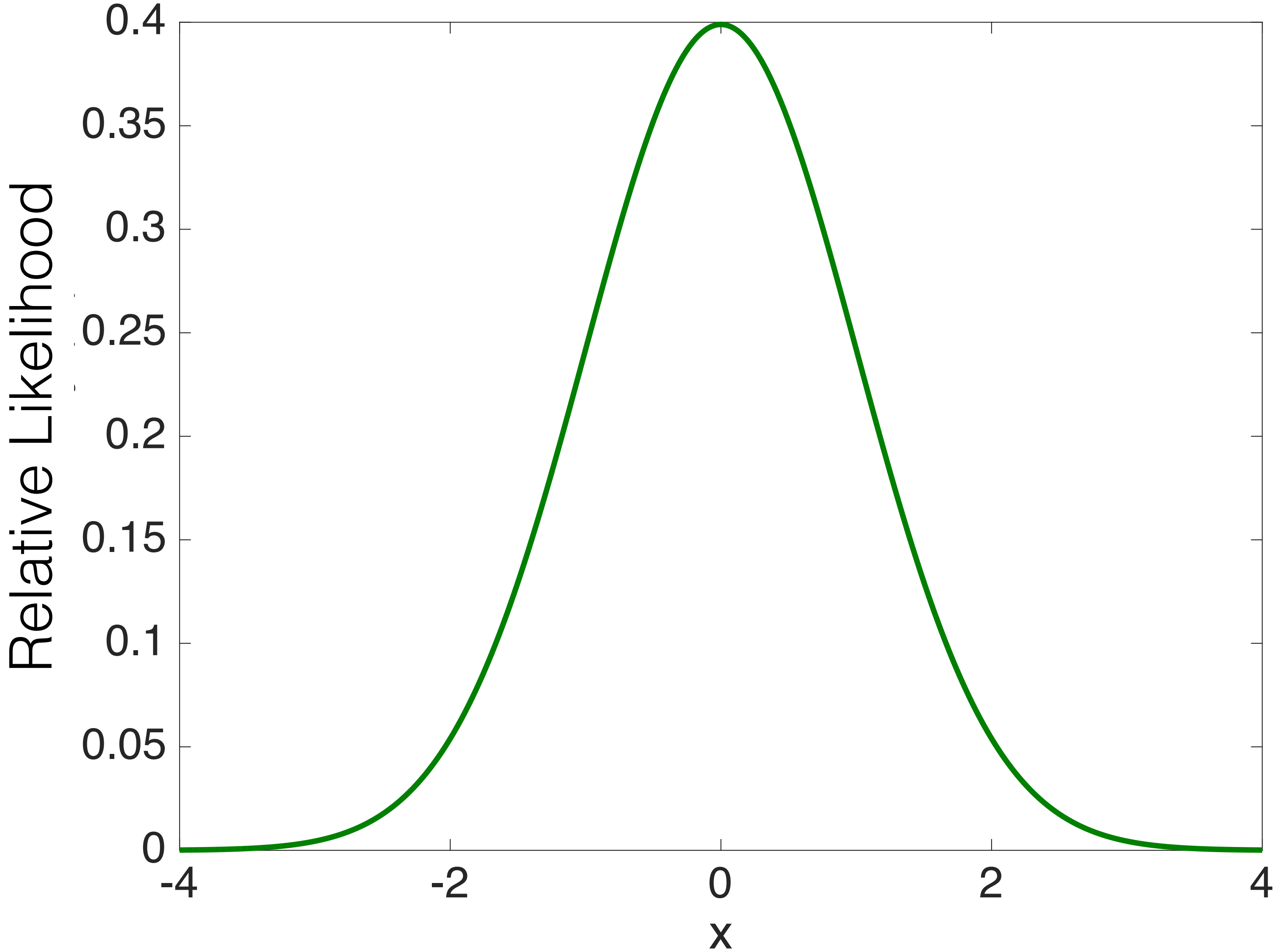
1. **Decide on a type of probability density function for each numeric input attribute.**

2. For each different output attribute value (class) and numeric input attribute, choose parameters for these functions.

3. Use these functions whenever we need to get a probability P(F|C) for the corresponding numeric input attributes.

1.Decide on a type of probability density function for each input attribute

Normal Distribution (μ, σ^2)



Typically normal (gaussian) distribution is adopted.

