Conditional independence, log-linear models

Marianna Bolla *

Institute of Mathematics, Budapest University of Technology and Economics

June 7, 2020

1 Basic definitions

The events A and B are conditionally independent, conditioned on C if

$$\mathbb{P}(AB|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C). \tag{1}$$

Show that this is equivalent to

$$\mathbb{P}(A|BC) = \mathbb{P}(A|C) \quad \text{and} \quad \mathbb{P}(B|AC) = \mathbb{P}(B|C).$$

So, knowing C, B does not give extra information for A; or knowing C, A does not give extra information for B. This fact is denoted by

$$A \perp\!\!\!\perp B \mid C.$$

Two random variables X and Y are conditionally independent, conditioned on Z if the events $X \in A$ and $Y \in B$ are are conditionally independent, conditioned on $Z \in C$, for any subsets A, B, C in their ranges. This fact is denoted by

$$X \perp\!\!\!\perp Y \mid Z,$$

which also means that given the value of Z, the rv's X and Y take on values independently.

The product rule (1) extends to the probability mass functions (in the discrete case) and to the probability density functions (in the continuous case).

2 Simpson paradox

Remark 1 *Consider three Gaussian variables X, Y, W, with X being the predictor variable, Y being the response variable, and W being the background variable. Cox and Wermuth investigated when the usual regression coefficient β_{YX}*

*marib@math.bme.hu

is the same as the partial regression coefficient $\beta_{YX|W}$ *(they called it 'no effect reversal'). For simplicity, assume that our variables are already standardized (have zero expectation and unit variance). Then* $\beta_{YX} = r_{YX}$ *is the Pearson correlation, and* $\beta_{YX|W} = r_{YX|W}$ *is the partial correlation coefficient of* X *on* Y *without or with conditioning on* W*. Then in view of*

$$r_{YX} = r_{YX|W} + \beta_{YW|X} \times r_{XW} \tag{2}$$

(Cochran formula), $r_{YX} = r_{YX|W}$ *if and only if either* $r_{XW} = 0$ *or* $\beta_{YW|X} = 0$*. This means that there is no effect reversal if either* X *and* W *are marginally independent or* Y *is independent of* W *given* X*, i.e.,*

$$X \perp\!\!\!\perp W \quad or \quad Y \perp\!\!\!\perp W|X. \tag{3}$$

Apply the Remark in the two examples (qualifying exercise).

- Murder case:

  $$Y = S(entence), \quad X = M(urderer), \quad W = V(ictim).$$

  Here the conditions in (3) do not hold. Therefore, W has 'reversing effect'.

- Danish women:

  $$\begin{aligned} Y &= \text{U(ses physical punishment)}, \\ X &= \text{E(xperience of p.p. in childhood)}, \\ W &= \text{A(ffiliation to political parties)}. \end{aligned}$$

  In wiew of the above, here the $Y \perp\!\!\!\perp W|X$ condition holds, i.e., $U \perp\!\!\!\perp A|E$. Therefore, A won't cause effect reversal as for the relation between U and E.

  We can check in the sample that approximately $U \perp\!\!\!\perp A|E$, or equivalently

  $$\mathbb{P}(U = u \,|\, A = a,\, E = e) = \mathbb{P}(U = u \,|\, E = e).$$

  Actually, when we made a U versus A cross-tabulation, and estimated the marginals, the 2×3 table manufactured as the product of the marginal probabilities, was very close to the original layers of the $2 \times 2 \times 3$ table, both in the $E = y$ and $E = n$ cases. We can also make a χ^2 test on the two 2×3 tables.

  Table 1: Results from the χ^2 tests for independence

Null Hypothesis (H_0)	$U \perp\!\!\!\perp A$	$U \perp\!\!\!\perp A\|E = \text{y}$	$U \perp\!\!\!\perp A\|E = \text{n}$
Degree of Freedom	2	2	2
χ^2 Value	5.25	4.08	0.42
Significance	0.072	0.130	0.810
Conclusion (7.5%)	Reject	Do not reject	Do not reject

  So we conclude that A and U are not marginally independent, but they are conditionally independent, conditioned on E.

We use the notation: $A \leftarrow E \rightarrow U$ and write

$$\mathbb{P}(U = u, A = a \mid E = e) = \mathbb{P}(U = u \mid A = a, E = e) \times \mathbb{P}(A = a \mid E = e) = \mathbb{P}(U = u \mid E = e) \times \mathbb{P}(A = a \mid E$$

or with probability mass functions:

$$p(u, a|e) = p(u|e) \times p(a|e)$$

for any values of u, a, e. Equivalently,

$$\frac{p(u, a, e)}{p(e)} = \frac{p(u, e)}{p(e)} \times \frac{p(a, e)}{p(e)}$$

or

$$p(u, a, e) = \frac{p(u, e) \times p(a, e)}{p(e)} = p(u, e) \times p(a|e).$$

Note that $U - E$ and $A - E$ are the *cliques* (maximal complete subgraphs) of the graph $U - E - A$, and E is the separator between them. We can as well state this as a log-linear model:

$$\log p(u, a, e) = f_{eu} + f_{ea}.$$

3 Constructing a discrete log-linear, decomposable model

Based on some expert knowledge, construct a Bayesian network, and make it an undirected, decomposable graph (see the Lauritzen–Spiegelhalter paper). With the help of conditional probability tables, find marginals, etc.

For example, COVID19. Possible variables: travel to Asia, travel to Latin-America, gender, age, social relations, previous illnesses, chronic illnesses, symptoms, previous vaccinations.