Efron–Stein, Jackknife, and Bootstrap: Sensitivity, Variance, and Resampling

November 10, 2025

Abstract

This short note collects the Efron–Stein inequality with intuition and two proofs (martingale and conditional-variance forms), then connects it to the jackknife, infinitesimal jackknife, delete-m jackknife, and bootstrap. Several worked examples show how the *leave-one-coordinate* perturbations control variance and how resampling-based estimators approximate the same sensitivity in practice.

1 Setup and Statement

Let $X_1, \ldots, X_n$ be independent random variables on a common probability space, and let $Z = f(X_1, \ldots, X_n)$ be square-integrable. Let $X_1', \ldots, X_n'$ be an independent copy, and define

$$Z^{(i)} \;=\; f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n).$$

Theorem 1.1 (Efron–Stein). *For $Z = f(X_1, \ldots, X_n)$ as above,*

$$\mathrm{Var}(Z) \;\leq\; \tfrac{1}{2} \sum_{i=1}^{n} \mathbb{E}\big[(Z - Z^{(i)})^2\big].$$

Equivalently,

$$\mathrm{Var}(Z) \;\leq\; \sum_{i=1}^{n} \mathbb{E}\big[\mathrm{Var}\big(Z \mid X_{-i}\big)\big], \quad \text{where } X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

Interpretation. The difference $Z - Z^{(i)}$ is a *leave-one-coordinate redraw*: it measures how much the statistic changes when we keep all inputs fixed except we resample X_i from its own law. Squaring and averaging gives the average *influence* of coordinate i; the inequality says that the sum of these influences (up to a factor $1/2$) controls the variance.

2 Two Short Proofs

2.1 Martingale (Doob) proof

Let $M_i = \mathbb{E}[Z \mid X_1, \ldots, X_i]$ be the Doob martingale with $M_0 = \mathbb{E}Z$ and $M_n = Z$. The orthogonality of martingale differences gives

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} \mathbb{E}\big[(M_i - M_{i-1})^2\big].$$

Introduce X_i' (an i.i.d. copy) and set $Z^{(i)}$ accordingly. By the usual symmetrization (swap X_i and X_i' conditioned on the rest),

$$\mathbb{E}\big[(M_i - M_{i-1})^2\big] \ \leq \ \tfrac{1}{2}\mathbb{E}\big[(Z - Z^{(i)})^2\big].$$

Summing over i yields Theorem 1.1.

2.2 Conditional-variance proof

Using the law of total variance iteratively,

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} \mathbb{E}\big[\mathrm{Var}\big(Z \mid X_1, \ldots, X_i\big) - \mathrm{Var}\big(Z \mid X_1, \ldots, X_{i-1}\big)\big] = \sum_{i=1}^{n} \mathbb{E}\big[\mathrm{Var}\big(Z \mid X_{-i}\big)\big].$$

Finally, note that

$$2\,\mathrm{Var}\big(Z \mid X_{-i}\big) = \mathbb{E}\big[(Z - Z^{(i)})^2 \mid X_{-i}\big],$$

and take expectations.

3 Corollaries and Quick Tools

Corollary 3.1 (Bounded differences $\Rightarrow$ variance bound). *If for each i, changing only X_i changes f by at most c_i (i.e. $|f(x) - f(x^{(i)})| \leq c_i$), then*

$$\mathrm{Var}(Z) \leq \tfrac{1}{2} \sum_{i=1}^{n} c_i^2.$$

Remark 3.2 (Tightness). For linear f (e.g. sample mean), Efron–Stein is tight (equality). For nonlinear f (max, median, thresholds), it remains informative but can be loose.

4 Worked Examples

Example 4.1 (Sample mean: equality). Let $Z = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ with $\mathrm{Var}(X_i) = \sigma^2$. Then

$$Z - Z^{(i)} = \frac{X_i - X_i'}{n}, \qquad \mathbb{E}\big[(Z - Z^{(i)})^2\big] = \frac{2\sigma^2}{n^2}.$$

Hence

$$\tfrac{1}{2}\sum_{i=1}^{n}\mathbb{E}\big[(Z - Z^{(i)})^2\big] = \tfrac{1}{2}\cdot n\cdot\frac{2\sigma^2}{n^2} = \frac{\sigma^2}{n} = \mathrm{Var}(\bar{X}).$$

Example 4.2 (Maximum of two Bernoulli variables)**.** Let $X_1, X_2 \sim \mathrm{Bernoulli}(p)$ i.i.d., and $Z = \max\{X_1, X_2\}$. Then $Z \sim \mathrm{Bernoulli}(2p - p^2)$ so

$$\mathrm{Var}(Z) = (2p - p^2)\,(1 - 2p + p^2).$$

For the RHS of Efron–Stein:

$$\mathbb{E}\big[(Z - Z^{(1)})^2\big] = \mathbb{P}(X_2 = 0)\,\mathbb{P}(X_1 \neq X_1') = (1 - p)\cdot 2p(1 - p) = 2p(1 - p)^2,$$

and by symmetry the same for $i = 2$. Thus

$$\tfrac{1}{2}\sum_{i=1}^{2}\mathbb{E}\big[(Z - Z^{(i)})^2\big] = 2p(1 - p)^2.$$

At $p = \tfrac{1}{2}$, the bound gives 0.25 whereas $\mathrm{Var}(Z) = 0.1875$; the bound holds but is not tight.