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\pi \approx 3.14159$
$e \approx 2.71828$

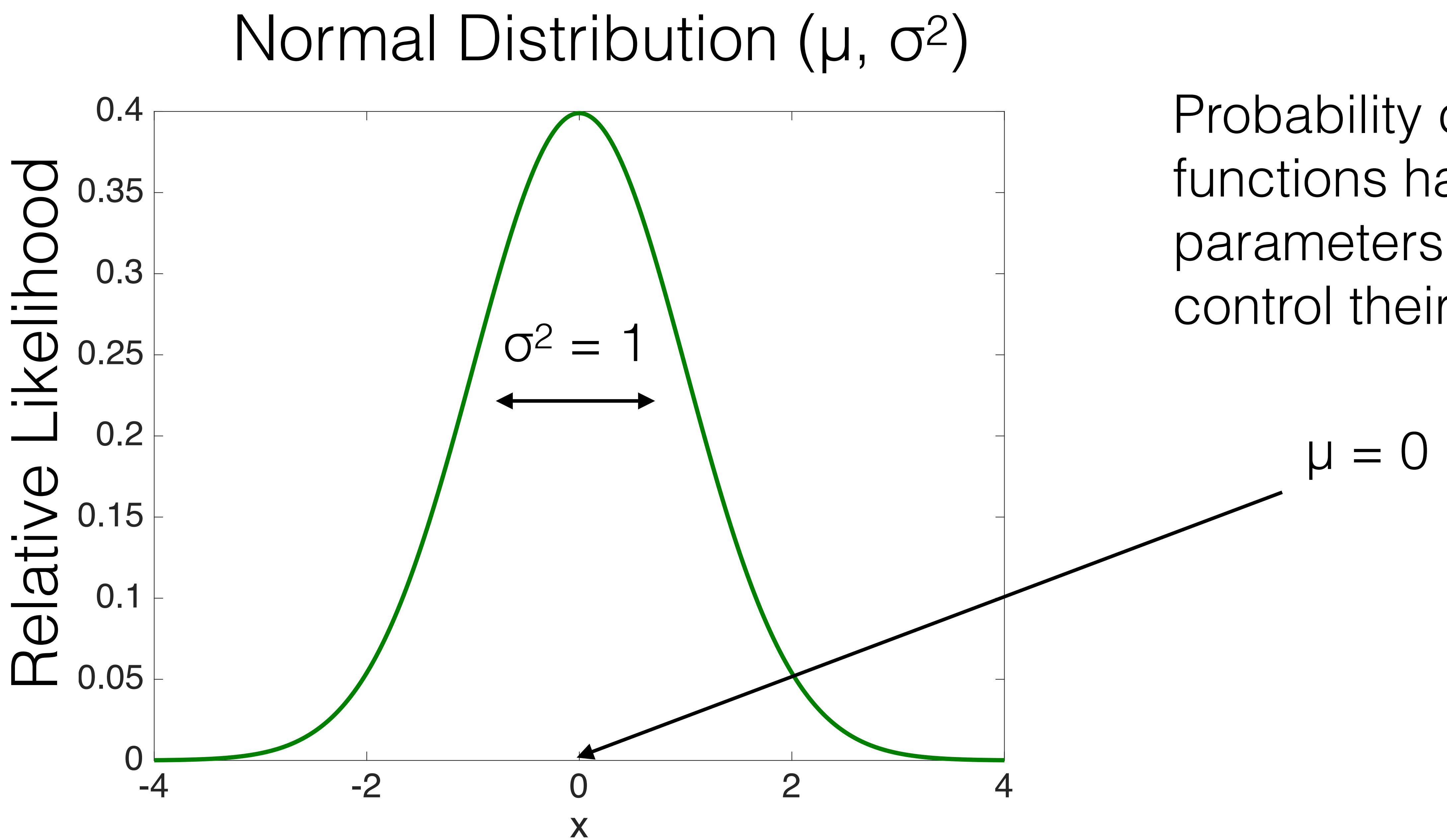
General Idea

1. Decide on a type of probability density function for each numeric input attribute.

2. **For each different output attribute value (class) and numeric input attribute, choose parameters for these functions.**

3. Use these functions whenever we need to get a probability $P(F|C)$ for the corresponding numeric input attributes.

Parameters of Probability Density Functions



Probability density functions have parameters that control their shape.

Example

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

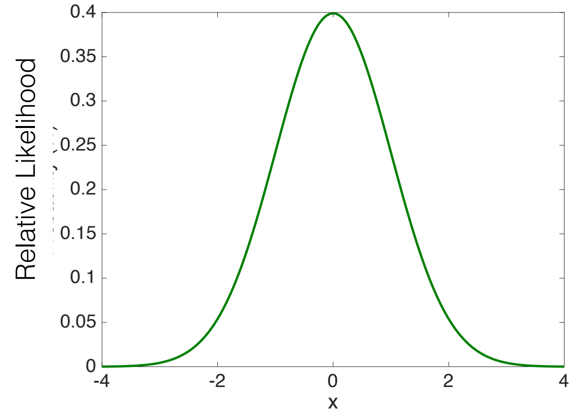
Consider that we have chosen to use a normal probability density function for Height.

