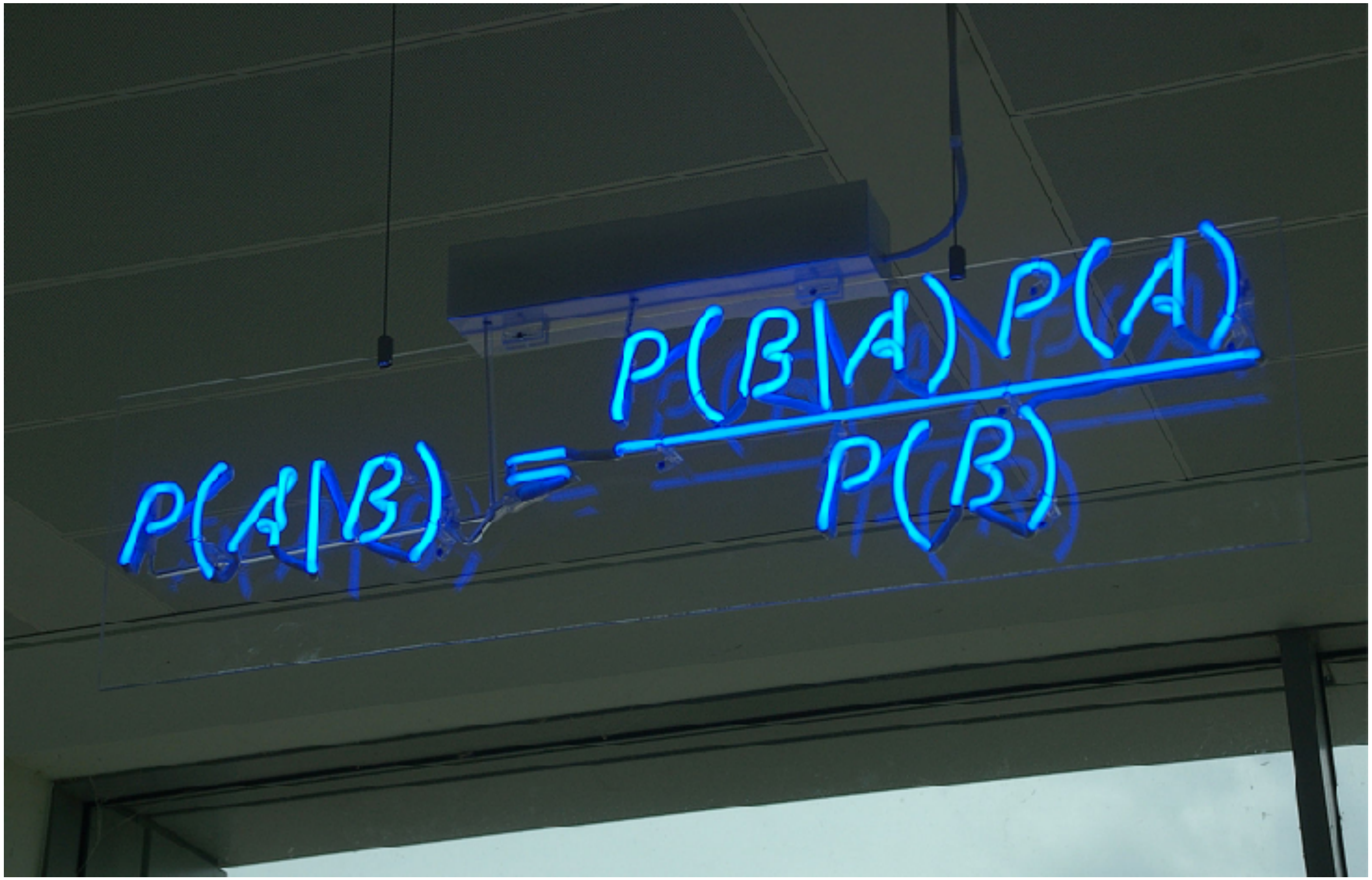


Naïve Bayes - Numeric Independent Variables

Leandro L. Minku

Naïve Bayes

$$P(c|a_1,...,a_n) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

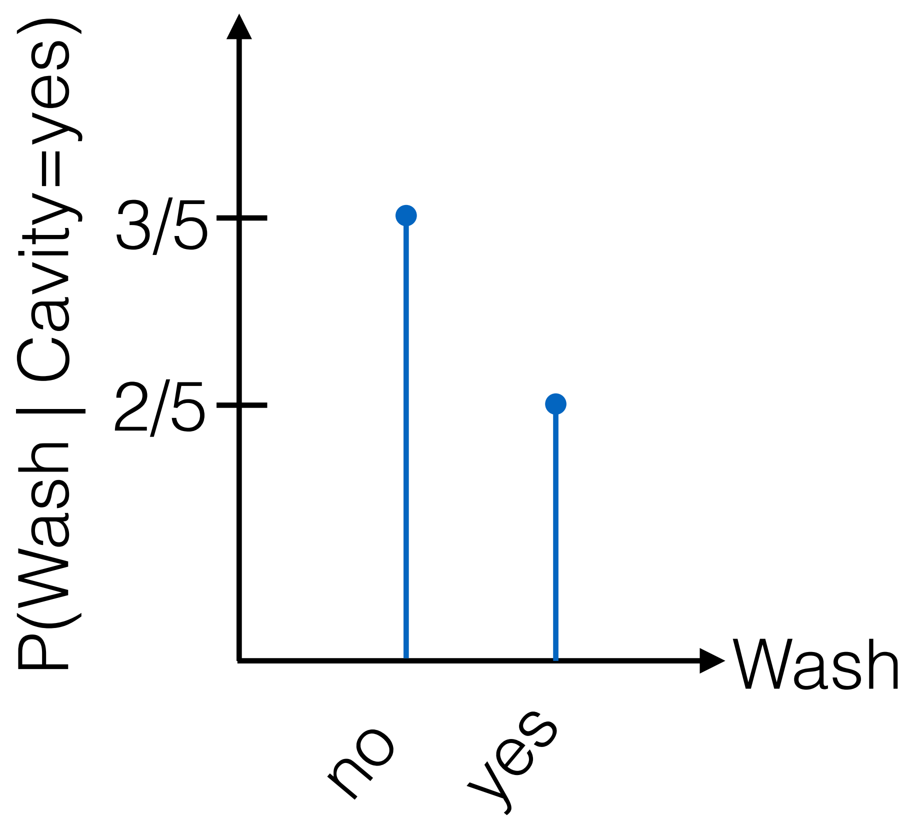
$$\text{where } \alpha = 1 / \beta \text{ and } \quad \beta = \sum_{c \in Y} \left(P(c) \prod_{i=1}^{d} P(a_i|c) \right)$$

Naïve bayes predicts the class with the maximum $P(c|a_1,...,a_n)$.

