An illustration of Conditional Independence

Martin Emms

September 20, 2018
Suppose you have some data on people concerning two possible variables sea, which is whether they live by the seaside, and hip which is whether they have hip problems:

\[
\begin{align*}
\text{sea} : + & \quad \text{sea} : - \\
\text{hip} : + & \quad 31 \quad 54 \\
\text{hip} : - & \quad 19 \quad 146
\end{align*}
\]
Suppose you have some data on people concerning two possible variables sea, which is whether they live by the seaside, and hip which is whether they have hip problems:

\[
\begin{array}{ccc}
\text{sea} : + & \text{sea} : - \\
\hline
\text{hip} : + & 31 & 54 \\
\text{hip} : - & 19 & 146 \\
\end{array}
\]  

one of the formulations of independence is \( P(X|Y) = P(X) \). Lets apply that to sea and hip, in fact to the ’+’ settings of these variables

\[
p(\text{hip} : +) = \frac{31 + 54}{250} = 0.34
\]
Suppose you have some data on people concerning two possible variables sea, which is whether they live by the seaside, and hip which is whether they have hip problems:

\[
\begin{array}{ccc}
\text{sea} : + & \text{sea} : - \\
\text{hip} : + & 31 & 54 \\
\text{hip} : - & 19 & 146 \\
\end{array}
\]  

(1)

one of the formulations of independence is \( P(X|Y) = P(X) \). Let's apply that to sea and hip, in fact to the '+' settings of these variables

\[ p(\text{hip} : +) = (31 + 54)/250 = 0.34 \]

\[ p(\text{hip} : +|\text{sea} : +) = 31/(31 + 19) = 0.62 \]

so hip : + and sea : + are not independent; in fact sea-side living seems to increase the chance of hip problems, which seems weird
suppose that digging into the data a little further you find there was one other variable: old for whether or not person was old. There were 50 old and 200 not old, and when the data is split into two sub-groups according to the value old you find:

\[
\begin{array}{c|cc|c|cc}
old & sea : + & sea : - & \neg old & sea : + & sea : - \\
\hline
hip : + & 27 & 18 & hip : + & 4 & 36 \\
hip : - & 3 & 2 & hip : - & 16 & 144 \\
\end{array}
\]
suppose that digging into the data a little further you find there was one other variable: old for whether or not person was old. There were 50 old and 200 not old, and when the data is split into two sub-groups according to the value old you find:

\[
\begin{array}{c|ccc}
\text{old} & \text{sea: +} & \text{sea: -} \\
\hline
\text{hip: +} & 27 & 18 \\
\text{hip: -} & 3 & 2 \\
\text{\neg old} & & \\
\hline
\text{hip: +} & 4 & 36 \\
\text{hip: -} & 16 & 144 \\
\end{array}
\]

we can show that hip: + is conditionally independent of sea: + given old: +

\[p(\text{hip: +} | \text{old: +}) = \]
suppose that digging into the data a little further you find there was one other variable: old for whether or not person was old. There were 50 old and 200 not old, and when the data is split into two sub-groups according to the value old you find:

<table>
<thead>
<tr>
<th>old</th>
<th>sea: +</th>
<th>sea: -</th>
</tr>
</thead>
<tbody>
<tr>
<td>hip: +</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>hip: -</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>¬old</th>
<th>sea: +</th>
<th>sea: -</th>
</tr>
</thead>
<tbody>
<tr>
<td>hip: +</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>hip: -</td>
<td>16</td>
<td>144</td>
</tr>
</tbody>
</table>

we can show that hip:+ is conditionally independent of sea:+ given old:+

\[
p(hip: +|old : +) = \frac{45}{50} = \frac{9}{10}
\]

\[
p(hip: +|old : +, sea : +) = \]
suppose that digging into the data a little further you find there was one other variable: old for whether or not person was old. There were 50 old and 200 not old, and when the data is split into two sub-groups according to the value old you find:

\[
\begin{array}{c|cc|c|cc}
\text{old} & \text{sea} : + & \text{sea} : - & \neg \text{old} & \text{sea} : + & \text{sea} : - \\
\text{hip} : + & 27 & 18 & \text{hip} : + & 4 & 36 \\
\text{hip} : - & 3 & 2 & \text{hip} : - & 16 & 144 \\
\end{array}
\]