Example

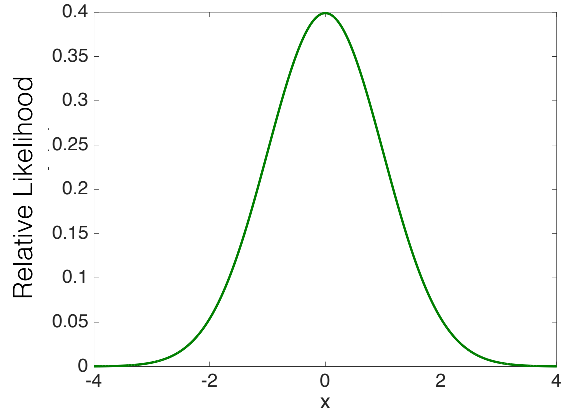
Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

We will create one normal probability density function for Height when Gender = Female and one for when Gender = Male.

Female



Male



We need to choose the parameters μ and σ^2 for each of them.

Example — Probability Density Function for Height when Gender = Female

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

μ = mean of female heights,
σ^2 = variance of female heights

$$\mu = \frac{1.65 + 1.85 + 1.70}{3} = 1.73$$

Example — Probability Density Function for Height when Gender = Female

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

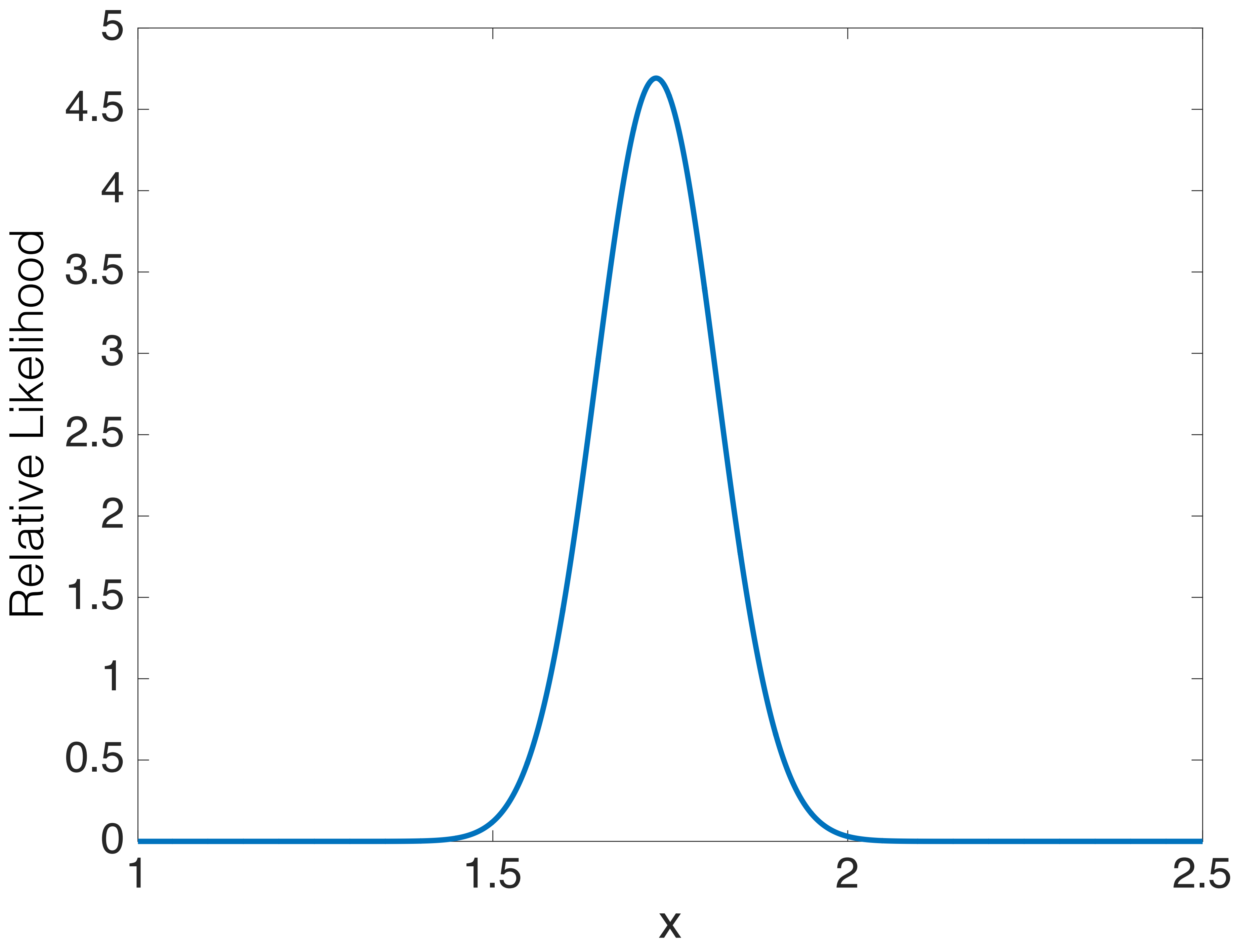
μ = mean of female heights,
σ^2 = variance of female heights

