BSM RES 2020 summer

Marianna Bolla *

July 12, 2020

This is an applied research, but we must use the theory behind it.

1 COVID-graph

We follow the theory of the Lauritzen–Spiegelhalter [5] and Bolla et al. [2] papers. You can cite page 165 of [5]. If the graph is triangulated and so, there is a JT structure (decomposable model), then the joint probability factorizes as

$$p(\mathbf{x}) = \frac{\prod_{j=1}^{k} p(\mathbf{x}_{C_j})}{\prod_{j=2}^{k} p(\mathbf{x}_{S_j})} = \prod_{j=1}^{k} p(\mathbf{x}_{R_j} \,|\, \mathbf{x}_{S_j}),$$

where p is the probability mass function in the discrete case, and $S_1 = \emptyset$ in the RIP ordering.

However, this factorization, in particular, the clique potentials are obtainable if we use the belief propagation with any starting potentials:

$$p \propto \prod_{j=1}^{k} \psi(\mathbf{x}_{C_j}).$$

This idea is well applicable in the current pandemic when we do not know exact probabilities or conditional probabilities, just some proportions between them.

In our causal graph, in the RIP ordering:

$$\begin{aligned} &C_1 = \{C, A, L\}, \quad R_1 = \{C, A, L\}, \quad S_1 = \emptyset \\ &C_2 = \{L, G, I\}, \quad R_2 = \{G, I\}, \quad S_2 = \{L\} \\ &C_3 = \{I, H\}, \quad R_3 = \{H\}, \quad S_3 = \{I\} \\ &C_4 = \{H, O\}, \quad R_4 = \{O\}, \quad S_4 = \{H\}. \end{aligned}$$

In this case, the factorization is

$$\frac{P(C, A, L) \cdot P(L, G, I) \cdot P(I, H) \cdot P(H, O)}{P(L) \cdot P(I) \cdot P(H)} = P(C, A, L) \cdot P(G, I|L) \cdot P(H|I) \cdot P(O|H).$$

But we do not need this factorization, we only need the potentials which have to be found for all state configurations:

$$\psi_1 = P(A) \cdot P(C|A) \cdot P(L|C, A), \quad \psi_2 = P(G) \cdot P(I|L, G), \quad \psi_3 = P(H|I), \quad \psi_4 = P(O|H).$$

*marib@math.bme.hu

Here ψ_1 is the same as $P(A, C, L)$ by the product rule (see STA course). ψ_3 and ψ_4 also correpond to the above factorization, but ψ_2 is not $P(G, I|L)$. You can put any product of probabilities or conditional probabilities containing all the clique variables and only those. RATHER USE LOWER CASE LETTERS FOR THE STATES! For example, $p(i|l, g)$ for all possible values of I, L, G.

For the cliques of the junction tree (in the reversed RIP = Sundberg's ordering), together with separators, we apply the so-called *belief propagation* algorithm so that we update the potentials in such a way, that at the end, they become the clique marginals. Let A and B be two consecutive cliques, and S be the separator between them. Starting with some potentials and denoting by * the newly updated potential, the algorithm is:

$$\begin{aligned}
\psi^*_S(\mathbf{x}_S) &= \sum_{\mathbf{x}_{A\setminus S}\in\mathcal{X}_{A\setminus S}} \psi_A(\mathbf{x}_S, \mathbf{x}_{A\setminus S}), && \forall \mathbf{x}_S \in \mathcal{X}_S \\
\psi^*_B(\mathbf{x}_B) &= \psi_B(\mathbf{x}_B) \cdot \frac{\psi^*_S(\mathbf{x}_S)}{\psi_S(\mathbf{x}_S)}, && \forall \mathbf{x}_B \in \mathcal{X}_B \\
\psi^{**}_S(\mathbf{x}_S) &= \sum_{\mathbf{x}_{B\setminus S}\in\mathcal{X}_{B\setminus S}} \psi^*_B(\mathbf{x}_S, \mathbf{x}_{B\setminus S}), && \forall \mathbf{x}_S \in \mathcal{X}_S \\
\psi^*_A(\mathbf{x}_A) &= \psi_A(\mathbf{x}_A) \cdot \frac{\psi^{**}_S(\mathbf{x}_S)}{\psi^*_S(\mathbf{x}_S)}, && \forall \mathbf{x}_A \in \mathcal{X}_A.
\end{aligned} \tag{1}$$

