THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

PhD Dissertation Proposal Presentation

Benedetta Bruni
Department of Statistics
The University of Chicago

"A Generalized Bayesian Approach to Tree Models for Densities"

December 10, 2025, at 9:00 AM
Room 322, 5460 S University Ave.

Abstract

Bayesian Tree models for probability distributions incorporate priors on the set of the recursive partitions of the sample space, to estimate an unknown target density as piece-wise constant over a partition. Splitting locations along the recursion are typically chosen within a set of fixed and data-agnostic candidate points, balancing the need to represent the local features of the target density with greater detail and the computational burden of deep trees. Regardless, this often results in partitions being too coarse in high-density regions of the sample space - causing underfitting - and too fine in low-density regions - causing overfitting. Although these issues are common to all tree models beyond the class considered here, the broader tree literature offers solutions. Data-dependent splitting locations allow partitions to adapt to the local concentration of datapoints, while tree ensembles can be designed to mitigate overfitting or underfitting. However, neither strategy is currently employed for density estimation with Bayesian Trees. Incorporating data-adaptive partitioning while maintaining coherent Bayesian inference remains difficult, while additive tree ensembles commonly used in supervised learning do not translate directly to density estimation without significant implementation effort. This work proposes to expand the class of Bayesian density trees via two Generalized Bayesian strategies, i.e. through inference with losses other than the negative log-likelihood. The 'Partial Self-Information Loss' (Cox's partial likelihood) provides a natural way to incorporate data-dependent splitting under a coherent Bayesian inference framework. A theoretical comparison between data-adaptive and data-agnostic trees could be possible, but it would require understanding how the partial likelihood behaves with density tree models. The 'Balancing Loss' (Awaya, Xu, Ma (2025)) could allow to construct additive tree ensembles for densities, exploiting a conjugate prior on the tree-step coefficients. Given the limited development of density tree ensembles, it is important to establish their theoretical properties in relation to this new loss function —both by analyzing

---

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/info/statseminars.

Gibbs posterior contraction rates and by determining the convergence rate of the Balancing Loss minimizer using empirical process theory.
