Jakob R. Juergens

University of Wisconsin – Madison
Department of Economics

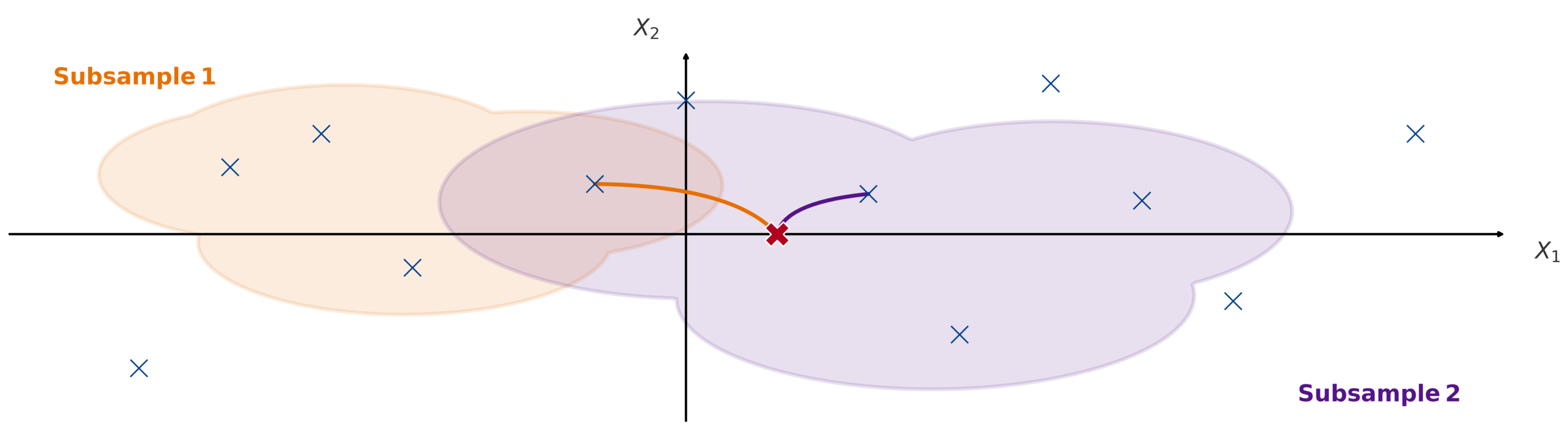
Robust Inference for Conditional Z-Estimation via Distributional Nearest Neighbors

Conditional Z-Estimation Setup

We observe an i.i.d. sample $\mathbf{D}_n = \{W_i = (X_i, Y_i, Z_i)\}_{i=1}^n$. For an interior point $\mathbf{x}$ and parameter $\alpha_0(\mathbf{x}) \in \mathbb{R}^{d_\alpha}$:

$$\mathbb{E}\left[\psi\left(W; \alpha_0(\mathbf{x}), \eta_0(Z)\right) | X = \mathbf{x}\right] = 0.$$

We want to estimate a local parameter $\alpha_0(\mathbf{x})$ and conduct inference at $\mathbf{x}$.



Rank Weights and Locality

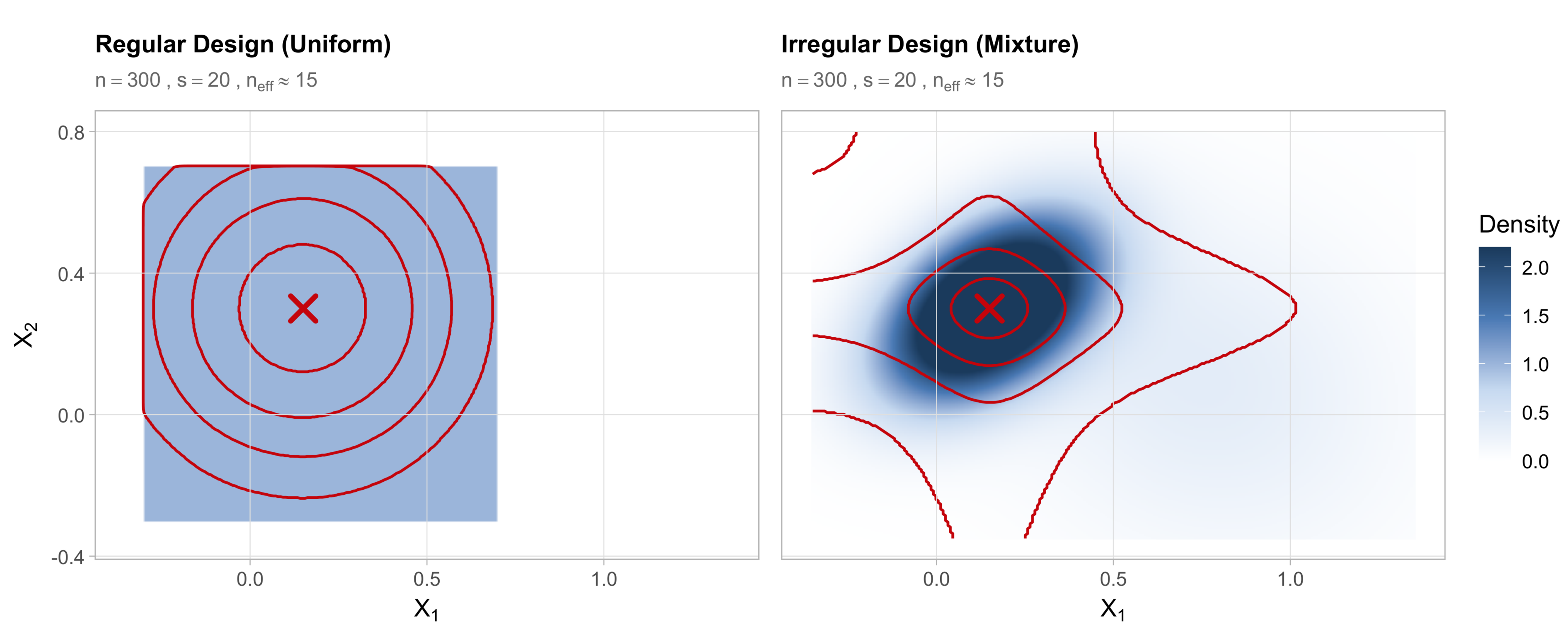
For ordered covariates $\|X_{(1)} - \mathbf{x}\| \leq \|X_{(2)} - \mathbf{x}\| \leq \ldots \leq \|X_{(n)} - \mathbf{x}\|$. around x,

