Optimal rates for k-NN density and mode estimation

Samory Kpotufe
ORFE, Princeton University

Joint work with Sanjoy Dasgupta, UCSD, CSE.
Goal:

Practical and Optimal estimator of all modes of $f$ from $X_{1:n} \sim F^n$.

Rate-Optimal: single mode case ([S. Tsybakov, 90] ...).
Practical: mean-shift (hard to analyze ... see [Genovese, ... Wasserman et.al., 13], [Arias-Castro et.al., 13] on consistency).

We derive a rate-optimal estimator based on $k$-NN graphs ...
Goal:

Practical and Optimal estimator of all modes of $f$ from $X_{1:n} \sim F^n$.

Rate-Optimal: single mode case ([S. Tsybakov, 90] ...).

Practical: mean-shift (hard to analyze ... see [Genovese, ... Wasserman et.al., 13], [Arias-Castro et.al., 13] on consistency)

We derive a rate-optimal estimator based on $k$-NN graphs ...
**Goal:**

Practical and Optimal estimator of all modes of $f$ from $X_{1:n} \sim F^n$.

Rate-Optimal: single mode case ([S. Tsybakov, 90] ...).
Practical: mean-shift (hard to analyze ... see [Genovesee, ... Wasserman et.al., 13], [Arias-Castro et.al., 13] on consistency)

We derive a rate-optimal estimator based on $k$-NN graphs ...
**Goal:**

Practical and Optimal estimator of all modes of $f$ from $X_{1:n} \sim F^n$.

**Rate-Optimal:** single mode case ([S. Tsybakov, 90] ...).

**Practical:** mean-shift (hard to analyze ... see [Genovesee, ... Wasserman et.al., 13], [Arias-Castro et.al., 13] on consistency)

We derive a rate-optimal estimator based on $k$-NN graphs ...
Goal:

Practical and Optimal estimator of all modes of $f$ from $X_{1:n} \sim F^n$.

Rate-Optimal: single mode case ([S. Tsybakov, 90] ...).
Practical: mean-shift (hard to analyze ... see [Genovesee, ... Wasserman et.al., 13], [Arias-Castro et.al., 13] on consistency)

We derive a rate-optimal estimator based on $k$-NN graphs ...
Program of construction

- **$k$-NN density rates:**
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]). We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg\sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg\max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers *just* modes at optimal rates!
Program of construction

• \textbf{\textit{k-NN density rates:}}
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]).
  We show high-prob. finite sample rates!

• \textbf{Single mode:}
  Common estimator in theory: $\hat{x} = \arg \sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg \max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

• \textbf{Multiple modes:}
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers \textit{just} modes at optimal rates!
Program of construction

- **$k$-NN density rates:**
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]). We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg \sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg \max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers *just* modes at optimal rates!
Program of construction

- **$k$-NN density rates:**
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]). We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg\sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg\max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers *just* modes at optimal rates!
Program of construction

- **$k$-NN density rates:** asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]). We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg \sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg \max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers *just* modes at optimal rates!
Program of construction

- **$k$-NN density rates:**
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]).
  We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg \sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg \max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers just modes at optimal rates!
Program of construction

- **k-NN density rates:**
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]). We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg \sup_{x \in \mathbb{R}^d} \hat{f}(x)$. 
  Practical estimator: $\tilde{x} = \arg \max_{x \in \mathcal{X}_{1:n}} \hat{f}(x)$. 
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04] 
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze. 
  Our procedure recovers just modes at optimal rates!
Program of construction

- **$k$-NN density rates:**
  asymptotic $1/\sqrt{k}$ rates (e.g. [Biau, ..., Devroye et.al., 11]).
  We show high-prob. finite sample rates!

- **Single mode:**
  Common estimator in theory: $\hat{x} = \arg \sup_{x \in \mathbb{R}^d} \hat{f}(x)$.
  Practical estimator: $\tilde{x} = \arg \max_{x \in X_{1:n}} \hat{f}(x)$.
  Consistency of $\tilde{x}$ [Abraham, Biau, Cadre, 04]
  We show that $\tilde{x}$ is also minimax-optimal!

