# Posterior Contraction Rate and Roadmap to Semiparametric Bernstein-von Mises Theorem for Bayesian Causal Forests

Andrej Srakar
International Research Centre on Artificial Intelligence under the auspices of UNESCO and Department for Artificial Intelligence
Jožef Stefan Institute, Ljubljana, Slovenia
andrej.srakar@ijs.si

## Abstract

Bayesian causal forest is a presently widely used and discussed approach in Bayesian causal inference. It is a machine learning model used to find how much a treatment or intervention affects different units using observational data. It consists of a Bayesian additive regression trees outcome regression with an estimated propensity score as predictor, which leads to a product-BART Bayesian framework. We derive posterior contraction rate for the model, as well as most of proofs for the corresponding BCF semiparametric Bernstein-von Mises theorem. We conclude by discussing some open research directions.

## Keywords

Bayesian causal forest, BART, posterior contraction rate, semiparametric BvM theorem, product-BART, product probability spaces, asymptotics, weak learners, artificial intelligence

## 1 Introduction

Bayesian causal inference is an interesting and mportant area of future development of both Bayesian and causal methods and probability theory in general. One of most common Bayesian priors for causal inference remains Bayesian additive regression trees (BART, see Chipman et al., 2010). Theoretical properties of BART have been subject of important recent research by Ročkova (Ročkova and van der Pas, 2020; Jeong and Ročkova, 2023). In her recent contribution on the topic Jeong and Ročkova propose a new class of sparse piecewise heterogeneous anisotropic Hölder functions and derive their minimax optimality properties, which lead to BART posterior contraction rate. But, they do not prove a Bernstein-von Mises theorem for BART. We extend their work to Bayesian Causal Forests (BCF), proposed by Hahn, Murray and Carvalho (Hahn et al., 2020).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
*Information Society 2026, 5–9 October 2026, Ljubljana, Slovenia*
© 2026 Copyright held by the owner/author(s).
DOI: https://doi.org/10.70314/is.2026.sikdd.16

BCF is a modular Bayesian approach, based on regression decomposition of Robinson into prognostic and treatment effect part of the outcome regression specification, which include propensity score parameter. In our article we extend theoretical derivations of Ročkova and van der Pas, and Jeong and Ročkova to accomodate for the modular and product-space nature of BCF. Posterior contraction rate and most proofs needed for semiparametric Bernstein-von Mises theorem for the BCF approach are derived. Our article delves into an important area of asymptotics for Bayesian nonparametric priors for causality, and we discuss also extensions to other BNP approaches for causality outside of the BART framework.

## 2 Derivation of posterior contraction rate for BCF

The BART prior is the hierarchical prior used in Bayesian Additive Regression Trees to regularize the ensemble of trees. It places priors on the tree structures; the terminal node (i.e. leaf) parameters; and the error variance.

Let the regression function be represented as a sum of m trees: $y_i = f(x_i) + \varepsilon_i$,where $f(x) = \sum_{j=1}^{m} g(x; T_j, M_j)$ $T_j$is the structure of the $j$-th tree, $M_j = \mu_{j1}, \ldots, \mu_{jb_j}$ are the terminal-node values, $g(x; T_j, M_j)$ maps $x$ to the leaf value corresponding to its terminal node, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The prior encourages small trees. For a node at depth d, the probability it is split is $P(\text{split at depth } d) = \alpha(1+d)^{-\beta}$, where typically $\alpha = 0.95, \beta = 2$. Given that a node splits, the splitting variable is chosen uniformly and the split point is chosen uniformly among available cutpoints. Thus, $p(T) = \prod_{\eta \in \mathcal{I}(T)} \alpha(1+d_\eta)^{-\beta} \prod_{\eta \in \mathcal{L}(T)} 1 - \alpha(1+d_\eta)^{-\beta} \times p(\text{split rules})$ where $\mathcal{I}(T)$ are internal nodes and $\mathcal{L}(T)$ are terminal nodes. If tree J has $b_j$ terminal nodes, $M_j = \mu_{j1}, \ldots, \mu_{jb_j}$, then $\mu_{jk} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau^2)$.

