Mean-Median gaps in real line and Inference in Infinite-dimensions

Arun Kumar Kuchibhotla

Carnegie Mellon University

13 November 2026

Table of contents

1. Introduction & Motivation

2. The Mean–Median Gap

3. Applications

4. Conclusions

Joint work with Abhimanyu Choudhary (CMU)
Joint work with Larry Wasserman and Sivaraman Balakrishnan (CMU)

Sign test & inversion

- Suppose $W_1, \ldots, W_m$ are independent real-valued random variables. The Scheffé–Tukey interval is a confidence interval for median.
- Let the augmented order statistics be

$$-\infty = W_{(0)} \le W_{(1)} \le \ldots \le W_{(m)} \le W_{(m+1)} = \infty,$$

and let $c_{m,\gamma}$ be the independently randomized $(1-\gamma)$-th binomial quantile:

$$\mathbb{P}\big(\mathrm{Binom}(m, 1/2) \le c_{m,\gamma}\big) = 1 - \gamma.$$

- The Scheffé–Tukey interval is

$$\widehat{\mathrm{CI}}^{\mathrm{sign}}_{m,\alpha} = [W_{(m-c_{m,\alpha/2})},\, W_{(c_{m,\alpha/2}+1)}].$$

- Define $\mathrm{Med}(W) = \{a : \min\{\mathbb{P}(W \ge a), \mathbb{P}(W \le a)\} \ge 1/2\}$. Then

$$\mathbb{P}\big(a \notin \widehat{\mathrm{CI}}^{\mathrm{sign}}_{m,\alpha}\big) \le \alpha, \qquad \text{for all } m \ge 1 \text{ and } a \in \bigcap_{j=1}^{m} \mathrm{Med}(W_j).$$

The HulC (Kuchibhotla et al., 2024, JRSS–B)

- More generally, for any $\mu \in \mathbb{R}$ and independent $W_1, \ldots, W_m$,

$$\mathbb{P}(\mu \notin [W_{(m-c_{m,\alpha/2})}, W_{(c_{m,\alpha/2}+1)}]) \;\leq\; \alpha \exp(2m^2\Delta^2),$$

  where the average median bias is

$$\Delta = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{2} - \min\{\mathbb{P}(W_j \geq \mu), \mathbb{P}(W_j \leq \mu)\}\right)_+ .$$

- If $\mu \in \mathrm{Med}(W_j)$ for every j, then $\Delta = 0$.
- The miscoverage upper bound tends to α if $m\Delta \to 0$.
- Taking $m \geq \lceil \log_2(2/\alpha) \rceil$ ensures finite endpoints almost surely.
- Below this threshold, the interval can equal $\mathbb{R}$.

The HulC: splitting, validity and width

- Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$, with target $\theta(P)$. Partition

$$\{1, \ldots, n\} = S_1 \cup \cdots \cup S_B, \qquad \widehat{\theta}_j = \widehat{\theta}(\{X_i : i \in S_j\}).$$

We assume $|S_j| = n/B$, an integer. Set

$$\widehat{\mathrm{CI}}^{\mathrm{HulC}}_{n,\alpha} = [\widehat{\theta}_{(B - c_{B,\alpha/2})}, \widehat{\theta}_{(c_{B,\alpha/2}+1)}].$$

- Then $\mathbb{P}\big(\theta(P) \notin \widehat{\mathrm{CI}}^{\mathrm{HulC}}_{n,\alpha}\big) \leq \alpha \exp(2B^2\Delta^2)$, where

$$\Delta = \frac{1}{B}\sum_{j=1}^{B}\left(\frac{1}{2} - \min\{\mathbb{P}(\widehat{\theta}_j \geq \theta(P)), \mathbb{P}(\widehat{\theta}_j \leq \theta(P))\}\right)_+ .$$

- For fixed $B \geq \lceil \log_2(2/\alpha) \rceil$, suppose

$$r_{n/B}(\widehat{\theta}_j - \theta(P)) \xrightarrow{d} L, \qquad L \text{ has median zero and is continuous at zero.}$$

Then $B\Delta \to 0$, coverage is asymptotically at least $1 - \alpha$, and the width is $O_P(1/r_{n/B})$. This includes asymptotically normal and many asymptotically symmetric estimators.

An example (Bernoulli probability)

- If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Ber}(p)$, compute

$$\widehat{\theta}_j = \frac{B}{n} \sum_{i \in S_j} X_i \overset{d}{=} \frac{\text{Binom}(n/B, p)}{n/B}.$$

- Normal approximation to binomial distribution implies that for fixed $p \in (0,1)$ and admissible B,

$$\limsup_{n \to \infty} \mathbb{P}_p\big(p \notin \widehat{\text{CI}}^{\text{HulC}}_{n,\alpha}\big) \leq \alpha.$$

- But

$$\limsup_{n \to \infty} \sup_{p \in (0,1)} \mathbb{P}_p\big(p \notin \widehat{\text{CI}}^{\text{HulC}}_{n,\alpha}\big) = 1.$$

- Wald, ordinary empirical bootstrap and subsampling intervals also fail in this uniform sense; empirical Bernstein gives a uniformly valid, conservative alternative.

The mean–median gap

- We want an interval to cover the mean of $\widehat{\theta}_j$, while HulC covers a median in finite samples. Hence, HulC can be inflated if the mean–median gap is upper bounded.
- Hamza (1995) proved that

$$\left| \mathbb{E}\left(\frac{B}{n} \sum_{i \in S_j} X_i \right) - \mathrm{Med}\left(\frac{B}{n} \sum_{i \in S_j} X_i \right) \right| \leq \frac{B \log 2}{n}, \qquad X_i \overset{\text{iid}}{\sim} \mathrm{Ber}(p).$$

- Consequently, for every $n \geq B \geq \log_2(2/\alpha)$,

$$\sup_{p \in [0,1]} \mathbb{P}_p \left(p \notin \left[\widehat{\theta}_{(B - c_{B,\alpha/2})} - \frac{B \log 2}{n}, \widehat{\theta}_{(c_{B,\alpha/2}+1)} + \frac{B \log 2}{n} \right] \right) \leq \alpha.$$

- Hamza's bound also applies to Poisson laws; related bounds exist for geometric, negative binomial, beta and gamma distributions.

The mean–median gap beyond the binomial

- Such bounds for nonparametric classes, such as bounded distributions, are less studied. Oleszkiewicz (2001) proved a bound for independent bounded random variables.
- For independent X_i with finite variances, boundedness can be weakened to

$$\mathbb{E}|X_i - \mathbb{E}X_i|^3 \le K\,\mathbb{E}|X_i - \mathbb{E}X_i|^2, \qquad \forall i. \tag{1}$$

- For every $n \ge 1$,

$$\boxed{\left|\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \operatorname{Med}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right| \le \frac{\mathfrak{C}K}{n}, \qquad \mathfrak{C} = 3C_{\mathrm{BE}}.} \tag{2}$$

- Note that the usual standard deviation bound only yields the bound

$$\frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathsf{Var}(X_i)\right)^{1/2} \le \frac{K}{\sqrt{n}}.$$

The inflated HulC and its constants

For independent observations $X_1, \ldots, X_n$ with common mean μ with

$$\mathbb{E}|X_i - \mathbb{E}X_i|^3 \le K\,\mathbb{E}|X_i - \mathbb{E}X_i|^2, \qquad \forall i,$$

put $\widehat{\theta}_j = (B/n)\sum_{i\in S_j} X_i$. Then

$$\boxed{\mathbb{P}\left(\mu \notin \left[\widehat{\theta}_{(B-c_{B,\alpha/2})} - \frac{\mathfrak{C}KB}{n},\, \widehat{\theta}_{(c_{B,\alpha/2}+1)} + \frac{\mathfrak{C}KB}{n}\right]\right) \le \alpha.}$$

Summands	Input	C_{BE}	$\mathfrak{C} = 3C_{\mathrm{BE}}$
Independent	Shevtsova (2013)	.5583	1.6749
IID	Shevtsova (2013)	.4690	1.4070
Independent	Bell (2026)	$\approx .4499$	≈ 1.3497
IID	Kołodziejek–Rodziewicz (2026)	$\approx .4097$	≈ 1.2292

Poincaré, log-Sobolev and the required moment ratio

- A probability measure P satisfies Poincaré with constant C_{PC}, or LSI with constant C_{LS}, if

$$\mathrm{Var}_P(g) \le C_{\mathrm{PC}}\, \mathbb{E}_P\|\nabla g\|_2^2, \qquad \mathrm{Ent}_P(g^2) \le 2C_{\mathrm{LS}}\, \mathbb{E}_P\|\nabla g\|_2^2,$$

  for all locally Lipschitz g, where

$$\mathrm{Ent}_P(Z) = \mathbb{E}_P[Z\log Z] - \mathbb{E}_P Z \log \mathbb{E}_P Z.$$

- LSI implies Poincaré with $C_{\mathrm{PC}} \le C_{\mathrm{LS}}$; Poincaré is the weaker condition.
- For every 1-Lipschitz g,

$$\boxed{\mathbb{E}|g(X) - \mathbb{E}g(X)|^3 \le K\, \mathbb{E}|g(X) - \mathbb{E}g(X)|^2, \qquad K = \sqrt{5}\min\{\sqrt{C_{\mathrm{PC}}}, \sqrt{C_{\mathrm{LS}}}\}.}$$

Weak log-concavity: an explicit supplied constant

- A probability measure P with Lebesgue density proportional to e^{-V} is said to be weakly log-concave if there exists $\kappa > 0$, $M \geq 0$ and

$$\frac{\langle \nabla V(x) - \nabla V(y), x - y \rangle}{\|x - y\|_2^2} \;\geq\; \kappa - \frac{2\sqrt{M}}{r} \tanh\left(\frac{\sqrt{M} r}{2}\right), \qquad r = \|x - y\|_2 > 0.$$

- Results of Conforti–Lacker–Pal (2025) imply that

$$\max\{C_{\mathrm{PC}}, C_{\mathrm{LS}}\} \;\leq\; \kappa^{-1} e^{M/\kappa}, \quad \Rightarrow \quad K = \sqrt{5/\kappa}\, e^{M/(2\kappa)}.$$

- Ordinary log-concavity means convex V. It implies Poincaré but need not imply LSI.

- Many distributions that are log-concave in the tails satisfy weak log-concavity. This includes mixtures of Gaussians.

Inflated HulC: Inference for Mean

- Recall the setting of independent $X_1, \ldots, X_n$ with mean μ satisfying the moment condition.
- Recall the inflated HulC interval

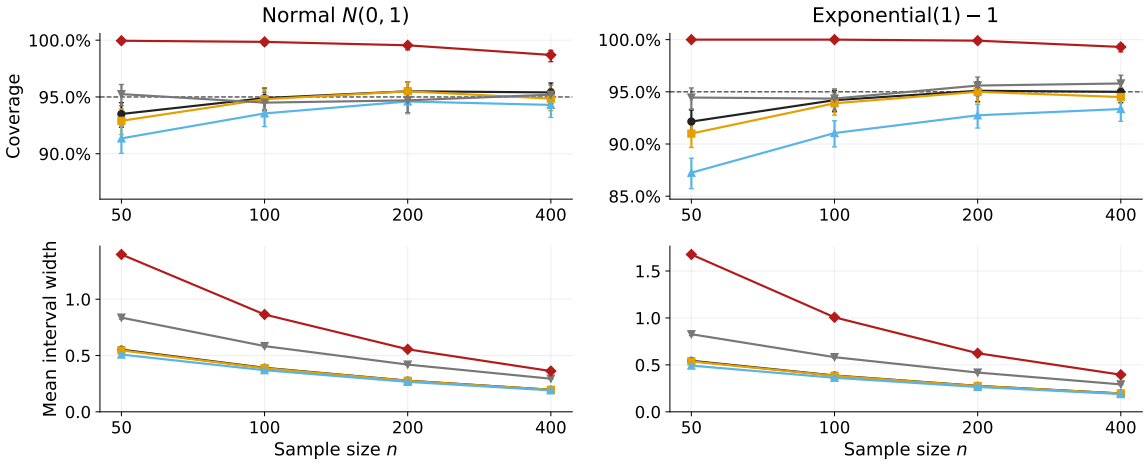
$$\widehat{\mathrm{CI}}_{n,\alpha}^{\mathrm{inf-HulC}} = \left[\widehat{\theta}_{(B-c_{B,\alpha/2})} - \frac{\mathfrak{C}KB}{n}, \widehat{\theta}_{(c_{B,\alpha/2}+1)} + \frac{\mathfrak{C}KB}{n}\right].$$

- Let $\sigma^2 \geq \max_i \mathrm{Var}(X_i)$. With fixed α and $B \asymp \log_2(2/\alpha)$, the width has the Bernstein-like form

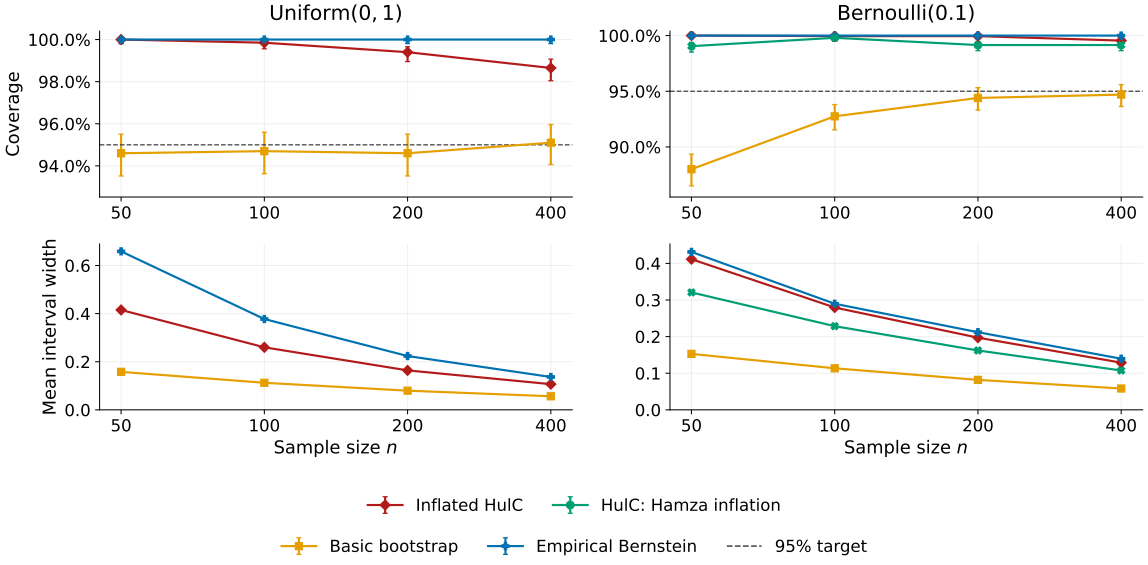
$$O_P\left(\sqrt{\frac{\sigma^2 \log(1/\alpha)}{n}} + \frac{K \log(1/\alpha)}{n}\right).$$

- Coverage is uniform and finite-sample over the moment class. Unlike empirical Bernstein bound, we do not require boundedness.
- Symmetry about the common mean already makes uninflated HulC valid.

