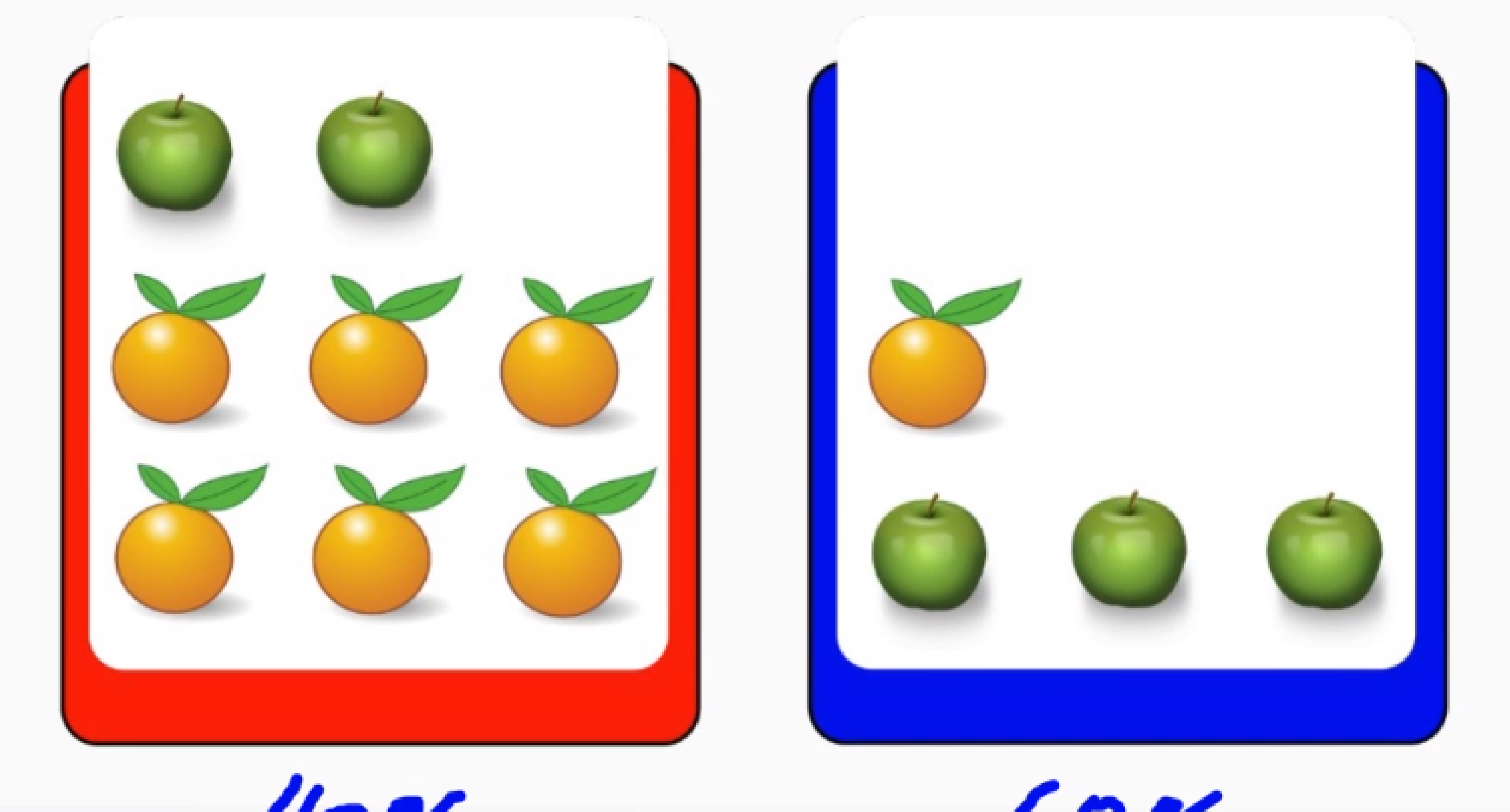
Naive Bayes

Probabilities refresher

$p(B=r) = \frac{4}{10}$ $p(B=b) = \frac{6}{10}$

$[0, 1]$ $\frac{4}{10} + \frac{6}{10} = \frac{10}{10} = 1$



40% 60%

*) Marginal probability - $p(B=r) = \frac{4}{10}$

*) Joint probability - $p(B=r, F=a) = ?$

*) Conditional probability - $p(F=a \mid B=r) = \frac{2}{8} = \frac{1}{4}$

Example from Christopher M. Bishop, *Pattern recognition and machine learning* (information science and statistics), Springer-Verlag New York Inc., Secaucus, NJ, USA, 2006.