$$\text{Variance(values)} = \frac{1}{|\text{values}|} \sum_{\text{value}_i \text{ in values}} [\text{value}_i - \text{mean(values)}]^2$$

$$\sigma^2 = \frac{1}{3} [(1.65 - 1.73)^2 + (1.85 - 1.73)^2 + (1.70 - 1.73)^2]$$

$$\approx \frac{1}{3} [0.0064 + 0.0144 + 0.0009] \approx 0.00723$$

Example — Probability Density Function for Height when Gender = Female



$\mu = 1.73$

$\sigma^2 = 0.00723$

Example — Probability Density Function for Height when Gender = Male

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

μ = mean of male heights,
σ^2 = variance of male heights

$\mu = ?, \sigma^2 = ?$

Example — Probability Density Function for Height when Gender = Male

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

μ = mean of male heights,
σ^2 = variance of male heights

$$\mu = \frac{1.70 + 1.90 + 1.95}{3} = 1.85$$

Example — Probability Density Function for Height when Gender = Male

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

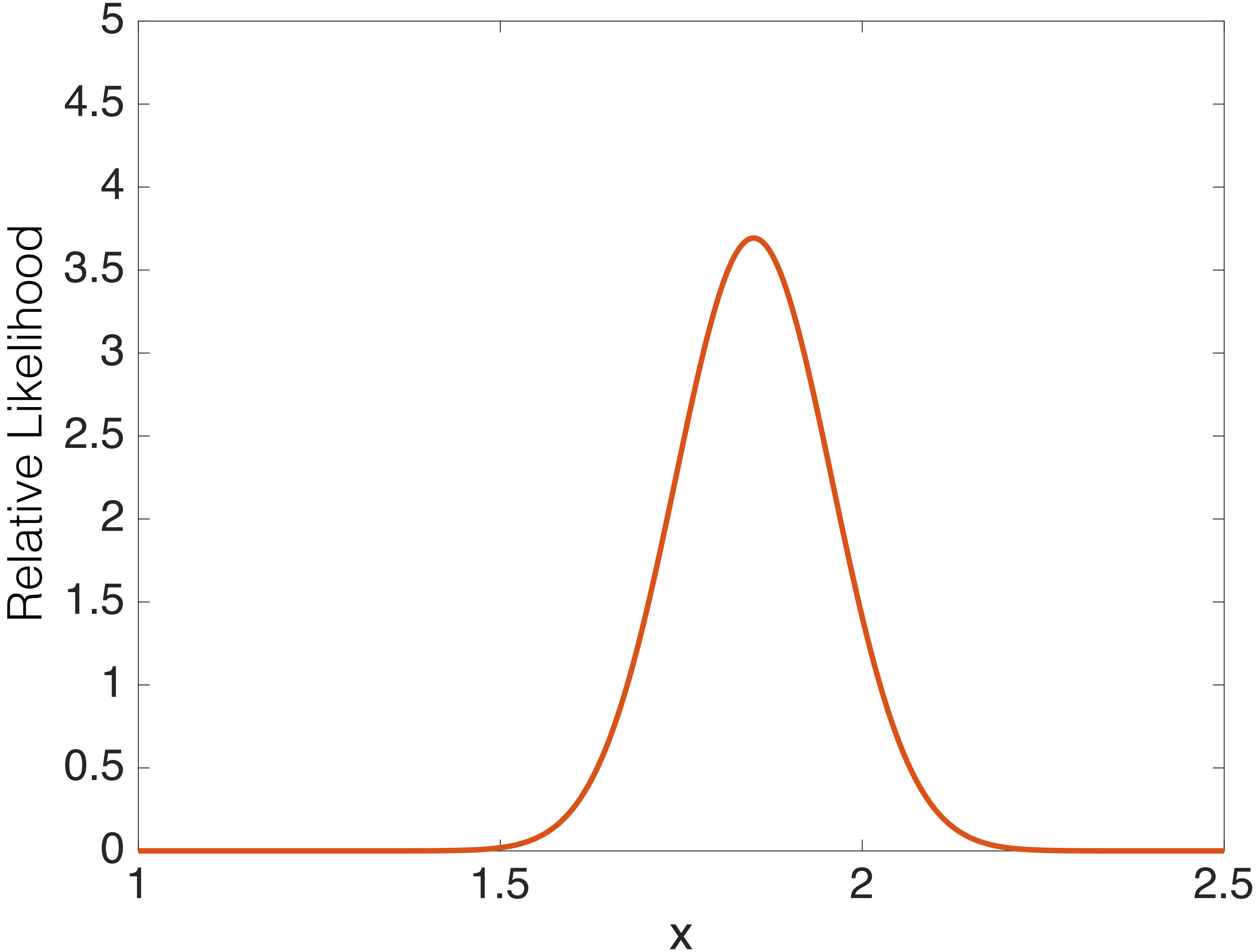
μ = mean of male heights,
σ^2 = variance of male heights