\[\text{we can show that hip:} + \text{ is conditionally independent of sea:} + \text{ given old:} + \]
\[
p(\text{hip} : + | \text{old} : +) = \frac{45}{50} = \frac{9}{10}
\]
\[
p(\text{hip} : + | \text{old} : +, \text{sea} : +) = \frac{27}{30} = \frac{9}{10}
\]

\[\text{we can show that hip:} + \text{ is conditionally independent of sea:} + \text{ given old:} - \]
\[
p(\text{hip} : + | \text{old} : -)
\]
suppose that digging into the data a little further you find there was one other variable: old for whether or not person was old. There were 50 old and 200 not old, and when the data is split into two sub-groups according to the value old you find:

\[
\begin{array}{c|cc|c|cc}
\text{old} & \text{sea : +} & \text{sea : -} & \neg \text{old} & \text{sea : +} & \text{sea : -} \\
\text{hip : +} & 27 & 18 & \text{hip : +} & 4 & 36 \\
\text{hip : -} & 3 & 2 & \text{hip : -} & 16 & 144 \\
\end{array}
\] (2)

- we can show that hip: + is conditionally independent of sea: + given old: +
  \[
p(\text{hip : +} | \text{old : +}) = \frac{45}{50} = \frac{9}{10}
\]
  \[
p(\text{hip : +} | \text{old : +, sea : +}) = \frac{27}{30} = \frac{9}{10}
\]
- we can show that hip: + is conditionally independent of sea: + given old: -
  \[
p(\text{hip : +} | \text{old : -}) = \frac{40}{200} = \frac{1}{5}
\]
  \[
p(\text{hip : +} | \text{old : -, sea : +}) =
\]
suppose that digging into the data a little further you find there was one other variable: \texttt{old} for whether or not person was old. There were 50 old and 200 not old, and when the data is split into two sub-groups according to the value \texttt{old} you find:

\begin{array}{c|cc|c|cc}
\text{old} & \text{sea} : + & \text{sea} : - & \neg \text{old} & \text{sea} : + & \text{sea} : - \\
\hline
\text{hip} : + & 27 & 18 & \text{hip} : + & 4 & 36 \\
\text{hip} : - & 3 & 2 & \text{hip} : - & 16 & 144 \\
\end{array}

\begin{itemize}
\item we can show that \texttt{hip}:+ is conditionally independent of \texttt{sea}:+ given \texttt{old}:+
\[ p(\text{hip} : + | \text{old} : +) = \frac{45}{50} = \frac{9}{10} \]
\[ p(\text{hip} : + | \text{old} : +, \text{sea} : +) = \frac{27}{30} = \frac{9}{10} \]
\item we can show that \texttt{hip}:+ is conditionally independent of \texttt{sea}:+ given \texttt{old}:-
\[ p(\text{hip} : + | \text{old} : -) = \frac{40}{200} = \frac{1}{5} \]
\[ p(\text{hip} : + | \text{old} : -, \text{sea} : +) = \frac{4}{20} = \frac{1}{5} \]
\item so zeroing in old people, sea-side living does not to increase the chance of hip problems; zeroing in on young people, it doesn’t either
\end{itemize}
An illustration of Conditional Independence

once you have a conditional independence it means that you can use the chain rule and use the conditional independence to simplify. We will see this in other examples; in the current case you could do this to get relatively simple formula for $p(\text{old}, \text{sea}, \text{hip})$

\[
p(\text{old}, \text{sea}, \text{hip}) = p(\text{hip}|\text{sea}, \text{old}) \times p(\text{sea}|\text{old}) \times p(\text{old}) \tag{3}
\]

\[
= p(\text{hip}|\text{old}) \times p(\text{sea}|\text{old}) \times p(\text{old}) \tag{4}
\]

(3) is just applying the chain rule and holds without any independence assumptions

(4) is the simplification which is possibly by putting in the conditional independence that $p(\text{hip}|\text{sea}, \text{old}) = p(\text{hip}|\text{old})$