Example 4.3 (Median: asymptotics and sensitivity)**.** Let $\tilde{X}_n$ be the sample median of i.i.d. data with continuous cdf F and density f positive at the population median $m = F^{-1}(1/2)$. Then

$$\sqrt{n}\,(\tilde{X}_n - m) \ \xrightarrow{d}\ \mathcal{N}\!\left(0,\ \frac{1}{4f(m)^2}\right),$$

so $\mathrm{Var}(\tilde{X}_n) \approx \frac{1}{4nf(m)^2}$. In terms of Efron–Stein, the terms $(\tilde{X}_n - \tilde{X}_n^{(i)})^2$ are usually 0 unless X_i is near the order statistics that determine the median, explaining looseness of the ES upper bound for this nonsmooth statistic.

Example 4.4 (U-statistic (sketch))**.** For a symmetric kernel h of order m, the U-statistic $U = \binom{n}{m}^{-1}\sum h(X_{i_1}, \ldots, X_{i_m})$ has Hoeffding decomposition $U = \theta + \sum_i \phi(X_i) + \deg \geq 2$. Efron–Stein bounds $\mathrm{Var}(U)$ by the average squared change when one coordinate is redrawn; for many kernels this recovers the classical $O(1/n)$ variance rate and can be surprisingly sharp when the linear component dominates.

5 Jackknife and Bootstrap in the Efron–Stein Light

Let $\hat{\theta} = T(X_1, \ldots, X_n)$ be any statistic.

5.1 Delete-1 Jackknife

Define leave-one-out estimates $\hat{\theta}_{(i)} = T(X_1, \ldots, \widehat{X_i}, \ldots, X_n)$ and their mean $\bar{\theta}_{(\cdot)} = \tfrac{1}{n}\sum_i \hat{\theta}_{(i)}$. The classic jackknife variance estimator is

$$\widehat{\mathrm{Var}}_{\mathrm{jack}}(\hat{\theta}) \ = \ \frac{n - 1}{n}\sum_{i=1}^{n}\big(\hat{\theta}_{(i)} - \bar{\theta}_{(\cdot)}\big)^2.$$

Why it relates to ES. For smooth T, a first-order expansion yields $\hat{\theta} - \hat{\theta}_{(i)} \approx \text{Inf}_i$, an empirical influence of X_i. The jackknife sums the *empirical* squared influences, while Efron–Stein controls variance by *population* squared influences $\mathbb{E}\big[(Z - Z^{(i)})^2\big]$.

Example 5.1 (Mean: jackknife equals truth). For $\hat{\theta} = \bar{X}$,

$$\widehat{\text{Var}}_{\text{jack}}(\bar{X}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{s^2}{n},$$

i.e. the unbiased sample variance s^2 divided by n, which equals $\text{Var}(\bar{X})$ under i.i.d. sampling (in expectation).

Remark 5.2 (When jackknife struggles). For nonsmooth or highly irregular functionals (e.g. sample maximum, hard-thresholded estimators), leave-one-out changes are zero most of the time and occasionally large, causing bias/instability. In such cases, consider the infinitesimal jackknife or delete-m jackknife below.

5.2 Infinitesimal Jackknife (IJ)

View $\hat{\theta}$ as a functional of the empirical measure $\hat{P} = \frac{1}{n} \sum_i \delta_{X_i}$. The (linearized) influence function ψ gives

$$\hat{\theta} - \theta \approx \frac{1}{n} \sum_{i=1}^{n} \psi(X_i), \qquad \mathbb{E}[\psi(X)] = 0.$$

Then

$$\text{Var}(\hat{\theta}) \approx \frac{1}{n^2} \sum_{i=1}^{n} \psi(X_i)^2 = \frac{1}{n} \widehat{\text{Var}}\big(\psi(X)\big),$$

providing a fast variance estimate once ψ (or an estimate thereof) is available. In many M-estimation problems, ψ arises from a score/estimating equation.

5.3 Delete-m Jackknife

For $m \to \infty$ with $m/n \to 0$, recompute $\hat{\theta}$ leaving out m points, average across all (or many) subsets, and rescale to estimate variance. This smooths the instability of delete-1 for nonsmooth statistics (e.g. median) and often improves finite-sample performance.

