Random Variables

Joint Distributions
Independence

Luigi Portinale

University of Piemonte Orientale, Italy

March 7-10, 2017

Probability Space

A probability space is a triple $(\Omega, \mathcal{F}, P)$ where

- a sample space Ω which is the set of possible outcomes of interest
- a σ-algebra $\mathcal{F} \subseteq 2^{\Omega}$ of events
- a probability function $P : \mathcal{F} \rightarrow [0, 1]$

Given a set X, a σ-algebra of X is a collection of subsets of X containing X, closed under complementation and closed under (countable) union.

Probability Axioms (Kolmogorov)

The probability function in a probability space statisfies the following axioms:

1. $\forall E \in \mathcal{F},\ P(E) \geq 0$
2. $P(\Omega) = 1$
3. let $E_1, E_2, \ldots$ a (possibly infinite) sequence of disjoint (mutually exclusive) events $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$

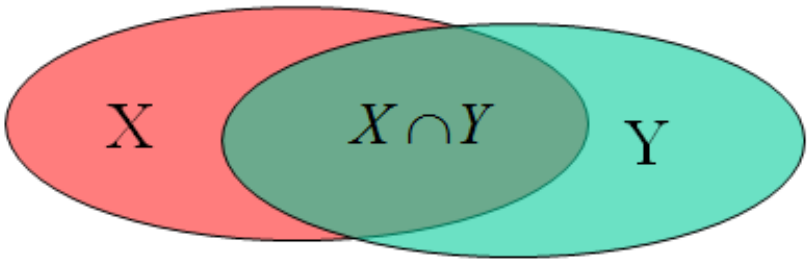
Some consequences:

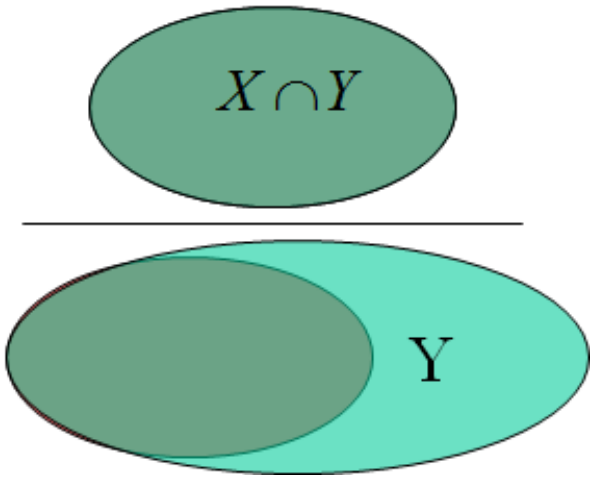
- $P(\emptyset) = 0$
- if $X \subseteq Y$, then $P(X) \leq P(Y)$
- $P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$

Conditional Probability

The probability of an event X to occur, given that event Y has occurred is given by

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





Law of Total Probability

Let $H_1 \dots H_n$ be a set of exhaustive (i.e. $\sum_{i=1}^{n} P(H_i) = 1$) and mutually exclusive events (*hypotheses*) and E a generic event.

$$P(E) = \sum_{i=1}^{n} P(H_i)P(E|H_i)$$

Example

Let R be the event *today's raining* and U the event *my friend Joe's carrying an umbrella*.

$$P(U) = P(R)P(U|R) + P(\overline{R})P(U|\overline{R})$$

Random Variable

A Random Variable is a function $X : \Omega \rightarrow \mathcal{K}$ from an atomic event on a *probability space* $(\Omega, \mathcal{F}, P)$ *to a given range* $\mathcal{K}$

$$X(\omega) = \kappa \qquad \text{for } \omega \in \Omega \text{ and } \kappa \in \mathcal{K} \tag{1}$$

e.g. Roll of a die

$$Even : \{1, 2, 3, 4, 5, 6\} \rightarrow \{\texttt{true}, \texttt{false}\}$$

$$Even(2) = Even(4) = Even(6) = \texttt{true}$$
$$Even(1) = Even(3) = Even(5) = \texttt{false}$$