The full BART prior can be summarized in five parts:

1) $y_i \mid f, \sigma^2 \sim \mathcal{N}(f(x_i), \sigma^2)$

2) $f(x) = \sum_{j=1}^{m} g(x; T_j, M_j)$

3) $T_j \sim p(T)$

4) $\mu_{jk} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau^2)$

5) $\sigma^2 \sim \text{Inverse-Gamma}(\frac{\nu}{2}, \frac{\nu\lambda}{2})$

The defining characteristic of the BART prior is the combination of a depth-penalizing prior on tree structures, Gaussian shrinkage on leaf values scaled by $1/\sqrt{m}$, and a prior on the residual variance. This induces strong prior toward many shallow, weak trees, with the sum flexibly approximating complex nonlinear regression functions.

Theorem of Rockova and van der Pas proves posterior contraction for BART. Assume $Y_i = f_0(x_i) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$, where $x_i \in [0,1]^p$, $f_0 \in C^{\alpha}([0,1]^p)$ and only q variables are relevant. Then under a suitably regularized BART prior, $\Pi(|f - f_0|_n > M\varepsilon_n \mid Y) \to 0$, where $\varepsilon_n = n^{-\alpha/(2\alpha+q)}(\log n)^c$. The proof is an application of the general Bayesian posterior contraction theorem. One must verify three ingredients (see van der Vaart and van Zanten, 2008): existence of exponentially consistent tests; sufficient prior mass near $f_0$; and exponentially small prior mass outside an appropriate sieve.

Bayesian causal forest (BCF) model developed in 2020 in a Bayesian Analysis article by Hahn, Murray and Carvalho is an extension of BART by including estimate of the propensity score in the outcome regression equation. It writes $Y_i = \mu(X_i, \hat{e}_i) + \tau(X_i)W_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$, Where $\mu \sim \Pi_\mu, \tau \sim \Pi_\tau$ are independent BART priors. The target is $\psi_0 = E_{P_X}[\tau_0(X)]$.

BCF are not a single BART prior but a joint prior on two regression functions with different roles: $Y_i = \mu(X_i, \hat{\pi}_i) + Z_i\tau(X_i) + \varepsilon_i$, where $\mu(\cdot)$i is the prognostic function, $\tau(\cdot)$ is the treatment effect function, each is given an independent BART prior and $\hat{\pi}$ is an estimated propensity score included as a covariate. Consequently, a Bernstein–von Mises (BvM) theorem for BCF is fundamentally a semiparametric BvM problem with nuisance functions, rather than simply a BvM theorem for BART. Semiparametric Bernstein-von Mises theorems have been proposed in research works from 2012-2014 by Ismael Castillo and Richard Nickl (see Castillo and Nickl, 2014) who proposed two broad solutions to develop them for general nonparmetric stituations.

Coherent proof strategy to develop a semiparametric Bernstein-von Mises theorem for BCF consists of six parts: 1) Extend the Rocková–van der Pas posterior contraction theory to the product BART prior $\Pi_\mu \otimes \Pi_\tau$; 2) Establish LAN for the joint likelihood in the directions $(h_\mu, h_\tau)$; 3) Characterize the nuisance tangent space associated with $\mu$ and derive the efficient influence function for the target functional of $\tau$; 4) Prove a no-bias condition using the posterior contraction rates for both BART components; 5) Linearize the target functional and show the remainder is $o_p(n^{-1/2})$; 6) Apply a semiparametric BvM theorem (in the spirit of results by van der Vaart, Rousseau, and Kleijn, see van der Vaart and van Zanten, 2008; Rousseau, 2010; Kleijn, 2021) to conclude asymptotic normality of the posterior for the treatment effect functional.

The main idea for the proof is that you are not proving a new contraction theorem for one BART prior, but for the product prior $\Pi = \Pi_\mu \otimes \Pi_\tau$, on the product space $\mathcal{F} = \mathcal{F}_\mu \times \mathcal{F}_\tau$. The proof follows the same high-level structure as the single-BART case, but ingredients must be lifted to a product of function spaces.