An illustration: unbounded distributions (Poincaré family)

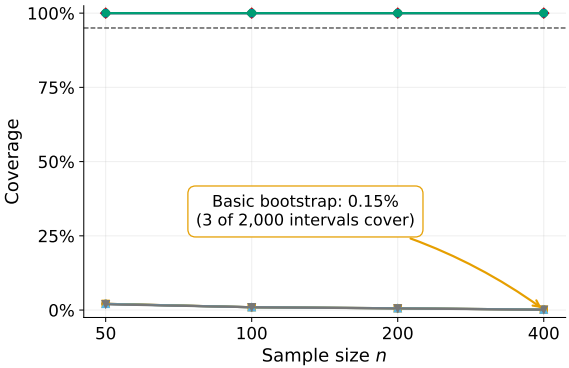


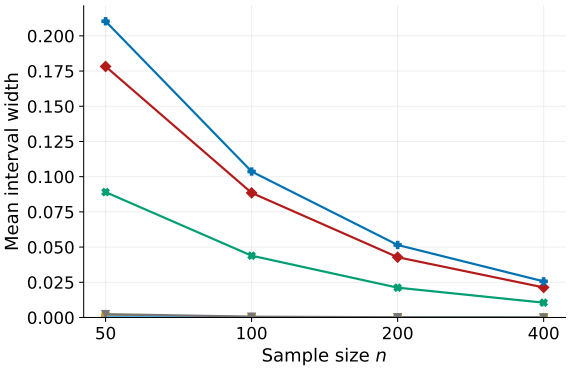
An illustration: bounded distributions



An illustration: rare events

Bernoulli(p_n), $p_n = n^{-2}$





HulC (no inflation) · Inflated HulC · HulC: Hamza inflation

Wald · Basic bootstrap · Subsampling · Empirical Bernstein · 95% target

Gap for general functions: sensitivity and interaction

- Let $T = f(X_1, \ldots, X_n)$ for independent inputs. Write x^i and x^{ij} for coordinate replacements, and define

$$\Delta_i f(x) = f(x) - f(x^i), \qquad \Delta_{ij} f(x) = f(x) - f(x^i) - f(x^j) + f(x^{ij}).$$

- Assume f is stable with first and second order influences, i.e.,

$$|\Delta_i f| \le c_i \quad \text{and} \quad |\Delta_{ij} f| \le c_{ij} \quad \text{for all} \quad 1 \le i \le j \le n, x.$$

Then

$$|\mathbb{E}T - \mathrm{Med}(T)| \;\le\; \min\left\{ \frac{1}{2}\sqrt{\sum_i c_i^2},\; 6\max_i c_i + 3\sqrt{\sum_{i<j} c_{ij}^2} \right\}.$$

- The first bound is sharp in that there is a T for which this bound is an equality.
- Moment versions can replace almost-sure conditions: Friedrich (1989), Chen–Shao (2007), and Shao–Zhang (2025).

Application 1 (U-statistics): through degeneracy

- Suppose the X_i are iid, h is symmetric and square-integrable, and

$$\theta(P) = \mathbb{E}_P h(X_1, X_2), \qquad U_n = \binom{n}{2}^{-1} \sum_{k<\ell} h(X_k, X_\ell).$$

  This is the UMVUE over the full nonparametric IID family.

- Put $g(x) = \mathbb{E}[h(x, X_2)] - \theta(P)$. If $\mathrm{Var}(g(X_1)) > 0$,

$$\sqrt{n}(U_n - \theta(P)) \xrightarrow{d} N(0, 4\,\mathrm{Var}(g(X_1))).$$

- If $g = 0$ almost surely,

$$n(U_n - \theta(P)) \xrightarrow{d} \sum_{r \geq 1} \lambda_r (Z_r^2 - 1),$$

  where the Z_r are independent $N(0,1)$ variables and the λ_r are eigenvalues of the centered kernel operator.

- Ths makes classical inference difficult, particularly near degeneracy.

U-statistics: an explicit finite-sample interval

- For $\|h\|_\infty \le K$ and $n/B \ge 2$, compute

$$\widehat{\theta}_j = \binom{n/B}{2}^{-1} \sum_{\substack{k<\ell \\ k,\ell \in S_j}} h(X_k, X_\ell), \qquad c_i = \frac{4KB}{n}, \quad c_{ij} = \frac{8KB^2}{n(n-B)}.$$

Enlarge the sign interval $[\widehat{\theta}_{(B-c_{B,\alpha/2})}, \widehat{\theta}_{(c_{B,\alpha/2}+1)}]$ at both ends by the explicit mean–median gap

$$K \min\left\{1,\, 2\sqrt{\frac{B}{n}},\, \frac{24B}{n} + \frac{12\sqrt{2}B}{\sqrt{n(n-B)}}\right\}.$$

- It has coverage at least $1-\alpha$ throughout the bounded-kernel class, without estimating the projection variance or deciding whether the kernel is degenerate.
- This can be compared with the empirical Bernstein bounds of Peel, Anthoine, and Ralaivola (2010).

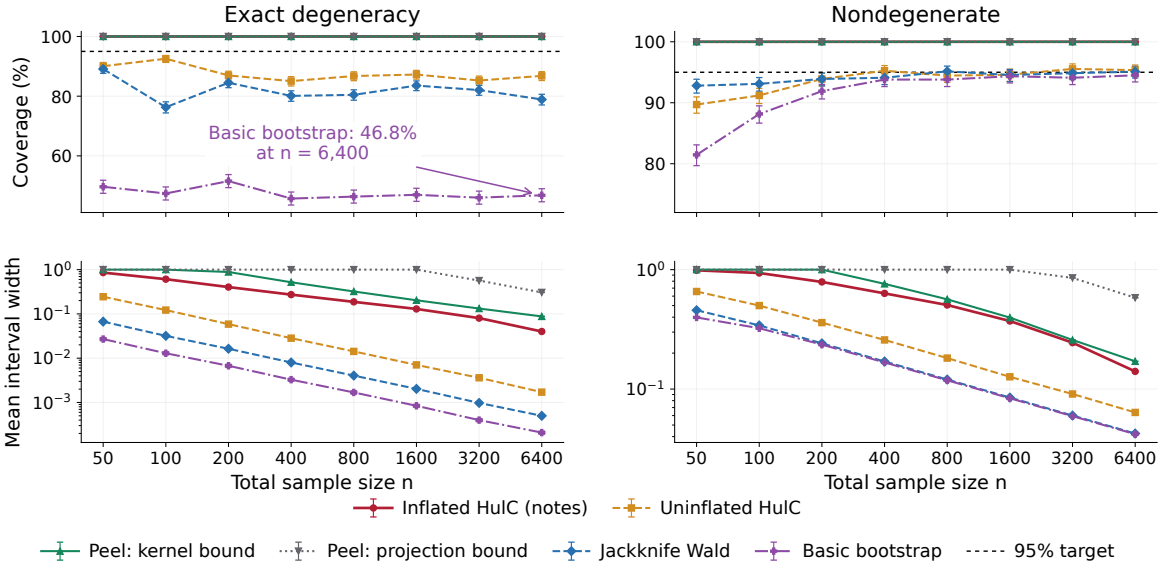
Illustration: two transitions to degeneracy

- **Product kernel:** $X \in \{-1, 1\}$, $\mathbb{E}X = \eta$, $h(x, y) = xy$. This estimates the squared mean, measuring signal strength for testing a zero mean.

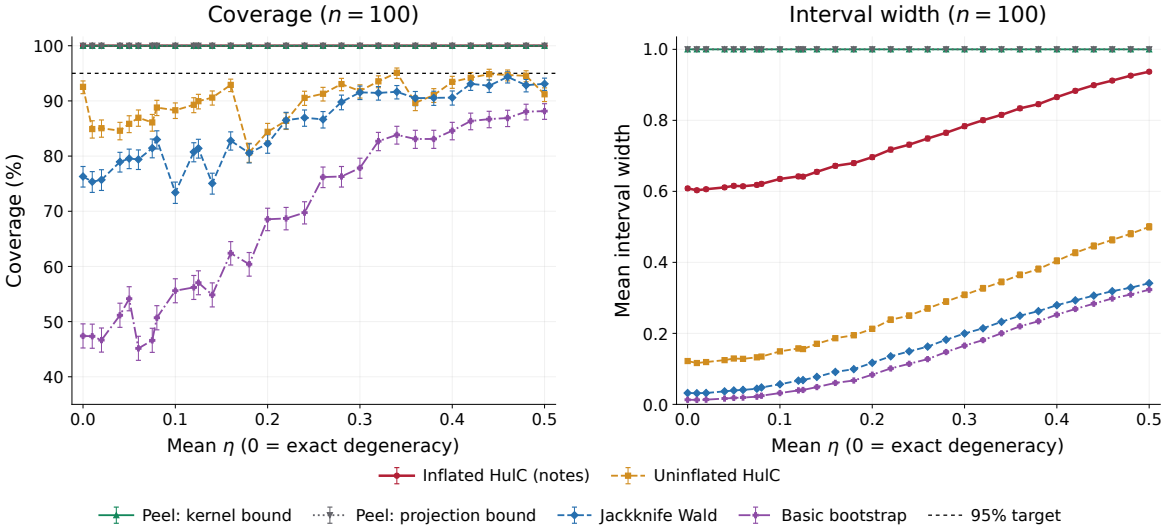
$$\theta = \eta^2, \qquad \mathrm{Var}(g(X)) = \eta^2(1 - \eta^2), \qquad \eta = 0,\ 1/\sqrt{n},\ 2/\sqrt{n},\ .5.$$

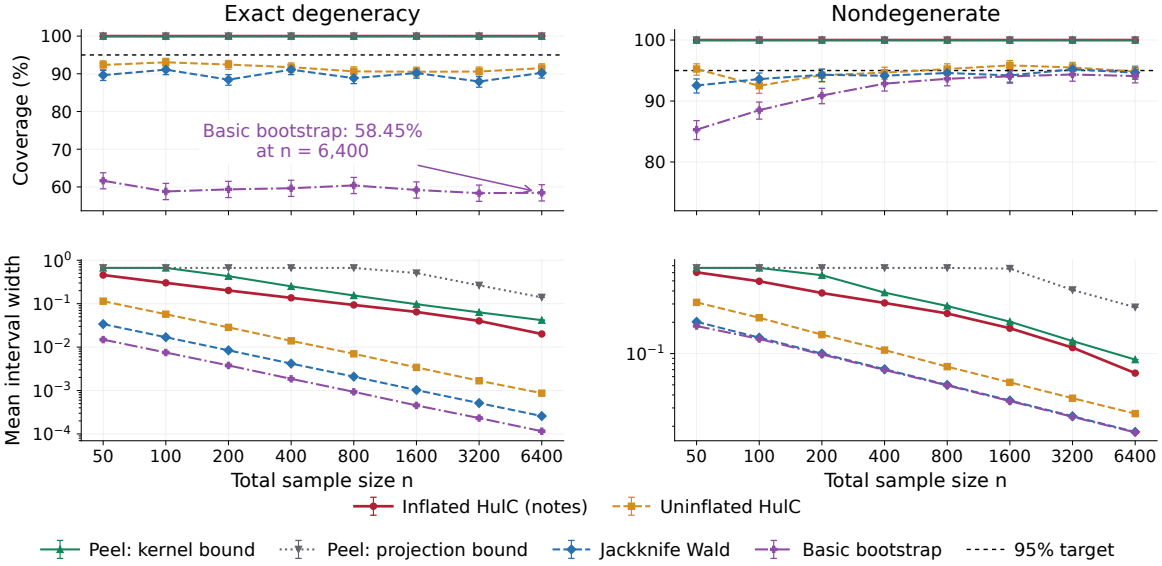
- This is degenerate when $\eta = 0$ and non-degenerate for $\eta \neq 0$.
- **Collision kernel:** $X \in \{1, 2, 3\}$, $h(x, y) = \mathbf{1}\{x = y\}$, $\theta = \sum_k p_k^2$. This estimates the match probability of two categorical draws, a measure of concentration.
- This is degenerate when $p = (1/3, 1/3, 1/3)$ and non-degenerate otherwise.
- Transitions from non-degeneracy to degeneracy cause failure of bootstrap and subsampling techniques.