Probability Distribution over Random Variables

$$P(X = \kappa) = \sum_{\{\omega : X(\omega) = \kappa\}} P(\omega)$$

$$P(Even = \texttt{true}) = \texttt{P}(2) + \texttt{P}(4) + \texttt{P}(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even = \texttt{false}) = \texttt{P}(1) + \texttt{P}(3) + \texttt{P}(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even) = \begin{matrix} \texttt{true} \\ \texttt{false} \end{matrix} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Probability Distribution over Random Variables

$$P(X = \kappa) = \sum_{\{\omega : X(\omega) = \kappa\}} P(\omega)$$

$$P(Even = \texttt{true}) = \texttt{P}(2) + \texttt{P}(4) + \texttt{P}(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even = \texttt{false}) = \texttt{P}(1) + \texttt{P}(3) + \texttt{P}(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even) = \begin{matrix} \texttt{true} \\ \texttt{false} \end{matrix} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Probability Distribution over Random Variables

$$P(X = \kappa) = \sum_{\{\omega : X(\omega) = \kappa\}} P(\omega)$$

$$P(Even = \texttt{true}) = \mathrm{P}(2) + \mathrm{P}(4) + \mathrm{P}(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even = \texttt{false}) = \mathrm{P}(1) + \mathrm{P}(3) + \mathrm{P}(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even) = \begin{matrix} \texttt{true} \\ \texttt{false} \end{matrix} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Probability Distribution over Random Variables

$$P(X = \kappa) = \sum_{\{\omega : X(\omega) = \kappa\}} P(\omega)$$

$$P(Even = \texttt{true}) = \mathrm{P}(2) + \mathrm{P}(4) + \mathrm{P}(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even = \texttt{false}) = \mathrm{P}(1) + \mathrm{P}(3) + \mathrm{P}(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

$$P(Even) = \begin{matrix} \texttt{true} \\ \texttt{false} \end{matrix} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Probabilistic Models with Random Variables

- **A probabilistic model can be built from a set of random variables $X_1, X_2, \ldots X_n$;**
- **The sample space is $\Omega = X_1 \times X_2 \times \ldots \times X_n$**

$A : \{a_1, a_2\}; B : \{b_1, b_2\}$
$\Omega = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

$$P(A, B) = \begin{pmatrix} P(a_1, b_1) \\ P(a_1, b_2) \\ P(a_2, b_1) \\ P(a_2, b_2) \end{pmatrix}$$ **Joint probability of A and B**

$$P(A) = \begin{pmatrix} P(a_1, b_1) + P(a_1, b_2) \\ P(a_2, b_1) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of A**

$$P(B) = \begin{pmatrix} P(a_1, b_1) + P(a_2, b_1) \\ P(a_1, b_2) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of B**

Probabilistic Models with Random Variables

- **A probabilistic model can be built from a set of random variables $X_1, X_2, \ldots X_n$;**
- **The sample space is $\Omega = X_1 \times X_2 \times \ldots \times X_n$**

$A : \{a_1, a_2\}; B : \{b_1, b_2\}$
$\Omega = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

$$P(A, B) = \begin{pmatrix} P(a_1, b_1) \\ P(a_1, b_2) \\ P(a_2, b_1) \\ P(a_2, b_2) \end{pmatrix}$$ **Joint probability of A and B**

$$P(A) = \begin{pmatrix} P(a_1, b_1) + P(a_1, b_2) \\ P(a_2, b_1) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of A**

$$P(B) = \begin{pmatrix} P(a_1, b_1) + P(a_2, b_1) \\ P(a_1, b_2) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of B**

Probabilistic Models with Random Variables

- **A probabilistic model can be built from a set of random variables $X_1, X_2, \ldots X_n$;**
- **The sample space is $\Omega = X_1 \times X_2 \times \ldots \times X_n$**

$A : \{a_1, a_2\}; B : \{b_1, b_2\}$
$\Omega = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

$$P(A, B) = \begin{pmatrix} P(a_1, b_1) \\ P(a_1, b_2) \\ P(a_2, b_1) \\ P(a_2, b_2) \end{pmatrix}$$ **Joint probability of A and B**

$$P(A) = \begin{pmatrix} P(a_1, b_1) + P(a_1, b_2) \\ P(a_2, b_1) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of A**

$$P(B) = \begin{pmatrix} P(a_1, b_1) + P(a_2, b_1) \\ P(a_1, b_2) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of B**

