My research: Total positivity in graphical models

Steffen Lauritzen

Background

Graphical models (Lauritzen, 1996; Maathuis et al., 2019) are probabilistic or statistical models which describe complex relationships between systems of variables in a modular fashion, exploiting conditional independence relations encoded with mathematical graphs. The modularity enables simple specification, interpretation, communication, and computation associated with application of the models.

The subject has reached maturity as a research area and graphical models are now applied in an abundance of contexts; for example in *digital communication*, *machine learning*, *causal inference*, *genetics*, *decision support systems*, *social sciences*, and *forensic science*.

Modern applications of graphical models, as exemplified above, have disclosed a variety of theoretical and methodological challenges that must be addressed to take full advantage of their versatility and expressiveness. I am working on a number of different topics within this general area, including the development of graphical models for extremes, graphical models for social networks, but this little note will focus on my study of *total positivity,* mostly done in collaboration with Caroline Uhler (MIT) and Piotr Zwiernik (UPF Barcelona).

Total positivity

A density on a product space $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$, where $\mathcal{X}_v \subseteq \mathbb{R}$, is *multivariate totally positive of order two* (MTP_2) if its density is log-supermodular, i.e. if

$$f(x)f(y) \quad \leq \quad f(x \wedge y)f(x \vee y) \qquad \text{for all } x, y$$

where $(x \wedge y)$ and $(x \vee y)$ denote coordinatewise minimum and maximum. Such distributions share an abundance of properties, one of them that they are closed under marginalization and conditioning, coarsening of the state spaces, and vanishing correlations always imply independence, which is otherwise only true in the Gaussian case.

In particular, if a distribution is MTP_2, the variables are always *positively associated* in the sense that for any pair of non-decreasing functions f, g, it holds that

$$\mathbf{V}(f(X), g(X)) \geq 0,$$

a result that is known in mathematical physics for the binary case as the *FKG inequality* (Fortuin et al., 1971). Indeed, a *ferromagnetic Ising model* is an example of a binary distribution which is MTP_2.

An interesting research avenue is to investigate how analysis in graphical models under the MTP_2 constraint can simplify both interpretation and structure estimation. In Fallat et al. (2017) it is shown how the constraint interacts with the

Markov property so that MTP_2 distributions are always *faithful* to their interaction graph, in the sense that graph separation yields exactly all valid independence relations.

Recent developments (Lauritzen et al., 2019a) show that estimating a Gaussian graphical model under the MTP_2 restriction also yields automatic sparsity, since the estimation problem can be formulated as a simple convex optimization problem. The solution can be shown to exist uniquely if the sample size is at least two and it is determined by the optimality conditions

$$\begin{aligned} \hat{\sigma}_{vv} - s_{vv} &= 0 \text{ for all } v \in V, \\ (\hat{\sigma}_{uv} - s_{uv}) &\geq 0 \text{ for all } uv, \\ (\hat{\sigma}_{uv} - s_{uv})\hat{k}_{uv} &= 0 \text{ for all } u \neq v, \end{aligned} \tag{1}$$

where $\hat{\Sigma} = \hat{K}^{-1} = \{\hat{\sigma}_{uv}\}_{u,v \in V}$ is the covariance matrix estimated under the assumption of multivariate total positivity.

The equation (1) ensures that K automatically becomes sparse. Many properties of the estimated graph that can be identified theoretically and form the basis of very fast algorithms for estimation under the MTP_2 constraint, making this computationally feasible for very large dimensions of the multivariate distribution considered.

It turns out that these phenomena generalize (with modifications) to the important case of ferromagnetic Ising models and other exponential families of distributions, as recently shown in Lauritzen et al. (2019b). For example, it holds that the maximum likelihood estimate of a multivariate binary distribution under the MTP_2 assumption exists if and only if the sample has both configurations $\{(1, -1), (-1, 1)\}$ represented for any pair of sites with a non-vanishing interaction potential. This fact enables estimation with only $O(d)$ observations rather than $O(2^d)$ observations where d is the number of sites in the Ising model.

References

Fallat, S., Lauritzen, S., Sadeghi, K., Uhler, C., Wermuth, N., and Zwiernik, P. (2017). Total positivity in Markov structures. *The Annals of Statistics*, 45:1152–1184.

Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22(2):89–103.

Lauritzen, S., Uhler, C., and Zwiernik, P. (2019a). Maximum likelihood estimation in Gaussian models under total positivity. *The Annals of Statistics*, 47(4):1835–1863.

Lauritzen, S., Uhler, C., and Zwiernik, P. (2019b). Total positivity in structured binary distributions. arXiv:1702.04031.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford, United Kingdom.

Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M., editors (2019). *Handbook of Graphical Models*. Handbooks of Modern Statistical Methods. Chapman and Hall/CRC, Boca Raton, FL.