Rules of probability

X, Y

1) Sum rule $\underline{p(X)} = \sum_Y p(X,Y)$

2) Product rule $p(\underline{X,Y}) = p(\underline{Y|X)p(X})$

Ex. 1 $p(B=r, F=a) = \underline{p(F=a|B=r)}\ \underline{p(B=r)} = \frac{1}{4} \cdot \frac{4}{10} = \frac{1}{10}$

Ex. 2 $p(F=a) = \sum_{B\in\{r,b\}} p(F=a, B) = p(F=a|B=r)p(B=r) +$

$p(F=a|B=b)\,p(B=b) =$

$= \frac{1}{4} \cdot \frac{4}{10} + \frac{3}{4} \cdot \frac{6}{10} = \frac{1}{10}$

r

Bayes Rule

$p(X,Y) = p(Y,X)$

$$p(X) = \sum_Y p(X,Y)$$

$$\underline{p(X,Y)} = \underline{p(Y|X)p(X)}$$

$p(Y,X) = \underline{p(X|Y)p(Y)}$

$$p(Y|X) = \frac{p(X,Y)}{p(X)} = \frac{p(Y,X)}{p(X)} = \frac{p(X|Y)p(Y)}{p(X)}$$

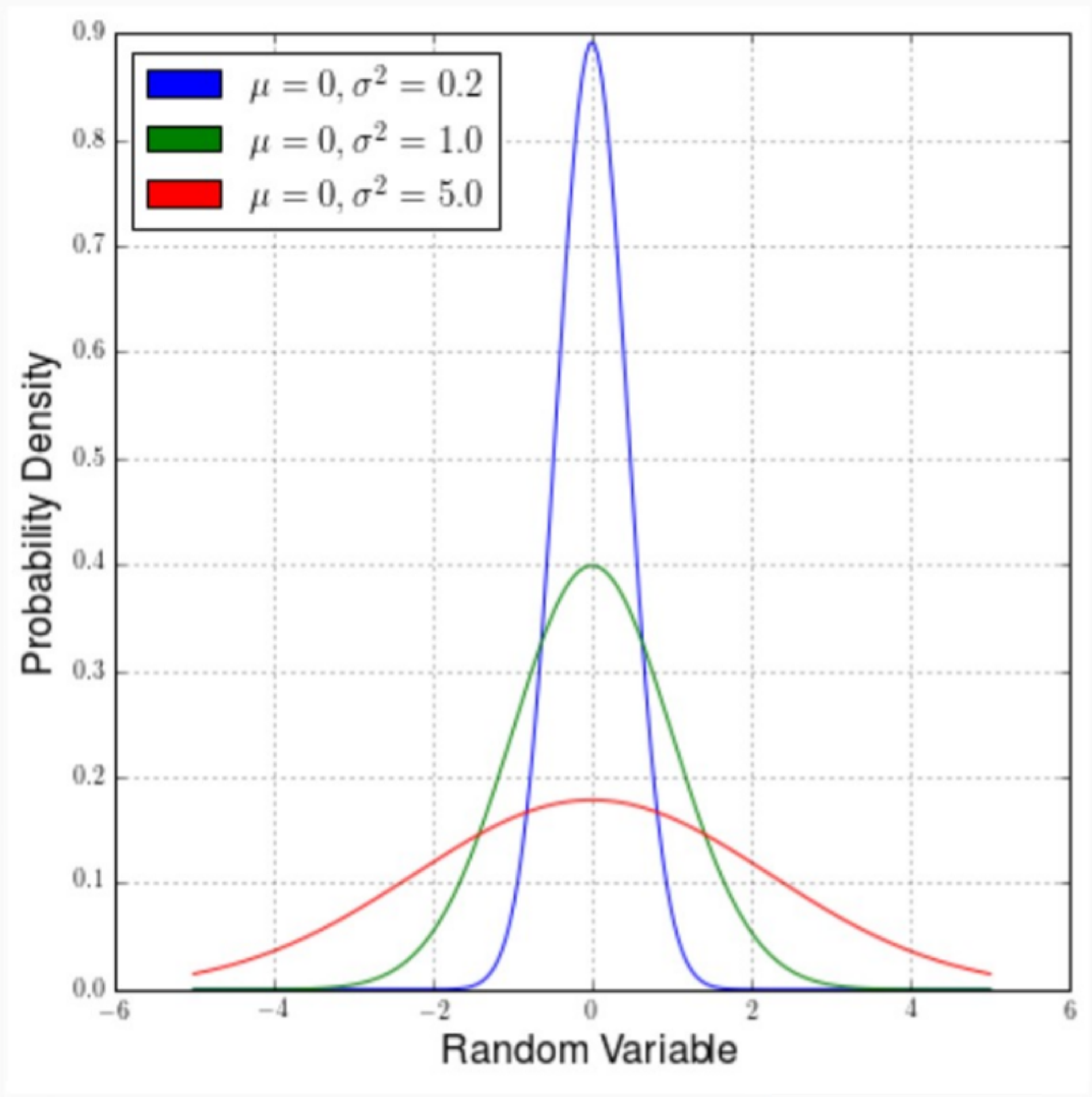
$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