Our main theorem is (Product BART contraction):

Assume $Y_i = \mu_0(X_i) + Z_i\tau_0(X_i) + \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2)$, where$Z_i \in 0, 1$. Suppose $\mu_0 \in C^{\alpha_\mu}, \tau_0 \in C^{\alpha_\tau}$, and $\Pi = \Pi_\mu \otimes \Pi_\tau$ is the independent product of two regularized BART priors. Assume overlap $0 < c < P(Z = 1|X = x) < 1 - c$ for some c>0. Then $\Pi(d((\mu, \tau), (\mu_0, \tau_0)) > M\varepsilon_n \mid Y) \to 0$, where $d^2 = |\mu - \mu_0|n^2 + |\tau - \tau_0|n^2$, and $\varepsilon_n^2 = \varepsilon\mu, n^2 + \varepsilon\tau, n^2$ with $\varepsilon_{\mu,n} = n^{-\alpha_\mu/(2\alpha_\mu+d)}(\log n)^{c_\mu}$, and $\varepsilon_{\tau,n} = n^{-\alpha_\tau/(2\alpha_\tau+d)}(\log n)^{c_\tau}$. Consequently, $\varepsilon_n = O(\max(\varepsilon_{\mu,n}, \varepsilon_{\tau,n}))$.

Complete proof exceeds length of this article and is accessible as a separate document. Above theorem proves that posterior contraction rate for BCF is of order of maximum of standard errors of treatment and intervention parts of the BCF regression equation. Additionally to the above, information operator coercivity proof is a special part not usual for such theorems in Bayesian asymptotics.

Strategy to derive the proof can be described shortly as follows. Firstly, we prove product approximation, product prior mass, product entropy, product sieve, and joint testing theorems which extend BART to a product-BART situation. This included specially considering and addressing problems of Kullback-Leibler (KL) likelihood geometry, approximation, prior concentration, entropy, sieve probability, testing, KL neighbourhoods, and metric equivalence. For the latter, we proved coercivity estimate by proving that information matrix is uniformly positive definite, and that its eigenvalues are bounded below.

## 3 Path to semiparametric BvM theorem for BCF

At this point, complete semiparametric Bernstein-von Mises theorem remains subject to our next work, as some part demand additional proofs and verifications. But we are able to prove some of its important parts.

Theorem (Conditional LAN Expansion over Product-BART Neighborhoods):

Consider $Y_i = \mu(X_i) + Z_i\tau(X_i) + \varepsilon_i, \varepsilon_i \overset{iid}{(0,\sigma^2)}$, with true functions $(\mu_0, \tau_0)$. Define $f(x,z) = \mu(x) + z\tau(x)$, and $f_0(x,z) = \mu_0(x) + z\tau_0(x)$. Suppose $|f - f_0|_n \leq M\varepsilon_n$, where $\varepsilon_n = n^{-\alpha/(2\alpha+p)}(\log n)^k$. Assume additionally product-BART posterior contracts uniformly over this neighborhood; tree depth is uniformly bounded by $D_n = o(\log n)$; leaf values satisfy $\sup_x |f(x)| < C$; the empirical design satisfies $\frac{1}{n}X^\top X \to \Sigma$.

Then $\sup_{|f-f_0|_n \leq M\varepsilon_n} \ell_n(f) - \ell_n(f_0)\sqrt{n}, W_n(h) + \frac{1}{2}n|h|n^2 = o_P(1)$, where $h = f - f_0$ and $W_n(h) = \frac{1}{\sigma^2\sqrt{n}} \sum i = 1^n \varepsilon_i h(X_i, Z_i)$.

Theorem (Prior Stability Under Least-Favorable Perturbations):

Let $f = (\mu, \tau)$ have an independent Product-BART prior $\Pi = \Pi_\mu \otimes \Pi_\tau$. Let $\eta = (\eta_\mu, \eta_\tau)$ denote the least-favorable direction associated with the efficient influence function. Define the shifted parameter $f_t = f + \frac{t}{\sqrt{n}}\eta t \in \mathbb{R}$. Suppose: posterior contraction holds; tree depth satisfies $D_n = o(\log n)$, leaf parameters satisfy $|\omega_j| \leq C$; every Product-BART density is locally Lipschitz; perturbations leave tree topology unchanged with probability $1 - o(1)$.

