The Graphical Nadaraya Watson Estimator on Latent Postion Models

Summary

- We derive **sample complexities** and **generalization bounds** for a signal averaging estimator on graphs
- Analysis valid for general **Latent Position Models**

Notation

All random variables belong to a joint probability space $(\Omega, \mathcal{F}, \mathbb{P})$

- $\mathbb{E}(F(X, X_1, .., X_n, X, U_1, U_2, ..., U_n, \epsilon_1, ..., \epsilon_n)$- expectation taken over every variable
- $\mathbb{E}_x(\cdot) = \mathbb{E}(\cdot | X = x)$ conditional expectation
- $d_n(x) = \mathbb{E}_x(\sum_{i=1}^n a(X, X_i))$-local expected degree at $x \in \mathbb{R}^d$
- $b_n(f, x) = \begin{cases} \frac{\int f(z) k_n(x,z) p(z) dz}{\int k_n(x,z) p(z) dz} & \text{if } d_n(x) > 0 \\ 0 & \text{otherwise} \end{cases}$
- $Q = \operatorname{supp} p$

The GNW Estimator

$$\hat{f}_{GNW}(X) = \begin{cases} \frac{\sum_{i=1}^n Y_i a(X, X_i)}{\sum_{i=1}^n a(X, X_i)} & \text{if } \sum a(X, X_i) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

- *How does $\mathbb{E}_x(\hat{f}_{GNW}(X) - b_n(f, X))^2$ depend on $d_n(x)$?*
- *For **convolutional kernels** and **smooth signals**, how does $\mathbb{E}(\hat{f}_{GNW}(X) - f(X))^2$ depend on the bandwith h_n?*