Dealing with Numeric Independent Variables — Probability Density Functions

We assume that examples are drawn from probability distributions.

Probability Mass Function

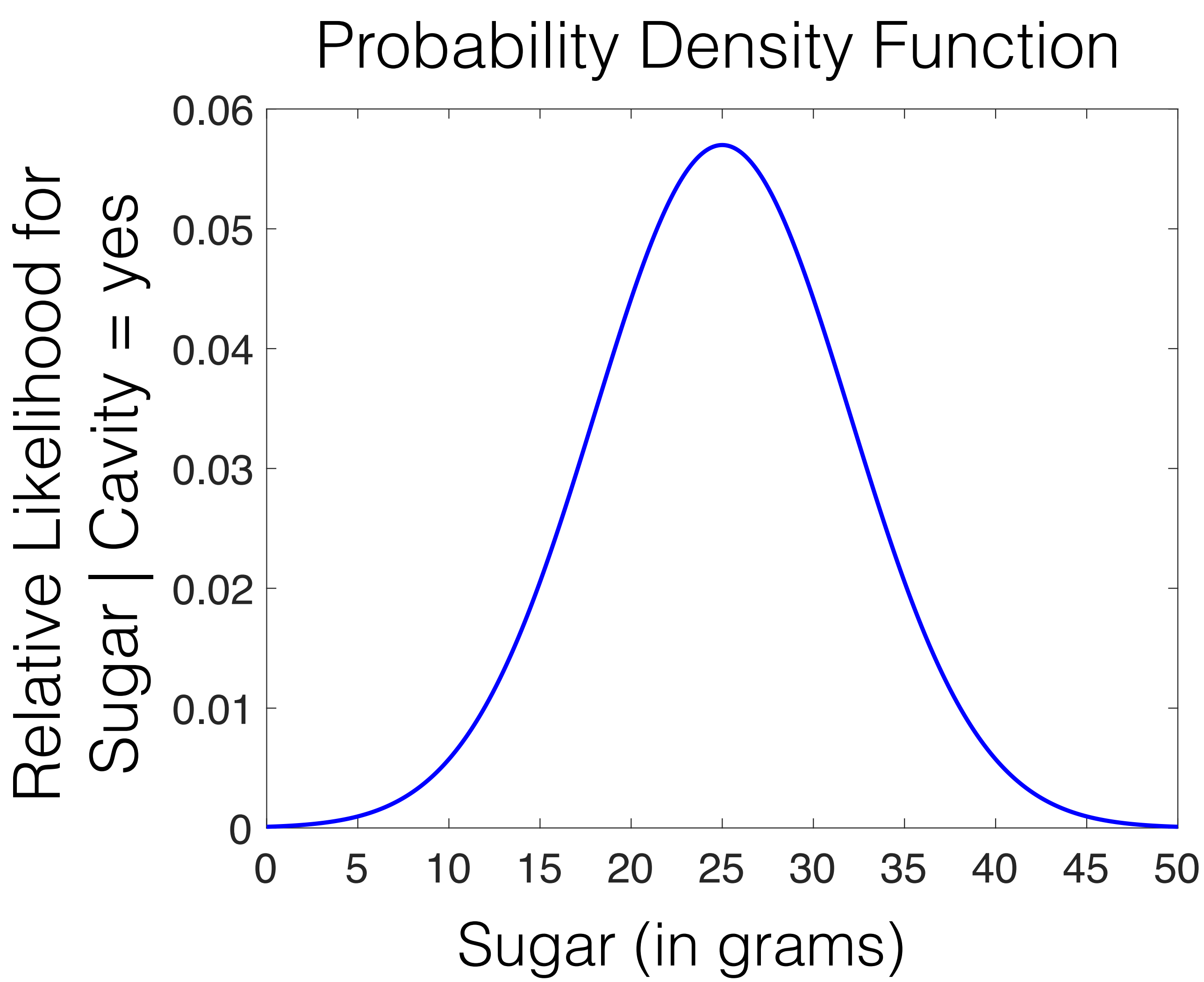


Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Our models were frequency tables that enabled us to compute $P(a_i|c)$ for categorical independent variables.

Dealing with Numeric Independent Variables — Probability Density Functions

We assume that examples are drawn from probability distributions.
Typically, a Gaussian distribution is adopted.



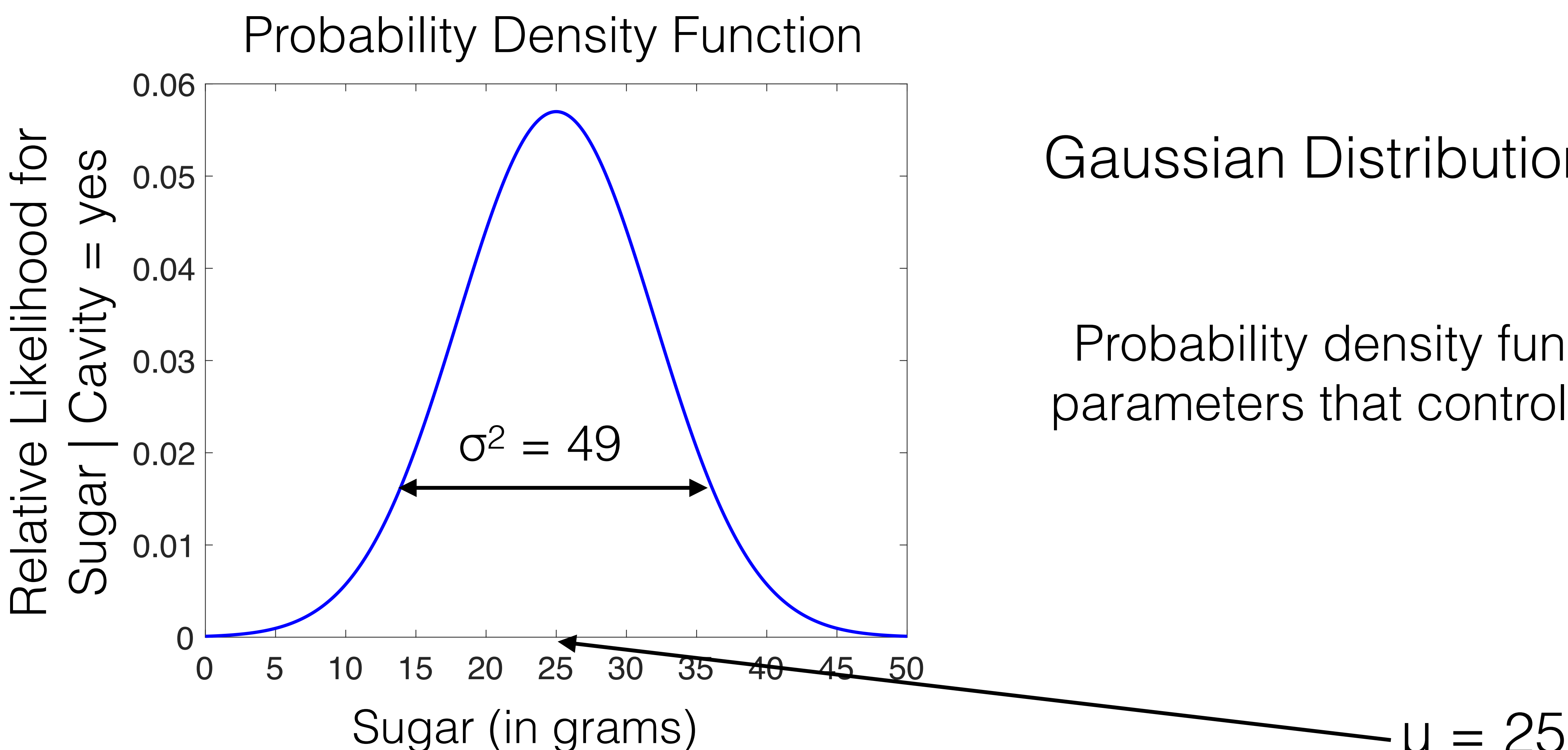
Gaussian Distribution $\mathcal{N}(\mu, \sigma^2)$