5.4 Bootstrap

Generate B resamples by sampling n points with replacement from $\{X_i\}$; compute $\hat{\theta}_b^* = T(X_1^{*(b)}, \ldots, X_n^{*(b)})$; estimate

$$\widehat{\text{Var}}_{\text{boot}}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} \big(\hat{\theta}_b^* - \bar{\hat{\theta}}^*\big)^2, \quad \bar{\hat{\theta}}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^*.$$

Intervals. Percentile, basic, studentized, and BCa (bias-corrected and accelerated) intervals offer increasing accuracy, with BCa often preferred for skewed or biased statistics.

Conceptual link to ES. Each bootstrap resample reweights (and replicates) coordinates, effectively randomizing their contributions. The resulting empirical variance of $\hat{\theta}^*$ estimates the same sensitivity structure that ES bounds theoretically.

6 Case Studies Revisited

6.1 Median

Asymptotically, $\mathrm{Var}(\tilde{X}_n) \approx 1/(4nf(m)^2)$. Practically:

- Delete-1 jackknife can be biased/unstable.

- Delete-m (moderate m) or IJ performs better.

- Bootstrap (with BCa) gives reliable variance and CIs, especially in small to moderate n.

6.2 Empirical risk (Lipschitz loss)

Let $Z = \frac{1}{n} \sum_{i=1}^n \ell(\theta; X_i)$ with ℓ L-Lipschitz in X. Then changing a single X_i by an independent redraw changes Z by at most L/n (heuristically), yielding the quick bound

$$\mathrm{Var}(Z) \;\lesssim\; \tfrac{1}{2}n \cdot (L/n)^2 \;=\; \frac{L^2}{2n}.$$

Jackknife/Bootstrap give data-driven refinements, often much tighter when ℓ has light tails.

6.3 Sample variance (sketch)

For $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$, ES yields $\mathrm{Var}(S^2) = O(1/n)$ under finite fourth moment. Jackknife variance of S^2 is consistent; the bootstrap is also consistent and convenient for CIs on σ^2 (with caution under heavy tails).

7 Practical Guidance

- **Need an analytic upper bound?** Use Efron–Stein; it is fast, assumption-lean, and insightful for sensitivity audits.

- **Smooth statistics, fast estimate?** Use jackknife or IJ.

- **Nonsmooth/complex statistics or full CIs?** Use bootstrap; prefer BCa for skew/bias.

- **Computational budget tight?** IJ or jackknife often give near-bootstrap accuracy at a fraction of the cost.

8 Minimal "Recipes" You Can Implement

Delete-1 Jackknife

1. For $i = 1, \ldots, n$, compute $\hat{\theta}_{(i)}$ on the sample with X_i removed.

2. Let $\bar{\hat{\theta}}_{(\cdot)} = \frac{1}{n} \sum_i \hat{\theta}_{(i)}$.

3. Report $\widehat{\mathrm{Var}}_{\mathrm{jack}} = \frac{n-1}{n} \sum_i (\hat{\theta}_{(i)} - \bar{\hat{\theta}}_{(\cdot)})^2$.

Infinitesimal Jackknife (conceptual)

1. Obtain (or estimate) the influence function ψ for T at the empirical distribution.

2. Compute $\widehat{\mathrm{Var}}_{\mathrm{IJ}}(\hat{\theta}) = \frac{1}{n^2} \sum_{i=1}^{n} \psi(X_i)^2$ (or its plug-in analog).

Bootstrap (nonparametric)

1. For $b = 1, \ldots, B$: sample with replacement n points from $\{X_i\}$; compute $\hat{\theta}_b^*$.

2. Let $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_b \hat{\theta}_b^*$; report $\widehat{\mathrm{Var}}_{\mathrm{boot}} = \frac{1}{B-1} \sum_b (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2$.

3. For CIs, use percentile or BCa rules.

9 Connections at a Glance

Method	Object perturbed	Core idea
Efron–Stein	Single $X_i \to$ independent copy X_i'	Sum of expected squared leave-one-coo changes bounds $\mathrm{Var}(Z)$.
Jackknife	Remove observed X_i	Empirical squared leave-one-out changes es variance (best for smooth stats).
Infinitesimal jackknife	Infinitesimal reweighting of each X_i	Influence function linearization yields fast, a variance estimates.
Bootstrap	Resample with replacement	Simulate the sampling distribution of $\hat{\theta}$; v and CIs from resamples.

Acknowledgments and Pointers

References

[1] B. Efron and C. Stein (1981). The jackknife estimate of variance. *Annals of Statistics* **9**(3): 586–596.

[2] B. Efron (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**(1): 1–26.

[3] B. Efron and R. J. Tibshirani (1993). *An Introduction to the Bootstrap.* Chapman & Hall/CRC.

[4] S. Boucheron, G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press.