These equations hold for any state-configurations $\mathbf{x}_A, \mathbf{x}_S, \mathbf{x}_B$ within the cliques.

Start with clique potentials obtained from conditional probability tables, whereas the separator potentials can be constantly 1's. To find all clique and separator marginals, we first run the algorithm in the reversed RIP, that is, in the Sundberg's ordering $C_k, \ldots, C_1$ of the cliques. In this forward step we start at C_k (called root), and via the separators, end at C_1. The so obtained potential of C_1 is already the clique potential. To obtain all the clique potentials, we have to run the algorithm again, that is to make a backward step (in the RIP ordering). The starting value for the separator potentials can be $\psi_S = 1$.

It is proved in [5, 7] that at the end, $\psi^{**}_{C_i}(\mathbf{x}_{C_i}) = p(\mathbf{x}_{C_i})$ and $\psi^{**}_{S_i}(\mathbf{x}_{S_i}) = p(\mathbf{x}_{S_i})$, $i = 1, \ldots, k$; so the iteration leads to the clique marginals. In other wording, in the forward steps, the cliques collect the information from all of its neighbors (parent cliques on the JT) recursively; whereas, in the backward steps, they distribute the information to them. Therefore, it is sometimes called message passing algorithm.

Formally, the forward steps are

$$\psi_{C_4} \to \psi^*_{S_4} \to \psi^*_{C_3} \to \psi^*_{S_3} \to \psi^*_{C_2} \to \psi^*_{S_2} \to \psi^*_{C_1}$$

and the backward steps are

$$\psi^*_{C_1} \to \psi^{**}_{S_2} \to \psi^{**}_{C_2} \to \psi^{**}_{S_3} \to \psi^{**}_{C_3} \to \psi^{**}_{S_4} \to \psi^{**}_{C_4}.$$

From the so obtained clique marginals you can marginalize for their variables or absorb evidences.

2 Sequences of regressions on the pain clinic data

Here the theory is the application of nonparametric regressions along a chain graph and comparison with linear regressions. It can also be done recursively (see ICE [1]). The data are available only for research purposes.

Past => Future: for each case in the test sample, where the nonpar. regr. ([3, 4, 6]) is based on the training sample.

- $$\hat{Z}_b = \mathbb{E}(Z_b \,|\, B), \quad \tilde{Z}_b = \mathbb{E}(Z_b \,|\, B, A, U, V)$$
  $$Z_b^{lin} = \text{lin. function}(B)$$
- $$\hat{X}_b = \mathbb{E}(X_b \,|\, U, V), \quad \tilde{X}_b = \mathbb{E}(X_b \,|\, B, A, U, V)$$
  $$X_b^{lin} = \text{lin. function}(U, V)$$
- $$\hat{X}_a = \mathbb{E}(X_a \,|\, X_b), \quad \tilde{X}_a = \mathbb{E}(X_a \,|\, X_b, B, A, U, V)$$
  $$X_a^{lin} = \text{lin. function}(X_b), \quad X_a^{ICE} = \mathbb{E}(X_a \,|\, \hat{X}_b)$$
- $$\hat{Z}_a = \mathbb{E}(Z_a \,|\, Z_b, A), \quad \tilde{Z}_a = \mathbb{E}(Z_a \,|\, Z_b, A, B, U, V)$$
  $$Z_a^{lin} = \text{lin. function}(Z_b, A), \quad Z_a^{ICE} = \mathbb{E}(Z_a \,|\, \hat{Z}_b, A)$$
- $$\hat{Y} = \mathbb{E}(Y \,|\, Z_a), \quad \tilde{Y} = \mathbb{E}(Y \,|\, Z_a, Z_b, A, B, U, V)$$
  $$Y^{lin} = \text{lin. function}(Z_a), \quad Y^{ICE} = \mathbb{E}(Y \,|\, \hat{Z}_a)$$

