Sample space

■ **Sample space (population) \( \Omega \):**
  ■ Set of possible outcomes of some experiment.

■ **Example:**
  ■ Experiment: randomly select a student among all UST postgraduate students.
  ■ Sample space \( \Omega \): the set of all UST postgraduate students.

The set of possible outcomes of an “experiment” is called the **sample space**

- Throwing a six sided die: \{1, 2, 3, 4, 5, 6\}.
- Will Denmark win the world cup: \{yes, no\}.
- The values in a deck of cards: \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}.

■ **Elements of the sample spaces are called** **samples**.

■ **Subsets of sample spaces are** **events**.

■ **Examples:**
  ■ Sample space \( \Omega \): the set of all UST postgraduate students.
  ■ \( E_{\text{female}} \) = \{female students\}
    the randomly selected student is a female.
  ■ \( E_{\text{male}} \) = \{male students\}
    the randomly selected student is a male.
  ■ \( E_{\text{MPhil}} \) = \{MPhil students\}
    the randomly selected student is an MPhil student.
  ■ \( E_{\text{PhD}} \) = \{PhD students\}
    the randomly selected student is a PhD student.

- The event that we will get an even number when throwing a die: \{2, 4, 6\}.
- The event that Denmark wins: \{yes\}.
- The event that we will get a 6 or below when drawing a card: \{2, 3, 4, 5, 6\}.
**Probability measure**

- A **probability measure** is a mapping from the set of events to $[0, 1]$
  \[ P : 2^\Omega \rightarrow [0, 1] \]

  that satisfies Kolmogorov’s axioms:

  1. $P(\Omega) = 1.$
  2. $P(A) \geq 0 \ \forall A \subseteq \Omega$
  3. **Additivity:** $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset.$

- **Example:**
  - Sample space $\Omega$: the set of all UST postgraduate students.
  - Define probability measure: $P(A) = |A|/|\Omega|.$
  - $P(E_{\text{female}})$ = ‘fraction of female postgraduate students

---

**Random Variables**

- **Random variable** $X$:
  - Function defined over sample space.
  - **Example:**
    - Gender of (randomly selected) student,
    - Programme of (randomly selected) student

- **Domain of a random variable** $\Omega_X$:
  - the set of possible states of $X$.
  - **Example:**
    \[ \Omega_{\text{Gender}} = \{f, m\} \]

  - For any state $x$ of a random variable $X$, let
    \[ \Omega_{X=x} = \{\omega \in \Omega | X(\omega) = x\} \]
    
    **This is an event**
Because of Kolmogorov’s axioms, a probability mass function completely determines a probability measure.
Frequentist interpretation

- **Frequentist interpretation:**
- Probability is long term frequency

- **Example:**
  - $X$ is result of coin tossing. $\Omega_X = \{H, T\}$
  - $P(X=H) = 1/2$ means that
    - *the frequency of getting heads* approaches $1/2$ as the number of tosses goes to infinite.
  - Justified by the Law of Large Numbers:
    - $X_i$: result of the i-th tossing; $1 - H, 0 — T$
    - Law of Large Numbers:
      \[
      \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1}{2} \quad \text{with probability 1}
      \]
  - The frequentist interpretation is meaningful only when experiment can be repeated.

Subjectivist interpretation

- **Probabilities are logically consistent degrees of beliefs.**

- Comes into play when experiment not repeatable.

- Depends on a person’s background knowledge.

- Subjective: another person with different background knowledge might have different probability.

- Experiment not repeatable. If I go to library and find out the truth, my background knowledge is no longer the same.

- The subjectivist interpretation was not widely accepted until 1970s
This is a major reason why probability theory did not play a big role in AI before 1980.

- Because probability was defined as statistical frequency and hence was seen as a technique that was appropriate only when statistical data were available.
- Not many interesting applications with statistical data at that time.
  Now, more common.

Now both interpretations are accepted. In practice, subjective beliefs and statistical data complement each other.

- We rely on subjective beliefs (prior probabilities) when data are scarce.
- As more and more data become available, we rely less and less on subjective beliefs.

As we will learn later, probability has a numerical aspect as well as a structural aspect.

- We will rely more on the subjectivity interpretation when it comes to building structures than estimating numbers. Our belief on “causality” often plays an important role when building structures.