$$P(X=a \mid \mu,\sigma^2) = \frac{1}{\sqrt{2\,\sigma^2\,\pi}}\, e^{\frac{-(a-\mu)^2}{2\sigma^2}}$$

$\pi \approx 3.14159$

$e \approx 2.71828$

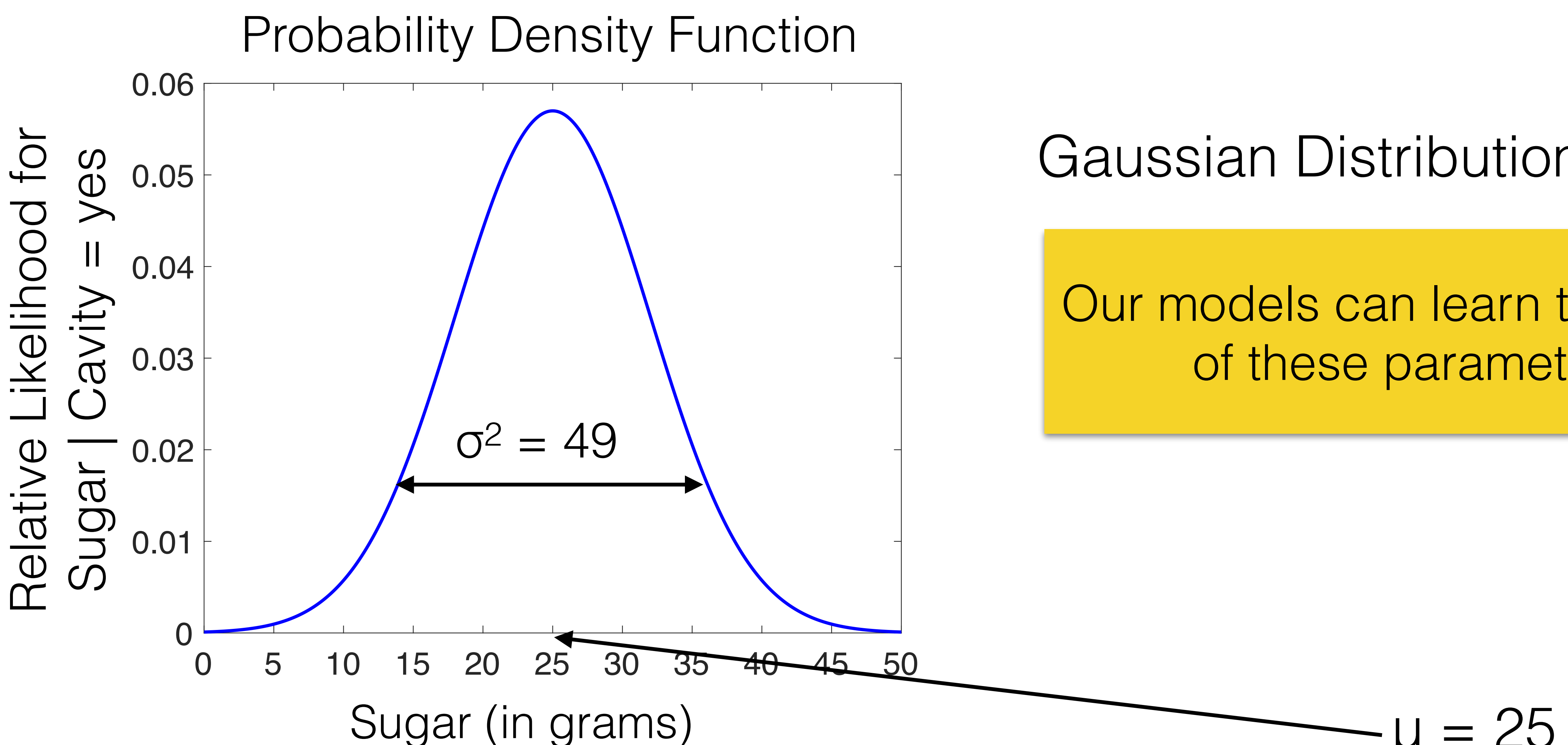
Dealing with Numeric Independent Variables — Probability Density Functions



Gaussian Distribution $\mathcal{N}(\mu, \sigma^2)$

Probability density functions have parameters that control their shape.