Then find the following MSE: square-root of (n is the size of the test sample)

$$\widehat{MSE}(Z_b) = \frac{1}{n} \sum_{\text{test sample}} (Z_b - \hat{Z}_b)^2, \quad \widetilde{MSE}(Z_b) = \frac{1}{n} \sum_{\text{test sample}} (Z_b - \tilde{Z}_b)^2, \ldots$$

$$\widehat{MSE}(Y) = \frac{1}{n} \sum_{\text{test sample}} (Y - \hat{Y})^2, \quad \widetilde{MSE}(Y) = \frac{1}{n} \sum_{\text{test sample}} (Y - \tilde{Y})^2,$$

$$MSE^{lin}(Y) = \frac{1}{n} \sum_{\text{test sample}} (Y - Y^{lin})^2, \quad MSE^{ICE}(Y) = \frac{1}{n} \sum_{\text{test sample}} (Y - Y^{ICE})^2.$$

For the pain clinic data see [9, 10, 12], possibly refer to the psychologist.

Acknowledgements

We are indebted to Nanny Wermuth for giving us the pain clinic data for research purposes.

References

[1] Baranyi, M., Bolla, M., Iterated Conditional Expectation Algorithm on DAGs and Regression Graphs, *Econometrics and Statistics* (2020), published online, https://doi.org/10.1016/j.ecosta.2020.05.003

[2] Bolla, M., Abdelkhalek, F., Baranyi, M., Graphical models, regression graphs, and recursive linear regression in a unified way, *Acta Scientiarum Mathematicarum (Szeged)* **85**, (2019), 9–57.

[3] Breiman, L. and Friedman, J. H., Estimating optimal transformations for multiple regression and correlation, *J. Am. Stat. Assoc.* **80** (1985), 580–619.

[4] Győrfi, L., et al., A distribution-free theory of nonparametric regression. Springer Science & Business Media, 2006.

[5] Lauritzen, S. L., Spiegelhalter, D., Local computations with probabilities on graphical structures and their application in expert systems, *J. R. Statist. Soc. B* **50** (1988), 157-224.

[6] Nadaraya, E. A., On nonparametric estimates of regression functions and regression curves, *Theory of Applied Probability* **10** (1965), 186–190.

[7] Wainwright, M. J., Graphical Models and Message-Passing Algorithms: Some Introductory Lectures. In: Mathematical Foundations of Complex Networked Iformation Systems, Lecture Notes in Mathematics 2141, F. Fagnani et al. (eds.), Springer (2015).

[8] Wermuth, N., Recursive equations, covariance selection, and path analysis, *J. Amer. Stat. Assoc.* **75** (1980), 963–972.

[9] Wermuth, N., Traceable regressions, *International Statistical Review* (2012) **80** (3), 415–438.

[10] Wermuth, N., Sadeghi, K., Sequences of regressions and their independences, *TEST* **21** (2012), 215-279.

[11] Wermuth, N., Graphical Markov models, unifying results and their interpretation, In: (Balakrishnan, N. et al. eds) Wiley StatsRef: Statistics Reference Online (2015), also ArXiv: 1505.02456.

[12] Wermuth, N., Cox, D. R., Graphical Markov models: Overview, In: (Wright, J., ed.) International Encyclopedia of the Social and Behavioral Sciences, 2nd ed., 10, Elsevier, Oxford, pp. 341–350. (2015), also ArXiv: 1407.7783.