Finite forests are not complete forests: missing ensemble-size conditions in two asymptotic theories

Askar Uval

25 July 2026

Abstract

Generalized random forests and orthogonal random forests define their estimators using a finite number B of randomized trees, but their consistency and asymptotic-normality theorems impose no condition on B. We give a one-line counterexample: with $B = 1$, independent Rademacher responses, and a fixed-leaf-size honest tree, the fitted local mean remains an average of a bounded number of Rademachers. It is therefore not even consistent, despite satisfying the remaining smoothness, support, honesty, regularity, variance, and subsample-rate assumptions. The proof reveals the source of the mismatch. The generalized-random-forest argument invokes a theorem whose random forest is explicitly the complete average over all subsamples and tree randomization; the orthogonal-random-forest proof likewise analyzes a complete U-statistic. The weakest general repair is to state the results for the complete forest or require the finite-to-complete discrepancy to be $o_p(\sigma_n)$. Under standard conditional Monte Carlo bounds, the transparent sufficient condition is $B_n^{-1/2} = o(\sigma_n)$, with stronger uniform versions when required by the proof.

1 The object mismatch

The generalized random forest (GRF) algorithm of Athey, Tibshirani, and Wager [1] declares all tuning parameters to be pre-specified, including the number of trees B, and returns weights

$$\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \alpha_{bi}(x).$$

Its Theorems 3 and 5 then assert consistency and asymptotic Gaussianity without a growth condition on B. The orthogonal random forest (ORF) of Oprescu, Syrgkanis, and Wu [2] is equally explicit: it builds B trees on B random subsamples and averages their weights. Its main-text error bounds and Theorem 4.4 likewise omit a condition on B. The published supplement does retain the finite-B error $O(\sqrt{\log B/B})$ and assumes $B \geq n/s$ in Theorem D.6 for consistency. Thus the issue is not that finite-ensemble error was nowhere recognized; it is that the printed headline statements omit it, and the asymptotic-normality proof never makes it negligible at the claimed central-limit scale.

This omission is substantive, not typographical decoration. A tuning parameter in a sequence of estimators may depend on n, but without a stated restriction the constant sequence $B_n \equiv 1$ remains admissible.

2 A common counterexample

Work, if desired, along the exact-integer subsequence $n_j = (2j)^{10}$ and $s_j = (2j)^9 = n_j^{9/10}$. Fix a target $x_0 \in (0,1)$ and let

$$X_i \stackrel{\text{iid}}{\sim} \text{Unif}[0,1], \qquad Y_i \stackrel{\text{iid}}{\sim} \text{Rad}(1/2), \qquad X_i \perp Y_i.$$

Thus $Y_i \in \{-1, 1\}$ with equal probabilities and the conditional mean is identically zero. Use a symmetric, response-free, honest, balanced one-dimensional tree whose estimation leaf containing x_0 has size

$$1 \leq M_n \leq 2k - 1$$

for a fixed k. Such fixed-leaf-size trees are precisely among the regular trees used by the cited base theory [3]; $k = 1$ gives the sharpest special case. Set $B_n = 1$.

For GRF use the least-squares score $\psi_\theta(y) = y - \theta$. For ORF use $\psi(y; \theta, h) = \theta - y$ and the vacuous nuisance model $h = g(w; \nu) = \nu$ with $\nu_0 = \widehat{\nu} = 0$. The sign difference only matches the respective curvature conventions. Both weighted estimating equations have the root

$$\widehat{\theta}_n(x_0) = \frac{1}{M_n} \sum_{i \in L_n(x_0)} Y_i.$$

Theorem 1 (Fixed forests need not be consistent)**.** *For the construction above,*

$$\inf_n \mathbb{P}\Big(\widehat{\theta}_n(x_0) = 1\Big) \geq 2^{-(2k-1)}.$$

Consequently $\widehat{\theta}_n(x_0)$ does not converge in probability to 0. In particular, it cannot satisfy a centered central limit theorem at any scale $\sigma_n \to 0$.

Proof. Honesty makes the selected index set $L_n(x_0)$ measurable without using its estimation responses. Conditional on the tree-building information, those responses remain independent Rademachers. Therefore

$$\mathbb{P}\Big(\widehat{\theta}_n(x_0) = 1 \mid L_n(x_0)\Big) = 2^{-M_n} \geq 2^{-(2k-1)}.$$

This fixed positive mass at 1 rules out convergence in probability to 0.

For the Gaussian claim, let $c = 2^{-(2k-1)}$ and choose $t > 0$ with $1 - \Phi(t) < c$. If $\sigma_n \to 0$, then eventually $1/\sigma_n > t$, and hence

$$\liminf_{n \to \infty} \mathbb{P}\left(\frac{\widehat{\theta}_n(x_0)}{\sigma_n} > t\right) \geq c > 1 - \Phi(t),$$

contradicting convergence to a standard normal law. $\square$

When $k = 1$, the statement becomes especially stark:

$$\widehat{\theta}_n(x_0) = Y_{I_n} \in \{-1, 1\} \quad \text{almost surely for every } n.$$

3 Assumption audit

The counterexample does not exploit a rough signal, a vanishing variance, an adaptive response leak, or a pathological covariate distribution.

For GRF, least squares is one of the paper's own examples. The expected score is $M_\theta(x) = -\theta$, its derivative is invertible, every second derivative and score variogram is zero, the score is bounded on the compact parameter set $[-1, 1]$, and the population loss is strongly convex. The covariate density is one and the conditional response variance is one. In one dimension the random-split condition holds with probability one. Choose balance $\omega = 1/5$, split probability $\pi = 1$, and subsample size $s_n = n^{0.9}$, with harmless integer rounding. The lower exponent in the paper is

$$\beta_{\min} = \frac{a}{1+a}, \qquad a = \frac{\log 5}{\log(5/4)} \approx 7.213,$$

so $\beta_{\min} \approx 0.878 < 0.9 < 1$.

