(Discrete) Conditional Independence as Non-Negative Matrix Factorization

Gonzalo Mena

November 6, 2017

Reichenbach's common cause principle states: If two random variables X and Y are statistically dependent ($X \not\perp\!\!\!\perp Y$), then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y). Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y \mid Z$.

One natural question is whether this principle derives, in some sense, from the usual axioms of probability, or if it rather a working hypothesis. Although these notes a far from providing a definite answer to this question, I hope at least they may provide insights. Let's focus on the statement "statistical dependence implies the existence of a variable Z that screens X and Y of each other". Further, let's assume that X, Y take a finite number of values; n and m, respectively.

The claim is the following: there exists a finite random variable Z taking k values that screens X, Y if an only if the joint density of X and Y, P, can be factored as $P = UV^\top$ with U, V non-negative matrices with dimensions $n \times k$ and $m \times k$, respectively.

Indeed, notice that conditional independence states as :

$$P(X = x, Y = y) = \sum_z P(X = x|Z = z)P(Y = y|Z = z)P(Z = z) \tag{1}$$

Also, suppose $P = UV^\top$ holds and define the following row sums:

$$S_z^U = \sum_x U_{x,z}, \quad S_z^V = \sum_y V_{y,z}.$$

Further, define

$$U'_{x,z} = \frac{U_{x,z}}{S_z^U}, \quad V'_{y,z} = \frac{V_{y,z}}{S_z^V}.$$

Then,

$$P_{x,y} = \sum_z U_{x,z}V_{y,z} = (U'_{x,z})(V'_{y,z})(S_z^U S_z^V).$$

Importantly, one has

$$\sum_z S_z^U S_z^V = \sum_{z,x,y} = U_{x,z}V_{y,z} = \sum_{x,y} P_{x,y} = 1$$

Therefore, it is rightful to define Z as having the law $P(Z = z) = S_z^U S_z^V$. Likewise, one can define the conditionals $P(X = x|Z = z) = U'_{x,z}$ and $P(Y = y|Z = z) = V'_{y,z}$. The algebra above shows these arbitrary choices are consistent with X, Y having a joint density P, and entail the conditional independence statement in equation (1).

Importantly, without loss of generality $S_z^U, S_z^V > 0$. Otherwise, one can safely reduce the dimension of the decomposition by ignoring problematic z's.

Discussion Now we can take a new look into Reichenbach's principle: first, notice that even vanilla independence $X \perp\!\!\!\perp Y$ frames as conditional independence with Z taking a single value; equivalently, as a rank-one factorization. Also, it is known that existence of a non-trivial non-negative matrix factorization (i.e., beyond $P = PI_m = I_nP$) is guaranteed in the general case (Vasiloglou et al., 2009), although the problem is NP-Hard and no general algorithms are known (Arora et al., 2016). Therefore, pairs of discrete random variables always either i) conditionally independent (rank-one factorization) or ii) there exists another variable that explains them away (greater rank).

Unfortunately, the causal statement in Reichenbach's principle appears disconnected from the conditional independence statement, so it is hard to draw a logical conclusion.

Finally, the above formulae may provide means to discover causal structures in Reichenbach's sense, by estimating the joint $P_{x,y}$ (how?) and then applying some (approximate) matrix factorization algorithm (Lee and Seung, 1999).

References

Arora, S., Ge, R., Kannan, R., and Moitra, A. (2016). Computing a nonnegative matrix factorization—provably. *SIAM Journal on Computing*, 45(4):1582–1611.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Vasiloglou, N., Gray, A. G., and Anderson, D. V. (2009). Non-negative matrix factorization, convexity and isometry. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pages 673–684. SIAM.