Gaussian Distribution

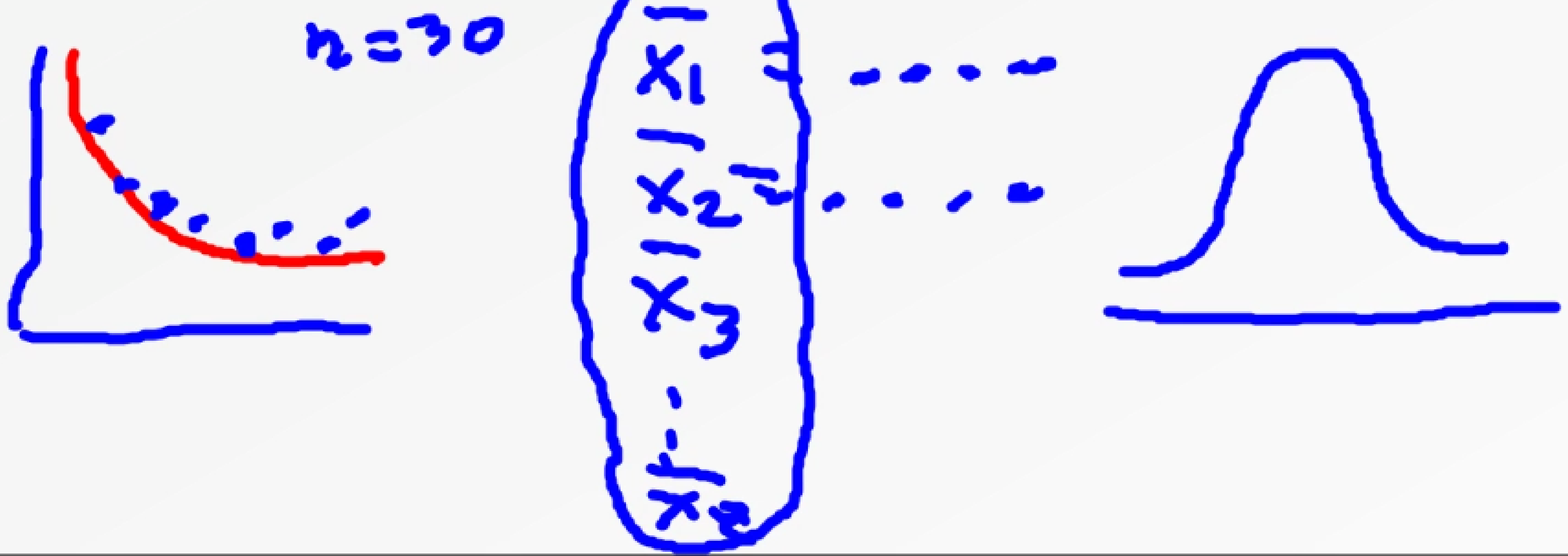
- A very common continuous distribution

$$\longrightarrow N(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Occurs naturally in many situations
  - height of people
  - salaries
  - blood pressure



CLT



Bayesian Inference

- A method of inference where the probability of a hypothesis is updated as new evidence becomes available.