For ORF, the moment has the unique root $\theta_0(x) = 0$, local orthogonality is automatic because the score ignores h, the nuisance error is exactly zero, the Jacobian equals one, and every required Lipschitz, Hessian, covariance, boundedness, and full-support condition holds. With $d = 1$, $\rho = 1/5$, $\pi = 1$, and the same $s_n = n^{0.9}$, the displayed subsample condition follows from

$$\frac{0.9}{2a} > (1 - 0.9)\left(\frac{1}{2} - \varepsilon\right)$$

for every sufficiently small $\varepsilon > 0$; larger ε only weakens the right side. The projected influence variance is one.

Thus the construction satisfies the substantive hypotheses while violating the conclusions. It refutes the literal finite-B statements, not the complete-forest results that the proofs were evidently intended to establish.

4 Where the proof changes objects

The imported Wager–Athey theory is explicit about its estimand. It defines a random forest as the average over *all* size-s subsamples, marginalized over auxiliary tree randomness [3]. Its finite B average is introduced only as a Monte Carlo approximation, and the paper states that its theory assumes B is large enough for Monte Carlo effects not to matter.

GRF nevertheless begins with the finite average in its algorithm, calls the resulting pseudo-forest a U-statistic, and invokes the complete-forest theorem. A finite random average is instead an incomplete U-statistic. The distinction and its extra variance term were already standard; Mentch and Hooker [4], for example, treat the complete and incomplete cases separately.

ORF repeats the same substitution more visibly. Equation (17) of its supplement defines the empirical score using B sampled trees, while equation (18) separately defines the complete U-statistic. Lemma D.5 bounds their difference by $O(\sqrt{\log B/B})$, and Theorem D.6 requires $B \geq n/s$ for consistency. The main-text Theorem 4.4 omits that assumption, and its central-limit argument applies the complete-U-statistic limit without establishing an $o_p(\sigma_n)$ transfer for the finite estimator. The fixed-B witness shows that this is not a harmless proof shortcut: the omitted Monte Carlo component can dominate the claimed sampling scale.

5 Sharp repair

Let $\widehat{\theta}_{\infty,n}$ denote the complete forest, i.e. the conditional expectation of a tree estimate over its subsample and internal randomization given the training data. Let $\widehat{\theta}_{B_n,n}$ be the implemented finite average.

Proposition 1 (Minimal theorem-level repair)**.** *Any complete-forest central limit theorem with scale σ_n transfers to the finite forest if*

$$\widehat{\theta}_{B_n,n} - \widehat{\theta}_{\infty,n} = o_p(\sigma_n).$$

Under conditional independence of the sampled trees and a uniformly bounded conditional second moment, a simple sufficient condition is

$$B_n^{-1/2} = o(\sigma_n), \qquad \textit{equivalently} \qquad B_n \sigma_n^2 \longrightarrow \infty.$$

Proof. The first statement is Slutsky's theorem. For the second, conditional on the training data, the finite forest is a Monte Carlo mean around the complete forest, so its conditional mean-square discrepancy is $O(B_n^{-1})$. Markov's inequality gives the required $o_p(\sigma_n)$ coupling. □

If a proof needs uniform control over parameters or test points, its covering argument may replace $B_n^{-1/2}$ by a bound such as $\sqrt{\log B_n / B_n}$. The correct condition is the coupling itself; a convenient displayed growth rate should be derived for the exact uniformity class rather than silently assumed.

For the GRF scale $\sigma_n^2 \asymp (s_n/n)/\operatorname{polylog}(n/s_n)$, the pointwise sufficient condition is roughly

$$B_n \gg \frac{n}{s_n} \operatorname{polylog}(n/s_n).$$

A requirement sufficient only for consistency must not be advertised as sufficient for the central limit theorem unless it also makes the Monte Carlo discrepancy negligible relative to σ_n.

6 Scope and priority boundary

This note does not claim that practical GRF or ORF intervals are automatically invalid. Current GRF software reports finite-forest excess error, recommends comparing it with sampling variance, and advises growing trees proportionally to the sample size [5]. Bootstrap or little-bag implementations may also account for some finite-ensemble randomness. Those implementation-level protections are separate from the printed fixed-B theorem statements.

Targeted searches through the cited incomplete-U-statistic literature, later random-forest asymptotics, software documentation, and public correction records found no correction of these exact theorem statements. The general mathematical distinction between complete and incomplete forests is old and is not a contribution. The contribution claimed here is narrower: two influential finite-forest estimators fall outside the theories printed for them, a single admissible Rademacher construction refutes the literal conclusions, and the missing coupling condition can be stated sharply. Domain-expert review and author contact remain necessary before asserting priority.

References

[1] S. Athey, J. Tibshirani, and S. Wager. Generalized random forests. *Annals of Statistics*, 47(2):1148–1178, 2019. arXiv:1610.01271.

[2] M. Oprescu, V. Syrgkanis, and Z. S. Wu. Orthogonal random forest for causal inference. *Proceedings of ICML*, 2019. https://proceedings.mlr.press/v97/oprescu19a.html.

[3] S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018. https://arxiv.org/abs/1510.04342.

[4] L. Mentch and G. Hooker. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *Journal of Machine Learning Research*, 17(26):1–41, 2016. https://jmlr.org/papers/v17/14-168.html.

[5] The GRF developers. The GRF algorithm: training options and excess error. https://grf-labs.github.io/grf/REFERENCE.html. Accessed 25 July 2026.