Illustration: product U-statistic



Mean-square inference: approaching degeneracy at $n = 100$




Illustration: collision U-statistic
Exact degeneracy
Nondegenerate
Coverage (%)
Basic bootstrap: 58.45%
at n = 6,400
Mean interval width
Total sample size n
Total sample size n
Inflated HulC (notes)
Uninflated HulC
Peel: kernel bound
Peel: projection bound
Jackknife Wald
Basic bootstrap
95% target

Application 2 (High-dimensional mean): COSI

- Suppose $X_1, \ldots, X_n$ are iid random vectors in $\mathbb{R}^{\mathcal{T}}$, with $\mu(t) = \mathbb{E}X_i(t)$.
- Partition

$$\{1, 2, \ldots, n\} = S_0 \cup S_1 \cup \cdots \cup S_B \quad \text{with} \quad |S_0| = \frac{n}{2},\ |S_j| = \frac{n}{2B}.$$

- Set $\vartheta(t) = |S_0|^{-1} \sum_{i \in S_0} X_i(t)$, and any positive scale $s(t)$. Set

$$\widehat{\theta}_j(t) = \frac{1}{|S_j|} \sum_{i \in S_j} X_i(t), \qquad R = \sup_{t \in \mathcal{T}} \left(\frac{|\widehat{\theta}_j(t) - \vartheta(t)|}{s(t)} : j = 1, \ldots, B \right)_{(c_{B,\alpha}+1)}.$$

- The COSI confidence set is

$$\widehat{\mathrm{CI}}^{\mathrm{COSI}} = \bigotimes_{t \in \mathcal{T}} [\vartheta(t) - Rs(t),\ \vartheta(t) + Rs(t)],$$

which maintains coverage if

$$\Delta = \sup_{t \in \mathcal{T}} \frac{1}{B} \sum_{j=1}^{B} \left(\frac{1}{2} - \min\{\mathbb{P}(\widehat{\theta}_j(t) \geq \mu(t)), \mathbb{P}(\widehat{\theta}_j(t) \leq \mu(t))\} \right)_+ = o(1/B) \quad \text{as} \quad n \to \infty.$$

- No dimension adjustment is needed for validity.

Inflated COSI and simultaneous-mean simulations

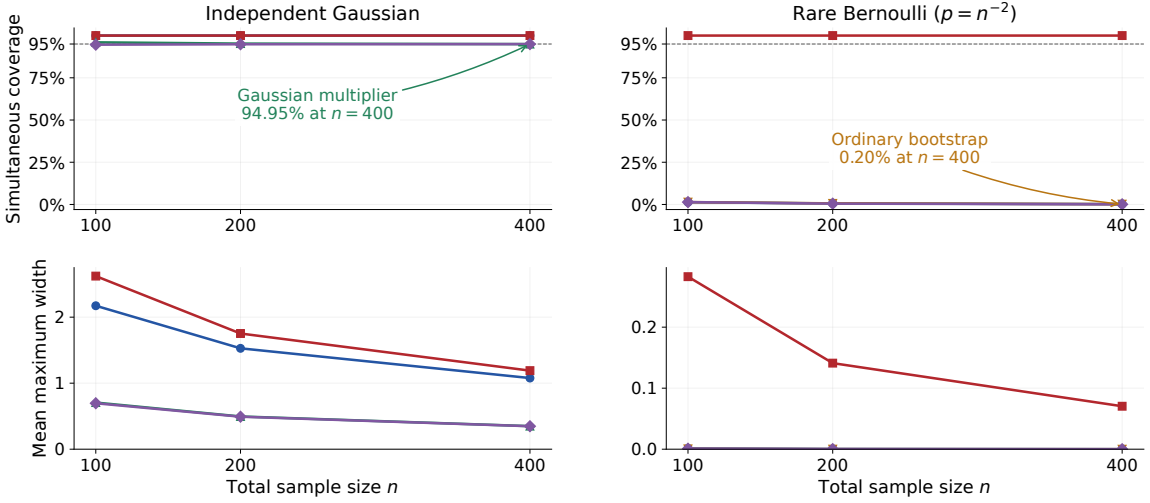
- If $\mathbb{E}|X_i(t) - \mu(t)|^3 \le K \operatorname{Var}(X_i(t))$ uniformly over $t \in \mathcal{T}$, then inflated COSI set is

$$\widehat{\mathrm{CI}}^{\mathrm{inf-COSI}} = \bigotimes_{t \in \mathcal{T}} \left[\vartheta(t) - Rs(t) - \frac{2\mathfrak{C}KB}{n},\ \vartheta(t) + Rs(t) + \frac{2\mathfrak{C}KB}{n} \right]$$

  covers μ simultaneously with probability at least $1 - \alpha$.

- Use $\mathfrak{C} = 1.407$ for IID observations; the general independent-summand constant is 1.6749.
- Note that COSI maintains finite-sample coverage over the moment class considered, with no restrictions on complexity of $\mathcal{T}$.
- These assumptions are insufficient to imply general validity of bootstrap or resampling or Gaussian approximation.

Illustration: regular and rare high-dimensional means, $d = 100$



Application 3 (Inference for Wasserstein distance)

- For independent samples from μ, ν with finite first moments,

$$W_1(\mu,\nu) = \sup_{f\in \mathrm{Lip}(1)} |\mathbb{E}_\mu f - \mathbb{E}_\nu f|.$$

- The empirical distance is biased; a useful limit law needs additional conditions.
- Poincaré inequality for μ, ν imply the moment comparison required for all $f \in \mathrm{Lip}(1)$.
- With the partition as before, set

$$\widehat{\mu}_j = \frac{1}{|S_j|}\sum_{i\in S_j} \delta_{X_i} \quad \text{and} \quad \widehat{\nu}_j = \frac{1}{|S_j|}\sum_{i\in S_j} \delta_{Y_i}, \quad \text{for} \quad 0 \le j \le B,$$

and let

$$R = \sup_{f\in \mathrm{Lip}(1)} \left(|(\widehat{\mu}_j - \widehat{\nu}_j - \widehat{\mu}_0 + \widehat{\nu}_0)f| : j = 1,\dots,B\right)_{(c_{B,\alpha}+1)}.$$

- A finite-sample $1-\alpha$ COSI based interval is

$$\left[W_1(\widehat{\mu}_0,\widehat{\nu}_0) \ \pm \left\{R + \frac{7.5B}{n}\sqrt{\max(C_{\mathrm{PC},\mu}, C_{\mathrm{PC},\nu})}\right\}\right].$$

Comments

- The proposed confidence interval makes no reference to the dimension of the data or the complexity of the probability measures.
- Interestingly, the bias of the empirical Wasserstein distance can be estimated consistently under Poincaré inequality or LSI. This implies an alternative correction using Chernoff inequality.
- Hence, we obtain two different confidence intervals for the Wasserstein distance under Poincaré inequality or LSI.
- Note that the COSI based interval only uses these constants in the second order and hence, specifying a conservative upper bound for C_{PC} or C_{LS} would not cost as much.
- This is not the case with the bias based CI.

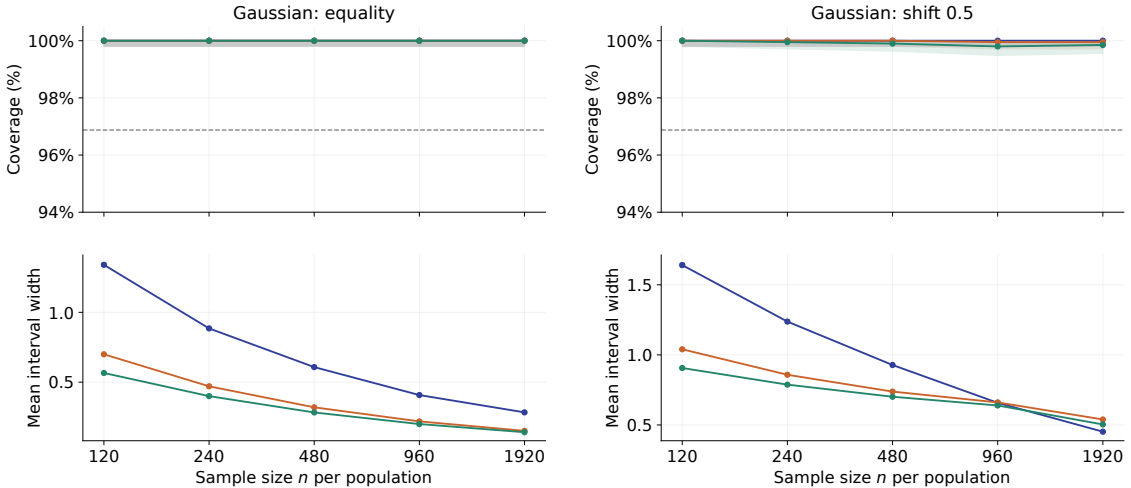
Illustration: Wasserstein simulation study

- We simulate the coverage and width of these confidence intervals in the following settings:

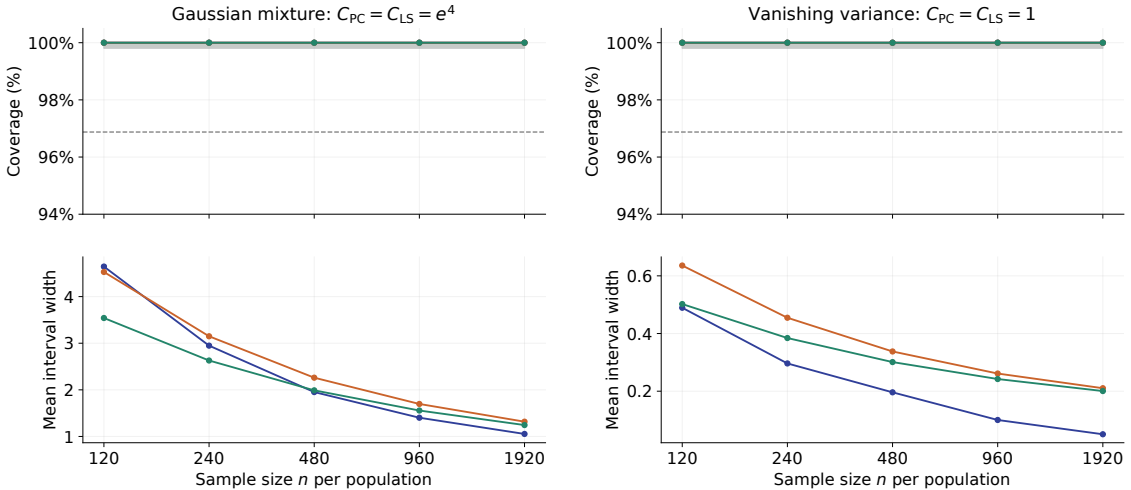
Populations (μ, ν)	True W_1	Supplied $C_{\mathrm{PC}} = C_{\mathrm{LS}}$
$N(0,1),\ N(0,1)$	0	1
$N(0,1),\ N(.5,1)$	$.5$	1
Mixture, shifted by .5	$.5$	e^4
$N(0,1/n),\ N(.1,1/n)$	$.1$	1

- In the last case, we supply $O(1)$ constants as upper bounds for the Poincaré constant when the true constant is $1/n$.
- Observed coverage across all eight settings and five sample sizes: 99.80%–100%.
- When the supplied constants are extremely conservative the bias estimation based methods suffer.

Wasserstein: equality and translation



Wasserstein: supplied constants and vanishing variation



Conclusions

- HulC turns median information into confidence intervals. A certified mean–median gap turns it into finite-sample inference for a mean.
- A third-to-second moment ratio yields an n^{-1} correction without variance estimation or a variance lower bound or bounded observations.
- This argument can be extended to other functionals of independent data including U-statistics through degeneracy.
- COSI extends the univariate construction to an arbitrary index set. Wasserstein–1 follows by projecting a simultaneous Lipschitz band.
- Validity in rare and degenerate regimes can cost width in regular problems.

Conclusions

- HulC turns median information into confidence intervals. A certified mean–median gap turns it into finite-sample inference for a mean.
- A third-to-second moment ratio yields an n^{-1} correction without variance estimation or a variance lower bound or bounded observations.
- This argument can be extended to other functionals of independent data including U-statistics through degeneracy.
- COSI extends the univariate construction to an arbitrary index set. Wasserstein–1 follows by projecting a simultaneous Lipschitz band.
- Validity in rare and degenerate regimes can cost width in regular problems.

Thank You!!

Detailed theoretical supplement

- Sign inversion, randomized cutoffs, ties and the median-bias factor.
- Mean–median bounds: Berry–Esseen, Poincaré and log-concavity.
- General functions: projection comparison, its numerical constant, and the small-interaction bound.
- U-statistics: Hoeffding decomposition, the explicit inflation, and both empirical Bernstein comparators.
- COSI: conditional validity, mean–median inflation, and the corrected general median-bias guarantee.
- Wasserstein inference: projection, signed transport, empirical-bias bounds, and concentration certificates.