- **Multiple modes:**
  Practical procedures (e.g. meanshift) are hard to analyze.
  Our procedure recovers *just* modes at optimal rates!
Outline:

- $k$-NN density rates
- Single mode rates
- Multiple modes rates
**k-NN density estimate:**

Define $r_k(x) \equiv$ distance from $x$ to its $k$th neighbor in $X_{1:n}$.

$$f_k(x) \triangleq \frac{k}{n \cdot \text{vol}(B(x, r_k(x)))} = \frac{k}{n \cdot v_d \cdot r_k(x)^d}.$$
**k-NN density estimate:**

Define $r_k(x) \equiv$ distance from $x$ to its $k$th neighbor in $X_{1:n}$.

$$f_k(x) \triangleq \frac{k}{n \cdot \text{vol}(B(x, r_k(x)))} = \frac{k}{n \cdot v_d \cdot r_k(x)^d}.$$
Devroye, Wagner, 77

Strong consistency.

Moore, Yackel, 76

\[ \sqrt{k} \cdot \frac{(f_k(x) - f(x))}{f(x)} \xrightarrow{D} \mathcal{N}(0,1), \]

provided \( \nabla f < \infty \) on some \( B(x) \), and \( k \to \infty, k/n^{2/(2+d)} \to 0 \).

Similar results by [Biau, Chazal, ... Devroye et. al., 2011]

We seek high-prob. finite sample rates ...
Devroye, Wagner, 77

Strong consistency.

Moore, Yackel, 76

\[ \sqrt{k} \cdot \frac{(f_k(x) - f(x))}{f(x)} \xrightarrow{D} \mathcal{N}(0, 1), \]

provided \( \nabla f < \infty \) on some \( B(x) \), and \( k \to \infty \), \( k/n^{2/(2+d)} \to 0 \).

Similar results by [Biau, Chazal, ... Devroye et. al., 2011]

We seek high-prob. finite sample rates ...
Devroye, Wagner, 77

Strong consistency.

Moore, Yackel, 76

\[ \sqrt{k} \cdot \frac{(f_k(x) - f(x))}{f(x)} \xrightarrow{D} \mathcal{N}(0, 1), \]

provided \( \nabla f < \infty \) on some \( B(x) \), and \( k \to \infty, k/n^{2/(2+d)} \to 0 \).

Similar results by [Biau, Chazal, ... Devroye et. al., 2011]

We seek high-prob. finite sample rates ...
**Devroye, Wagner, 77**

Strong consistency.

**Moore, Yackel, 76**

\[
\sqrt{k} \cdot \frac{(f_k(x) - f(x))}{f(x)} \overset{D}{\to} \mathcal{N}(0, 1),
\]

provided \( \nabla f < \infty \) on some \( B(x) \), and \( k \to \infty, \, k/n^{2/(2+d)} \to 0. \)

Similar results by [Biau, Chazal, ... Devroye et. al., 2011]

We seek high-prob. finite sample rates ...
Express rates generally in terms of mod. of continuity at $x$:

$$\hat{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{||x-x'|| \leq r} f(x') \leq f(x) + \epsilon \right\}$$

$$\check{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{||x-x'|| \leq r} f(x') \geq f(x) - \epsilon \right\}$$

Why not just $r(\epsilon, x)$?
For $x = \arg\max f(x)$, $\hat{r}(\epsilon, x) = \infty$ while $\check{r}(\epsilon, x) < \infty$. 
Express rates generally in terms of mod. of continuity at \( x \):

\[
\hat{r}(\epsilon, x) \triangleq \sup \left \{ r : \sup_{\|x-x'\| \leq r} f(x') \leq f(x) + \epsilon \right \}
\]

\[
\check{r}(\epsilon, x) \triangleq \sup \left \{ r : \sup_{\|x-x'\| \leq r} f(x') \geq f(x) - \epsilon \right \}
\]

Why not just \( r(\epsilon, x) \)?

For \( x = \arg \max f(x) \), \( \hat{r}(\epsilon, x) = \infty \) while \( \check{r}(\epsilon, x) < \infty \).
Express rates generally in terms of mod. of continuity at $x$:

$$\hat{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{\|x-x'\| \leq r} f(x') \leq f(x) + \epsilon \right\}$$

$$\tilde{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{\|x-x'\| \leq r} f(x') \geq f(x) - \epsilon \right\}$$

Why not just $r(\epsilon, x)$?