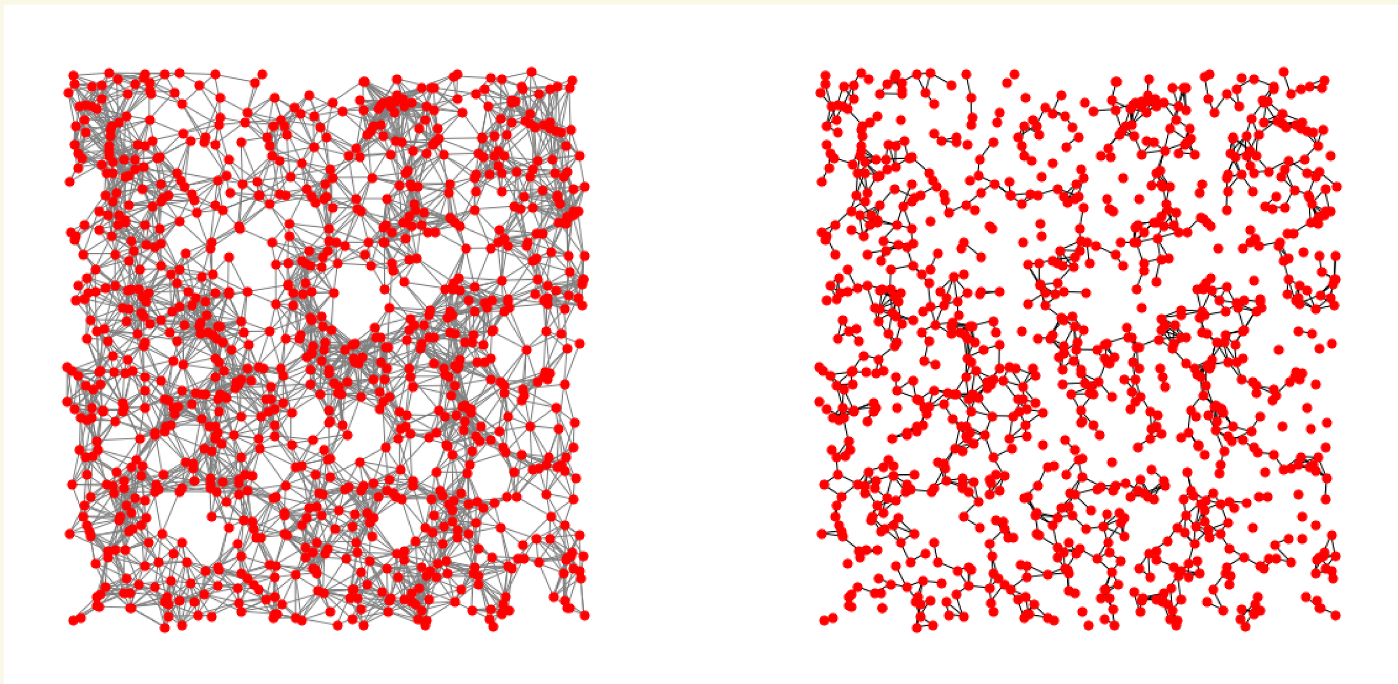


Figure: Left- Graph with $\sim \log n$ average degree, Right- Graph with $\sim \log(\log(n))$ average degree

MISE bound for convolutional kernels

Convolutional kernels $k_n(x, z) = K(\frac{x-z}{h_n})$ with $K \colon \mathbb{R}^d \to [0, 1]$, $h_n > 0$. Going from $\mathbb{E}_x(\cdot)$ to $\mathbb{E}(\cdot)$:

- Bounds on $|b_n(f, x) - f(x)|$ (the **bias** term)
- Asymptotically $d_n(x) \sim n h_n^d p(x)$ (**Lebesgue Density theorem**)

Nonasimptotically, the following holds:

- K compactly supported
- $p(x) \geqslant p_0 > 0$ on Q and Q *satisfies interior cone con dition*
- f is α Hölder continous on Q

then for sufficiently small bandwiths h_n

$$\mathbb{E}(\hat{f}_{GNW}(X) - f(X))^2 \leqslant C_1(\alpha) h_n^\alpha + \frac{C(B, \sigma)}{n h_n^d}$$

Framework

Latent Position Models:

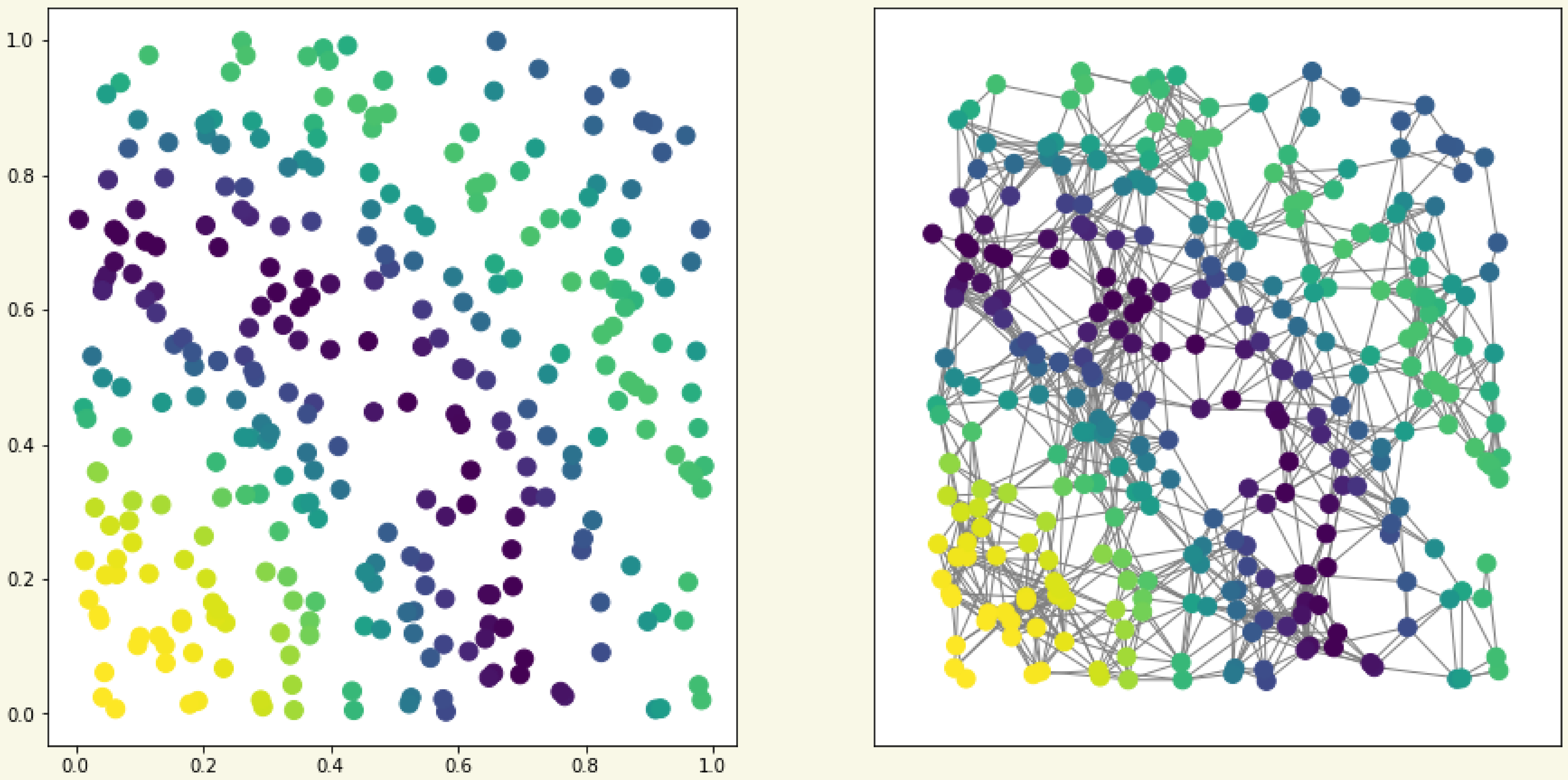


Figure: Left- latent positions, Right - Latent Graph Random Graph

- $X_1, ...X_n, X$ i.i.d. $\sim p$, p a density on $\mathbb{R}^d$ ***not observed***
- $k_n : \mathbb{R}^d \to [0, 1]$ probability kernel
- $a(X_i, X_j) = bern(k_n(X_i, X_j)) : = \mathbb{I}(k_n(X_i, X_j) \leqslant U_{i,j})$
- $Y_i - f(X_i) + \epsilon_i$, $\epsilon = (\epsilon_1, ..., \epsilon_n)$ independent from $(X_1, ..., X_n)$, $\mathbb{E}\epsilon_i = 0$, $\mathbb{E}\epsilon_i^2 = \sigma^2 < \infty$

Sharp Variance Bound

Suppose that $f \colon \mathbb{R}^d \to \mathbb{R}$ is such that $||f||_\infty \leqslant B$ and $\mathbb{E}(\epsilon_1^2) = \sigma^2 > 0$. Then

$$\frac{\sigma^2(1 - e^{-d_n(x)})}{d_n(x)} \leqslant \mathbb{E}(\hat{f}_{GNW}(x) - b_n(f, x))^2 \leqslant \frac{C(B, \sigma^2)}{d_n(x)}$$

Proof Sketch - the Decoupling trick

Let $I \subseteq [n]$. Define[a]