Supplement: the randomized binomial cutoff

Let F_B and p_B be the CDF and PMF of $\mathrm{Binom}(B, 1/2)$. For $0 < \gamma < 1$, define

$$q = \min\{k : F_B(k) \geq 1 - \gamma\}, \qquad \tau = \frac{1 - \gamma - F_B(q-1)}{p_B(q)}, \qquad c_{B,\gamma} = \begin{cases} q, & \text{with probability } \tau, \\ q - 1, & \text{with probability } 1 - \tau. \end{cases}$$

The cutoff is independent of the observations. Since $\tau F_B(q) + (1 - \tau)F_B(q-1) = 1 - \gamma$,

$$\mathbb{P}\big(\mathrm{Binom}(B, 1/2) > c_{B,\gamma}\big) = \gamma,$$

where the probability includes the cutoff randomization.

- Scalar sign inversion uses $\gamma = \alpha/2$ and the *same* cutoff at both endpoints: $B - c_{B,\alpha/2}$ and $c_{B,\alpha/2} + 1$. At $\alpha = .05, B = 6$, the full range is used with probability .9 and ranks $(2, 5)$ with probability .1.
- COSI uses $\gamma = \alpha$ and residual rank $c_{B,\alpha} + 1$. At $\alpha = .05, B = 5$, rank 5 is used with probability .88 and rank 4 with probability .12.
- Finite scalar endpoints require $B \geq \lceil \log_2(2/\alpha) \rceil$; a finite COSI rank requires $B \geq \lceil \log_2(1/\alpha) \rceil$. These follow from $F_B(B-1) = 1 - 2^{-B}$.
- Below the relevant threshold, the randomized cutoff can select an infinite endpoint. Use $W_{(0)} = -\infty$, $W_{(B+1)} = \infty$ and the $(B+1)$st residual $+\infty$.

Supplement: sign inversion, exactness and ties

Let $W_1, \ldots, W_m$ be independent and let μ be a median of every W_j:

$$\min\{\mathbb{P}(W_j \geq \mu), \mathbb{P}(W_j \leq \mu)\} \geq \tfrac{1}{2}.$$

Consider the closed interval

$$I = \left[W_{(m-c_{m,\alpha/2})},\, W_{(c_{m,\alpha/2}+1)}\right].$$

- If μ is below its lower endpoint, at least $c_{m,\alpha/2}+1$ observations are strictly above μ. If μ is above its upper endpoint, at least that many are strictly below μ.
- The independent strict-positive and strict-negative indicators each have success probabilities at most $1/2$. Their sums are stochastically dominated by $\mathrm{Binom}(m, 1/2)$, so

$$\mathbb{P}(\mu \notin I) \leq \mathbb{P}\Big(\sum\nolimits_j \mathbf{1}\{W_j > \mu\} > c_{m,\alpha/2}\Big) + \mathbb{P}\Big(\sum\nolimits_j \mathbf{1}\{W_j < \mu\} > c_{m,\alpha/2}\Big) \leq \alpha.$$

- If $\mathbb{P}(W_j > \mu) = \mathbb{P}(W_j < \mu) = 1/2$ for every j, the signs are independent and fair. The two rejection tails are disjoint, and each has probability $\alpha/2$; coverage is exactly $1-\alpha$.
- Atoms at μ can only make the closed interval conservative. Nonunique medians cause no problem: the proof holds for every common median separately.

Supplement: the scalar median-bias factor

For independent W_j and any target μ, put

$$\delta_j = \left(\tfrac{1}{2} - \min\{\mathbb{P}(W_j \geq \mu), \mathbb{P}(W_j \leq \mu)\}\right)_+, \qquad \Delta = \frac{1}{m}\sum\nolimits_{j=1}^{m} \delta_j.$$

First suppose there are no ties and write $\mathbb{P}(W_j > \mu) = 1/2 + \varepsilon_j$; then $|\varepsilon_j| = \delta_j$.

- For each fixed cutoff, two-sided sign rejection is invariant under $s \mapsto -s$. Relative to independent fair signs, the average likelihood ratio of a pair is

$$\frac{1}{2}\left\{\prod\nolimits_{j=1}^{m}(1 + 2\varepsilon_j s_j) + \prod\nolimits_{j=1}^{m}(1 - 2\varepsilon_j s_j)\right\} \leq \sum\nolimits_{r\geq 0} \frac{(2\sum_j |\varepsilon_j|)^{2r}}{(2r)!} = \cosh\Big(2\sum\nolimits_j |\varepsilon_j|\Big).$$

  The inequality expands the even products and bounds each elementary symmetric sum by the corresponding power divided by its factorial.

- Since $\cosh x \leq e^{x^2/2}$ and the fair-sign rejection probability averages to α,

$$\boxed{\mathbb{P}\big(\mu \notin [W_{(m-c_{m,\alpha/2})}, W_{(c_{m,\alpha/2}+1)}]\big) \leq \alpha e^{2m^2\Delta^2}.}$$

- With ties, independently assign each tied observation a sign so that its positive-sign probability is as close to $1/2$ as possible. The resulting absolute sign bias is exactly δ_j. Closed-interval exclusion implies rejection for these auxiliary signs, so the same bound applies.

Supplement: HulC validity and its pointwise width

For the scalar HulC partition, $|S_j| = n/B$ and the block estimators are independent. Fix an admissible B and suppose, for a fixed distribution P,

$$r_{n/B}\{\widehat{\theta}_j - \theta(P)\} \xrightarrow{d} L, \qquad r_{n/B} \longrightarrow \infty,$$

where L has median zero and its CDF is continuous at zero.

- Continuity at zero and the median property imply $\mathbb{P}(L \le 0) = \mathbb{P}(L \ge 0) = 1/2$. Convergence of the sign probabilities gives $\Delta \to 0$, hence $B\Delta \to 0$ and

$$\liminf_{n\to\infty} \mathbb{P}_P\Big(\theta(P) \in [\widehat{\theta}_{(B-c_{B,\alpha/2})}, \widehat{\theta}_{(c_{B,\alpha/2}+1)}]\Big) \ge 1-\alpha.$$

- Tightness of each scaled block error and fixed B give

$$\max_{1\le j\le B} |\widehat{\theta}_j - \theta(P)| = O_P(r_{n/B}^{-1}), \qquad \text{width}(I) \le 2\max_j |\widehat{\theta}_j - \theta(P)| = O_P(r_{n/B}^{-1}).$$

- These conclusions include nonnormal zero-median limits. Continuity at zero is essential: weak convergence to a law with an atom at zero need not determine the limiting sign probabilities.
- This is a pointwise statement in P. A uniform coverage conclusion needs uniform control of the median biases; a separate rare-event sequence can defeat pointwise normal approximation.

Supplement: Bernoulli uniform failure and Hamza inflation

For $X_i \overset{\text{iid}}{\sim} \mathrm{Ber}(p)$, $\widehat{\theta}_j \overset{d}{=} (B/n)\,\mathrm{Binom}(n/B, p)$. At fixed $p \in (0,1)$, the block CLT gives pointwise asymptotic validity.

- Take instead $p_n = n^{-2}$ and a fixed admissible B. Then
$$\mathbb{P}_{p_n}(X_1 = \cdots = X_n = 0) = (1 - n^{-2})^n \longrightarrow 1.$$
On this event every block mean is zero, and the finite-rank uninflated HulC is $[0,0]$. Thus
$$\lim_{n\to\infty} \sup_{p\in(0,1)} \mathbb{P}_p\Big(p \notin [\widehat{\theta}_{(B-c_{B,\alpha/2})}, \widehat{\theta}_{(c_{B,\alpha/2}+1)}]\Big) = 1.$$
- The specified Wald, ordinary empirical-bootstrap and subsampling intervals also collapse on this all-zero sample. For rare-event sequences with finite nonzero intensity, a Poisson limit can replace the Gaussian approximation.
- Hamza (1995) supplies a median of each binomial block mean satisfying
$$\left|\mathbb{E}\widehat{\theta}_j - \mathrm{Med}(\widehat{\theta}_j)\right| \le \frac{B\log 2}{n}.$$
The inflation argument therefore gives, uniformly over $p \in [0,1]$,
$$\mathbb{P}_p\Big(p \in \Big[\widehat{\theta}_{(B-c_{B,\alpha/2})} - \frac{B\log 2}{n}, \widehat{\theta}_{(c_{B,\alpha/2}+1)} + \frac{B\log 2}{n}\Big] \cap [0,1]\Big) \ge 1 - \alpha.$$

Hamza's bound also covers Poisson laws. The counterexample concerns the displayed procedures, not every possible protected resampling method.

Supplement: from normal approximation to a median bound

Let Y have finite mean μ, variance $\sigma^2 > 0$, and $d_K((Y-\mu)/\sigma, N(0,1)) \le \varepsilon$. Here $\mathrm{Med}(Y)$ denotes any fixed choice of median.

- If $\mathrm{Med}(Y) > \mu$, Cantelli's inequality gives

$$\tfrac{1}{2} \le \mathbb{P}\big(Y - \mu \ge \mathrm{Med}(Y) - \mu\big) \le \frac{\sigma^2}{\sigma^2 + (\mathrm{Med}(Y) - \mu)^2}.$$

  Apply the same argument to $-Y$ when the median is below μ. Consequently $|\mathrm{Med}(Y) - \mu| \le \sigma$ for every median.

- The two median inequalities, using a left limit of the CDF when needed, imply

$$\Phi\left(\frac{|\mathrm{Med}(Y) - \mu|}{\sigma}\right) - \tfrac{1}{2} \le \varepsilon.$$

  If $\varepsilon \le \Phi(1) - 1/2$, concavity of Φ on $[0,1]$ yields

$$\frac{|\mathrm{Med}(Y) - \mu|}{\sigma} \le \frac{\varepsilon}{\Phi(1) - 1/2}.$$

- If $\varepsilon > \Phi(1) - 1/2$, use the Cantelli bound instead. Both cases give

$$\boxed{|\mathrm{Med}(Y) - \mu| \le \frac{\sigma\varepsilon}{\Phi(1) - 1/2} < 3\sigma\varepsilon}$$

  whenever $\varepsilon > 0$; for $\varepsilon = 0$ the gap is zero.

Supplement: independent sums and the numerical constants

Let the X_i be independent, and suppose $\mathbb{E}|X_i - \mathbb{E}X_i|^3 \le K \operatorname{Var}(X_i)$. For $S = \sum_i (X_i - \mathbb{E}X_i)$ and $\sigma_S^2 = \sum_i \operatorname{Var}(X_i) > 0$, Berry–Esseen gives

$$d_K(S/\sigma_S, N(0,1)) \le C_{\mathrm{BE}} \frac{\sum_i \mathbb{E}|X_i - \mathbb{E}X_i|^3}{\sigma_S^3}.$$

The preceding transfer, followed by division by n, implies

$$\left|\mathbb{E}\overline{X}_n - \operatorname{Med}(\overline{X}_n)\right| \le \frac{C_{\mathrm{BE}}}{n\{\Phi(1) - 1/2\}} \frac{\sum_i \mathbb{E}|X_i - \mathbb{E}X_i|^3}{\sum_i \operatorname{Var}(X_i)} \le \frac{3C_{\mathrm{BE}}K}{n}.$$

If $\sigma_S = 0$, every summand is constant almost surely and the gap is zero.

Summands	Input	C_{BE}	$\mathfrak{C} = 3C_{\mathrm{BE}}$
Independent	Shevtsova (2013)	.5583	1.6749
IID	Shevtsova (2013)	.4690	1.4070
Independent	Bell (2026)	$\approx .4499$	≈ 1.3497
IID	Kołodziejek–Rodziewicz (2026)	$\approx .4097$	≈ 1.2292

The simulations retain the published Shevtsova inputs. The 2026 entries are rounded displays; their external certificate suites are not replayed here. An IID improvement applies only when the relevant summands are identically distributed. Bounded centered observations satisfy the ratio condition directly, recovering the type of n^{-1} gap studied by Oleszkiewicz (2001).

Supplement: why inflation preserves coverage

Suppose independent T_j have selected medians with $|\operatorname{Med}(T_j) - \theta| \le b_j$, where the allowances b_j are deterministic. Define $V_j = T_j - \operatorname{Med}(T_j) + \theta$. Each V_j has median θ, and

$$T_j - b_j \le V_j \le T_j + b_j.$$

Monotonicity of every order statistic therefore gives

$$(T_j - b_j)_{(B - c_{B,\alpha/2})} \le V_{(B - c_{B,\alpha/2})}, \qquad V_{(c_{B,\alpha/2}+1)} \le (T_j + b_j)_{(c_{B,\alpha/2}+1)}.$$

Thus the interval

$$\boxed{\left[(T_j - b_j)_{(B - c_{B,\alpha/2})}, (T_j + b_j)_{(c_{B,\alpha/2}+1)}\right]}$$

contains the sign interval for the V_j's and covers θ with probability at least $1 - \alpha$.

- ⋆ For equal scalar mean blocks, $b_j = \mathfrak{C}KB/n$, so the two order statistics shift by the same displayed amount. The argument also permits unequal known block sizes and different block medians.
- ⋆ If $\operatorname{Var}(X_i) \le \sigma^2$ and B, α are fixed, block means have variance at most $\sigma^2 B/n$, giving

$$\operatorname{width}(\widehat{\mathrm{CI}}^{\mathrm{inf-HulC}}) = O_P\left(\sigma\sqrt{\frac{B}{n}} + \frac{KB}{n}\right).$$

- ⋆ With $B = \lceil \log_2(2/\alpha) \rceil$, this has the stated Bernstein-like form for fixed α. It is not a sharp uniform assertion as $\alpha \downarrow 0$: increasing-block ranges can introduce an additional logarithmic factor.

The moment-ratio bound is supplied in advance; an unprotected plug-in estimate need not preserve the guarantee.

