Consistency of Robust Kernel Density Estimators

Robert A. Vandermeulen & Clayton D. Scott

University of Michigan, Dept. of Electrical Engineering and Computer Science

June 13, 2013

Robust Kernel Density Estimation

- $X_1, \cdots X_n \overset{iid}{\sim} f$ in $\mathbb{R}^d$

$$\bar{f}_\sigma^n := \frac{1}{n} \sum_1^n k_\sigma(\cdot, X_i)$$

$$\left[\begin{array}{l} \Phi_\sigma : \mathbb{R}^d \to \mathcal{H}_\sigma \\ \Phi_\sigma(x) := k_\sigma(\cdot, x) \end{array} \right]$$

$$= \frac{1}{n} \sum_1^n \Phi_\sigma(X_i)$$

$$= \arg\min_{g \in \mathcal{H}_\sigma} \sum_1^n \|\Phi_\sigma(X_i) - g\|_{\mathcal{H}_\sigma}^2$$

- Replacing the squared loss with a robust loss ρ yields a Robust Kernel Density Estimator (Kim and Scott 2012)

$$f_\sigma^n := \arg\min_{g \in \mathcal{H}_\sigma} \sum_1^n \rho\left(\|\Phi_\sigma(X_i) - g\|_{\mathcal{H}_\sigma}\right)$$

Theorem

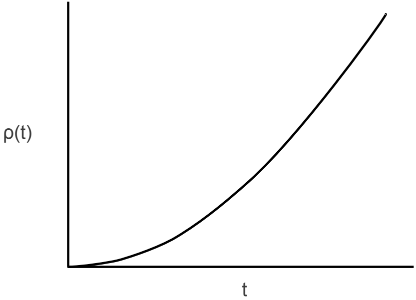
Given certain properties on ρ:

Theorem

Let $f \in L^2\left(\mathbb{R}^d\right)$. If $n\sigma^d \to \infty$ and $\sigma \to 0$ as $n \to \infty$ then $\|f_\sigma^n - f\|_1 \xrightarrow{p} 0$.

Example Losses

- $\rho(t) = \sqrt{t^2+1} - 1$
- $\rho(t) = t \arctan(t)$
- $\rho(t) = t - \log(1+t)$



Proof Strategy

- Traditional bias-variance decompositon presents challenges
  - No closed form expression for f_σ^n or its infinite-sample counterpart
  - $\mathcal{H}_\sigma$ gets larger as $n \to \infty$ since $\sigma \to 0$
- Alternate approach based on IRWLS algorithm
  - $R_\sigma^n : \mathcal{H}_\sigma \to \mathcal{H}_\sigma$
  - $R_\sigma^n(g) := \sum w_i(g)\, k_\sigma(\cdot, X_i)$
  - $w_i(g) := \frac{\varphi\left(\|\Phi_\sigma(X_i) - g\|_{\mathcal{H}_\sigma}\right)}{\sum_1^n \varphi\left(\|\Phi_\sigma(X_j) - g\|_{\mathcal{H}_\sigma}\right)}$
  - $\varphi(t) := \frac{\rho'(t)}{t}$
  - $\left(R_\sigma^n\right)^k(g) \to f_\sigma^n$ as $k \to \infty$

Proof Sketch

- $R_\sigma^n(0) = \bar{f}_\sigma^n$ (the traditional KDE)
- $R_\sigma(f_\sigma^n) = f_\sigma^n$

$$\begin{aligned} \left\| f_\sigma^n - \bar{f}_\sigma^n \right\|_{\mathcal{H}_\sigma} &= \left\| R_\sigma^n(f_\sigma^n) - R_\sigma^n(0) \right\|_{\mathcal{H}_\sigma} \\ &\leq C\sigma^{d/2} \left\| f_\sigma^n - 0 \right\|_{\mathcal{H}_\sigma} \end{aligned}$$

References

- J. Kim and C. Scott, “Robust kernel density estimation,” Journal of Machine Learning Research, vol. 13, pp. 2529-2565, 2012.