$p(H)$

  - Begin with a prior distribution processes

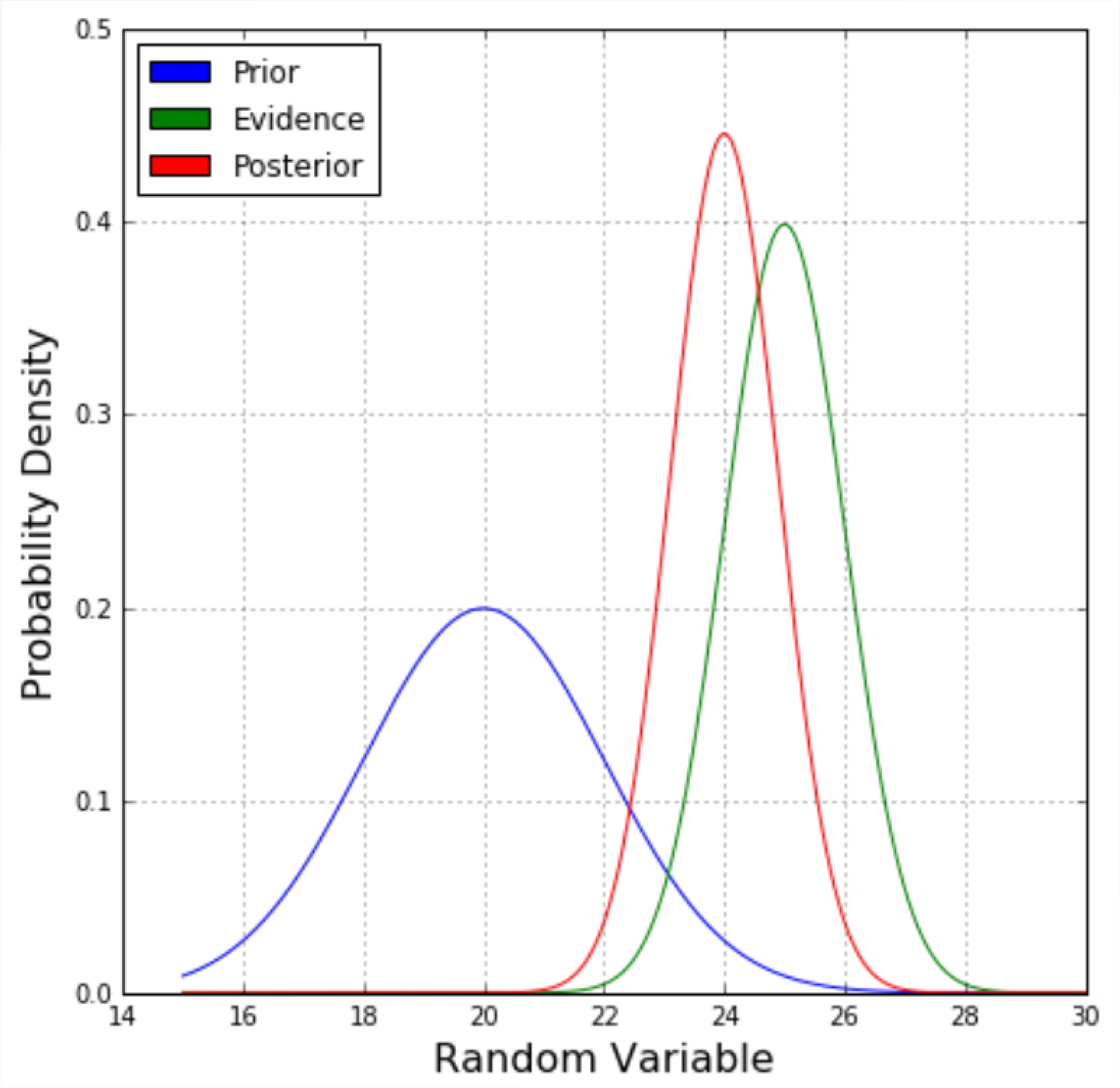
E

  - Collect data to obtain the observed distribution

$p(E|H)$

  - Calculate the likelihood – how compatible is the evidence with the hypothesis

  - Obtain the posterior – the probability of our hypothesis given the observed evidence