Supplement: the Poincaré fourth-moment argument

The normalizations are

$$\mathrm{Var}_P(g) \le C_{\mathrm{PC}}\,\mathbb{E}_P\|\nabla g\|^2, \qquad \mathrm{Ent}_P(g^2) \le 2C_{\mathrm{LS}}\,\mathbb{E}_P\|\nabla g\|^2.$$

Linearizing LSI at $g = 1 + \varepsilon f$ gives Poincaré with $C_{\mathrm{PC}} \le C_{\mathrm{LS}}$. For a 1-Lipschitz f, put $Y = f(X) - \mathbb{E}f(X)$ and $v = \mathbb{E}Y^2$.

- Poincaré first gives $v \le C_{\mathrm{PC}}$. To justify the fourth-moment step before assuming integrability, apply it to $g_M = \min\{Y^2, M\}$, using Lipschitz approximation at the corners:

  $$\mathbb{E}g_M^2 \le (\mathbb{E}g_M)^2 + C_{\mathrm{PC}}\,\mathbb{E}\|\nabla g_M\|^2 \le v^2 + 4C_{\mathrm{PC}}\,v.$$

  Monotone convergence as $M \to \infty$ proves

  $$\mathbb{E}Y^4 \le v^2 + 4C_{\mathrm{PC}}\,v \le 5C_{\mathrm{PC}}\,v.$$

- Cauchy–Schwarz now gives the required variance-sensitive ratio:

  $$\boxed{\mathbb{E}|Y|^3 \le \sqrt{\mathbb{E}Y^2\,\mathbb{E}Y^4} \le \sqrt{5C_{\mathrm{PC}}}\,\mathbb{E}Y^2.}$$

- LSI supplies the same result with C_{LS}. If both constants are available, one may use $K = \sqrt{5}\min\{\sqrt{C_{\mathrm{PC}}}, \sqrt{C_{\mathrm{LS}}}\}$. The variance factor is retained even when v is very small.

This is the moment comparison of Bobkov–Ledoux (1997).

Supplement: weak and ordinary log-concavity

For a positive density proportional to e^{-V}, the supplied weak-convexity profile is

$$\frac{\langle \nabla V(x) - \nabla V(y), x - y \rangle}{\|x - y\|^2} \geq \kappa - \frac{2\sqrt{M}}{r} \tanh\left(\frac{\sqrt{M}\, r}{2}\right), \qquad r = \|x - y\|, \quad x \neq y, \quad \kappa > 0,\ M \geq 0.$$

- Conforti–Lacker–Pal (2025) supplies a transport from a standard Gaussian with Lipschitz constant at most $\kappa^{-1/2} e^{M/(2\kappa)}$. The chain rule transfers both Gaussian functional inequalities, multiplying their constants by the squared Lipschitz constant:

$$C_{\mathrm{PC}}, C_{\mathrm{LS}} \leq \kappa^{-1} e^{M/\kappa}, \qquad K = \sqrt{5/\kappa}\, e^{M/(2\kappa)}.$$

- For $\frac{1}{2}N(-a, \sigma^2) + \frac{1}{2}N(a, \sigma^2)$, the profile has $\kappa = \sigma^{-2}$ and $M = a^2\sigma^{-4}$, hence $C_{\mathrm{PC}}, C_{\mathrm{LS}} \leq \sigma^2 e^{a^2/\sigma^2}$.
- Ordinary log-concavity means convex V. It implies a Poincaré inequality but need not imply LSI: an exponential-tail law need not have the Gaussian concentration implied by LSI.
- In one dimension, Bobkov (1999) gives

$$\mathrm{Var}(X) \leq C_{\mathrm{PC,opt}} \leq 12\, \mathrm{Var}(X).$$

  Thus a known $\mathrm{Var}(X) \leq \sigma^2$ permits the supplied choice $C_{\mathrm{PC}} = 12\sigma^2$ and $K = \sqrt{60}\,\sigma$. The upper comparison is for the *optimal* constant, not every deliberately loose supplied upper bound.

Supplement: bounded differences and the factor $1/2$

Let $T = f(X_1, \dots, X_n)$ for independent inputs, and suppose replacing coordinate i changes f by at most c_i. Write $M_i = \mathbb{E}(T \mid X_1, \dots, X_i)$ and $D_i = M_i - M_{i-1}$.

- ⋆ Conditional on $X_1, \dots, X_{i-1}$, independence and averaging over the remaining coordinates show that M_i, as a function of X_i, has range length at most c_i.
- ⋆ A variable in $[a, b]$ has $\mathrm{Var}(Y) \le (\mathbb{E}Y - a)(b - \mathbb{E}Y) \le (b-a)^2/4$. Martingale orthogonality consequently gives

$$\mathrm{Var}(T) = \sum\nolimits_i \mathbb{E}D_i^2 \le \frac{1}{4} \sum\nolimits_i c_i^2.$$

- ⋆ Applying the mean–median Cantelli bound proves, for every median,

$$\boxed{|\mathbb{E}T - \mathrm{Med}(T)| \le \frac{1}{2}\sqrt{\sum\nolimits_i c_i^2}.}$$

  For a single equally likely two-point variable with range c, either endpoint is a median and its distance from the mean is $c/2$; the universal factor cannot be smaller.
- ⋆ Coordinate oscillations of order n^{-1} alone do not yield an n^{-1} gap. For even n and independent Rademacher X_i, $T_n = (n^{-1}\sum_i X_i)_+$ has median zero and $c_i = 2/n$, whereas $\mathbb{E}T_n = \mathbb{E}|\sum_i X_i|/(2n) \sim (2\pi n)^{-1/2}$.

Supplement: projection, remainder and mixed differences

For square-integrable T, set

$$Z_i = \mathbb{E}(T \mid X_i) - \mathbb{E}T, \quad W = \sum_i Z_i, \quad R = T - \mathbb{E}T - W, \quad v^2 = \mathbb{E}W^2, \quad \tau^2 = \mathbb{E}R^2.$$

Let X_{-i} denote all inputs except X_i. For independent coordinate replacements, write $\Delta_i T = T - T^{(i)}$ and $\Delta_{ij} T = T - T^{(i)} - T^{(j)} + T^{(ij)}$. Define

$$D_i = R - \mathbb{E}(R \mid X_{-i}), \qquad G = \sum_i \mathbb{E}D_i^2, \qquad J = \frac{1}{4}\sum_{i<j} \mathbb{E}(\Delta_{ij}T)^2.$$

- The orthogonal Hoeffding expansion $T - \mathbb{E}T = \sum_{\varnothing \neq A} T_A$, with $v_A = \mathbb{E}T_A^2$, gives

$$\tau^2 = \sum_{|A|\geq 2} v_A, \quad G = \sum_{|A|\geq 2} |A| v_A, \quad J = \sum_{|A|\geq 2} \binom{|A|}{2} v_A, \qquad \tau^2 \leq G/2 \leq J.$$

  Each Hoeffding component containing i, j contributes $4v_A$ to $\mathbb{E}(\Delta_{ij}T)^2$.
- If $|\Delta_{ij} f| \leq c_{ij}$, condition on X_i, X_i' and all remaining coordinates except X_j, X_j'. The first difference as a function of X_j has range at most c_{ij}. Two independent draws therefore give mixed-difference second moment at most $c_{ij}^2/2$, hence

$$J \leq \frac{1}{8}\sum_{i<j} c_{ij}^2.$$

- If $|\Delta_i f| \leq c_i$, the centered projection Z_i has range at most c_i and $|Z_i| \leq c_i$. For $v > 0$, $\sum_i \mathbb{E}|Z_i|^3/v^2 \leq \max_i c_i$.

Supplement: the random-interval comparison

Let $\xi_i = g_i(X_i)$ be independent and centered, with $\sum_i \mathbb{E}\xi_i^2 = 1$. Put $W = \sum_i \xi_i$. Let R, R_i be finite, with X_i independent of $(R_i, W - \xi_i)$. Assume the following quantities are finite:

$$L = \sum_i \mathbb{E}|\xi_i|^3, \qquad A = \mathbb{E}|WR| + \sum_i \mathbb{E}|\xi_i(R - R_i)|, \qquad T = W + R.$$

For either $E_+ = \{z \le W \le z + |R|\}$ or $E_- = \{z - |R| \le W \le z\}$, let $p = \mathbb{P}(E_\pm)$.

- For $\delta > 0$, use

$$f_c^{\pm}(w) = \mathrm{clip}(w - z \mp c/2, -c/2 - \delta, c/2 + \delta), \qquad \eta_i = |\xi_i| \min\{\delta, |\xi_i|\}.$$

  These increasing functions have derivative one on the corresponding interval padded by δ, $|f_c^{\pm}| \le c/2 + \delta$ and $\|f_c^{\pm} - f_d^{\pm}\|_\infty \le |c - d|/2$.

- Independence and centering give the exact decomposition

$$\mathbb{E}[W f_{|R|}(W)] = \sum_i \mathbb{E}\{\xi_i[f_{|R|}(W) - f_{|R|}(W - \xi_i)]\} + \sum_i \mathbb{E}\{\xi_i[f_{|R|}(W - \xi_i) - f_{|R_i|}(W - \xi_i)]\}.$$

Supplement: bounding the random crossing interval

Retain the independent normalized summands and remainder from the preceding comparison. For either crossing event $E_{\pm}$, let $p = \mathbb{P}(E_{\pm})$.

- In the exact decomposition, the first sum is at least $\mathbb{E}[\mathbf{1}_{E_{\pm}} \sum_i \eta_i]$. To see this, write each difference as the integral of the derivative of $f_{|R|}$ against the nonnegative kernel $\xi_i\{\mathbf{1}_{[-\xi_i,0]}(t) - \mathbf{1}_{(0,-\xi_i]}(t)\}$, and restrict to $|t| \le \delta$.
- The absolute value of the second sum is at most $\frac{1}{2}\sum_i \mathbb{E}|\xi_i(R - R_i)|$. Also $|\mathbb{E}[W f_{|R|}(W)]| \le \delta\mathbb{E}|W| + \frac{1}{2}\mathbb{E}|WR|$. Thus
$$\mathbb{E}[\mathbf{1}_{E_{\pm}} \sum_i \eta_i] \le \delta\mathbb{E}|W| + A/2.$$
- With $q = \sum_i \mathbb{E}\eta_i$ and $v_\delta^2 = \sum_i \mathrm{Var}(\eta_i)$, centered Cauchy–Schwarz gives $\mathbb{E}[\mathbf{1}_{E_{\pm}} \sum_i \eta_i] \ge qp - v_\delta\sqrt{p(1-p)}$. Therefore
$$qp - v_\delta\sqrt{p(1-p)} \le \delta\mathbb{E}|W| + A/2.$$
The pointwise inequalities $u\min\{\delta, u\} \ge u^2 - u^3/(4\delta)$ and $\eta_i^2 \le \delta^2\xi_i^2$ give $q \ge 1 - L/(4\delta)$ and $v_\delta \le \delta$. Taking $\delta = L/2$ yields
$$\boxed{p \le L\{\mathbb{E}|W| + \sqrt{p(1-p)}\} + A.}$$

Apply a common truncation to R, R_i first if needed, then pass to the limit.

Supplement: the projection-comparison constant

Use the published independent Berry–Esseen input $d_K(W, N(0,1)) \le .5583L$. Choose $a = .543$, $r = 1.075$, and let $I = \int_a^r (r-t)\{2\Phi(t) - 1\}\, dt$. Put $a_0 = (r^2+1)/(2r) - I/r$ and $b_0 = .5583(r-a)^2/r$.

- ⋆ Since $\mathbb{E}W^2 = 1$ and $\mathbb{P}(|W| \le t) \ge 2\Phi(t) - 1 - 2(.5583)L$,
$$\mathbb{E}|W| = \{r^2 + 1 - \mathbb{E}(|W| - r)^2\}/(2r) \le a_0 + b_0 L.$$
This uses the lower-tail integral for $\mathbb{E}(r - |W|)_+^2$ over $[a, r]$.
- ⋆ Take $k = 1.485403469$. The explicit inequality
$$k^2 \ge a_0 k + \frac{b_0 + \sqrt{b_0^2 + k^2}}{2}$$
and $b_0 p + k\sqrt{p(1-p)} \le \{b_0 + \sqrt{b_0^2 + k^2}\}/2$ show that $p > kL + A$ contradicts the preceding random-interval bound:
$$k^2 < a_0 k + b_0 p + k\sqrt{p(1-p)} \le k^2.$$
- ⋆ The two crossing intervals control $d_K(W + R, W)$. The triangle inequality therefore gives
$$\boxed{d_K(W + R, N(0,1)) \le 2.043703469\, L + A.}$$

The numerical inequality is certified by integrating the alternating expansion $2\Phi(t) - 1 = \sqrt{2/\pi} \sum_{j \ge 0} (-1)^j t^{2j+1}/\{2^j j!(2j+1)\}$. Forty terms and the next-term remainder enclose I; $r^2 < 2$ makes the term magnitudes decrease. The supplied draft's certificate and the independent audit verify the displayed rounded value of k.

Supplement: transferring the projection bound to medians

Apply the comparison to $\xi_i = Z_i/v$, normalized remainder R/v, and $R_i = \mathbb{E}(R \mid X_{-i})/v$, for $v > 0$. Cauchy–Schwarz gives

$$\mathbb{E}|WR|/v^2 \le \tau/v, \qquad \sum_i \mathbb{E}|Z_i D_i|/v^2 \le \sqrt{G}/v.$$

Put $d = \Phi(1) - 1/2$ and

$$A_0 = 2.043703469\,\frac{\sum_i \mathbb{E}|Z_i|^3}{v^2} + \tau + \sqrt{G}.$$

The comparison therefore gives $d_K((T - \mathbb{E}T)/v, N(0,1)) \le A_0/v$.