Probabilistic Models with Random Variables

- **A probabilistic model can be built from a set of random variables $X_1, X_2, \ldots X_n$;**
- **The sample space is $\Omega = X_1 \times X_2 \times \ldots \times X_n$**

$A : \{a_1, a_2\}; B : \{b_1, b_2\}$
$\Omega = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

$$P(A, B) = \begin{pmatrix} P(a_1, b_1) \\ P(a_1, b_2) \\ P(a_2, b_1) \\ P(a_2, b_2) \end{pmatrix}$$ **Joint probability of A and B**

$$P(A) = \begin{pmatrix} P(a_1, b_1) + P(a_1, b_2) \\ P(a_2, b_1) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of A**

$$P(B) = \begin{pmatrix} P(a_1, b_1) + P(a_2, b_1) \\ P(a_1, b_2) + P(a_2, b_2) \end{pmatrix}$$ **Marginal probability of B**

Probabilistic Models with Random Variables

- **A probabilistic model can be built from a set of random variables $X_1, X_2, \ldots X_n$;**
- **The sample space is $\Omega = X_1 \times X_2 \times \ldots \times X_n$**

$A : \{a_1, a_2\}; B : \{b_1, b_2\}$
$\Omega = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

$$P(A,B) = \begin{pmatrix} P(a_1,b_1) \\ P(a_1,b_2) \\ P(a_2,b_1) \\ P(a_2,b_2) \end{pmatrix}$$ **Joint probability of A and B**

$$P(A) = \begin{pmatrix} P(a_1,b_1) + P(a_1,b_2) \\ P(a_2,b_1) + P(a_2,b_2) \end{pmatrix}$$ **Marginal probability of A**

$$P(B) = \begin{pmatrix} P(a_1,b_1) + P(a_2,b_1) \\ P(a_1,b_2) + P(a_2,b_2) \end{pmatrix}$$ **Marginal probability of B**

Joint Probability

Given a set of r.v. $X_1, \ldots X_n$ their joint probability is a function $P : X_1 \times \ldots \times X_n \rightarrow [0, 1]$ such that

$$0 \leq P(X_1 = x_1, \ldots X_n = x_n) \leq 1$$

$$\sum_{x_1 \ldots x_n} P(X_1 = x_1, \ldots X_n = x_n) = 1$$

Marginal Probability

The marginal probability of X_i is given by

$$P(X_i = x_i) = \sum_{x_k (k \neq i)} P(X_1 = x_1, \ldots X_i = x_i, \ldots X_n = x_n)$$

Important

Given the joint probability of a set $X = X_1, \ldots X_n$ of r.v., any probability of the model can be computed. Let $Q \subset X$ be a set of query variables and $E \subset X - Q$ be a set of evidence variables:

$$P(Q|E) = \frac{\sum_{X-(Q \cup E)} P(X_1, \ldots X_n)}{\sum_{X-E} P(X_1, \ldots X_n)}$$

Definition of independence

Two random variables X and Y are independent given a set of random variables E $(X \perp Y|E)$iff

$$P(X, Y|E) = P(X|E)P(Y|E)$$

Given a set of marginal probability of independent variables is it possible to obtain their joint probability (this is not possible if variables are not independent)

Bayes Theorem

Given a set of mutually exclusive and exhaustive hypotheses $h_1, \ldots h_n$ and an evidence e, for every $h_i (1 \leq i \leq n)$:

$$P(h_i|e) = \frac{P(h_i)P(e|h_i)}{P(e)} = \frac{P(h_i)P(e|h_i)}{\sum_{j=1}^{n} P(h_j)P(e|h_j)}$$

$P(h_i|e)$ is the posterior probability of the hypothesis;

$P(h_i)$ is the prior probability of the hypothesis;

$P(e|h_i)$ is the likelihood of the evidence;

$\sum_{j=1}^{n} P(h_j)P(e|h_j)$ is the marginal likelihood of the evidence (a normalization factor)

$$P(H) = \begin{pmatrix} P(\bar{h}) \\ P(h) \end{pmatrix}$$

$$P(E|H) = \begin{pmatrix} P(\bar{e}|\bar{h}) & P(\bar{e}|h) \\ P(e|\bar{h}) & P(e|h) \end{pmatrix}$$