The subjectivist interpretation makes concepts such as conditional independence easy to understand.
Joint probability mass function

- **Probability mass function** of a random variable $X$:
  \[ P(X) : \Omega_X \to [0, 1] \]

- Suppose there are $n$ random variables $X_1, X_2, \ldots, X_n$.
- A **joint probability mass function**, $P(X_1, X_2, \ldots, X_n)$, over those random variables is:
  - a probability mass function defined on the Cartesian product of their state spaces:
  \[ \prod_{i=1}^{n} \Omega_{X_i} \to [0, 1] \]

Joint probability distribution

- The joint distribution $P(X_1, X_2, \ldots, X_n)$ contains information about all aspects of the relations among the $n$ random variables.
- In theory, one can answer any query about relations among the variables based on the joint probability.
Example:

- Population: Apartments in Hong Kong rental market.
- Random variables: (of a random selected apartment)
  - Monthly Rent: \{low ($\leq 1k$), medium ((1k, 2k]), upper medium((2k, 4k]), high ($\geq 4k$)}
  - Type: \{public, private, others\}
- Joint probability distribution $P(\text{Rent, Type})$:

<table>
<thead>
<tr>
<th></th>
<th>public</th>
<th>private</th>
<th>others</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>.17</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>medium</td>
<td>.44</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td>upper medium</td>
<td>.09</td>
<td>.07</td>
<td>.01</td>
</tr>
<tr>
<td>high</td>
<td>0</td>
<td>.14</td>
<td>.1</td>
</tr>
</tbody>
</table>

What is the probability of a randomly selected apartment being a public one?

\[
P(\text{Type}=\text{public}) = P(\text{Type}=\text{public, Rent}=\text{low})+P(\text{Type}=\text{public, Rent}=\text{medium})+P(\text{Type}=\text{public, Rent}=\text{upper medium})+P(\text{Type}=\text{public, Rent}=\text{high}) = .7
\]

\[
P(\text{Type}=\text{private}) = P(\text{Type}=\text{private, Rent}=\text{low})+P(\text{Type}=\text{private, Rent}=\text{medium})+P(\text{Type}=\text{private, Rent}=\text{upper medium})+P(\text{Type}=\text{private, Rent}=\text{high}) = .25
\]

<table>
<thead>
<tr>
<th></th>
<th>public</th>
<th>private</th>
<th>others</th>
<th>$P(\text{Rent})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>.17</td>
<td>.01</td>
<td>.02</td>
<td>.2</td>
</tr>
<tr>
<td>medium</td>
<td>.44</td>
<td>.03</td>
<td>.01</td>
<td>.48</td>
</tr>
<tr>
<td>upper medium</td>
<td>.09</td>
<td>.07</td>
<td>.01</td>
<td>.17</td>
</tr>
<tr>
<td>high</td>
<td>0</td>
<td>.14</td>
<td>.1</td>
<td>.15</td>
</tr>
</tbody>
</table>

- Called marginal probability because written on the margins.
Marginal probability

\[ P(\text{Type}) = \sum_{\text{Rent}} P(\text{Type, Rent}) \]

- The operation is called **marginalization**: Variable "Rent" is marginalized from the joint probability \( P(\text{Type, Rent}) \).

- Notations for more general cases:
  - \( P(X, Y) = \sum_{U,V} P(X, Y, U, V) \).
  - \( Y \subset \{X_1, X_2, \ldots, X_n\} \), \( Z = \{X_1, X_2, \ldots, X_n\} - Y \),
  - \( P(Y) = \sum_{Z} P(X_1, X_2, \ldots, X_n) \)

- A joint probability gives us a full picture about how random variables are related.
- Marginalization lets us to focus one aspect of the picture.
The probabilistic approach to reasoning under uncertainty

- A problem domain is modeled by a list of variables $X_1, X_2, \ldots, X_n$.
- Knowledge about the problem domain is represented by a joint probability $P(X_1, X_2, \ldots, X_n)$.