$$\text{Variance(values)} = \frac{1}{|\text{values}|} \sum_{\text{value}_i \text{ in values}} [\text{value}_i - \text{mean(values)}]^2$$

$$\sigma^2 = \frac{1}{3} [(1.70 - 1.85)^2 + (1.90 - 1.85)^2 + (1.95 - 1.85)^2]$$

$$\approx \frac{1}{3} [0.0225 + 0.0025 + 0.01] \approx 0.01167$$

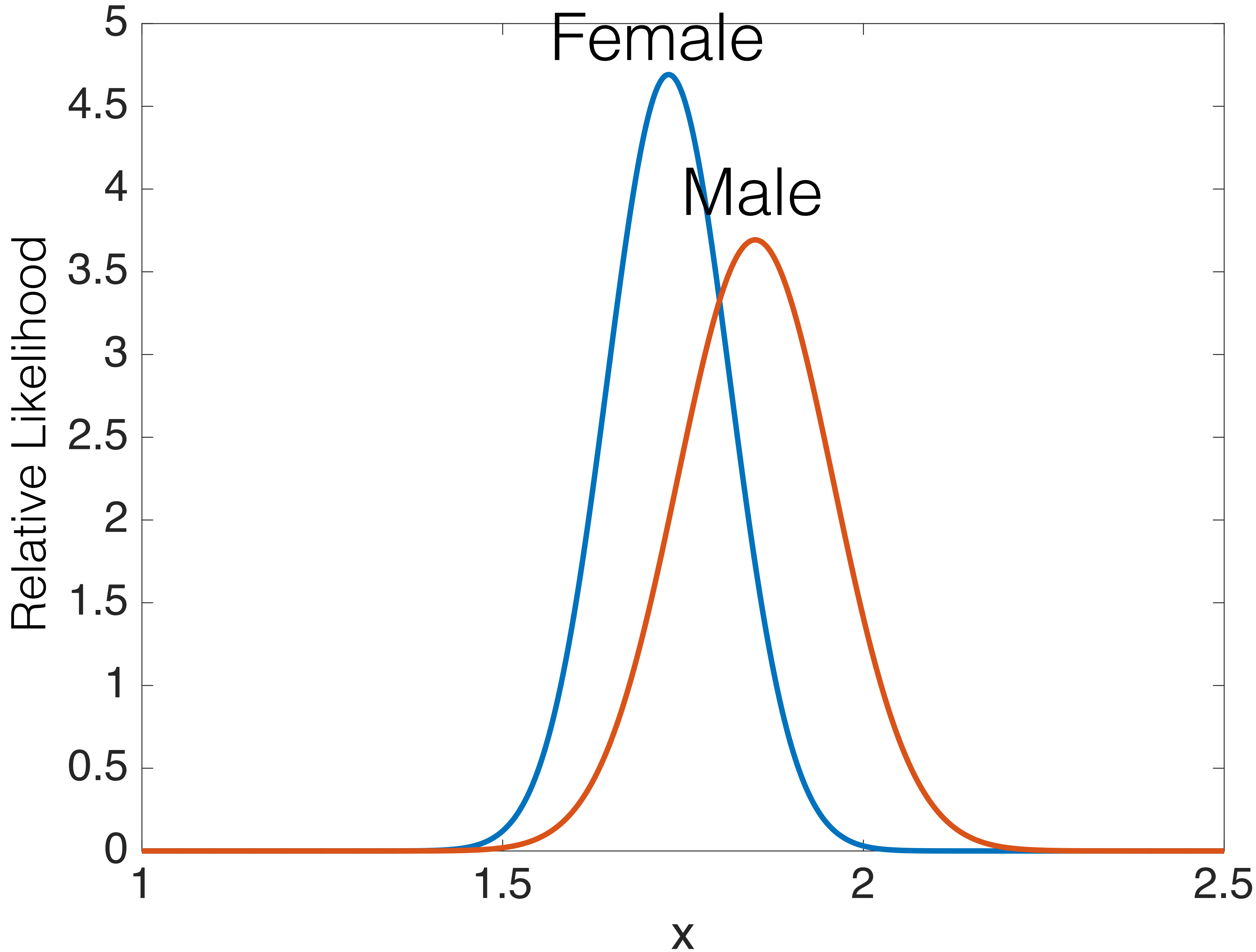
Probability Density Function for Height When Gender = Male



