Kernel von Mises Formula of the Influence Function

Yaroslav Mukhin (Cornell CDSES/Econ/Stats)

NeurIPS2025 @ San Diego

Problem 1: statistical inference

Functional $\theta(P)$ estimated with $\hat{\theta}_n : X_{1:n} \mapsto \mathbb{R}$ in the setting

a given functional

unknown probability measure

parametric rate

$$\sqrt{n}\Big(\hat{\theta}_n(X_1, \ldots, X_n) - \theta(P)\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_P(X_i) + o_P$$

observed random sample from P

influence function of θ

What is a 95% confidence interval for $\theta(P)$?

Solution:

$$\hat{\theta} \overset{\text{approx}}{\sim} N\Big(\theta(P), \frac{\mathrm{E}\psi_P^2(X)}{n}\Big) \qquad \text{so} \qquad \Big[\hat{\theta} \pm 1.96\sqrt{\frac{\frac{1}{n}\sum \hat{\psi}^2(X_i)}{n}}\Big]$$

Problem 2: data attribution

What data points have the most influence on $\hat{\theta}_n$?

$$\sqrt{n}\Big(\hat{\theta}_n(X_1, \ldots, X_n) - \theta(P)\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_P(X_i) + o_P$$

Perturbing the probability weights:

$$\max_{i} \big|\hat{\theta}_{n-1}(X_{1:n\setminus i}) - \hat{\theta}_n(X_{1:n})\big|$$

Perturbing the spatial location:

$$\max_{i,\ \Delta_i} \big|\hat{\theta}_n(X_1, \ldots, X_i + \Delta_i, \ldots, X_n) - \hat{\theta}_n(X_{1:n})\big|$$

Solution (approximate):

$$\arg\max_{i} \psi(X_i) \qquad \text{and} \qquad \arg\max_{i} \nabla\psi(X_i)$$

Problem 3: efficient estimator

How to construct estimator $\hat{\theta}_n$ with smallest asymptotic variance?

$$\sqrt{n}\Big(\hat{\theta}_n(X_1, \ldots, X_n) - \theta(P)\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_P(X_i) + o_P$$

Solution: Bickel's one-step estimator:

$$\hat{\theta}_{\text{1-step}} = \underset{\substack{\nearrow \\ \text{preliminary estimate}}}{\tilde{\theta}} + \underbrace{\frac{1}{n} \sum_{i=1}^{n} \psi(X_i; \tilde{\eta}, \tilde{\theta})}_{\text{correction term}}.$$

Other approaches:

- targeted maximum likelihood of van der Laan
- Neyman orthogonal estimating equations of Chernozhukov

also require the influence function ψ as input.

Problem 4: distributionally robust optimization

Consider optimization over a probability measure P:

$$\boxed{\min_P \kappa\Big(\theta(P), P\Big)}$$

How to find the minimizer $P^\star$ and minimum value $\kappa(\theta(P^\star), P^\star)$?

Solution: gradient descent. What is the gradient of κ?

Gradient for $\theta \in \mathbb{R}^p$ is a vector:

$$\nabla_\theta \kappa = \Big(\frac{\partial \kappa}{\partial \theta_1}, \dots, \frac{\partial \kappa}{\partial \theta_p}\Big).$$

Gradient of θ with respect to P is a function: $\nabla_P \theta = \psi$, known as

- influence function
- first variation
- Fisher-Rao gradient

In this paper I propose an rkHs estimator of this function.

Spectral von Mises Formula

How to compute the influence function ψ?

State of the art: **von Mises formula**

mollification bandwidth

toward point-mass at z

$$\psi(z) = \lim_{\delta \to 0} \left[\frac{d}{dt} \theta(P_t^{\delta, z}) \right]_{|t=0}$$

perturbation of P

Limitations: (i) separate calculation for each z (ii) ill-conditioned computation

I propose **Spectral von Mises formula**

compute cost

toward e_j, well-conditioned

$$\psi_P(x) \approx \sum_{j=1}^{r} \frac{1}{1 + 2\lambda/\sigma_j} \left[\frac{d}{dt} \theta(P_t^{j}) \right]_{|t=0} e_j(x)$$

regularization, bias-variance trade-off

kPCA basis

kernel von Mises Estimator

$$\hat{\psi}^r_\lambda(x) := \sum_{j=1}^{r} \frac{1}{1 + 2\lambda/\hat{\sigma}} \left[\frac{d}{dt}\hat{\theta}(\hat{P}^j_t)\right]_{|t=0} \hat{e}_j(x)$$

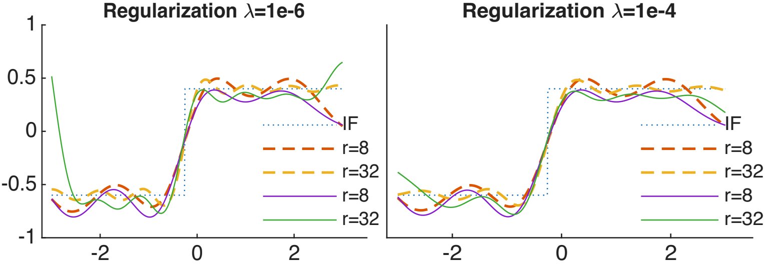

Figure: IF ψ_q of quantile θ_q, regularized surrogate ψ^r_λ (dashed) and estimate $\hat{\psi}^r_{\lambda,n}$ (solid).

Theorem Under assumptions, $\hat{\psi}^r_\lambda$ is consistent in $L^2(P)$ and rkHs norms.