

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:

	<i>sea</i> : +	<i>sea</i> : -	(1)
<i>hip</i> : +	31	54	
<i>hip</i> : -	19	146	

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:

	<i>sea</i> : +	<i>sea</i> : -	
<i>hip</i> : +	31	54	(1)
<i>hip</i> : -	19	146	

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

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

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:

	<i>sea</i> : +	<i>sea</i> : -	
<i>hip</i> : +	31	54	(1)
<i>hip</i> : -	19	146	

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

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

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

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:

<i>old</i>	<i>sea</i> : +	<i>sea</i> : -	\neg <i>old</i>	<i>sea</i> : +	<i>sea</i> : -	(2)
<i>hip</i> : +	27	18	<i>hip</i> : +	4	36	
<i>hip</i> : -	3	2	<i>hip</i> : -	16	144	

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:

<i>old</i>	<i>sea</i> : +	<i>sea</i> : -	\neg <i>old</i>	<i>sea</i> : +	<i>sea</i> : -
<i>hip</i> : +	27	18	<i>hip</i> : +	4	36
<i>hip</i> : -	3	2	<i>hip</i> : -	16	144

(2)

- ▶ 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:

<i>old</i>	<i>sea</i> : +	<i>sea</i> : -	\neg <i>old</i>	<i>sea</i> : +	<i>sea</i> : -
<i>hip</i> : +	27	18	<i>hip</i> : +	4	36
<i>hip</i> : -	3	2	<i>hip</i> : -	16	144

(2)

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

$$p(\text{hip} : + | \text{old} : +) = 45/50 = 9/10$$

$$p(\text{hip} : + | \text{old} : +, \text{sea} : +) =$$

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:

<i>old</i>	<i>sea</i> : +	<i>sea</i> : -	\neg <i>old</i>	<i>sea</i> : +	<i>sea</i> : -
<i>hip</i> : +	27	18	<i>hip</i> : +	4	36
<i>hip</i> : -	3	2	<i>hip</i> : -	16	144

(2)

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

$$p(\text{hip} : + | \text{old} : +) = 45/50 = 9/10$$

$$p(\text{hip} : + | \text{old} : +, \text{sea} : +) = 27/30 = 9/10$$

- ▶ 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:

<i>old</i>	<i>sea</i> : +	<i>sea</i> : -	\neg <i>old</i>	<i>sea</i> : +	<i>sea</i> : -
<i>hip</i> : +	27	18	<i>hip</i> : +	4	36
<i>hip</i> : -	3	2	<i>hip</i> : -	16	144

(2)

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

$$p(\text{hip} : + | \text{old} : +) = 45/50 = 9/10$$

$$p(\text{hip} : + | \text{old} : +, \text{sea} : +) = 27/30 = 9/10$$

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

$$p(\text{hip} : + | \text{old} : -) = 40/200 = 1/5$$

$$p(\text{hip} : + | \text{old} : -, \text{sea} : +) =$$

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:

<i>old</i>	<i>sea</i> : +	<i>sea</i> : -	\neg <i>old</i>	<i>sea</i> : +	<i>sea</i> : -
<i>hip</i> : +	27	18	<i>hip</i> : +	4	36
<i>hip</i> : -	3	2	<i>hip</i> : -	16	144

(2)

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

$$p(\text{hip} : + | \text{old} : +) = 45/50 = 9/10$$

$$p(\text{hip} : + | \text{old} : +, \text{sea} : +) = 27/30 = 9/10$$

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

$$p(\text{hip} : + | \text{old} : -) = 40/200 = 1/5$$

$$p(\text{hip} : + | \text{old} : -, \text{sea} : +) = 4/20 = 1/5$$

- ▶ 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

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 $p(old, sea, hip)$

$$p(old, sea, hip) = p(hip|sea, old) \times p(sea|old) \times p(old) \quad (3)$$

$$= p(hip|old) \times p(sea|old) \times p(old) \quad (4)$$

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