- ⋆ If $A_0/v \le d$, the normal-median argument gives $|\operatorname{Med}(T) - \mathbb{E}T| \le A_0/d$.
- ⋆ If $A_0/v > d$, Cantelli and projection orthogonality give $|\operatorname{Med}(T) - \mathbb{E}T| \le \sqrt{v^2 + \tau^2} \le \sqrt{(A_0/d)^2 + \tau^2}$. Both cases imply

$$\boxed{|\operatorname{Med}(T) - \mathbb{E}T| \le \sqrt{(A_0/d)^2 + \tau^2}.}$$

- ⋆ When $v = 0$, use $|\operatorname{Med}(T) - \mathbb{E}T| \le \tau \le \sqrt{J}$ directly. No division by a vanishing projection variance is needed.

This combines the normal approximation to the linear projection with a variance bound for the nonlinear remainder.

Supplement: deriving the rounded interaction bound

Use $d = \Phi(1) - 1/2$, $\sum_i \mathbb{E}|Z_i|^3/v^2 \le \max_i c_i$ and $\tau^2 \le G/2 \le J \le \sum_{i<j} c_{ij}^2/8$. These imply

$$A_0 \le 2.043703469 \max_i c_i + (1+\sqrt{2})\sqrt{\frac{1}{8}\sum\nolimits_{i<j} c_{ij}^2}.$$

The median bound and the triangle inequality in $\mathbb{R}^2$ now give

$$|\operatorname{Med}(T) - \mathbb{E}T| \le \frac{2.043703469}{d} \max_i c_i + \sqrt{\frac{(1+\sqrt{2})^2/d^2 + 1}{8}} \sqrt{\sum\nolimits_{i<j} c_{ij}^2}.$$

The two coefficients are below 5.988 and 2.526, respectively, so

$$\boxed{|\operatorname{Med}(T) - \mathbb{E}T| \le 6 \max_i c_i + 3\sqrt{\sum\nolimits_{i<j} c_{ij}^2}.}$$

- ⋆ The zero-projection case is already covered by Cantelli, since $\tau \le \sqrt{J}$. The result therefore holds through degeneracy.
- ⋆ If $c_i \le a/n$ and $c_{ij} \le b/n^2$, the displayed allowance is at most $(6a + 3b/\sqrt{2})/n$; the first-difference-only allowance is $a/(2\sqrt{n})$.
- ⋆ The comparison lemma and the inequalities above establish the rounded constants used in the talk. Friedrich (1989), Chen–Shao (2007) and Shao–Zhang (2025) provide related moment-based normal-approximation inputs.

Supplement: the Hoeffding normalization and degeneracy

For a symmetric square-integrable kernel and IID observations, set

$$\theta = \mathbb{E}h(X_1, X_2), \quad g(x) = \mathbb{E}h(x, X_2) - \theta, \quad h_0(x, y) = h(x, y) - \theta - g(x) - g(y).$$

The canonical kernel satisfies $\mathbb{E}(h_0(X_1, X_2) \mid X_1) = 0$. For $U_n = \binom{n}{2}^{-1} \sum_{i<j} h(X_i, X_j)$,

$$U_n - \theta = \tfrac{2}{n} \sum_{i=1}^{n} g(X_i) + \binom{n}{2}^{-1} \sum_{i<j} h_0(X_i, X_j).$$

- Orthogonality of the projection and canonical terms gives the exact variance

$$\mathrm{Var}(U_n) = \frac{4}{n} \mathrm{Var}(g(X_1)) + \frac{2}{n(n-1)} \mathbb{E}h_0(X_1, X_2)^2.$$

  If $\mathrm{Var}(g(X_1)) > 0$, the canonical remainder is $o_P(n^{-1/2})$ and $\sqrt{n}(U_n - \theta) \Rightarrow N(0, 4\,\mathrm{Var}(g(X_1)))$. The factor 4 is the square of the projection coefficient 2.

- If $g = 0$ almost surely, the $n^{-1/2}$ term vanishes. The Hilbert–Schmidt expansion of the centered kernel yields

$$n(U_n - \theta) \Rightarrow \sum\nolimits_{r \geq 1} \lambda_r (Z_r^2 - 1),$$

  where the Z_r are independent standard normals and λ_r are the kernel-operator eigenvalues. Square integrability gives $\sum_r \lambda_r^2 < \infty$.

- The scale and limiting shape change through degeneracy. A uniform median allowance avoids choosing a limiting regime.

Supplement: the bounded-kernel allowance and interval

Suppose $\|h\|_\infty \le K$ and a block contains $m \ge 2$ observations.

- Replacing one observation changes $m-1$ kernel terms, each by at most $2K$. In a mixed replacement, only the pair containing the two changed indices remains, with four terms of magnitude at most K. Therefore

$$c_i = \frac{(m-1)2K}{\binom{m}{2}} = \frac{4K}{m}, \qquad c_{ij} = \frac{4K}{\binom{m}{2}} = \frac{8K}{m(m-1)}.$$

- The range bound gives a median gap at most K; bounded differences give $2K/\sqrt{m}$. The rounded interaction bound gives

$$6\frac{4K}{m} + 3\sqrt{\binom{m}{2}\left(\frac{8K}{m(m-1)}\right)^2} = \frac{24K}{m} + \frac{12\sqrt{2}K}{\sqrt{m(m-1)}}.$$

All three deterministic bounds hold, hence

$$\boxed{|\mathbb{E}U_m - \mathrm{Med}(U_m)| \le \min\left\{K, \frac{2K}{\sqrt{m}}, \frac{24K}{m} + \frac{12\sqrt{2}K}{\sqrt{m(m-1)}}\right\}.}$$

- With scalar blocks $m = n/B$, inflate the lower and upper HulC endpoints by this amount. The monotonicity proof gives finite-sample coverage at least $1-\alpha$, uniformly through degeneracy.

For $h \in [a,b]$, center at $(a+b)/2$ and use $K = (b-a)/2$; the mean–median gap is unchanged. For rounded block sizes, the code inserts each block's own allowance before sorting.

Supplement: the two empirical Bernstein statistics

For $h \in [a, b]$, put $D = b - a > 0$ and $q = (h - a)/D \in [0, 1]$. Write $q_{ij} = q(X_i, X_j)$ and let $\mathcal{A}_n^r$ be the ordered distinct r-tuples, of cardinality $n!/(n-r)!$. Peel–Anthoine–Ralaivola (2010) uses the complete statistics

$$\widehat{Q} = \frac{1}{|\mathcal{A}_n^4|} \sum\nolimits_{(i,j,k,\ell) \in \mathcal{A}_n^4} (q_{ij} - q_{k\ell})^2, \qquad \widehat{A} = \frac{1}{|\mathcal{A}_n^3|} \sum\nolimits_{(i,j,k) \in \mathcal{A}_n^3} q_{ij} q_{ik}.$$

- For distinct $1, 2, 3, 4$, q_{12} and q_{34} are IID and independent. Consequently
$$\mathbb{E}\widehat{Q} = \mathbb{E}(q_{12} - q_{34})^2 = 2\,\mathrm{Var}(q(X_1, X_2)).$$
- Conditional independence of X_2 and X_3 given X_1 gives
$$\mathbb{E}\widehat{A} = \mathbb{E}\{\mathbb{E}(q(X_1, X_2) \mid X_1)^2\} = \mathrm{Var}\{\mathbb{E}(q \mid X_1)\} + (\mathbb{E}q)^2.$$
- The literal first statistic has no factor $1/2$; the second is uncentered. The two displayed formulas preserve those conventions.
- The center and both statistics use all observations. The implementation evaluates exact count identities; its validation compares them with literal ordered-tuple enumeration.

If $D = 0$, the kernel and its target are known constants and no concentration bound is needed.

Supplement: the two empirical Bernstein radii

For $n \geq 12$, use $M = 6\lfloor n/12 \rfloor$ in the concentration denominators. The two separate intervals from Peel–Anthoine–Ralaivola (2010) are

$$U_n \;\pm\; D\left\{\sqrt{\frac{2\widehat{Q}\log(4/\alpha)}{M}} + \frac{5\log(4/\alpha)}{M}\right\}, \tag{12}$$

$$U_n \;\pm\; D\left\{\sqrt{\frac{4\widehat{A}\log(8/\alpha)}{M}} + \frac{(5\sqrt{2}+64+1/6)\log(8/\alpha)}{M}\right\}. \tag{13}$$

- Each is individually a $1-\alpha$ comparator. Their intersection is not used without an error-level adjustment.
- The lower effective sample size corresponds to $12\lfloor n/12 \rfloor$ observations only in the bounds' denominators. It conservatively accommodates the floor terms arising from complete U-statistics of orders $2, 3, 4$. The estimates $U_n, \widehat{Q}, \widehat{A}$ continue to use the full sample.
- The simulations call the first method *Peel: kernel bound* and the second *Peel: projection bound*. The source equation tags appear on this defining frame only.
- Where a known target range is available, all methods use the same intersection with that range. Empty intersections remain empty and count as coverage failures.

These inequalities are external theorem inputs; the normalization, complete-statistic expectations and implemented denominators are specified here.

Supplement: COSI with a half-sample pilot

Let $X_1, \ldots, X_n$ be IID random elements in $\mathbb{R}^{\mathcal{T}}$, with mean $\mu(t)$. Here n is the total sample size: $n/2$ observations form the pilot and the other half form B blocks of size $n/(2B)$, assumed integral.

$$\vartheta(t) = \frac{2}{n} \sum\nolimits_{i \in S_0} X_i(t), \qquad \widehat{\theta}_j(t) = \frac{2B}{n} \sum\nolimits_{i \in S_j} X_i(t).$$

The pilot may instead supply any other center. Choose a positive pilot-measurable scale $s(t)$, and set

$$R = \sup\nolimits_{t \in \mathcal{T}} \left(\frac{|\widehat{\theta}_j(t) - \vartheta(t)|}{s(t)} : j = 1, \ldots, B \right)_{(c_{B,\alpha}+1)}.$$

- The independent cutoff uses the one-sided level α. The $(B+1)$st residual is $+\infty$. Assume measurable/separable suprema and pilot-measurable approximate maximizing coordinates.
- If the block medians equal $\mu(t)$ coordinatewise, the COSI set is

$$\widehat{\mathrm{CI}}^{\mathrm{COSI}} = \bigtimes_{t \in \mathcal{T}} [\vartheta(t) - Rs(t),\, \vartheta(t) + Rs(t)].$$

- Under the uniform moment condition $\mathbb{E}|X_i(t) - \mu(t)|^3 \le K \operatorname{Var}(X_i(t))$, every block median is within $2\mathfrak{C}KB/n$ of $\mu(t)$. The inflated set is

$$\boxed{\widehat{\mathrm{CI}}^{\mathrm{inf-COSI}} = \bigtimes_{t \in \mathcal{T}} \left[\vartheta(t) - Rs(t) - \frac{2\mathfrak{C}KB}{n},\, \vartheta(t) + Rs(t) + \frac{2\mathfrak{C}KB}{n} \right].}$$

Supplement: the conditional COSI proof

Condition on the actual pilot data $(X_i)_{i\in S_0}$ and its scales. Suppose selected block medians satisfy $|\operatorname{Med}(\widehat{\theta}_j(t)) - \mu(t)| \le b$ uniformly in j, t. For inflated means, $b = 2\mathfrak{C}KB/n$; for median-unbiased blocks, $b = 0$.

- If $\mu(t) > \vartheta(t) + b$, then $\operatorname{Med}(\widehat{\theta}_j(t)) \ge \mu(t) - b > \vartheta(t)$. With conditional probability at least $1/2$, $\widehat{\theta}_j(t)$ is at least its median. The opposite direction is treated similarly, giving
$$\mathbb{P}\Big(|\widehat{\theta}_j(t) - \vartheta(t)| \ge |\mu(t) - \vartheta(t)| - b \;\Big|\; (X_i)_{i\in S_0}\Big) \ge \tfrac{1}{2}$$
whenever the right-hand threshold is positive.
- Conditional independence across blocks implies, for any residual rank k,
$$\mathbb{P}\left(\left(\frac{|\widehat{\theta}_j(t) - \vartheta(t)|}{s(t)}\right)_{(k)} < \frac{(|\mu(t) - \vartheta(t)| - b)_+}{s(t)} \;\middle|\; (X_i)_{i\in S_0}\right) \le \mathbb{P}\big(\mathrm{Binom}(B, 1/2) \ge k\big).$$
- Let $D = \sup_{t\in\mathcal{T}}(|\mu(t) - \vartheta(t)| - b)_+/s(t)$. If $0 < D < \infty$, choose a pilot-measurable t_ε whose value exceeds $D - \varepsilon$. The event $\{R < D - \varepsilon\}$ is contained in rank failure at t_ε.
- Average over $k = c_{B,\alpha} + 1$ and let $\varepsilon \downarrow 0$. Then $\mathbb{P}(R < D \mid \text{pilot}) \le \alpha$. For $D = 0$ the claim is automatic; for $D = \infty$, choose coordinates exceeding any fixed level and take that level to infinity.

The event $R \ge D$ implies all band inequalities simultaneously. No union bound over $\mathcal{T}$ is used; dependence among coordinates is unrestricted.

