Natural Language Processing

Joint, Conditional and Marginal Probability

Joakim Nivre

Uppsala University
Department of Linguistics and Philology
joakim.nivre@lingfil.uu.se
Conditional Probability

- Given events $A$ and $B$ in $\Omega$, with $P(B) > 0$, the conditional probability of $A$ given $B$ is:

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

- $P(A \cap B)$ or $P(A, B)$ is the joint probability of $A$ and $B$.
  - The prob that a person is rich and famous – joint
  - The prob that a person is rich if they are famous – conditional
  - The prob that a person is famous if they are rich – conditional
Conditional Probability

\[ P(A) = \text{size of } A \text{ relative to } \Omega \]

\[ P(A, B) = \text{size of } A \cap B \text{ relative to } \Omega \]

\[ P(A|B) = \text{size of } A \cap B \text{ relative to } B \]
Example

- We sample word bigrams (pairs) from a large text $T$
- Sample space and events:
  - $\Omega = \{(w_1, w_2) \in T\} =$ the set of bigrams in $T$
  - $A = \{(w_1, w_2) \in T \mid w_1 = \text{run}\} =$ bigrams starting with run
  - $B = \{(w_1, w_2) \in T \mid w_2 = \text{amok}\} =$ bigrams starting with amok
- Probabilities:
  - $P(\text{run}_1) = P(A) = 10^{-3}$
  - $P(\text{amok}_2) = P(B) = 10^{-6}$
  - $P(\text{run}_1, \text{amok}_2) = (A, B) = 10^{-7}$
- Probability of amok following run? Of run preceding amok?
Example

- We sample word bigrams (pairs) from a large text $T$
- Sample space and events:
  - $\Omega = \{(w_1, w_2) \in T\} =$ the set of bigrams in $T$
  - $A = \{(w_1, w_2) \in T \mid w_1 = \text{run}\} =$ bigrams starting with run
  - $B = \{(w_1, w_2) \in T \mid w_2 = \text{amok}\} =$ bigrams starting with amok
- Probabilities:
  - $P(\text{run}_1) = P(A) = 10^{-3}$
  - $P(\text{amok}_2) = P(B) = 10^{-6}$
  - $P(\text{run}_1, \text{amok}_2) = (A, B) = 10^{-7}$
- Probability of amok following run? Of run preceding amok?
  - $P(\text{run before amok}) = P(A \mid B) = \frac{10^{-7}}{10^{-6}} = 0.1$
  - $P(\text{amok after run}) = P(B \mid A) = \frac{10^{-7}}{10^{-3}} = 0.0001$
Multiplication Rule for Joint Probability

- Given events $A$ and $B$ in $\Omega$, with $P(B) > 0$:
  \[ P(A, B) = P(B)P(A|B) \]

- Since $A \cap B = B \cap A$, we also have:
  \[ P(A, B) = P(A)P(B|A) \]

- The multiplication rule is also known as the chain rule
Quiz 1

- The probability of winning the Nobel Prize if you have a PhD in Physics is 1 in a million \[ P(A|B) = 0.000001 \]
- Only 1 in 10,000 people have a PhD in Physics \[ P(B) = 0.0001 \]
- What is the probability of a person both having a PhD in Physics and winning the Nobel Prize? \[ P(A, B) = ? \]
  1. Smaller than 1 in a million
  2. Greater than 1 in a million
  3. Impossible to tell
Marginal Probability

- Marginalization, or the law of total probability
- If events $B_1, \ldots, B_k$ constitute a partition of the sample space $\Omega$ (and $P(B_i) > 0$ for all $i$), then for any event $A$ in $\Omega$:

\[
P(A) = \sum_{i=1}^{k} P(A, B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i)
\]

- Partition = pairwise disjoint and $B_1 \cup \cdots \cup B_k = \Omega$
Joint, Marginal and Conditional

- Joint probabilities for rain and wind:

<table>
<thead>
<tr>
<th></th>
<th>no wind</th>
<th>some wind</th>
<th>strong wind</th>
<th>storm</th>
</tr>
</thead>
<tbody>
<tr>
<td>no rain</td>
<td>0.1</td>
<td>0.2</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>light rain</td>
<td>0.05</td>
<td>0.1</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>heavy rain</td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Joint, Marginal and Conditional

- Joint probabilities for rain and wind:

<table>
<thead>
<tr>
<th></th>
<th>no wind</th>
<th>some wind</th>
<th>strong wind</th>
<th>storm</th>
</tr>
</thead>
<tbody>
<tr>
<td>no rain</td>
<td>0.1</td>
<td>0.2</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>light rain</td>
<td>0.05</td>
<td>0.1</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>heavy rain</td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
<td>0.05</td>
</tr>
</tbody>
</table>

- Marginalize to get simple probabilities:
  - \( P(\text{no wind}) = 0.1 + 0.05 + 0.05 = 0.2 \)
  - \( P(\text{light rain}) = 0.05 + 0.1 + 0.15 + 0.04 = 0.34 \)
Joint, Marginal and Conditional

Joint probabilities for rain and wind:

<table>
<thead>
<tr>
<th></th>
<th>no wind</th>
<th>some wind</th>
<th>strong wind</th>
<th>storm</th>
</tr>
</thead>
<tbody>
<tr>
<td>no rain</td>
<td>0.1</td>
<td>0.2</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>light rain</td>
<td>0.05</td>
<td>0.1</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>heavy rain</td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Marginalize to get simple probabilities:
- \( P(\text{no wind}) = 0.1 + 0.05 + 0.05 = 0.2 \)
- \( P(\text{light rain}) = 0.05 + 0.1 + 0.15 + 0.04 = 0.34 \)

Combine to get conditional probabilities:
- \( P(\text{no wind}|\text{light rain}) = \frac{0.05}{0.34} = 0.147 \)
- \( P(\text{light rain}|\text{no wind}) = \frac{0.05}{0.2} = 0.25 \)
Bayes Law

- Given events $A$ and $B$ in sample space $\Omega$:
  \[ P(A|B) = \frac{P(A)P(B|A)}{P(B)} \]

  - Follows from definition using chain rule
  - Allows us to “invert” conditional probabilities

- Denominator can be computed using marginalization:
  \[ P(B) = \sum_{i=1}^{k} P(B, A_i) = \sum_{i=1}^{k} P(B|A_i)P(A_i) \]

  - Special case of partition: $P(A)$, $P(\overline{A})$
Independence

- Two events \( A \) and \( B \) are independent if and only if:

\[
P(A, B) = P(A)P(B)
\]

- Equivalently:

\[
P(A) = P(A|B) \\
P(B) = P(B|A)
\]
Independence

- Two events $A$ and $B$ are independent if and only if:

\[ P(A, B) = P(A)P(B) \]

- Equivalently:

\[ P(A) = P(A|B) \]
\[ P(B) = P(B|A) \]

- Example:
  - $P(\text{run}_1) = P(A) = 10^{-3}$
  - $P(\text{amok}_2) = P(B) = 10^{-6}$
  - $P(\text{run}_1, \text{amok}_2) = (A, B) = 10^{-7}$

- $A$ and $B$ are not independent
Quiz 2

- Research has shown that people with disease $D$ exhibit symptom $S$ with 0.9 probability
- A doctor finds that a patient has symptom $S$
- What can we conclude about the probability that the patient has disease $D$
  1. The probability is 0.1
  2. The probability is 0.9
  3. Nothing