$\mu = 1.85$

$\sigma^2 = 0.01167$

Probability Density Function for Height When Gender = Male



Naïve Bayes Model — Categorical + Numerical Input Attributes

Training Set

Person	x_1 (Flowers)	x_2 (Height)	y (Gender)
P1	Likes	1.65	Female
P2	Likes	1.85	Female
P3	!Like	1.70	Female
P4	!Like	1.70	Male
P5	!Like	1.90	Male
P6	Like	1.95	Male

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	1+1	5
!Like	1+1	2+1	5
Total:	5	5	10

Parameters Table for Height	Gender = Female	Gender = Male
μ	1.73	1.85
σ^2	0.00723	0.01167

Gender = Female	Gender = Male	Total:
3	3	6

General Idea

1. Decide on a type of probability density function for each numeric input attribute.

2. For each different output attribute value (class) and numeric input attribute, choose parameters for these functions.

3. **Use these functions whenever we need to get a probability P(F|C) for the corresponding numeric input attributes.**

Making Predictions

Naïve bayes predicts the class with the maximum $P(C|F_1,\ldots,F_n)$.

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

Take $P(F_i|C)$ for the categorical input attributes from a frequency table.

Take $P(F_i|C)$ for the numerical input attributes from a probability density function.

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	1+1	5
!Like	1+1	2+1	5
Total:	5	5	10

Parameters Table for Height	Gender = Female	Gender = Male
μ	1.73	1.85
σ^2	0.00723	0.01167

Gender = Female	Gender = Male	Total:
3	3	6

Making Predictions — Numerical Input Attributes

Example (Likes, 1.87, Gender = ?)

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

$P(\text{Female}|\text{Likes}, 1.87) =$

$P(\text{Female})\, P(\text{Likes}|\text{Female})\, P(1.87|\text{Female})$

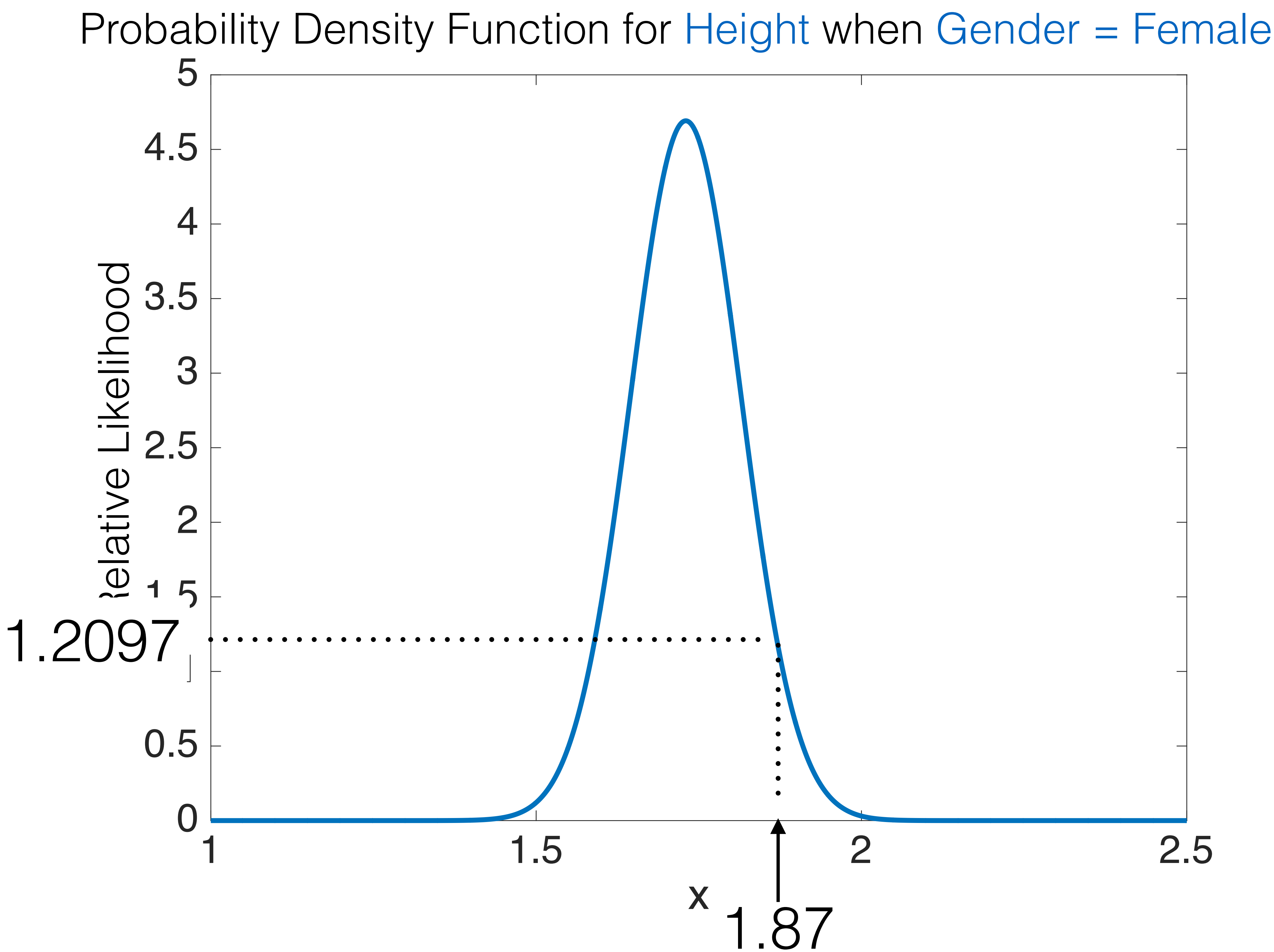
$= 3/6 \; * \; 3/5 \; * \; P(1.87|\text{Female})$