For $x = \arg \max f(x)$, $\hat{r}(\epsilon, x) = \infty$ while $\tilde{r}(\epsilon, x) < \infty$. 
Express rates generally in terms of mod. of continuity at $x$:

$$\hat{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{\|x-x'\| \leq r} f(x') \leq f(x) + \epsilon \right\}$$

$$\check{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{\|x-x'\| \leq r} f(x') \geq f(x) - \epsilon \right\}$$

Why not just $r(\epsilon, x)$?

For $x = \arg \max f(x)$, $\hat{r}(\epsilon, x) = \infty$ while $\check{r}(\epsilon, x) < \infty$. 
Express rates generally in terms of mod. of continuity at $x$:

\[
\hat{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{\|x-x'\| \leq r} f(x') \leq f(x) + \epsilon \right\}
\]

\[
\tilde{r}(\epsilon, x) \triangleq \sup \left\{ r : \sup_{\|x-x'\| \leq r} f(x') \geq f(x) - \epsilon \right\}
\]

Why not just $r(\epsilon, x)$?

For $x = \arg \max f(x)$, $\hat{r}(\epsilon, x) = \infty$ while $\tilde{r}(\epsilon, x) < \infty$. 

Theorem 1.

W.p > 1 − \(e^{-C}\), simult. \(\forall x \in \text{supp}(f), \forall \epsilon > 0,\)

\[
\left(1 - \frac{C}{\sqrt{k}}\right) (f(x) - \epsilon) \leq f_k(x) \leq \left(1 + \frac{C}{\sqrt{k}}\right) (f(x) + \epsilon),
\]

provided \(\ln n/n \lesssim k/n \lesssim v_d \cdot r(\epsilon, x)^d \cdot (f(x) - \epsilon).\)

\[\therefore\] optimal (local) rates under smoothness conditions.

If \(f\) is \(\alpha\)-Hölder at \(x\), i.e. \(\forall x', |f(x') - f(x)| \leq L \|x - x'\|^{\alpha}\), then

\[
|f_k(x) - f(x)| = O \left(n^{-\alpha/(2\alpha+d)}\right), \quad \text{for } k = \Theta(n^{2\alpha/(2\alpha+d)}).\]
Theorem 1.

W.p $> 1 - e^{-C}$, simult. $\forall x \in \text{supp}(f)$, $\forall \epsilon > 0$,

$$\left(1 - \frac{C}{\sqrt{k}}\right)(f(x) - \epsilon) \leq f_k(x) \leq \left(1 + \frac{C}{\sqrt{k}}\right)(f(x) + \epsilon),$$

provided $\ln n/n \lesssim k/n \lesssim v_d \cdot r(\epsilon, x)^d \cdot (f(x) - \epsilon)$.

$\therefore$ optimal (local) rates under smoothness conditions.

If $f$ is $\alpha$-Hölder at $x$, i.e. $\forall x', |f(x') - f(x)| \leq L \|x - x'\|^\alpha$, then

$$|f_k(x) - f(x)| = O\left(n^{-\alpha/(2\alpha+d)}\right), \quad \text{for } k = \Theta\left(n^{2\alpha/(2\alpha+d)}\right).$$
Proof idea:

\[ f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}. \]

Express \( r_k(x) \) in terms of \( f(x) \):

- For \( r < r(\epsilon, x) \) s.t. \( F(B(x, r)) \approx f(x) \cdot r^d \).
- If \( F(B(x, r)) \approx k/n \) then \( r \approx (k/n \cdot f(x))^{1/d} \).
- W.h.p. \( F_n(B(x, r)) \approx F(B(x, r)) + 1/n \), so \( r \approx r_k(x) \).

Show that \( r \) exists, done!
Proof idea:

\[ f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}. \]

Express \( r_k(x) \) in terms of \( f(x) \):

- For \( r < r(\epsilon, x) \) s.t. \( F(B(x, r)) \approx f(x) \cdot r^d \).
- If \( F(B(x, r)) \approx k/n \) then \( r \approx (k/n \cdot f(x))^{1/d} \).
- W.h.p. \( F_n(B(x, r)) \approx F(B(x, r)) + 1/n \), so \( r \approx r_k(x) \).