$$R_I(x) = \frac{1}{|I| + \sum_{j \notin I} a(x, X_i)}$$

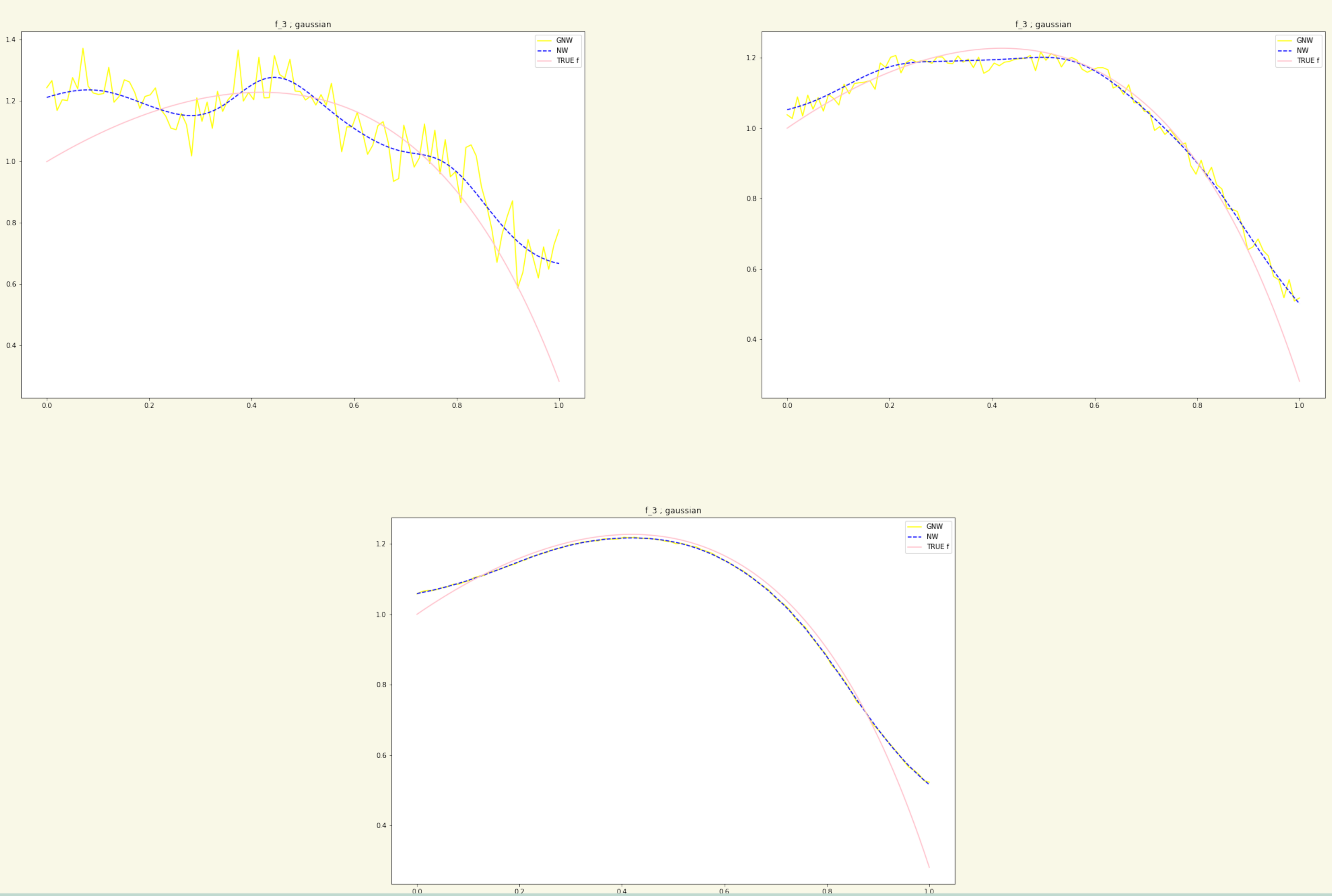
For all pairs of **disjoint** subsets $I, J \subseteq$[n] we have

$$R_J(x) \prod_{i \in I} a(x, X_i) = R_{I \cup J}(x) \prod_{i \in I} a(x, X_i)$$

and $R_{I \cup J}(x)$ is **independent** from $\{a(x, X_i) | i \in I\}$.

- "Linearized" representation $\hat{f}_{GNW}(x) = \sum_{i=1}^n Y_i a(x, X_i) R_i(x)$
- Difficult computations become tractable e.g. $\mathbb{E}(\hat{f}_{GNW}(x)) = b_n(f, x)(1 - d_n(x)/n)^n$

Simulations



Martin Gjorgjevski

Martin.Gjorgjevski@gipsa-lab.grenoble-inp.fr

GIPSA-lab, Grenoble