Supplement: the general COSI median-bias bound

For uninflated COSI, define

$$\delta_j(t) = \left(\tfrac{1}{2} - \min\{\mathbb{P}(\widehat{\theta}_j(t) \geq \mu(t)), \mathbb{P}(\widehat{\theta}_j(t) \leq \mu(t))\}\right)_+, \qquad \Delta = \sup_{t\in\mathcal{T}} \frac{1}{B}\sum_{j=1}^{B} \delta_j(t).$$

- Condition on the pilot and select an approximate maximizing coordinate for $|\mu(t) - \vartheta(t)|/s(t)$. Insufficient residual implies $\widehat{\theta}_j(t) < \mu(t)$ when $\mu(t) > \vartheta(t)$, and the reversed inequality otherwise. Its probability is at most $1/2 + \delta_j(t)$.
- Rank failure increases in the insufficient-residual indicators. It is dominated by independent Bernoulli indicators with probabilities exactly $1/2 + \delta_j(t)$.
- For these dominating variables, the likelihood ratio of any configuration relative to independent fair Bernoulli variables is at most

$$\prod_{j=1}^{B}(1 + 2\delta_j(t)) \leq \left(1 + \frac{2}{B}\sum_{j=1}^{B}\delta_j(t)\right)^B \leq (1 + 2\Delta)^B.$$

  AM–GM gives the first inequality. This likelihood comparison concerns the dominating variables, whose success probabilities are at least $1/2$.
- Under fair signs, averaging over $c_{B,\alpha} + 1$ gives failure probability α. The same limiting-coordinate argument proves

$$\boxed{\mathbb{P}(\mu \notin \widehat{\mathrm{CI}}^{\mathrm{COSI}}) \leq \alpha(1 + 2\Delta)^B \leq \alpha e^{2B\Delta}.}$$

At $\Delta = 0$ the bound is nominal; the one-sided event lacks scalar sign-reversal symmetry.

Supplement: why COSI does not have the quadratic factor

Consider one coordinate, $B = 5$, $\alpha = 1/32$, unit scale and an independent deterministic pilot equal to zero. Let the independent block estimates have the bounded two-point distribution

$$\widehat{\theta}_j = \begin{cases} 0, & \text{with probability } .51, \\ 100/49, & \text{with probability } .49. \end{cases} \qquad \mathbb{E}\widehat{\theta}_j = 1, \quad \delta_j = .01.$$

- ⋆ At this level the COSI residual rank is the maximum. Its interval is
$$\left[-\max_j |\widehat{\theta}_j|,\ \max_j |\widehat{\theta}_j|\right].$$
It misses the mean exactly when all five estimates equal zero.
- ⋆ The actual failure probability is
$$.51^5 = .0345025251.$$
The scalar quadratic-factor expression would instead give
$$\alpha e^{2B^2\Delta^2} = \frac{1}{32} e^{2 \cdot 25 \cdot .01^2} = .031406641\ldots,$$
which is too small.
- ⋆ The valid polynomial bound is attained:
$$\alpha(1 + 2\Delta)^B = \frac{1}{32}(1.02)^5 = .51^5.$$

The example can be realized by one IID observation per inference block; a constant center is an admissible independent pilot. The correction changes the theorem, not the one-sided COSI calibration.

Supplement: projecting the Lipschitz band onto W_1

Let μ, ν have finite first moments, with independent samples of total size n from each law. Anchor the 1-Lipschitz critics by $f(x_\star) = 0$ and assume measurable/separable suprema. The target and pilot center are

$$W_1(\mu,\nu) = \sup_{f\in \mathrm{Lip}(1)} |(\mu-\nu)f|, \qquad T_0 = W_1(\widehat{\mu}_0, \widehat{\nu}_0).$$

Use half of each population sample as pilot and B inference blocks of size $n/(2B)$.

- For unit scales the COSI residual radius is
$$R = \sup_{f\in \mathrm{Lip}(1)} \big(|(\widehat{\mu}_j - \widehat{\nu}_j - \widehat{\mu}_0 + \widehat{\nu}_0)f| : j = 1, \dots, B\big)_{(c_{B,\alpha}+1)}.$$
- If every block contrast has a median within a deterministic b of its mean, the conditional COSI argument gives, simultaneously for all critics, $|(\mu - \nu - \widehat{\mu}_0 + \widehat{\nu}_0)f| \le R + b$ with probability at least $1-\alpha$.
- The inequality $|\sup_f |a_f| - \sup_f |d_f|| \le \sup_f |a_f - d_f|$ then gives
$$\boxed{W_1(\mu,\nu) \in [(T_0 - R - b)_+,\, T_0 + R + b]}$$
on that same event.
- The next frame derives the common allowance used in the simulations:
$$b = \frac{2B}{n} \min\left\{1.6749\sqrt{5\max(C_{\mathrm{PC},\mu}, C_{\mathrm{PC},\nu})},\quad 1.407\sqrt{5(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})}\right\}.$$

This projects a simultaneous band for all population critic expectations; it imposes no translation model on the confidence set.

Supplement: the two Wasserstein moment certificates

Fix a 1-Lipschitz critic and write $Y_\mu = f(X) - \mathbb{E}_\mu f$, $Y_\nu = f(Y) - \mathbb{E}_\nu f$. The Poincaré argument gives

$$\mathbb{E}|Y_\lambda|^3 \le \sqrt{5C_{\mathrm{PC},\lambda}}\,\mathbb{E}Y_\lambda^2, \qquad \lambda \in \{\mu,\nu\}.$$

- A block contrast of size $m = n/(2B)$ per population is a sum of $2m$ independent centered terms divided by m. The aggregate moment ratio is at most $\sqrt{5\max(C_{\mathrm{PC},\mu}, C_{\mathrm{PC},\nu})}$. The independent-summand coefficient therefore gives the allowance

$$\frac{1.6749}{m}\sqrt{5\max(C_{\mathrm{PC},\mu}, C_{\mathrm{PC},\nu})}.$$

- Alternatively pair the observations into IID contrasts $Z = Y_\mu - Y_\nu$. Product Poincaré and the coordinatewise gradient bounds give, with $v = \mathbb{E}Z^2$,

$$v \le C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu}, \qquad \mathrm{Var}(Z^2) \le 4(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})v.$$

  Thus $\mathbb{E}Z^4 \le 5(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})v$ and

$$\mathbb{E}|Z|^3 \le \sqrt{5(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})}\,v.$$

  The IID coefficient gives the allowance $1.407\sqrt{5(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})}/m$.

- Both certificates hold uniformly over critics. They are deterministic functions of the supplied constants, so their minimum is valid without another probability budget. Substituting $m = n/(2B)$ explains the factor $2B/n$.

The product-Poincaré application to Z^2 is justified by the same truncation argument used for a single critic.

Supplement: exact signed transport for the COSI radius

With the maximal residual rank, finite maximum and supremum commute:

$$R = \max_{1\le j\le B} \sup_{f\in \mathrm{Lip}(1)} |(\widehat{\mu}_j - \widehat{\nu}_j - \widehat{\mu}_0 + \widehat{\nu}_0)f|.$$

The signed measure in each residual is twice the difference of the probability measures $(\widehat{\mu}_j + \widehat{\nu}_0)/2$ and $(\widehat{\nu}_j + \widehat{\mu}_0)/2$. Kantorovich duality gives

$$\boxed{R = \max_j 2W_1\left(\frac{\widehat{\mu}_j + \widehat{\nu}_0}{2}, \frac{\widehat{\nu}_j + \widehat{\mu}_0}{2}\right).}$$

- In one dimension, each residual norm is exactly

$$\int_{\mathbb{R}} \left| F_{\widehat{\mu}_j}(x) - F_{\widehat{\nu}_j}(x) - F_{\widehat{\mu}_0}(x) + F_{\widehat{\nu}_0}(x) \right| dx.$$

  For sorted pooled knots $x_1 < \cdots < x_N$, evaluate it as $\sum_{r=1}^{N-1} |A(x_r)|(x_{r+1} - x_r)$, where A is the cumulative signed mass.
- Each empirical measure has mass one. Pilot atoms have weight $2/n$ and inference atoms have weight $2B/n$; these unequal weights are retained in the exact integral.
- For $B = 5$ and maximal rank, the one-sided failure bound is 2^{-5}, giving target coverage $31/32$ without cutoff randomization.
- For nonmaximal ranks, generally only $\sup_f (r_j(f))_{(k)} \le (\sup_f r_j(f))_{(k)}$ holds. The maximal-rank experiment evaluates the entire Lipschitz class.

Supplement: the empirical-bias inequality

For total size $n = 2\ell$ per population, set

$$\widehat{T} = W_1(\widehat{\mu}_{2\ell}, \widehat{\nu}_{2\ell}), \qquad H = W_1(\widehat{\mu}_\ell^{(1)}, \widehat{\mu}_\ell^{(2)}) + W_1(\widehat{\nu}_\ell^{(1)}, \widehat{\nu}_\ell^{(2)}),$$

and write $e_\lambda(q) = \mathbb{E}W_1(\widehat{\lambda}_q, \lambda)$.

- ⋆ Joint convexity of W_1 and $\mathbb{E}\widehat{\lambda}_q = \lambda$ imply $W_1(\mu, \nu) \le \mathbb{E}\widehat{T}$. The triangle inequality gives the upper bound
$$\mathbb{E}\widehat{T} - W_1(\mu, \nu) \le e_\mu(2\ell) + e_\nu(2\ell).$$
- ⋆ Since the full empirical law is the average of its two half-sample laws, convexity implies
$$e_\lambda(2\ell) \le \frac{1}{2}\mathbb{E}W_1(\widehat{\lambda}_\ell^{(1)}, \lambda) + \frac{1}{2}\mathbb{E}W_1(\widehat{\lambda}_\ell^{(2)}, \lambda) = e_\lambda(\ell).$$
- ⋆ Conditional Jensen, averaging the independent second half, gives
$$W_1(\widehat{\lambda}_\ell^{(1)}, \lambda) \le \mathbb{E}\Big(W_1(\widehat{\lambda}_\ell^{(1)}, \widehat{\lambda}_\ell^{(2)}) \mid \widehat{\lambda}_\ell^{(1)}\Big).$$
Taking expectations and summing the two populations proves
$$\boxed{0 \le \mathbb{E}\widehat{T} - W_1(\mu, \nu) \le \mathbb{E}H.}$$

Finite first moments ensure that all these expectations are finite. The argument concerns the population distance itself, including equality and shape changes.

Supplement: bias-interval coverage and its width

Choose deterministic margins c, t such that

$$\mathbb{P}(\mathbb{E}H > H + c) \le \eta, \qquad \mathbb{P}(|\widehat{T} - \mathbb{E}\widehat{T}| > t) \le \delta.$$

On the intersection of the two good events,

$$\widehat{T} - H - c - t \le \mathbb{E}\widehat{T} - \mathbb{E}H \le W_1(\mu, \nu) \le \mathbb{E}\widehat{T} \le \widehat{T} + t.$$

Since the target is nonnegative, the union bound proves

$$\boxed{I_B = [(\widehat{T} - H - c - t)_+, \widehat{T} + t], \qquad \mathbb{P}(W_1(\mu, \nu) \in I_B) \ge 1 - \eta - \delta.}$$

- ⋆ The same observations may be used for $\widehat{T}$ and H. Their independence is not required. The comparisons use $\eta = \delta = 1/64$, so the target is $31/32$.
- ⋆ Before any known-support truncation, the interval width is exactly
$$|I_B| = \widehat{T} + t - \max\{0, \widehat{T} - H - c - t\} = \min\{\widehat{T} + t,\ H + c + 2t\}.$$
Its one-sided geometry can be favorable near equality, where the first term is small.
- ⋆ COSI's explicit median allowance is of order $\sqrt{C_{\mathrm{PC}}/n}$ for fixed B; the bias-concentration margins are of order $\sqrt{C_{\mathrm{PC}}/n}$ at fixed error levels. Empirical transport terms remain in both widths, so these orders do not imply a universal width ordering.
- ⋆ Both procedures use the same supplied constants: 1 in the vanishing-variance study and e^4 in the mixture study. Sharper supplied bounds can change the comparison.

Supplement: matched LSI and bounded-support margins

Changing one observation in $\widehat{T}$ changes it by at most its displacement divided by $n = 2\ell$. For H, each observation occurs in exactly one half-sample discrepancy and has Lipschitz coefficient $1/\ell$.

- ⋆ Under tensorized LSI, a statistic with coordinatewise bounds a_i has

$$\log \mathbb{E} e^{s(F-\mathbb{E}F)} \le \frac{s^2}{2} \sum\nolimits_i C_{\mathrm{LS},i} a_i^2.$$

  The corresponding sums are $(C_{\mathrm{LS},\mu} + C_{\mathrm{LS},\nu})/n$ for $\widehat{T}$ and $2(C_{\mathrm{LS},\mu} + C_{\mathrm{LS},\nu})/\ell$ for H. One lower tail for H and two tails for $\widehat{T}$ give

$$\boxed{c = \sqrt{\frac{4(C_{\mathrm{LS},\mu} + C_{\mathrm{LS},\nu})}{\ell} \log(1/\eta)}, \quad t = \sqrt{\frac{2(C_{\mathrm{LS},\mu} + C_{\mathrm{LS},\nu})}{n} \log(2/\delta)}.}$$

- ⋆ With support diameters D_μ, D_ν, McDiarmid uses squared oscillation sums $(D_\mu^2 + D_\nu^2)/n$ for $\widehat{T}$ and $2(D_\mu^2 + D_\nu^2)/\ell$ for H. Its one-sided bound is $\exp(-2u^2/\sum_i c_i^2)$, yielding

$$\boxed{c = \sqrt{\frac{D_\mu^2 + D_\nu^2}{\ell} \log(1/\eta)}, \qquad t = \sqrt{\frac{D_\mu^2 + D_\nu^2}{2n} \log(2/\delta)}.}$$

The LSI margins use functional inequalities; the McDiarmid margins use actual bounded supports. Each method is compared only under its stated supplied assumptions.