**Example:** Alarm (Pearl 1988)

- **Story:** In LA, burglary and earthquake are not uncommon. They both can cause alarm. In case of alarm, two neighbors John and Mary may call.
- **Problem:** Estimate the probability of a burglary based who has or has not called.
- **Variables:** Burglary (B), Earthquake (E), Alarm (A), JohnCalls (J), MaryCalls (M).
- **Knowledge required by the probabilistic approach in order to solve this problem:**

$$P(B, E, A, J, M)$$

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>J</th>
<th>M</th>
<th>Prob</th>
<th>B</th>
<th>E</th>
<th>A</th>
<th>J</th>
<th>M</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>.00001</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>.000025</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>.0004</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>.000025</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>.0004</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>.000000</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>.00001</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>.0000015</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>.000015</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>.000000</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>.00001</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>.0001</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>.000025</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>.000025</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>.0002</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>.000000</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>.0001</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>.000001</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>.0001</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>.000001</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>.0001</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>.000001</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>.0001</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>.000000</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>.996</td>
</tr>
</tbody>
</table>
Inference with joint probability distribution

- What is the probability of burglary given that Mary called, \( P(B=y|M=y) \)?
- Compute marginal probability:

\[
P(B, M) = \sum_{E, A, J} P(B, E, A, J, M)
\]

<table>
<thead>
<tr>
<th>B</th>
<th>M</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>y</td>
<td>.000115</td>
</tr>
<tr>
<td>y</td>
<td>n</td>
<td>.000075</td>
</tr>
<tr>
<td>n</td>
<td>y</td>
<td>.000150</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>.99971</td>
</tr>
</tbody>
</table>

- Compute answer (reasoning by conditioning):

\[
P(B=y|M=y) = \frac{P(B=y, M=y)}{P(M=y)} = \frac{.000115}{.000115 + 000075} = 0.61
\]

Conditional probability

- For events \( A \) and \( B \):

\[
P(A|B) = \frac{P(A, B)}{P(B)}
\]

- Meaning:
  - \( P(A) \): my probability on \( A \) (without any knowledge about \( B \))
  - \( P(A|B) \): My probability on event \( A \) assuming that I know event \( B \) is true.

- What is the probability of a randomly selected private apartment having “low” rent?

\[
P(Rent=low|Type=private) = \frac{P(Rent=Low, Type=private)}{P(Type=private)} = 0.01/.25 = .04
\]

In contrast:

\[
P(Rent=low) = 0.2.
\]
Properties of Conditional Probability

- The conditional probability of an event $A$, given an event $B$ with $P(B) > 0$, is defined by
  \[ P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \]

  and specifies a new (conditional) probability law on the same sample space $\Omega$. In particular, all properties of probability laws remain valid for conditional probability laws.

- Conditional probabilities can also be viewed as a probability law on a new universe $B$, because all of the conditional probability is concentrated on $B$.

- If the possible outcomes are finitely many and equally likely, then
  \[ P(A \mid B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B}. \]

---

Example 1.9. **Radar Detection.** If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present, the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of no aircraft presence and a false alarm? What is the probability of aircraft presence and no detection?

$A = \{ \text{an aircraft is present} \}$,

$B = \{ \text{the radar generates an alarm} \}$,

$A^c = \{ \text{an aircraft is not present} \}$,

$B^c = \{ \text{the radar does not generate an alarm} \}$. 
\( \mathbb{P}(\text{not present, false alarm}) = \mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B \mid A^c) = 0.95 \cdot 0.10 = 0.095, \)

\( \mathbb{P}(\text{present, no detection}) = \mathbb{P}(A \cap B^c) = \mathbb{P}(A)\mathbb{P}(B^c \mid A) = 0.05 \cdot 0.91 = 0.0005. \)

- \( \mathbb{P}(\text{Rent} \mid \text{Type}) \)

<table>
<thead>
<tr>
<th>Type</th>
<th>public</th>
<th>private</th>
<th>others</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>.17/.7</td>
<td>.01/.25</td>
<td>.02/.05</td>
</tr>
<tr>
<td>medium</td>
<td>.44/.7</td>
<td>.03/.25</td>
<td>.01/.05</td>
</tr>
<tr>
<td>upper medium</td>
<td>.09/.7</td>
<td>.07/.25</td>
<td>.01/.05</td>
</tr>
<tr>
<td>high</td>
<td>0/.7</td>
<td>0.14/.25</td>
<td>0.1/.05</td>
</tr>
</tbody>
</table>

- Note that

\[
\sum_{\text{Rent}} \mathbb{P}(\text{Rent} \mid \text{Type}) = 1.
\]
Marginal independence

- Two random variables $X$ and $Y$ are marginally independent, written $X \perp Y$, if
  - for any state $x$ of $X$ and any state $y$ of $Y$,
    \[ P(X=x|Y=y) = P(X=x), \text{ whenever } P(Y=y) \neq 0. \]

- Meaning: Learning the value of $Y$ does not give me any information about $X$ and vice versa. $Y$ contains no information about $X$ and vice versa.