Bayes Theorem

Given a set of mutually exclusive and exhaustive hypotheses $h_1, \ldots h_n$ and an evidence e, for every $h_i (1 \leq i \leq n)$:

$$P(h_i|e) = \frac{P(h_i)P(e|h_i)}{P(e)} = \frac{P(h_i)P(e|h_i)}{\sum_{j=1}^{n} P(h_j)P(e|h_j)}$$

$P(h_i|e)$ is the posterior probability of the hypothesis;

$P(h_i)$ is the prior probability of the hypothesis;

$P(e|h_i)$ is the likelihood of the evidence;

$\sum_{j=1}^{n} P(h_j)P(e|h_j)$ is the marginal likelihood of the evidence (a normalization factor)

$$P(H) = \begin{pmatrix} P(\bar{h}) \\ P(h) \end{pmatrix}$$

$$P(E|H) = \begin{pmatrix} P(\bar{e}|\bar{h}) & P(\bar{e}|h) \\ P(e|\bar{h}) & P(e|h) \end{pmatrix}$$

Bayes Theorem

Given a set of mutually exclusive and exhaustive hypotheses $h_1, \ldots h_n$ and an evidence e, for every $h_i (1 \leq i \leq n)$:

$$P(h_i|e) = \frac{P(h_i)P(e|h_i)}{P(e)} = \frac{P(h_i)P(e|h_i)}{\sum_{j=1}^{n} P(h_j)P(e|h_j)}$$

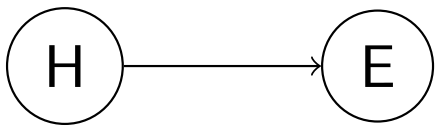
$P(h_i|e)$ is the posterior probability of the hypothesis;

$P(h_i)$ is the prior probability of the hypothesis;

$P(e|h_i)$ is the likelihood of the evidence;

$\sum_{j=1}^{n} P(h_j)P(e|h_j)$ is the marginal likelihood of the evidence (a normalization factor)

$$P(H) = \begin{pmatrix} P(\bar{h}) \\ P(h) \end{pmatrix}$$



$$P(E|H) = \begin{pmatrix} P(\bar{e}|\bar{h}) & P(\bar{e}|h) \\ P(e|\bar{h}) & P(e|h) \end{pmatrix}$$

How to reason with Bayes
Example: a medical test (HIV)

T^+=positive test; I=AIDS infection

False positive rate: 1.5%; no false negative
$P(T^+|I) = 1$ $P(T^+|\bar{I}) = 0.015$

Question: if a patient is positive to the test, what is the probability he/she is infected?

$$P(I|T^+) = ??$$

Common mistake

Since the false positives occur with probability 1.5%, then the probability of infection if tested positive is $1 - 0.015 = 0.985$ (98.5%)

The value depends on the size of infected population

$$P(I|T^+) = \frac{P(I)P(T^+|I)}{P(I)P(T^+|I) + P(\bar{I})P(T^+|\bar{I})}$$

The posterior probability strongly depends on the prior in this case.

AIDS frequency in Italy: 0.4%
AIDS frequency in South Africa: 18.1%

$P(I|T^+) = \frac{0.004 \times 1}{0.004 \times 1 + 0.996 \times 0.015} \approx 21.1\%$ **ITALY**
$P(I|T^+) = \frac{0.181 \times 1}{0.181 \times 1 + 0.819 \times 0.015} \approx 93.6\%$ **SOUTH AFRICA**

The value depends on the size of infected population

$$P(I|T^+) = \frac{P(I)P(T^+|I)}{P(I)P(T^+|I) + P(\bar{I})P(T^+|\bar{I})}$$

The posterior probability strongly depends on the prior in this case.