$$w_{(i)}(\mathbf{x}) = \binom{n}{s}^{-1}\binom{n-i}{s-1}, \qquad i = 1, \ldots, n-s+1.$$

- $w_{(i)}(\mathbf{x})$ is the share of s-subsamples where rank-i is nearest to $\mathbf{x}$
- Distributional Nearest Neighbor weights act like a design-adaptive kernel
- The kernel order s is the single tuning parameter acting as an inverse bandwidth

The effective sample size is

$$n_{\text{eff}} := \left(\sum_{i=1}^{n-s+1} w_{(i)}(\mathbf{x})^2\right)^{-1} \asymp \frac{n}{s}.$$



DNN weight level sets adapt to local covariate density, acting as a design-adaptive kernel.

DNN-Localized Z-Estimator

1. Estimate nuisance components once on the full sample to obtain $\hat{\eta}$.
2. Order observations by distance to $\mathbf{x}$ and compute DNN rank weights $w_{(i)}(\mathbf{x})$.
3. Solve the localized moment equation

$$\sum_{i=1}^{n-s+1} w_{(i)}(\mathbf{x})\, \psi\left(W_{(i)}; \hat{\alpha}(\mathbf{x}), \hat{\eta}(Z_{(i)})\right) = 0$$

to obtain $\hat{\alpha}(\mathbf{x})$.

Why No Cross-Fitting?

When only n_{eff} observations matter locally, fold-splitting can cause first-order precision loss. Same-sample nuisance fits are justified by two ideas: orthogonality removes first-order nuisance bias, and LOO stability says no single observation can materially change the nuisance fit.

$$\partial_r \mathbb{E}\left[\psi\left(W; \alpha_0(\mathbf{x}), [\eta_0 + r(\eta - \eta_0)](Z)\right) | X = \mathbf{x}\right]\big|_{r=0} = 0$$

$$\max_{l \in [n]} \left\| \hat{\eta}_j - \hat{\eta}_j^{(-l)} \right\|_\eta = o\left(n^{-1/2}\right).$$

For DNN weights, the moment level weight-stability conditions are automatic. Practically, this preserves full local signal in both stages.

Main Results

CLT. Under local smoothness, orthogonality, LOO stability, and DNN undersmoothing

$$s \to \infty, \qquad s = o(n), \qquad s \gg n^{k_x/(k_x+2)},$$

we obtain a pointwise CLT at $\mathbf{x}$:

$$\sqrt{n_{\text{eff}}}\left(\hat{\alpha}(\mathbf{x}) - \alpha_0(\mathbf{x})\right) \rightsquigarrow \mathcal{N}\left(0, \Omega(\mathbf{x})\right), \qquad \Omega(\mathbf{x}) = \Gamma(\mathbf{x})^{-1}\Sigma(\mathbf{x})\Gamma(\mathbf{x})^{-T}.$$

$$\Gamma(\mathbf{x}) = \partial_a \mathbb{E}[\psi(W; a, \eta_0) \mid X = \mathbf{x}]|_{a=\alpha_0(\mathbf{x})},$$
$$\Sigma(\mathbf{x}) = \mathrm{Var}(\psi(W; \alpha_0(\mathbf{x}), \eta_0) \mid X = \mathbf{x}).$$

Variance Estimation. Plug-in sandwich at $(\hat{\alpha}(\mathbf{x}), \hat{\eta})$:

$$\hat{\Omega}(\mathbf{x}) = \hat{\Gamma}(\mathbf{x})^{-1}\hat{\Sigma}(\mathbf{x})\hat{\Gamma}(\mathbf{x})^{-T}.$$

$\hat{\Sigma}(\mathbf{x})$: delete-1 jackknife meat using DNN LOO rank weights (computed in $O(n)$ time),
$\hat{\Gamma}(\mathbf{x})$: localized finite-difference bread (or analytic derivative).

Feasible Inference. Standard Wald tests and confidence sets are asymptotically valid:

$$\hat{\Omega}(\mathbf{x})^{-1/2}\sqrt{n_{\text{eff}}}\left(\hat{\alpha}(\mathbf{x}) - \alpha_0(\mathbf{x})\right) \rightsquigarrow \mathcal{N}\left(0, I_{d_\alpha}\right).$$

Applied Diagnostics at $\mathbf{x}$.

- Report nuisance LOO sensitivity $\max_i \left\| \hat{\eta}^{(-i)} - \hat{\eta} \right\|_\eta$ and conditioning of $\hat{\Gamma}(\mathbf{x})$.
- Check sensitivity to s and Treat n_{eff} as a tuning summary implied by (n, s)

WWW.JAKOBJUERGENS.COM
JRJUERGENS@WISC.EDU