Computing the posterior

likelihood

prior

$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)}$$

posterior

marginal likelihood

$$p(H|E) = \frac{p(E|H)\,p(H)}{p(E)}$$

Example

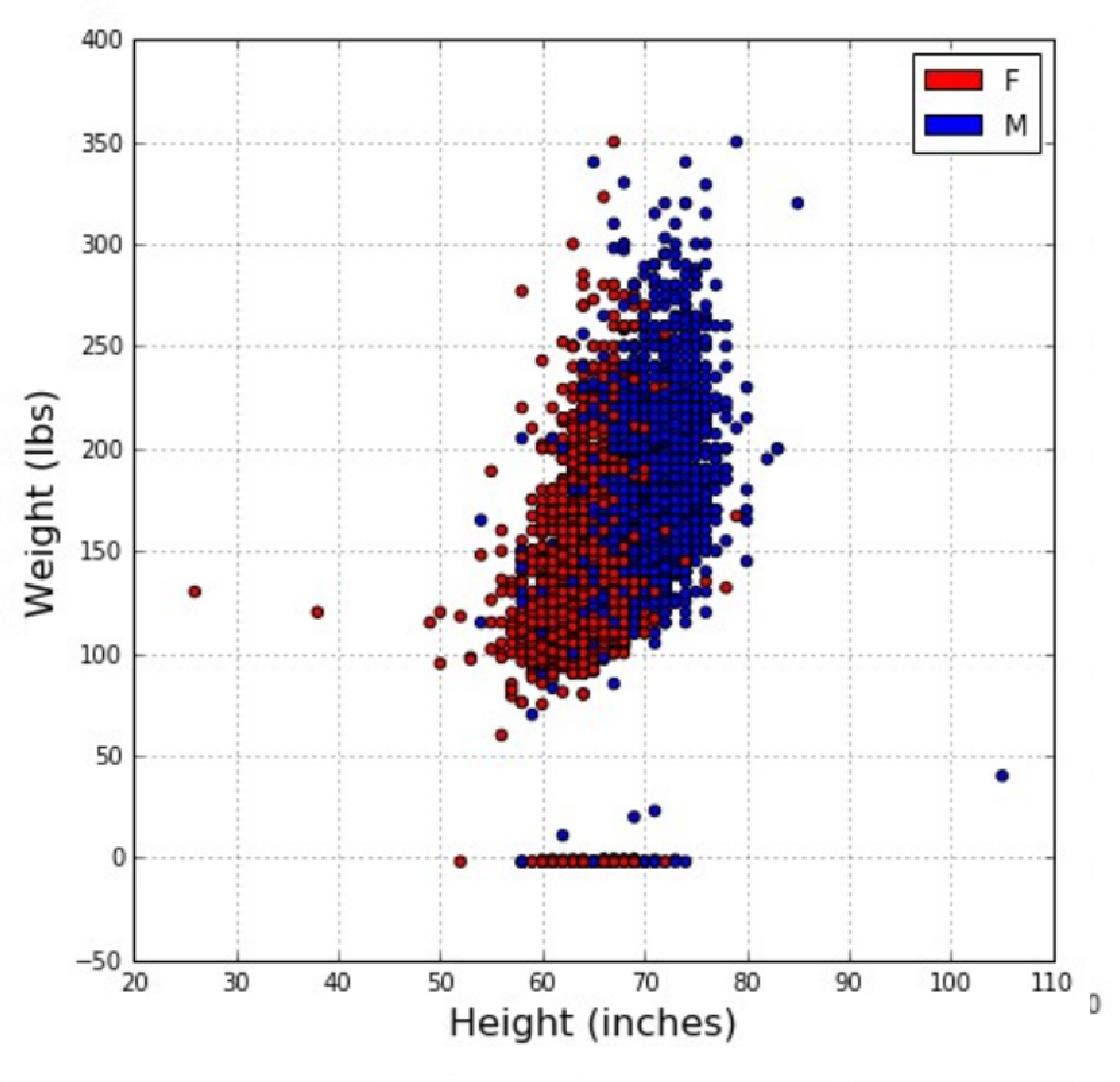
0 - male
1 - female

Sex	Height	Weight
1	67	150
0	67	140
1	67	100
1	62	185
0	69	145
1	68	140
...	...	...

$C \in \{M, F\}$

$\rightarrow p(\text{height} = \underline{\underline{69}} \mid C_M) = ?$

$p(\text{height} = 69 \mid C_F) = ?$



Data from National Longitudinal Youth Survey, Bureau of Labor Statistics, United States Department of Labor

Example

$N = 10$

males

Sex	Height	Weight
0	67	140
0	69	145
0	69	183
0	71	175
0	66	108

5

females

Sex	Height	Weight
1	67	150
1	67	100
1	62	185
1	68	140
1	64	123

priors: $p(C = M) = \frac{5}{10} = 0.5$

$p(C = F) = 1 - 0.5 = 0.5$

$$\mu_M = \frac{\sum_i x_i}{n} = \frac{67 + 69 + \dots + 66}{5} = 68.4$$

$$\sigma_M = \sqrt{\frac{\sum_i (x_i - \mu_M)^2}{n}} = 1.74$$

new data : 69

$$p(x=69 \mid c_M) = \frac{1}{\sqrt{2\pi\sigma_M^2}} e^{-\frac{(69-\mu_M)^2}{2\sigma_M^2}} = \underset{-----}{A_1}$$

