Inference with discrete data via score matching

Eduardo García-Portugués • Department of Statistics, UC3M

Score matching is an important method for fitting parametric distribution and regression models without needing to compute their normalizing constant — a common challenge in high-dimensional or complex models and a requirement for applying maximum likelihood estimation. Originally introduced by Aapo Hyvärinen in 2005 (Hyvärinen, 2005) for continuous distributions, score matching has since evolved, with many generalizations aimed at handling data coming from different supports. One particular situation where score matching has required a rethinking is distributions for discrete data, where the original formulation, based on derivatives of density functions, does not directly apply. Recent advancements (Hyvärinen, 2007; Matsubara et al., 2024; Meng et al., 2022; Xu et al., 2025) have extended score matching to discrete setups, both univariate and multivariate, where issues with untractable normalizing constants and non-closed form maximum likelihood estimates are also present. Discrete distributions serve to model count data and ordinal data, either univariate or multivariate.

This thesis will focus on reviewing the main developments in score matching for discrete data modelling, digesting the mathematical foundations of these contributions, performing numerical studies to compare with the maximum likelihood inference framework and method of moments, and proposing new real data applications.

References

Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, *6*(Apr), 695–709. http://jmlr.org/papers/v6/hyvarinen05a.html

Hyvärinen, A. (2007). Some extensions of score matching. *Computational Statistics & Data Analysis*, *51*(5), 2499–2512. https://doi.org/10.1016/j.csda.2006.09.003

Matsubara, T., Knoblauch, J., Briol, F.-X., & Oates, Chris. J. (2024). Generalized bayesian inference for discrete intractable likelihood. *Journal of the American Statistical Association*, *119*(547), 2345–2355. https://doi.org/10.1080/01621459.2023.2257891

Meng, C., Choi, K., Song, J., & Ermon, S. (2022). Concrete score matching: Generalized score matching for discrete data. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (Vol. 35, pp. 34532–34545). Curran Associates, Inc. https://arxiv.org/abs/2211.00802

Xu, J., Scealy, J. L., Wood, A. T. A., & Zou, T. (2025). Generalized score matching. *Journal of Multivariate Analysis*, *210*, 105473. https://doi.org/10.1016/j.jmva.2025.105473