Show that \( r \) exists, done!
Proof idea:

\[ f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}. \]

Express \( r_k(x) \) in terms of \( f(x) \):

- For \( r < r(\epsilon, x) \) s.t. \( F(B(x, r)) \approx f(x) \cdot r^d \).
- If \( F(B(x, r)) \approx k/n \) then \( r \approx (k/n \cdot f(x))^{1/d} \).
- W.h.p. \( F_n(B(x, r)) \approx F(B(x, r)) + 1/n \), so \( r \approx r_k(x) \).

Show that \( r \) exists, done!
Proof idea:

$$f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}.$$ 

Express $r_k(x)$ in terms of $f(x)$:

- For $r < r(\epsilon, x)$ s.t. $F(B(x, r)) \approx f(x) \cdot r^d$.
- If $F(B(x, r)) \approx \frac{k}{n}$ then $r \approx (\frac{k}{n} \cdot f(x))^{1/d}$.
- W.h.p. $F_n(B(x, r)) \approx F(B(x, r)) + 1/n$, so $r \approx r_k(x)$.

Show that $r$ exists, done!
Proof idea:

\[ f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}. \]

Express \( r_k(x) \) in terms of \( f(x) \):

- For \( r < r(\epsilon, x) \) s.t. \( F(B(x, r)) \approx f(x) \cdot r^d \).
- If \( F(B(x, r)) \approx k/n \) then \( r \approx (k/n \cdot f(x))^{1/d} \).
- W.h.p. \( F_n(B(x, r)) \approx F(B(x, r)) + 1/n \), so \( r \approx r_k(x) \).

Show that \( r \) exists, done!
Proof idea:

\[ f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}. \]

Express \( r_k(x) \) in terms of \( f(x) \):

- For \( r < r(\epsilon, x) \) s.t. \( F(B(x, r)) \approx f(x) \cdot r^d \).
- If \( F(B(x, r)) \approx k/n \) then \( r \approx (k/n \cdot f(x))^{1/d} \).
- W.h.p. \( F_n(B(x, r)) \approx F(B(x, r)) + 1/n \), so \( r \approx r_k(x) \).

Show that \( r \) exists, done!
Proof idea:

\[ f_k(x) = \frac{k}{n \cdot v_d \cdot r_k(x)^d}. \]

Express \( r_k(x) \) in terms of \( f(x) \):

- For \( r < r(\epsilon, x) \) s.t. \( F(B(x, r)) \approx f(x) \cdot r^d \).
- If \( F(B(x, r)) \approx k/n \) then \( r \approx (k/n \cdot f(x))^{1/d} \).
- W.h.p. \( F_n(B(x, r)) \approx F(B(x, r)) + 1/n \), so \( r \approx r_k(x) \).

Show that \( r \) exists, done!
Outline:

- $k$-NN density rates
- **Single mode rates**
- Multiple modes rates
Most commonly studied

\[ \hat{x} = \arg \sup_{x \in \mathbb{R}^d} f_n(x) \]

Recursive estimates (One sample at the time)

[Devroye 79, Tsybakov, 90 (optimal for Hölder classes.)]

Direct estimates

\[ \tilde{x} = \arg \max_{x \in X_{1:n}} f_k(x) = \arg \min_{x \in X_{1:n}} r_k(x). \]

(Consistency, [Abraham, Biau, Cadre, ])

Most commonly studied
\[ \hat{x} = \arg \sup_{x \in \mathbb{R}^d} f_n(x) \]

Recursive estimates (One sample at the time)
[L. Devroye 79], [S. Tsybakov, 90 (optimal for Hölder classes.)]

Direct estimates
\[ \tilde{x} = \arg \max_{x \in X_{1:n}} f_k(x) = \arg \min_{x \in X_{1:n}} r_k(x). \]
(Consistency, [Abraham, Biau, Cadre, ]
Most commonly studied
\[ \hat{x} = \arg \sup_{x \in \mathbb{R}^d} f_n(x) \]

Recursive estimates (One sample at the time)
[L. Devroye 79], [S. Tsybakov, 90 (optimal for Hölder classes.)]

Direct estimates
\[ \tilde{x} = \arg \max_{x \in X_{1:n}} f_k(x) = \arg \min_{x \in X_{1:n}} r_k(x). \]
(Consistency, [Abraham, Biau, Cadre, ]
A.1 (local): single mode $x = \arg \max f(x), \nabla^2 f(x) < 0$.
A.2 (global): level sets of $f$ have single CC.