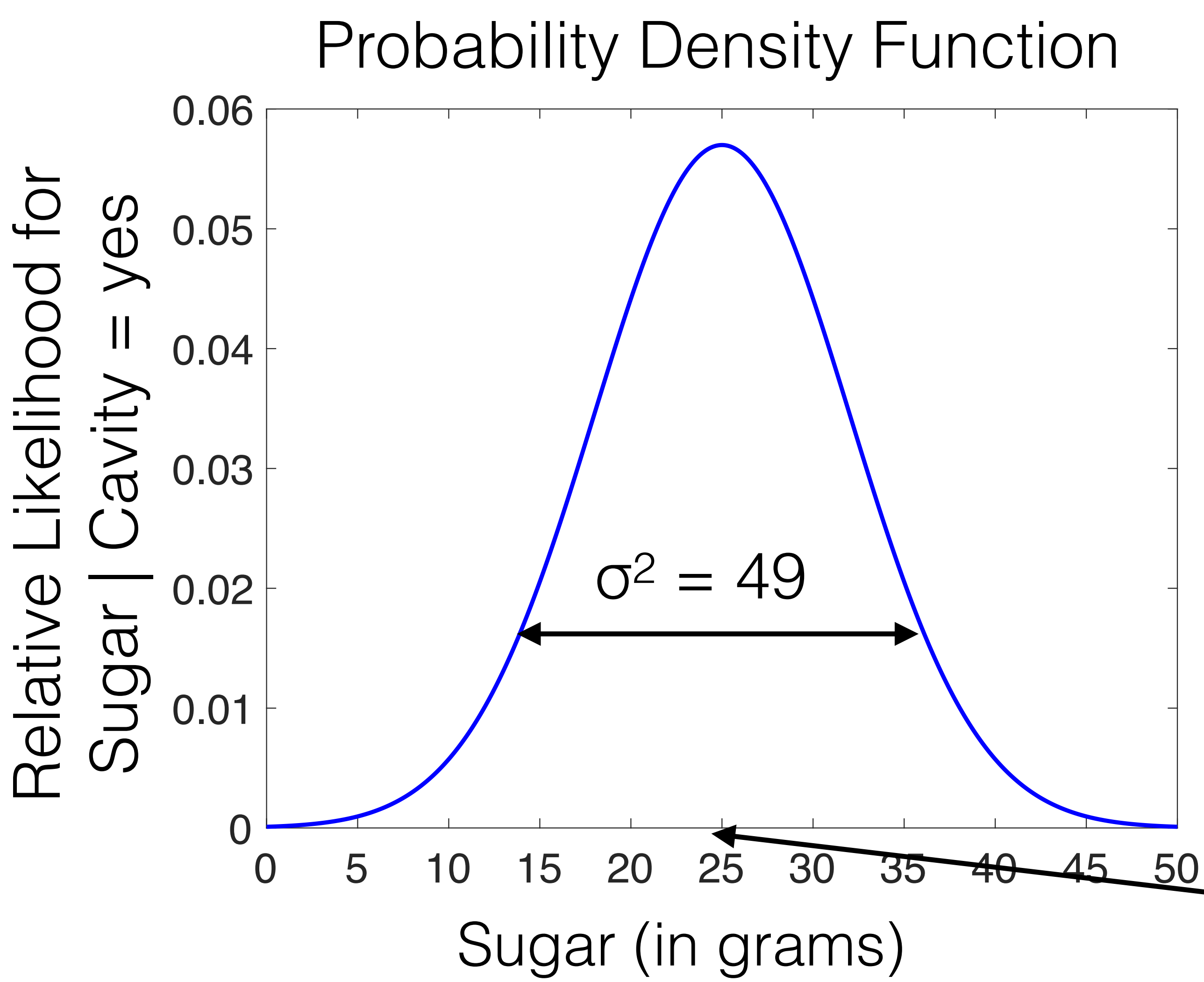
Dealing with Numeric Independent Variables — Probability Density Functions



Gaussian Distribution $\mathcal{N}(\mu, \sigma^2)$

Our models can learn the values of these parameters.

Dealing with Numeric Independent Variables — Probability Density Functions



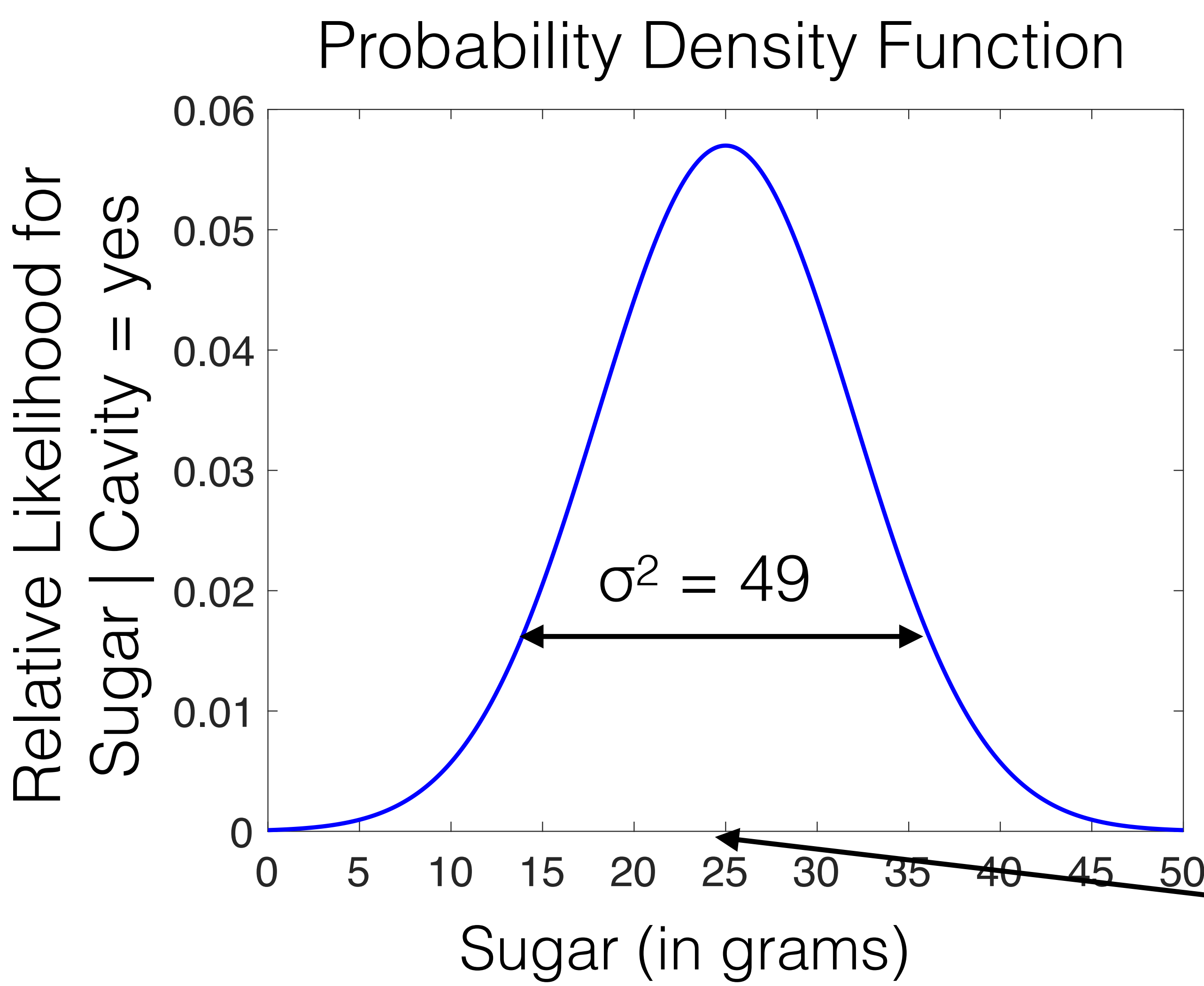
$\mu = 25$

Gaussian Distribution $\mathcal{N}(\mu, \sigma^2)$

We can then use the relative likelihoods as $P(a_i|c)$.