AIDS frequency in Italy: 0.4%
AIDS frequency in South Africa: 18.1%

$P(I|T^+) = \frac{0.004 \times 1}{0.004 \times 1 + 0.996 \times 0.015} \approx 21.1\%$ **ITALY**
$P(I|T^+) = \frac{0.181 \times 1}{0.181 \times 1 + 0.819 \times 0.015} \approx 93.6\%$ **SOUTH AFRICA**

The value depends on the size of infected population

$$P(I|T^+) = \frac{P(I)P(T^+|I)}{P(I)P(T^+|I) + P(\bar{I})P(T^+|\bar{I})}$$

The posterior probability strongly depends on the prior in this case.

AIDS frequency in Italy: 0.4%
AIDS frequency in South Africa: 18.1%

$$P(I|T^+) = \frac{0.004 \times 1}{0.004 \times 1 + 0.996 \times 0.015} \approx 21.1\% \quad \textbf{ITALY}$$
$$P(I|T^+) = \frac{0.181 \times 1}{0.181 \times 1 + 0.819 \times 0.015} \approx 93.6\% \quad \textbf{SOUTH AFRICA}$$

Finding a positive result in a repeated test:

$$P(I|T_1^+, T_2^+) \propto P(I|T_1^+)P(T_2^+|I, T_1^+)$$

If we assume that each test result is independent from previous tests given the hypothesis, we have $P(T_2^+|I, T_1^+) = P(T_2^+|I)$ and $P(T_2^+|\bar{I}, T_1^+) = P(T_2^+|\bar{I})$; then we can simply apply Bayes rule using the previous posterior $P(I|T_1^+)$ as the new prior.

$P(I|T_1^+, T_2^+) \approx 94.7\%$ **ITALY**
$P(I|T_1^+, T_2^+) \approx 99.9\%$ **SOUTH AFRICA**

Naive Bayes

Given a set of mutually exclusive and exhaustive hypotheses $h_1, \ldots h_n$ and a set of evidences $e_1 \ldots e_m$ such that $(e_j \perp e_k | h_i)$ for $(j \neq k)$, $(1 \leq j, k \leq m)$ and $(1 \leq i \leq n)$:

$$P(h_i|e_1, \ldots e_m) = \frac{P(h_i)\prod_{j=1}^{m} P(e_j|h_i)}{\sum_{l=1}^{n} P(h_l)\prod_{j=1}^{m} P(e_j|h_l)}$$

If evidences are conditionally independent given the hypotheses, one can provide a linear number of parameters $(P(e_j|h_i)(j = 1..m))$ instead of an exponential one $(P(e_1, \ldots e_m|h_i)$ for every combination of values of $e_1 \ldots e_m)$.

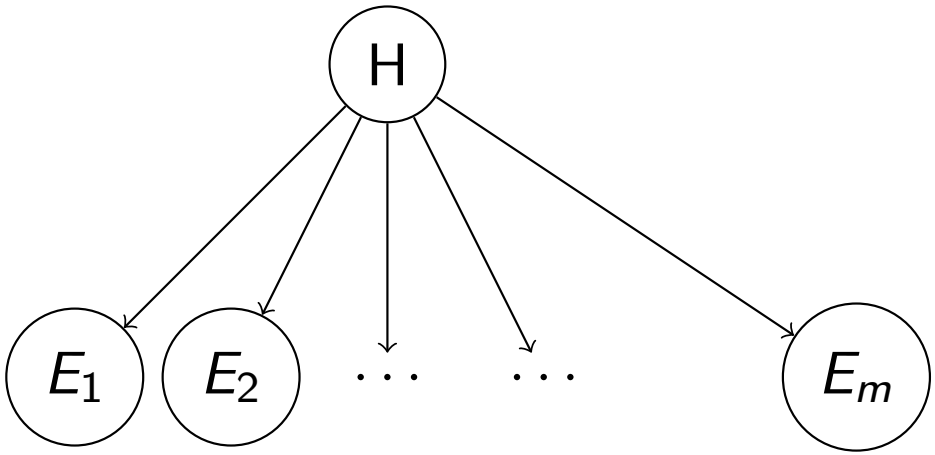
Naive Bayes

Given a set of mutually exclusive and exhaustive hypotheses $h_1, \ldots h_n$ and a set of evidences $e_1 \ldots e_m$ such that $(e_j \perp e_k | h_i)$ for $(j \neq k)$, $(1 \leq j, k \leq m)$ and $(1 \leq i \leq n)$:

$$P(h_i|e_1, \ldots e_m) = \frac{P(h_i)\prod_{j=1}^{m} P(e_j|h_i)}{\sum_{l=1}^{n} P(h_l)\prod_{j=1}^{m} P(e_j|h_l)}$$

If evidences are conditionally independent given the hypotheses, one can provide a linear number of parameters $(P(e_j|h_i)(j = 1..m))$ instead of an exponential one $(P(e_1, \ldots e_m|h_i)$ for every combination of values of $e_1 \ldots e_m)$.