Then $\frac{d\Pi(f_t)}{d\Pi(f)}1 + o(1)$ uniformly over $|f - f_0|_n \leq M\varepsilon_n$. This is the no-bias condition.

In order to finalize also the proof of Bernstein-von Mises theorem, one needs the following:

- An inverse approximation inequality relating empirical $L_2$ and $L_\infty$ errors for Product-BART function classes.

- High-probability bounds on the maximum diameter of random Product-BART partition cells.

- A margin theorem for split stability, showing that scale perturbations do not alter selected partitions.

- A proof that the effective number of leaves grows slow enough under the Product-BART prior.

- Uniform versions of these bounds over the posterior contraction neighborhood.

This forms a still missing final roadmap toward a complete semiparametric Bernstein–von Mises theorem for Bayesian Causal Forests.

## 4 Discussion and open problems

Our article derives posterior contraction rate for Bayesian causal forests which has not been derived thus far. It also derives most parts in order to prove the semiparametric Bernstein-von Mises theorem for BCF. Analysis takes part using product-BART structure of the Bayesian causal forests regression problem.

Our proof derivation opens several interesting research directions of future work. Firstly, one is now able to prove semiparametric BvM theorem for BART which previous authors have not done. Next, additional usages of tools from analysis, such as piecewise harmonic functions and piecewise polyhedral multifunctions will probably help to prove the final version of both the BvM theorem for BART and BCF as well as their future extensions to Bayesian causal models. Finally, analysis in our article will prove beneficial in considering future needed extensions and development of area of Bayesian causal inference where a most of the research questions are still nonresponded.

## Acknowledgments

Article has been firstly presented in the session Advances in Bayesian causal inference at the 2026 World Meeting of the International Society of Bayesian Analysis (ISBA). Author thanks participants of the session for discussion. This work was supported by the Slovenian Research Agency under grant number N5-0402 (ERC Perspectives).

## References

[1] Ismaël Castillo and Richard Nickl. 2014. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* 42 (5) 1941-1969. https://doi.org/10.1214/14-AOS1246.

[2] Hugh A. Chipman. Edward I. George and Robert E. McCulloch. 2010. BART: Bayesian additive regression trees. *Ann. Appl. Stat.,* 4 (1) 266-298, March 2010. https://doi.org/10.1214/09-AOAS285.

[3] P. Richard Hahn. Jared S. Murray and Carlos M. Carvalho, 2020.. Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion). *Bayesian Anal.,* 15 (3) 965-1056, September 2020. https://doi.org/10.1214/19-BA1195.

[4] Seonghyun Jeong and Veronika Rockova, 2023. The Art of BART: Minimax Optimality over Nonhomogeneous Smoothness in High Dimension. *Journal of Machine Learning Research*, 24, 337, 1-65, URL: http://jmlr.org/papers/v24/22-0382.htm.

[5] B. J. K. Kleijn, 2021. Frequentist validity of Bayesian limits. *Ann. Statist.*, 49 (1) 182-202, February 2021. https://doi.org/10.1214/20-AOS1952.

[6] Veronika Ročková. Stéphanie van der Pas, 2020. Posterior concentration for Bayesian regression trees and forests. *Ann. Statist.*, 48 (4) 2108-2131, August 2020. https://doi.org/10.1214/19-AOS1879.

[7] Judith Rousseau, 2010. Rates of convergence for the posterior distributions of mixtures of Betas and adaptive nonparametric estimation of the density. *Ann. Statist.*, 38 (1) 146-180, February 2010. https://doi.org/10.1214/09-AOS703.

[8] A. W. van der Vaart. J. H. van Zanten, 2008. Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, 36 (3) 1435-1463, June 2008. https://doi.org/10.1214/009053607000000613.