- Equivalent definition:
  \[ P(X=x, Y=y) = P(X=x)P(Y=y) \]

- Shorthand for the equations:
  \[ P(X|Y) = P(X), P(X, Y) = P(X)P(Y). \]

- Examples:
  - $X$: result of tossing a fair coin for the first time,
    $Y$: result of second tossing of the same coin.
  - $X$: result of US election, $Y$: your grades in this course.

- Counter example: $X$ – oral presentation grade, $Y$ – project report grade.
Conditional independence

- Two random variables $X$ and $Y$ are conditionally independent given a third variable $Z$, written $X \perp Y \mid Z$, if

$$P(X=x \mid Y=y, Z=z) = P(X=x \mid Z=z) \text{ whenever } P(Y=y, Z=z) \neq 0$$

- Meaning:
  - If I know the state of $Z$ already, then learning the state of $Y$ does not give me additional information about $X$.
  - $Y$ might contain some information about $X$.
  - However all the information about $X$ contained in $Y$ are also contained in $Z$.

- Shorthand for the equation:

$$P(X \mid Y, Z) = P(X \mid Z)$$

- Equivalent definition:

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

- There is a bag of 100 coins. 10 coins were made by a malfunctioning machine and are biased toward head. Tossing such a coin results in head 80% of the time. The other coins are fair.

- Randomly draw a coin from the bag and toss it a few times.

- $X_i$: result of the $i$-th tossing, $Y$: whether the coin is produced by the malfunctioning machine.

- The $X_i$'s are not marginally independent of each other:
  - If I get 9 heads in first 10 tosses, then the coin is probably a biased coin. Hence the next tossing will be more likely to result in a head than a tail.
  - Learning the value of $X_i$ gives me some information about whether the coin is biased, which in turn gives me some information about $X_j$. 

[Diagram of a tree with nodes labeled Coin Type, Toss 1 Result, Toss 2 Result, ..., Toss n Result]
However, they are conditionally independent given $Y$:

- If the coin is not biased, the probability of getting a head in one toss is $1/2$ regardless of the results of other tosses.
- If the coin is biased, the probability of getting a head in one toss is $80\%$ regardless of the results of other tosses.
- If I already knows whether the coin is biased or not, learning the value of $X_i$ does not give me additional information about $X_j$.

Total Probability Theorem

Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in exactly one of the events $A_1, \ldots, A_n$) and assume that $P(A_i) > 0$, for all $i$. Then, for any event $B$, we have

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)$$

$$= P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n).$$
Example 1.13. You enter a chess tournament where your probability of winning a game is 0.3 against half the players (call them type 1), 0.4 against a quarter of the players (call them type 2), and 0.5 against the remaining quarter of the players (call them type 3). You play a game against a randomly chosen opponent. What is the probability of winning?

Let \( A_i \) be the event of playing with an opponent of type \( i \). We have

\[
P(A_1) = 0.5, \quad P(A_2) = 0.25, \quad P(A_3) = 0.25.
\]

Also, let \( B \) be the event of winning. We have

\[
P(B \mid A_1) = 0.3, \quad P(B \mid A_2) = 0.4, \quad P(B \mid A_3) = 0.5.
\]

Thus, by the total probability theorem, the probability of winning is

\[
P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3)
\]

\[
= 0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5
\]

\[
= 0.375.
\]

**Prior, posterior, and likelihood**

- **Prior probability**: belief about a hypothesis \( h \) before obtaining observations, \( P(h) \).

  - Example: Suppose 10% of people suffer from Hepatitis B. A doctor’s prior probability about a new patient suffering from Hepatitis B is 0.1.