likelihood

$$p(x=69 \mid c_F) = \frac{1}{\sqrt{2\pi\sigma_F^2}} e^{-\frac{(69-\mu_F)^2}{2\sigma_F^2}} = \underset{-----}{A_2}$$

Maximum a posteriori estimation

Our prediction is the value of C, which maximizes the posterior distribution

$A_1 A_2 \ldots$ 0.5

$$C_{MAP} = \underset{c \in \{M, F\}}{\text{argmax}}\ p(C|x) = \underset{c \in C}{\text{argmax}} \boxed{\frac{p(x|c)\,p(c)}{\cancel{p(x)}}}$$

$$\frac{V_1}{\cancel{c}} \quad \frac{V_2}{\cancel{c}} \quad \frac{V_3}{\cancel{c}} \quad \frac{V_4}{\cancel{c}}$$

$$= \text{argmax}\ p(X|c)\,p(C)$$

$x = \{x_1, x_2, \ldots, x_D\}$

$$p(x_i | C_j) = p(x_1 | x_2, x_3, \ldots, x_D, C_j) \cdot p(x_2 | x_3, \ldots, C_j) \cdots p(x_D | C_j)\, p(C_j)$$

Assume $x_1, x_2 \ldots x_D$

$$p(x_1, x_2, \ldots, x_D | C_j) = p(x_1 | C_j)\, p(x_2 | C_j) \cdots p(x_D | C_j)$$

WIN MONEY $| C_j$

WIN 

Not just Gaussian

- Gaussian Naive Bayes
- Binomial Naive Bayes

$$p(x|c_j) = \prod_{i=1}^{n} p_{ji}^{x_i} (1 - p_{ji})^{(1-x_i)}$$

- Multinomial Naive Bayes