Supplement: Poincaré variance and exponential margins

For independent inputs, tensorization gives $\mathrm{Var}(F) \le K_F := \sum_i C_{\mathrm{PC},i} a_i^2$ when $\|\nabla_i F\| \le a_i$. For the bias statistics,

$$K_{\widehat{T}} = \frac{C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu}}{n}, \qquad K_H = \frac{2(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})}{\ell}.$$

- Cantelli for the lower tail of H and Chebyshev for $\widehat{T}$ give
$$c = \sqrt{K_H(1-\eta)/\eta}, \qquad t = \sqrt{K_{\widehat{T}}/\delta}.$$
- For a centered statistic, write $M(s) = \mathbb{E}e^{s(F-\mathbb{E}F)}$. For $|s| < 2/\sqrt{K_F}$, Poincaré applied to $e^{sF/2}$ yields $M(s) - M(s/2)^2 \le (K_F s^2/4) M(s)$. Iteration gives
$$\log M(s) \le \sum\nolimits_{j\ge 0} 2^j \left[-\log\left(1 - \frac{K_F s^2}{4^{j+1}}\right)\right].$$
At $s = K_F^{-1/2}$, use $-\log(1-u) \le 4u/3$ for $u \le 1/4$ to bound the series by $2/3$. Exponential Markov then bounds $\mathbb{P}(F - \mathbb{E}F \ge \sqrt{K_F}\{\log(1/p) + 2/3\}) \le p$.
- Applying this to $-H$ and both signs of $\widehat{T}$ gives the second pair
$$c = \sqrt{K_H}\{\log(1/\eta) + 2/3\}, \qquad t = \sqrt{K_{\widehat{T}}}\{\log(2/\delta) + 2/3\}.$$

The moment-generating-function argument is first applied to bounded truncations. For $K_F = 0$, F is constant almost surely and the margin is zero.

Supplement: tensorization and the product-Poincaré MGF

For $\|\nabla g\| \le u < 2/\sqrt{C_{\mathrm{PC}}}$, Bobkov–Ledoux (1997) gives

$$\mathrm{Ent}(e^g) \le C_{\mathrm{PC}}\, q(\sqrt{C_{\mathrm{PC}}}\, u)\, \mathbb{E}[e^g \|\nabla g\|^2], \qquad q(v) = \frac{1}{2}\left(\frac{2+v}{2-v}\right)^2 e^{\sqrt{5}v}.$$

Here q is defined on $0 \le v < 2$, with $q(0) = 1/2$. For independent inputs, $\|\nabla_i F\| \le a_i$, and $K_F = \sum_i C_{\mathrm{PC},i} a_i^2$, put $\beta_F = \max_i \sqrt{C_{\mathrm{PC},i}}\, a_i$.

- ⋆ Apply the entropy inequality conditionally in each coordinate at $g = sF$. Entropy tensorization and the increasing function q give, for $0 < s < 2/\beta_F$,

$$\begin{aligned}\mathrm{Ent}(e^{sF}) &\le s^2 \sum_i C_{\mathrm{PC},i}\, q(s\sqrt{C_{\mathrm{PC},i}}\, a_i)\, \mathbb{E}[e^{sF}\|\nabla_i F\|^2] \\ &\le K_F s^2 q(s\beta_F)\, \mathbb{E}e^{sF}.\end{aligned}$$

- ⋆ Let $\psi(s) = \log \mathbb{E}e^{s(F-\mathbb{E}F)}$. Dividing by the moment-generating function yields $s\psi'(s) - \psi(s) \le K_F s^2 q(s\beta_F)$, or $(\psi(s)/s)' \le K_F q(s\beta_F)$.
- ⋆ Since $\psi(s)/s \to 0$ as $s \downarrow 0$, integrate to obtain

$$\boxed{\psi(s) \le K_F s \int_0^s q(u\beta_F)\, du \le K_F s^2 q(s\beta_F).}$$

Apply the argument to bounded truncations first; the gradient bounds are unchanged. Fatou then supplies the same moment-generating-function bound for the original statistic.

Supplement: evaluating the product-Poincaré margins

The preceding MGF bound and exponential Markov show that every deterministic $0 < s < 2/\beta_F$ certifies the one-sided failure-p margin

$$\boxed{r_p = \frac{\log(1/p)}{s} + K_F s\, q(s\beta_F).}$$

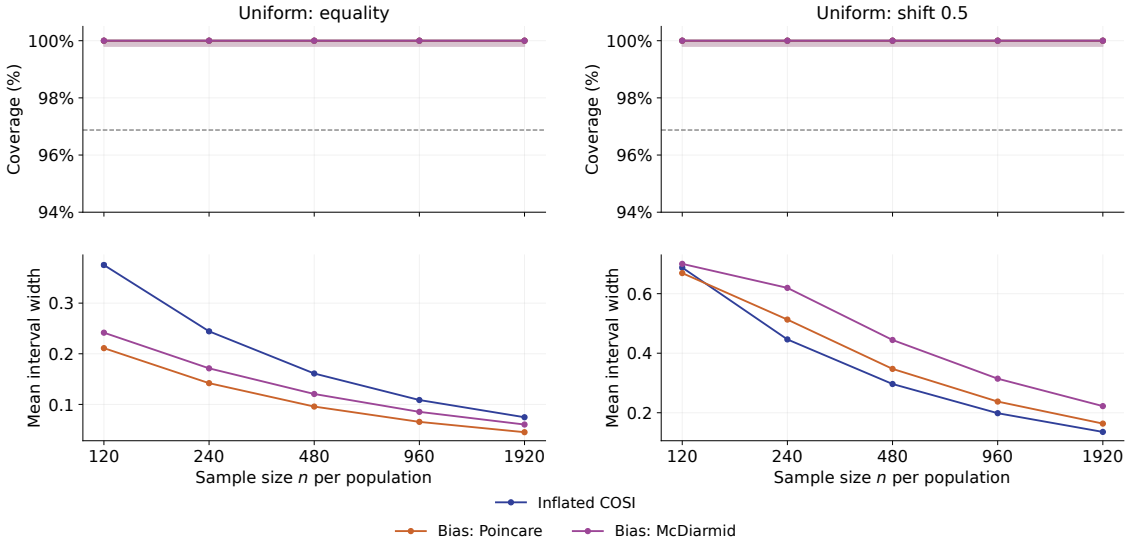
Equivalently, with $v = s\beta_F$, $r_p = \beta_F \log(1/p)/v + K_F v q(v)/\beta_F$. For the empirical-bias statistics, the coordinatewise coefficients give

$$(K_{\widehat{T}}, \beta_{\widehat{T}}) = \left(\frac{C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu}}{n}, \frac{\sqrt{\max(C_{\mathrm{PC},\mu}, C_{\mathrm{PC},\nu})}}{n} \right),$$

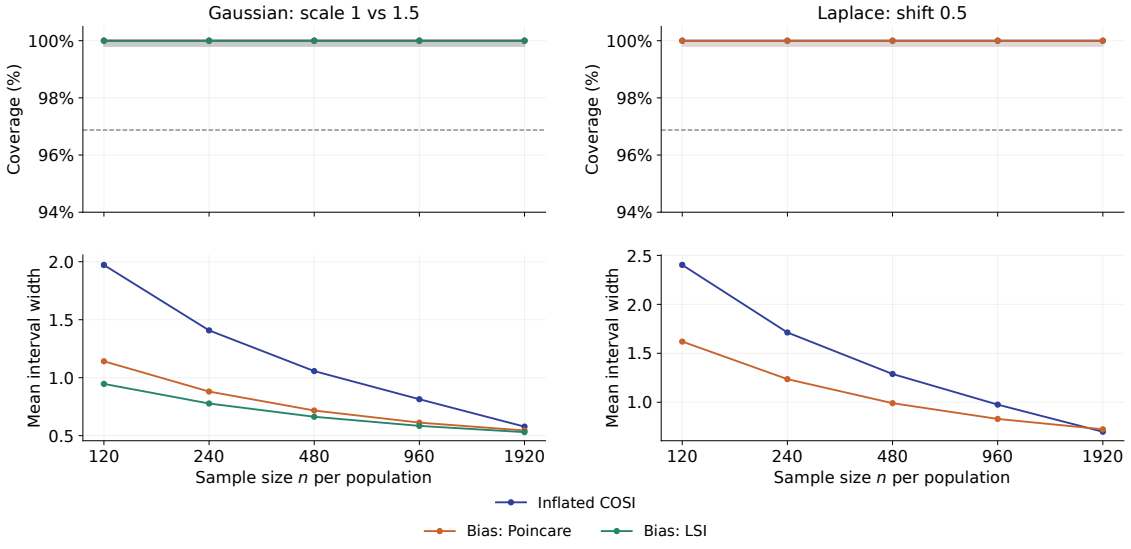
$$(K_H, \beta_H) = \left(\frac{2(C_{\mathrm{PC},\mu} + C_{\mathrm{PC},\nu})}{\ell}, \frac{\sqrt{\max(C_{\mathrm{PC},\mu}, C_{\mathrm{PC},\nu})}}{\ell} \right).$$

- Use $p = \eta$ for the lower-tail margin c and $p = \delta/2$ for the two-sided margin t. When the variance certificate is zero, the statistic is constant and its margin is zero.
- The code selects admissible trial points and evaluates the displayed certificate with an outward numerical guard. Finding its exact optimizer is unnecessary for validity.
- Each margin is the minimum of the three deterministic certificates using the supplied constants; no extra probability budget is needed.

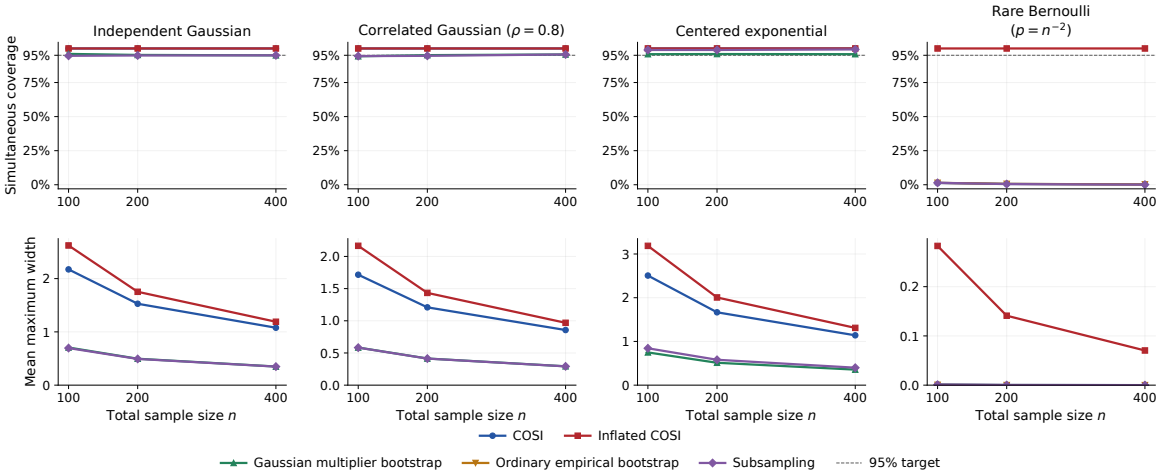
Supplement: Wasserstein on bounded supports



Supplement: Wasserstein scale and tail changes



Supplement: all high-dimensional settings, $d = 100$



References

- Bobkov (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* 27, 1903–1921.
- Bobkov–Ledoux (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *PTRF* 107, 383–400.
- Conforti–Lacker–Pal (2025). Projected Langevin dynamics and a gradient flow for entropic optimal transport. *JEMS*, online first; arXiv:2309.08598.
- Hamza (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *SPL* 23, 21–25.
- Kuchibhotla–Balakrishnan–Wasserman (2024). The HulC: confidence regions from convex hulls. *JRSS–B* 86, 586–622.
- Maurer–Pontil (2009). Empirical Bernstein bounds and sample variance penalization. *COLT*.
- Oleszkiewicz (2001). Concentration of capital—the product form of the Law of Large Numbers in L_1. *SPL* 55, 159–162.
- Peel–Anthoine–Ralaivola (2010). Empirical Bernstein inequalities for U-statistics. *NeurIPS* 23.

References (contd.)

- Shevtsova (2013). On the absolute constants in the Berry–Esseen inequality and its structural and nonuniform improvements. *Inform. Primen.* 7(1), 124–125. Publisher record.
- Bell (2026). An upper bound on the Berry–Esseen constant for independent summands. Research collection.
- Kołodziejek–Rodziewicz (2026). The sharp iid Berry–Esseen inequality: analytic reductions and exact certificate proof. Hexagon 2610.00590v1.
- Friedrich (1989). A Berry–Esseen bound for functions of independent random variables. *Ann. Statist.* 17, 170–183.
- Chen–Shao (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* 13, 581–599.
- Shao–Zhang (2025). Berry–Esseen bounds for functionals of independent random variables. *Stochastic Process. Appl.* 183, 104574.
- Research drafts: *COSI: Scale-invariant Inference*; *Mean–Median Bounds under Moment Constraints and Finite-Sample Confidence Intervals for Means and Nonlinear Statistics*; *COSI and empirical-bias intervals for Wasserstein distances*, 12 November 2026.