- **Posterior probability**: belief about a hypothesis after obtaining observations.

- **Likelihood** of hypothesis given observation:
  - Conditional probability of observation given hypothesis \( L(h \mid o) = P(o \mid h) \)
  - Example: \( o \): eye-color=yellow; \( h_1 \): Hepatitis B; \( h_2 \): no Hepatitis B
    \[
P(o \mid h_1) > P(o \mid h_2)
\]

If we observe \( o \), \( h_1 \) is more likely than \( h_2 \).

As a function of \( h \), \( P(o \mid h) \) measures the likelihood of \( h \).
**Bayes’ Theorem**

- **Bayes’ Theorem**: relates prior probability, likelihood, and posterior probability:

\[
P(h|o) = \frac{P(h)P(o|h)}{P(o)} \propto P(h)P(o|h) = P(h)L(h|o)
\]

where \( P(o) \) is normalization constant to ensure \( \sum_h P(h|o) = 1 \).

In words: \( \text{posterior} \propto \text{prior} \times \text{likelihood} \)

- **Example**:

\[
P(\text{disease}|\text{symptoms}) = \frac{P(\text{disease})P(\text{symptoms}|\text{disease})}{P(\text{symptoms})}
\]

- \( P(\text{symptom}|\text{disease}) \) from understanding of disease,
- \( P(\text{disease}|\text{symptoms}) \) needed in clinical diagnosis.
Let us return to the radar detection problem

\[ A = \{ \text{an aircraft is present} \}, \]
\[ B = \{ \text{the radar generates an alarm} \}. \]

We are given that

\[ P(A) = 0.05, \quad P(B \mid A) = 0.99, \quad P(B \mid A^c) = 0.1. \]

Applying Bayes’ rule, with \( A_1 = A \) and \( A_2 = A^c \), we obtain

\[
P(\text{aircraft present} \mid \text{alarm}) = P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)} = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^c)P(B \mid A^c)}
\]
\[
= \frac{0.05 \cdot 0.99}{0.05 \cdot 0.99 + 0.95 \cdot 0.1}
\]
\[
\approx 0.3426.
\]

Let us return to the chess problem

\[ P(A_1) = 0.5, \quad P(A_2) = 0.25, \quad P(A_3) = 0.25. \]

Also, \( B \) is the event of winning, and

\[ P(B \mid A_1) = 0.3, \quad P(B \mid A_2) = 0.4, \quad P(B \mid A_3) = 0.5. \]

Suppose that you win. What is the probability \( P(A_1 \mid B) \) that you had an opponent of type 1?

Using Bayes’ rule, we have

\[
P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3)}
\]
\[
= \frac{0.5 \cdot 0.3}{0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5}
\]
\[
= 0.4.
\]
In 1964 an interracial couple was convicted of robbery in Los Angeles, largely on the grounds that they matched a highly improbable profile, a profile which fit witness reports [272]. In particular, the two robbers were reported to be

- A man with a mustache
- Who was black and had a beard
- And a woman with a ponytail
- Who was blonde

- The couple was interracial
- And were driving a yellow car

The prosecution suggested that these characteristics had the following probabilities of being observed at random in the LA area:

1. A man with a mustache 1/4
2. Who was black and had a beard 1/10
3. And a woman with a ponytail 1/10
4. Who was blonde 1/3
5. The couple was interracial 1/1000
6. And were driving a yellow car 1/10

\[ P(e | \neg h) = \prod_{i} P(e_{i} | \neg h) = \frac{1}{12000000} \]

\[ e_{i} (i = 1, \ldots, 6), \text{ the joint evidence } e \]

A Much better estimate

\[ P(e_{2} | \neg h)P(e_{3} | \neg h)P(e_{4} | \neg h)P(e_{6} | \neg h) = \frac{1}{3000}. \]

The Bayesian approach

\[ P(h | e) = \frac{P(e | h)P(h)}{P(e | h)P(h) + P(e | \neg h)P(\neg h)} \]

\[ P(h | e) = \frac{P(h)}{P(h) + P(\neg h)/3000} \]
6.5 million people

this gives us 1,625,000 eligible males and as many females

\[
P(h\mid e) = \frac{1/1625000}{1/1625000 + (1 - 1/1625000)/3000} \approx 0.002
\]