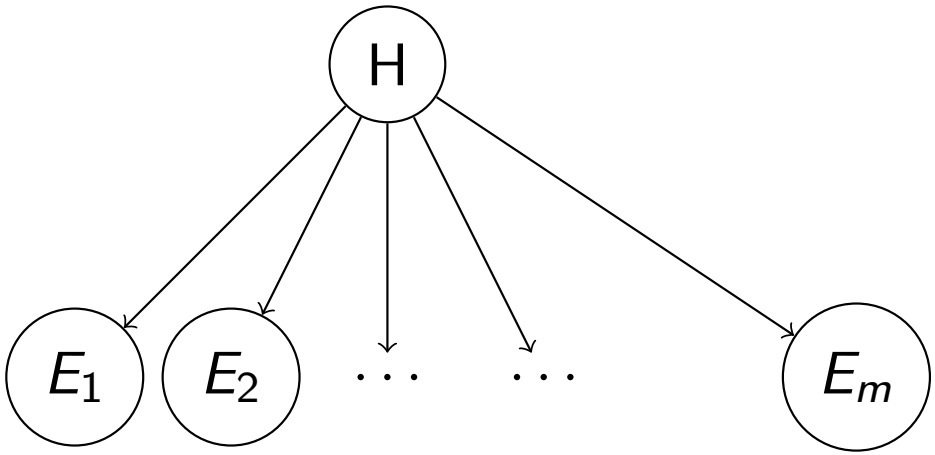
Naive Bayes Graphical Representation



- Parameters:

$$P(H) = \begin{pmatrix} P(\bar{h}) \\ P(h) \end{pmatrix} \; P(E_j|H) = \begin{pmatrix} P(\bar{e}_j|\bar{h}) & P(\bar{e}_j|h) \\ P(e_j|\bar{h}) & P(e_j|h) \end{pmatrix} (j = 1 \ldots m)$$