$$P(c|a_1,\ldots,a_n) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

Dealing with Numeric Independent Variables — Probability Density Functions



Gaussian Distribution $\mathcal{N}(\mu, \sigma^2)$

So, we need to compute the parameters of $P(a_i|c)$ for all possible classes.

$$P(c|a_1,\ldots,a_n) = \alpha P(c) \prod_{i=1}^{d} P(a_i|c)$$

$\mu = 25$

Example

Person	x_1 (Wash)	x_2 (Sugar)	y (Cavity)
P1	no	40	yes
P2	no	35	yes
P3	yes	60	yes
P4	yes	20	no
P5	yes	30	no
P6	no	17	no

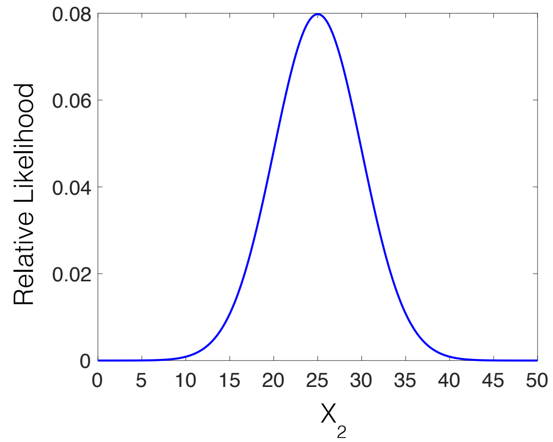
Consider that we have chosen to use a Gaussian probability density function for Sugar.

Example - Computing Parameters for P(Sugar|Cavity)

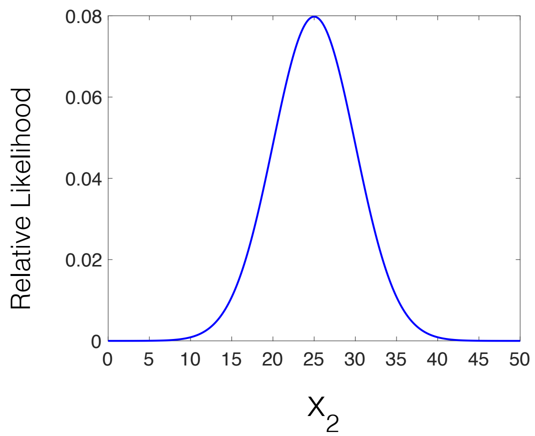
Person	x_1 (Wash)	x_2 (Sugar)	y (Cavity)
P1	no	40	yes
P2	no	35	yes
P3	yes	60	yes
P4	yes	20	no
P5	yes	30	no
P6	no	17	no

We will create one Gaussian probability density function for Sugar when Cavity = yes and one for when Cavity = no.

P(Sugar|Cavity=yes)



P(Sugar|Cavity=no)



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

Example - Computing Parameters for P(Sugar|Cavity=yes)

Person	x_1 (Wash)	x_2 (Sugar)	y (Cavity)
P1	no	40	yes
P2	no	35	yes
P3	yes	60	yes
P4	yes	20	no
P5	yes	30	no
P6	no	17	no

μ = mean of sugar for Cavity=yes
σ^2 = sample variance of sugar for Cavity=yes

$$\mu = \frac{40 + 35 + 60}{3} = 45$$

Example - Computing Parameters for P(Sugar|Cavity=yes)

Person	x_1 (Wash)	x_2 (Sugar)	y (Cavity)
P1	no	40	yes
P2	no	35	yes
P3	yes	60	yes
P4	yes	20	no
P5	yes	30	no
P6	no	17	no

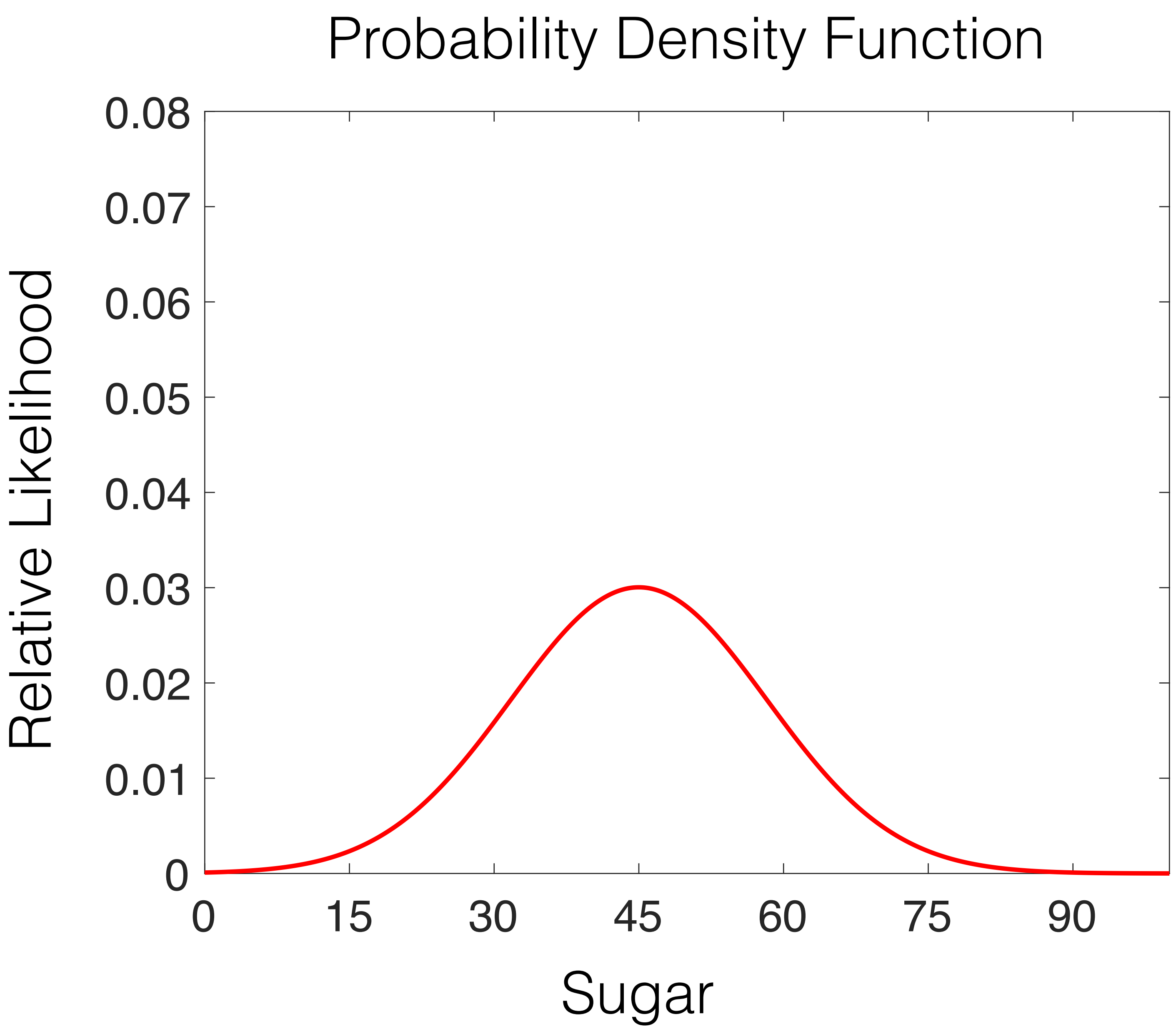
μ = mean of sugar for Cavity=yes
σ^2 = sample variance of sugar for Cavity=yes

