Botzmann-Gibbs distributions and applications to data-driven modeling

<u>Dionissios Hristopulos</u>
Technical University Of Crete, Chania, Greece

Boltzmann-Gibbs (BG) models are a cornerstone of statistical physics. They are used to describe the probability functions of discrete physical systems as well as continuum field theories. In addition, they provide a useful framework for the representation of data-driven spatial or spatiotemporal processes. The structure of statistical dependence in such processes is enforced by means of near-neighbor interactions in space and/or time. The interactions contribute to a scalar energy (appearing in the exponent of the exponential BG density), and they are controlled by data-driven interaction parameters.

The BG representation can lead to sparsely connected space-time networks comprising edges that are determined according to suitable distance concepts. For Gaussian BG models, the spatial/spatiotemporal structure of the interactions determines the model's precision (inverse covariance) matrix. The latter is sparse by construction if the interactions are local. The sparsity property leads to efficient prediction algorithms for the representation of random functions. Hence, the sparsity of the precision matrix provides considerable computational gains compared to standard approaches which are based on the covariance function and require inverting the (potentially very large) covariance matrix.

This presentation focuses on recent progress in the application of Boltzmann-Gibbs models to spatial and space-time geo-referenced data. A topic that will be discussed is the construction of new temporal/spatial/spatiotemporal covariance kernels which are based on continuum-space BG models and can be used in Gaussian process regression tasks. In addition, we will discuss the extension of BG models to network-based stochastic local interaction (SLI) models which are suitable for the interpolation of scattered spatiotemporal datasets without using covariance kernels. In SLI models the interactions are implemented via compactly supported kernel functions (not to be confused with the covariance kernels). SLI models provide a flexible statistical framework for spatiotemporal datasets; they are computationally efficient, due to the explicit precision matrix construction, and therefore useful for applications to big datasets. In addition, it is straightforward to ensure that the precision matrices of SLI models are non-negative definite. This feature enables the use of different data types and distance measures, topics which will be further investigated in future studies.

References

[1] D.T. Hristopulos, Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers. Springer Nature, Dordrecht, Netherlands (2020).
[2] D. Allard, D.T. Hristopulos, T. Opitz, Linking Physics and Spatial Statistics: A New Family of Boltzmann-Gibbs Random Fields, Electron. J. Stat., 15(2), 4085-4116 (2021).
[3] D.T. Hristopulos, V.D. Agou, Stochastic local interaction model with sparse precision matrix for space–time interpolation. Spat. Stat. 40, 100403 (2020).