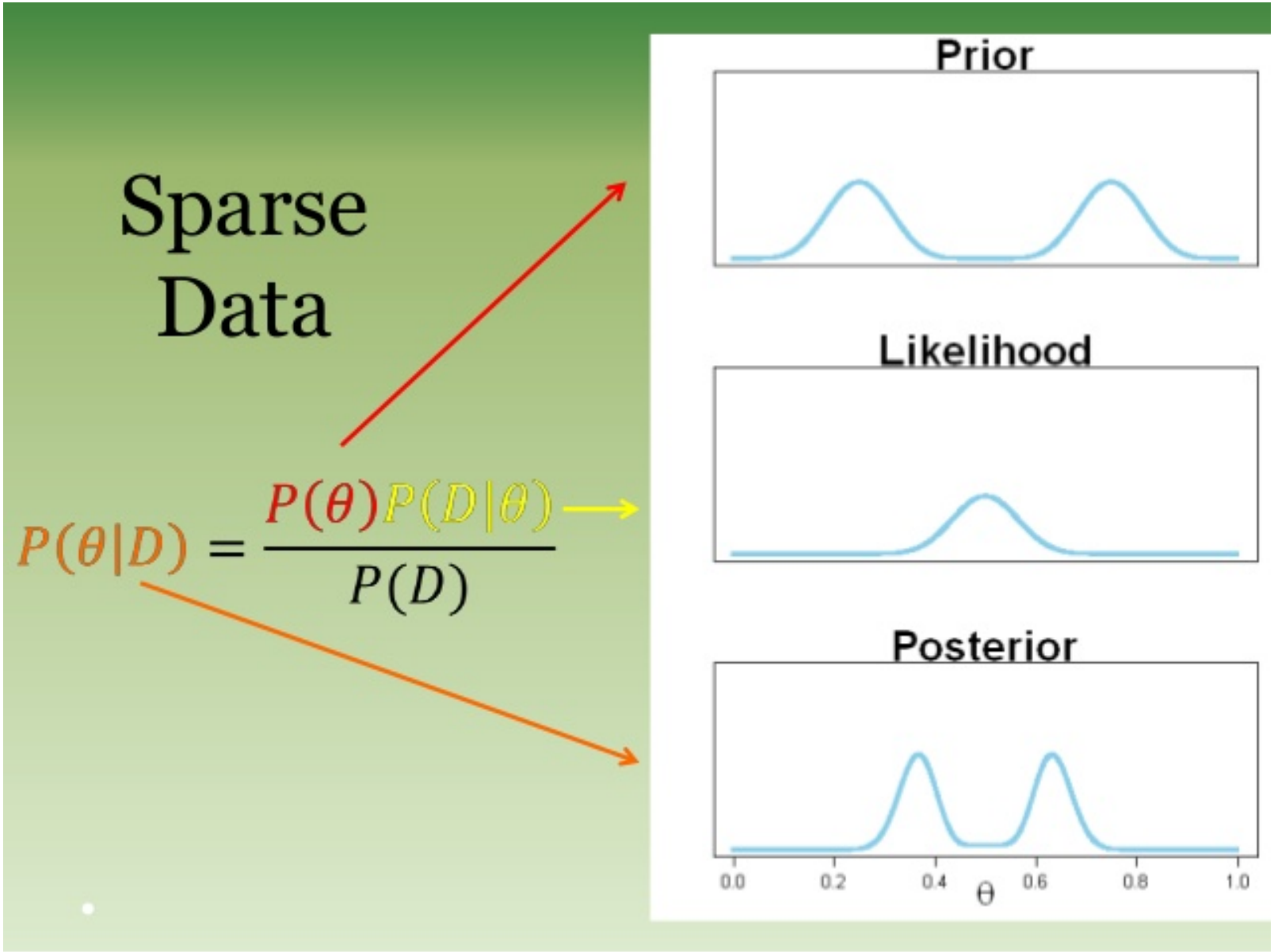
Naive Bayes Graphical Representation

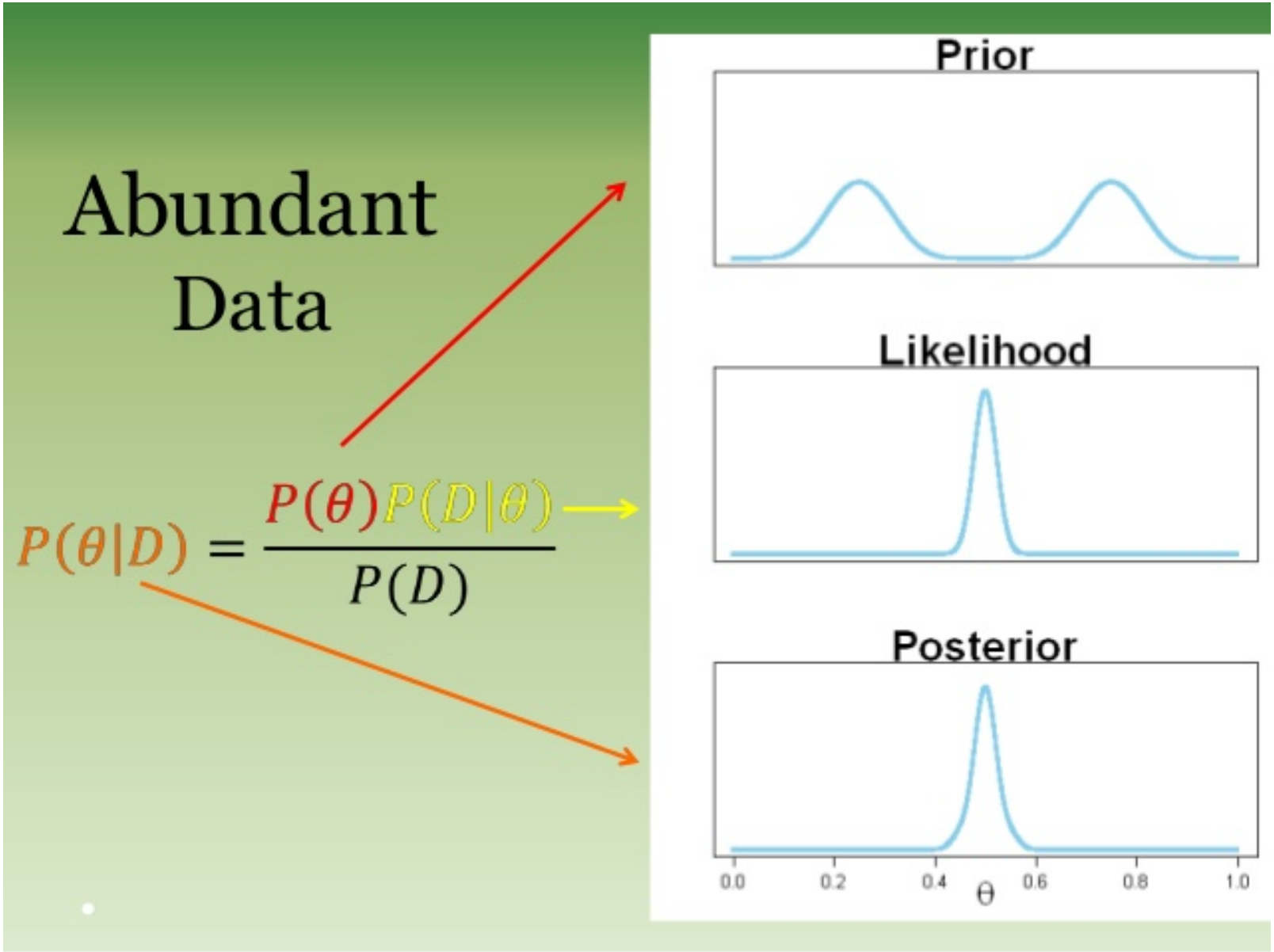


- Parameters:

$$P(H) = \begin{pmatrix} P(\bar{h}) \\ P(h) \end{pmatrix} \; P(E_j|H) = \begin{pmatrix} P(\bar{e}_j|\bar{h}) & P(\bar{e}_j|h) \\ P(e_j|\bar{h}) & P(e_j|h) \end{pmatrix} (j = 1 \ldots m)$$

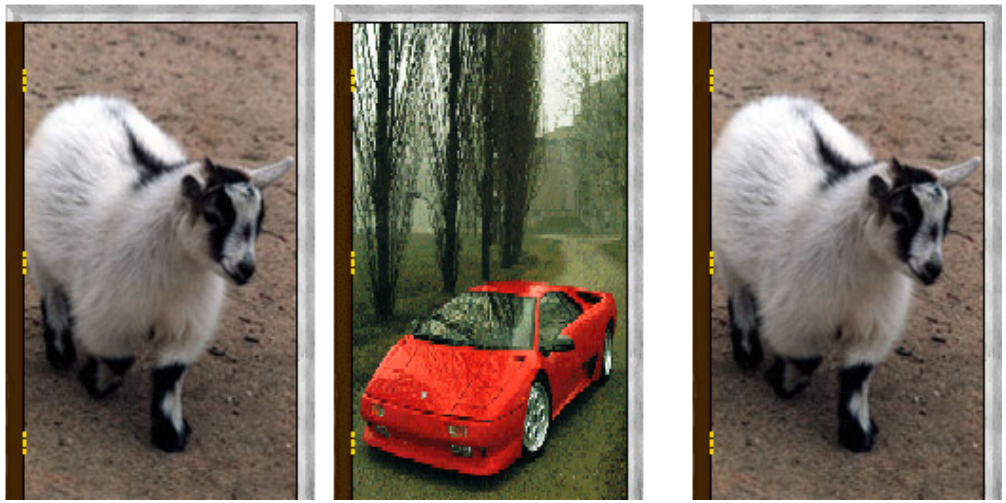
What is the prior's influence?





Monty Hall Puzzle (Let's make a deal TV show)

Monty Hall



Stay or Switch?

Play at: *http://www.stayorswitch.com*