An illustration of Conditional Independence

Martin Emms

September 18, 2019

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:

$$\begin{array}{lll} & sea:+ & sea:- \\ hip:+ & 31 & 54 \\ hip:- & 19 & 146 \end{array} \tag{1}$$

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:

$$\begin{array}{lll} & sea:+ & sea:- \\ hip:+ & 31 & 54 \\ hip:- & 19 & 146 \end{array} \tag{1}$$

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

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:

$$\begin{array}{lll} & sea:+ & sea:- \\ hip:+ & 31 & 54 \\ hip:- & 19 & 146 \end{array} \tag{1}$$

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

$p(hip:+) = (31+54)/250 = 0.34$

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:

$$\begin{array}{lll} & sea:+ & sea:- \\ hip:+ & 31 & 54 \\ hip:- & 19 & 146 \end{array} \tag{1}$$

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

$p(hip:+) = (31+54)/250 = 0.34$

$p(hip:+|sea:+) = 31/(31+19) = 0.62$

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:

$$\begin{array}{lll} & sea:+ & sea:- \\ hip:+ & 31 & 54 \\ hip:- & 19 & 146 \end{array} \tag{1}$$

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

$p(hip:+) = (31+54)/250 = 0.34$

$p(hip:+|sea:+) = 31/(31+19) = 0.62$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

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

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) =$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) =$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$

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:

old	*sea* : +	*sea* : −
hip : +	27	18
hip : −	3	2

¬*old*	*sea* : +	*sea* : −
hip : +	4	36
hip : −	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$
- we can show that **hip:+ is conditionally independent of sea:+ given 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:

old	*sea* : +	*sea* : −
hip : +	27	18
hip : −	3	2

¬*old*	*sea* : +	*sea* : −
hip : +	4	36
hip : −	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$
- we can show that **hip:+ is conditionally independent of sea:+ given old:-**
  $p(hip:+|old:-) =$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$
- we can show that **hip:+ is conditionally independent of sea:+ given old:-**
  $p(hip:+|old:-) = 40/200 = 1/5$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$
- we can show that **hip:+ is conditionally independent of sea:+ given old:-**
  $p(hip:+|old:-) = 40/200 = 1/5$
  $p(hip:+|old:-, sea:+) =$

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:

old	*sea* : +	*sea* : −
hip : +	27	18
hip : −	3	2

¬*old*	*sea* : +	*sea* : −
hip : +	4	36
hip : −	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$
- we can show that **hip:+ is conditionally independent of sea:+ given old:-**
  $p(hip:+|old:-) = 40/200 = 1/5$
  $p(hip:+|old:-, sea:+) = 4/20 = 1/5$

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:

old	$sea:+$	$sea:-$
$hip:+$	27	18
$hip:-$	3	2

$\neg old$	$sea:+$	$sea:-$
$hip:+$	4	36
$hip:-$	16	144

(2)

- we can show that **hip:+ is conditionally independent of sea:+ given old:+**
  $p(hip:+|old:+) = 45/50 = 9/10$
  $p(hip:+|old:+, sea:+) = 27/30 = 9/10$
- we can show that **hip:+ is conditionally independent of sea:+ given old:-**
  $p(hip:+|old:-) = 40/200 = 1/5$
  $p(hip:+|old:-, sea:+) = 4/20 = 1/5$
- so zeroing in old people, sea-side living does **not** increase the chance of hip problems; zeroing in on young people, it doesn't either

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 $p(old, sea, hip)$

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 $p(old, sea, hip)$

$$
\begin{aligned}
p(old, sea, hip) &= p(hip|sea, old) \times p(sea|old) \times p(old) && (3)\\
&= p(hip|old) \times p(sea|old) \times p(old) && (4)
\end{aligned}
$$

(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 $p(hip|sea, old) = p(hip|old)$