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

$$\sigma^2 = \frac{1}{2} [(40 - 45)^2 + (35 - 45)^2 + (60 - 45)^2]$$

$$= \frac{1}{2} [25 + 100 + 225] = 175$$

Example - Computing Parameters for P(Sugar|Cavity=Yes)



$$P(X=a \mid \mu,\sigma^2) = \frac{1}{\sqrt{2\,\sigma^2\,\pi}}\, e^{\frac{-(a-\mu)^2}{2\sigma^2}}$$

$\mu = 45$

$\sigma^2 = 175$

Example - Computing Parameters for P(Sugar|Cavity=no)

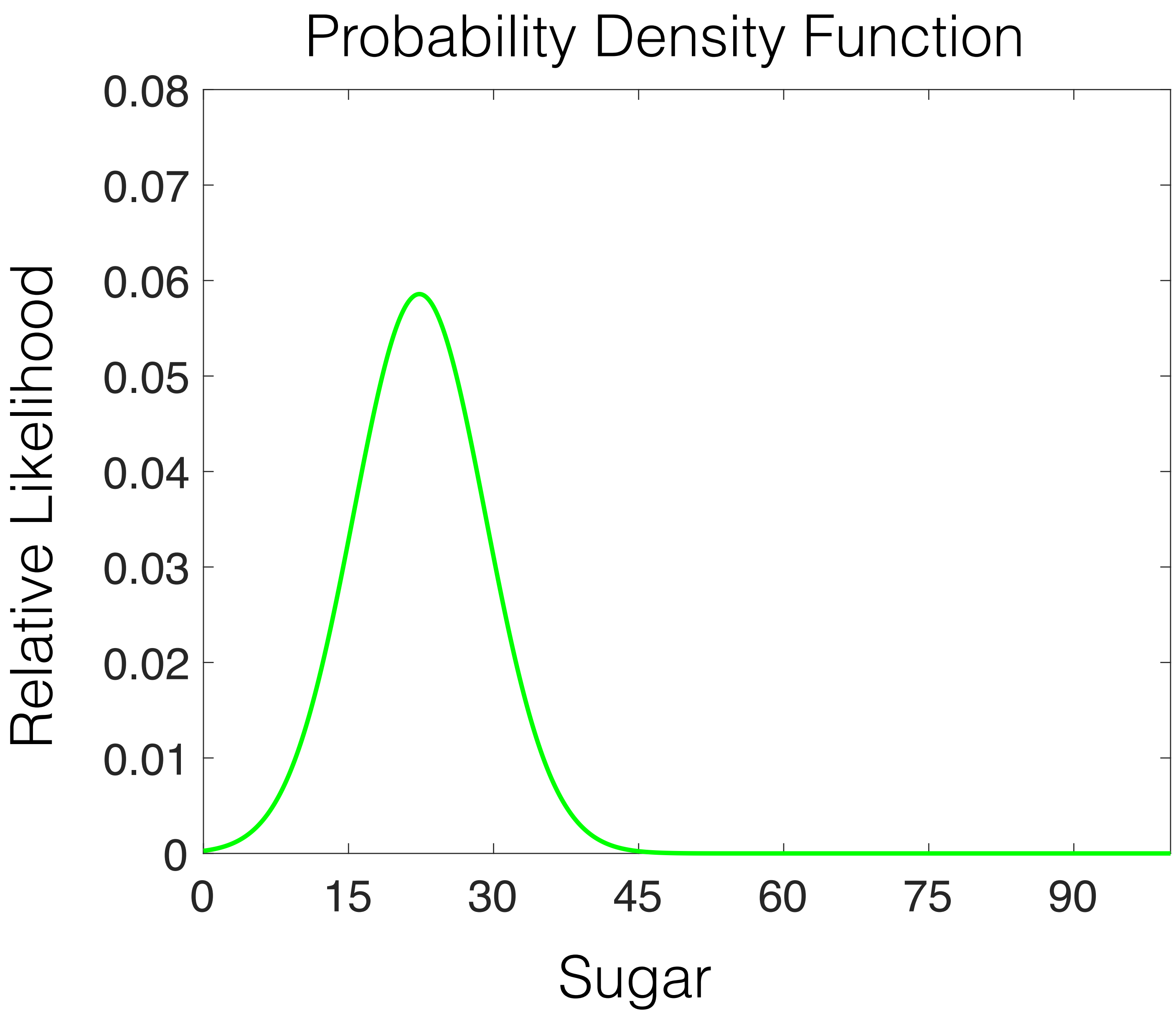
Person	x_1 (Wash)	x_2 (Sugar)	y (Cavity)
P1	no	40	yes
P2	no	35	yes
P3	yes	60	yes
P4	yes	20	no
P5	yes	30	no
P6	no	17	no

μ = mean of sugar for Cavity=no, σ^2 = variance of sugar for Cavity=no

$\mu = 22.33$

$\sigma^2 = 46.34$

Example - Computing Parameters for P(Sugar|Cavity=no)