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	1+1	5
!Like	1+1	2+1	5
Total:	5	5	10

Parameters Table for Height	Gender = Female	Gender = Male
μ	1.73	1.85
σ^2	0.00723	0.01167

Gender = Female	Gender = Male	Total:
3	3	6

Getting P(1.87|Female)



Making Predictions — Numerical Input Attributes

Example (Likes, 1.87, Gender = ?)

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

P(Female|Likes, 1.87) =

P(Female) P(Likes|Female) P(1.87|Female)

≈ 3/6 * 3/5 * 1.2097

≈ 0.36291

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	1+1	5
!Like	1+1	2+1	5
Total:	5	5	10

Parameters Table for Height	Gender = Female	Gender = Male
μ	1.73	1.85
σ^2	0.00723	0.01167

Gender = Female	Gender = Male	Total:
3	3	6

Making Predictions — Numerical Input Attributes

Example (Likes, 1.87, Gender = ?)

$$P(C|F_1,\dots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

$P(Male|Likes, 1.87) =$

$= P(Male)\ P(Likes|Male)\ P(1.87|Male)$

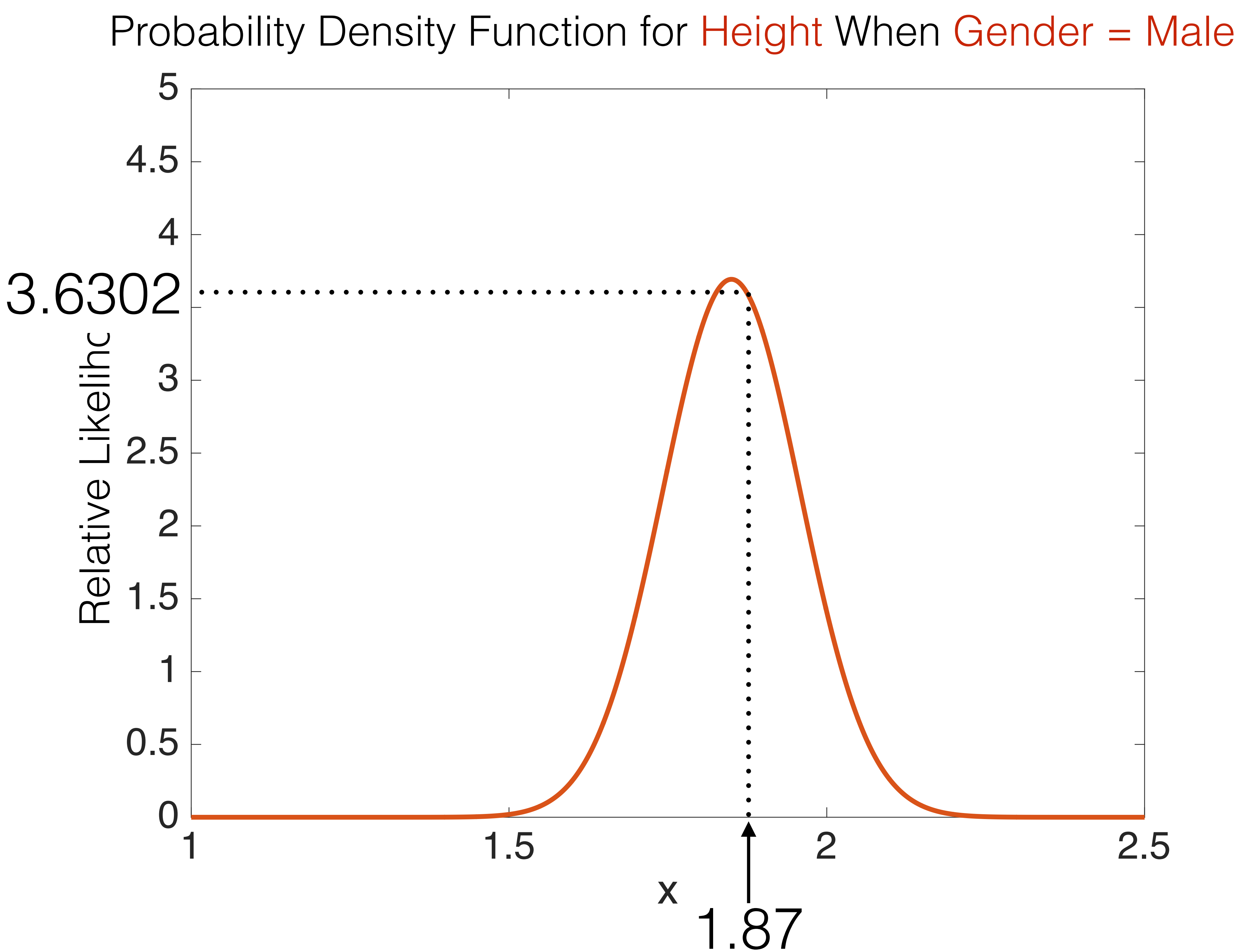
$= 3/6 \ * 2/5 \ * P(1.87|Male)$

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	1+1	5
!Like	1+1	2+1	5
Total:	5	5	10

Parameters Table for Height	Gender = Female	Gender = Male
μ	1.73	1.85
σ^2	0.00723	0.01167

Gender = Female	Gender = Male	Total:
3	3	6

Getting P(1.87|Male)



Making Predictions — Numerical Input Attributes

Example (Likes, 1.87, Gender = ?)

$$P(C|F_1,\dots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

$P(Male|Likes, 1.87) =$

$= P(Male)\ P(Likes|Male)\ P(1.87|Male)$

$\approx 3/6 \quad * 2/5 \quad * 3.6302$

≈ 0.72604

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	1+1	5
!Like	1+1	2+1	5
Total:	5	5	10

Parameters Table for Height	Gender = Female	Gender = Male