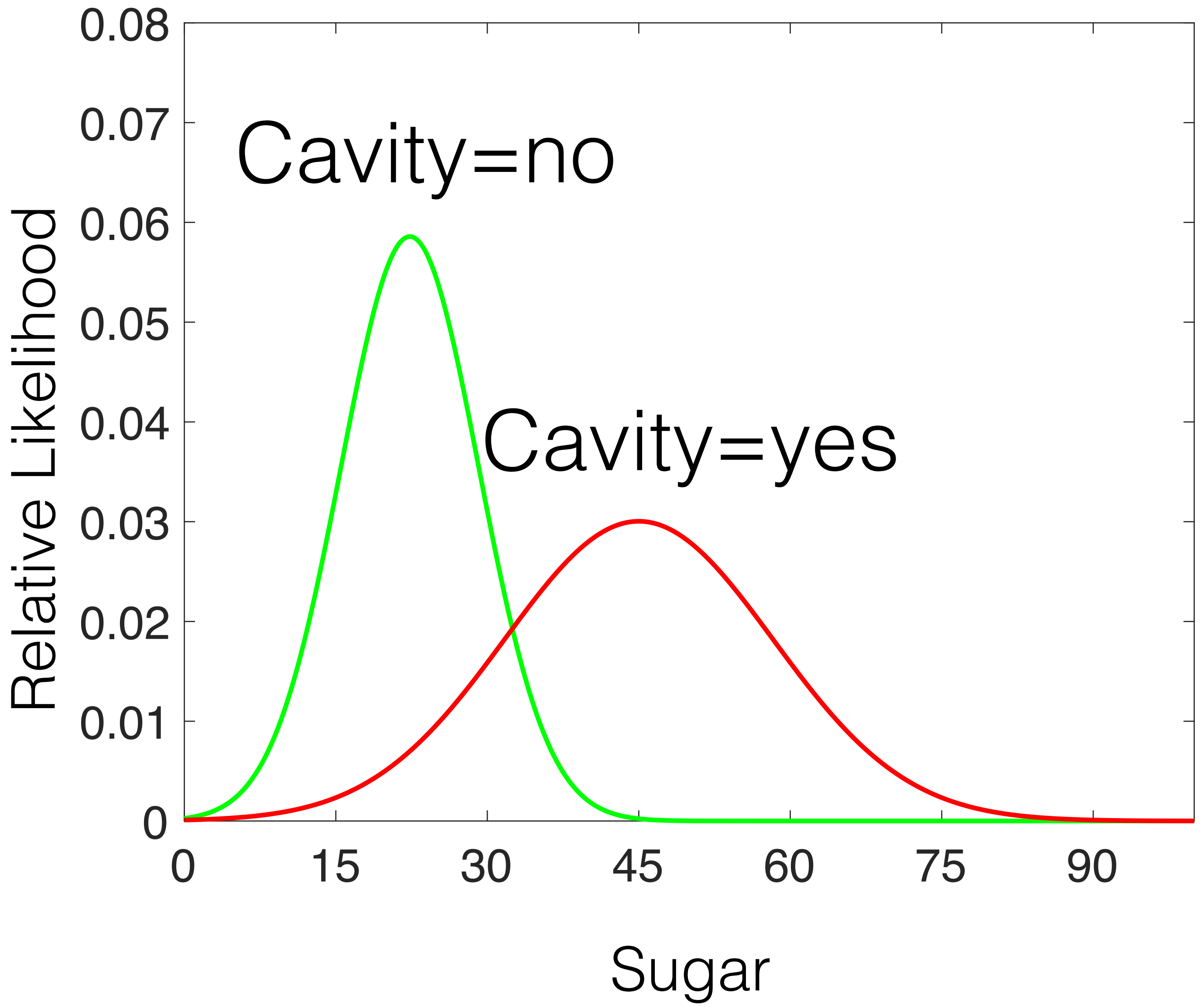


$$P(X=a \mid \mu,\sigma^2) = \frac{1}{\sqrt{2\,\sigma^2\,\pi}}\, e^{\frac{-(a-\mu)^2}{2\sigma^2}}$$

$\mu = 22.33$

$\sigma^2 = 46.34$

Probability Density Functions for Sugar Given Cavity=yes and Cavity=no



Naïve Bayes Model — Categorical + Numerical Independent Variables

Training Set

Person	x_1 (Wash)	x_2 (Sugar)	y (Cavity)
P1	no	40	yes
P2	no	35	yes
P3	yes	60	yes
P4	yes	20	no
P5	yes	30	no
P6	no	17	no

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Parameters Table for Sugar	Cavity = no	Cavity = yes
μ	22.33	45
σ^2	46.34	175

Cavity = no	Cavity = yes	Total:
3	3	6

Model

Making Predictions

Naïve bayes predicts the class with the maximum $P(c|a_1,\dots,a_n)$.

$$P(c|a_1,\dots,a_n) = \alpha P(c) \prod_{i=1}^{d} P(a_i|c)$$

Compute $P(a_i|c)$ for the categorical independent variables from a frequency table.

Compute $P(a_i|c)$ for the numeric independent variables from a probability density function with the learnt parameters.

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Parameters Table for Sugar	Cavity = no	Cavity = yes
μ	22.33	45
σ^2	46.34	175

Cavity = no	Cavity = yes	Total:
3	3	6

Model

Predicting (Wash=no, Sugar=20, Cavity=?)

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Parameters Table for Sugar	Cavity = no	Cavity = yes
μ	22.33	45
σ^2	46.34	175

Cavity = no	Cavity = yes	Total:
3	3	6

Model

$$P(c|a_1,\ldots,a_n) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

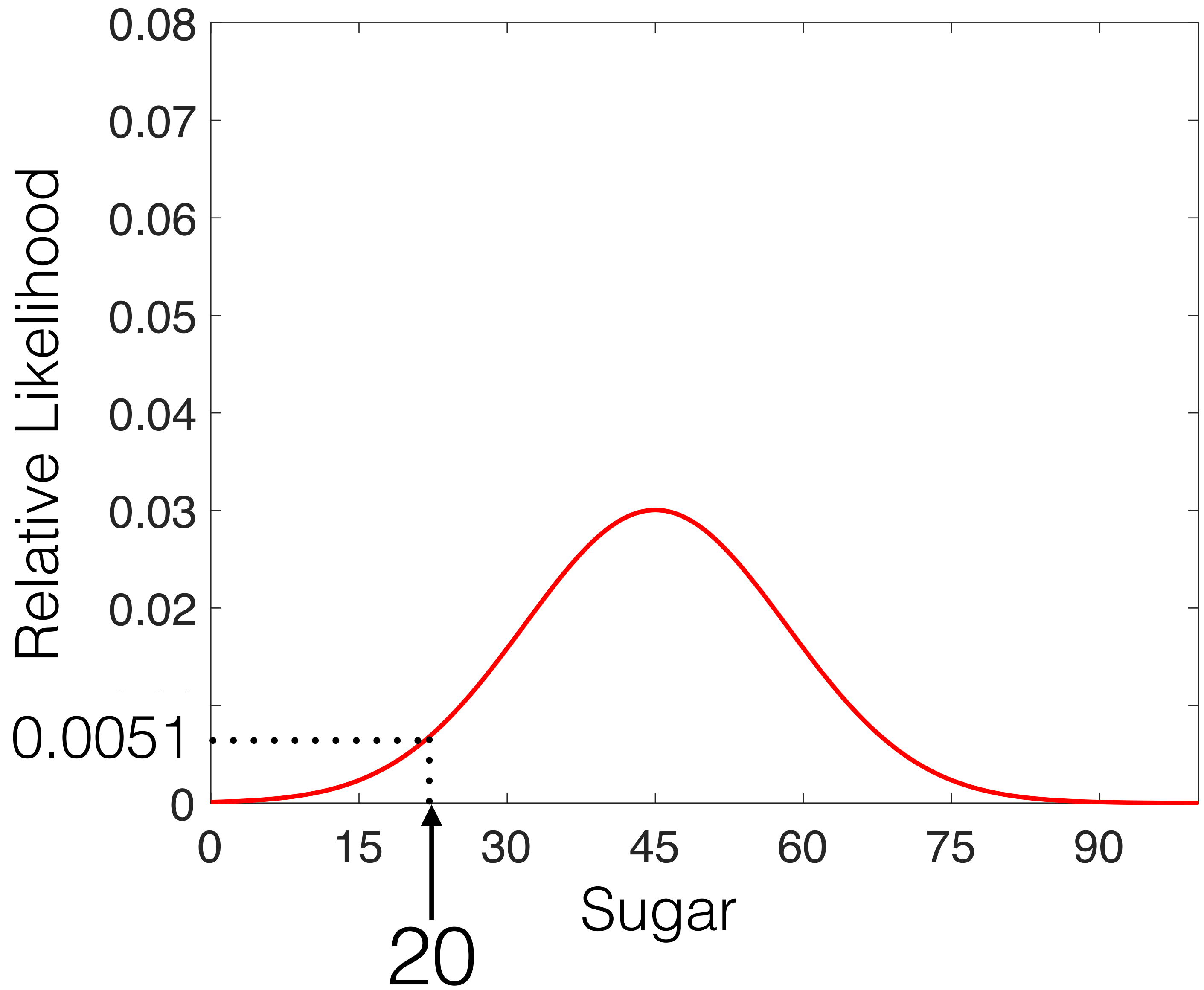
$$P(Cavity=yes|Wash=no,Sugar=20) =$$

$$= \alpha\, P(Cavity=yes) P(Wash=no|Cavity=yes) P(Sugar=20|Cavity=yes)$$

$$= \alpha * 3/6 * 3/5 * P(Sugar=20|Cavity=yes)$$

Getting P(Sugar=20|Cavity=yes)

Probability Density Function for Sugar given Cavity = yes



$$P(X=a \mid \mu,\sigma^2) = \frac{1}{\sqrt{2\,\sigma^2\,\pi}}\, e^{\frac{-(a-\mu)^2}{2\sigma^2}}$$

$\mu = 45$

$\sigma^2 = 175$

Predicting (Wash=no, Sugar=20, Cavity=?)

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Parameters Table for Sugar	Cavity = no	Cavity = yes
μ	22.33	45
σ^2	46.34	116.67

Cavity = no	Cavity = yes	Total:
3	3	6

Model

$$P(c|a_1,\ldots,a_n) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

$$P(Cavity=yes|Wash=no,Sugar=20) =$$

$$= \alpha\, P(Cavity=yes) P(Wash=no|Cavity=yes)\, P(Sugar=20|Cavity=yes)$$

$$= \alpha * 3/6 * 3/5 * 0.0051 = \alpha * 0.00153$$

Predicting (Wash=no, Sugar=20, Cavity=?)

Frequency Table for Wash	Cavity = no	Cavity = yes	Total:
Wash=no	1+1	2+1	3+2
Wash=yes	2+1	1+1	3+2
Total:	3+2	3+2	6+4

Parameters Table for Sugar	Cavity = no	Cavity = yes
μ	22.33	45
σ^2	46.34	116.67

Cavity = no	Cavity = yes	Total:
3	3	6

Model

$$P(c|a_1,\ldots,a_n) = \alpha\, P(c) \prod_{i=1}^{d} P(a_i|c)$$

$$P(Cavity=no|Wash=no,Sugar=20) = \alpha * 0.01106$$

Predicting (Wash=no, Sugar=20, Cavity=?)

Example (Wash=no, Sugar=20, Cavity=?)

P(Cavity=yes|Wash=no, Sugar=20) = α * 0.00153

P(Cavity=no|Wash=no, Sugar=20) = $\boxed{\alpha * 0.01106}$

Predicted class: Cavity = no

Predicting (Wash=no, Sugar=20, Cavity=?)

Example (Wash=no, Sugar=20, Cavity=?)

P(Cavity=yes|Wash=no, Sugar=20) = α * 0.00153 ≈ 12.15%

P(Cavity=no|Wash=no, Sugar=20) = α * 0.01106 ≈ 87.85%

where $\alpha = 1 / \beta$ and $\beta = \sum_{c \in Y} \left(P(Y=c) \prod_{i=1}^{d} P(X=a_i|Y=c) \right)$

$\alpha = 1 / (0.00153 + 0.01106) \approx 79.43$

Naïve Bayes Approach for Classification Problems with Categorical and Numerical Independent Variables

- Naïve Bayes Learning Algorithm:
  - Create frequency tables for categorical independent variables.
  - Apply Laplace Smoothing.
  - Determine probability density function parameters for numerical independent variables.
- Naïve Bayes Model:
  - Frequency tables (with and without Laplace Smoothing) and tables of probability density function parameters, associated with the Bayes Theorem using the conditional independence assumption for predictions.
- Naïve Bayes prediction for an instance (**x**,?):
  - Use Bayes Theorem with conditional independence assumption.

Advantages and Disadvantages of Naïve 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 numeric independent variables.
  - Does not work very well for regression.

Applications

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

Quiz

- Compute step-by-step by hand the mean and variance values asked in slide 14 and the probability from slide 22.

Further Reading

- Essential:
    - Leandro Minku’s notes on “Naïve Bayes for Numeric Independent Variables”.

- Background:
    - Russell and Norvig’s “Artificial Intelligence: A Modern Approach”
        - Section 13.2 (Basic Probability Notation) up to the end of section 13.6 (The Wumpus World Revisited).