μ	1.73	1.85
σ^2	0.00723	0.01167

Gender = Female	Gender = Male	Total:
3	3	6

Making a Prediction

Example (Likes, 1.87, Gender = ?)

$P(\text{Female}|\text{Likes}, 1.87) \approx 0.36291$

$P(\text{Male}|\text{Likes}, 1.87) \approx 0.72604$

Predicted class: gender = male

Naïve Bayes Approach for Classification Problems with Categorical and Numerical Input Attributes

- Naïve Bayes Learning Algorithm:
  - Create frequency tables for categorical input attributes.
  - Apply Laplace Smoothing.
  - Determine density function parameters for numerical input attributes.

- Naïve Bayes Model:
  - Frequency tables (with and without Laplace Smoothing) and tables of density function parameters.

- Naïve Bayes prediction for an instance (**x**,?):
  - Use Bayes Theorem with conditional independence assumption.

Advantages and Disadvantages of Naive Bayes

- Advantages:
  - Training is fast. It needs only one pass through the data, i.e., online learning.
  - Relative probabilities are good for making predictions for many applications.
- Disadvantage:
  - Assumes conditional independence.
  - Assumes a certain probability distribution for numerical input attributes.
  - Does not work very well for regression.

Applications

- Text categorisation, e.g., spam or not spam.
- Medical diagnosis.
- Software defect prediction.
- Etc.

Further Reading

On the relative value of cross-company and within-company data for defect prediction

Burak Turhan, Tim Menzies, Ayşe B. Bener, Justin Di Stefano

Journal of Empirical Software Engineering

Volume 14 Issue 5, October 2009

Section 3.2 (Naïve Bayes Classifier)

http://readinglists.le.ac.uk/lists/D888DC7C-0042-C4A3-5673-2DF8E4DFE225.html