**Theorem 2.** Let $\tilde{x} = \arg \max_{x \in X_{1:n}} f_k(x)$. W.h.p. we have

$$\|\tilde{x} - x\| \lesssim k^{-1/4}, \quad \text{provided } \ln n \lesssim k \lesssim n^{4/(4+d)}.$$

Constants depend on $f(x)$ and $\nabla^2 f(x)$. (OPTIMAL, see Tsyb.90)
A.1 (local): single mode $x = \arg\max f(x), \nabla^2 f(x) < 0$.  
A.2 (global): level sets of $f$ have single CC.

**Theorem 2.** Let $\tilde{x} = \arg\max_{x \in X_{1:n}} f_k(x)$. W.h.p. we have

$$
\|\tilde{x} - x\| \lesssim k^{-1/4}, \quad \text{provided } \ln n \lesssim k \lesssim n^{4/(4+d)}.
$$

Constants depend on $f(x)$ and $\nabla^2 f(x)$. (OPTIMAL, see Tsyb.90)
Proof idea:

\[ r_n \triangleq \text{dis}(x, X_{1:n}) \lesssim_{\text{w.h.p.}} n^{-1/d} = o(n^{-1/(4+d)}) \triangleq \bar{r} \]

\[ \nabla^2 f(x) \prec 0 : \exists \text{ a level set } A_x : \]

\[ c \|x - x'\|^2 \leq f(x) - f(x') \leq C \|x - x'\|^2. \]

Theorem 1 allows for different rates near or far from \( x \):

\[ \min_{B(x, r_n(x))} f_k > \max_{X \setminus B(x, \bar{r})} f_k \]
Proof idea:

\[ r_n \triangleq \text{dis}(x, X_{1:n}) \lesssim_{\text{w.h.p.}} n^{-1/d} = o(n^{-1/(4+d)}) \triangleq \bar{r} \]

\[ \nabla^2 f(x) \prec 0 : \exists \text{ a level set } A_x: \]

\[ c \|x - x'\|^2 \leq f(x) - f(x') \leq C \|x - x'\|^2. \]

Theorem 1 allows for different rates near or far from \( x \):

\[ \min_{B(x, r_n(x))} f_k > \max_{x \in X \setminus B(x, \bar{r})} f_k \]
Proof idea:

\[ r_n \triangleq \text{dis}(x, X_{1:n}) \lesssim_{\text{w.h.p.}} n^{-1/d} = o(n^{-1/(4+d)}) \triangleq \bar{r} \]

\[ \nabla^2 f(x) \prec 0 : \exists \text{ a level set } A_x : \]

\[ c \| x - x' \|^2 \leq f(x) - f(x') \leq C \| x - x' \|^2. \]

Theorem 1 allows for different rates near or far from \( x \):

\[ \min_{B(x, r_n(x))} f_k > \max_{\mathcal{X} \setminus B(x, \bar{r})} f_k \]
Proof idea:

\[ r_n \triangleq \text{dis}(x, X_{1:n}) \lesssim_{\text{w.h.p.}} n^{-1/d} = o(n^{-1/(4+d)}) \triangleq \bar{r} \]

\[ \nabla^2 f(x) \prec 0 : \exists \text{ a level set } A_x: \]

\[ c \|x - x'\|^2 \leq f(x) - f(x') \leq C \|x - x'\|^2. \]

Theorem 1 allows for different rates near or far from \( x \):

\[ \min_{B(x, r_n(x))} f_k > \max_{x \setminus B(x, \bar{r})} f_k \]
Outline:

• $k$-NN density rates
• Single mode rates
• **Multiple modes rates**
Setup:

Modes: \( \mathcal{M} \equiv \{ x : \exists r > 0, \forall x' \in B(x, r), f(x') < f(x) \} \).

A.1 (local) \( \forall x \in \mathcal{M}, \nabla^2 f(x) \prec 0 \).

A.2 (global) Any CC of any level set of \( f \) contains a mode in \( \mathcal{M} \).
ALGO: As $f_k$ goes down, pick a new mode as a new *bump* appears.

**Identifying CCs of level sets:**
CCs of subgraphs of a $k$-NN graph [Chau., Das., Kpo., v Lux., 14]

**How to identify false modes in $f_k$?**
Remove all *bumps* of height $\lesssim |f_k - f| \approx 1/\sqrt{k}$. 
ALGO: As $f_k$ goes down, pick a new mode as a new *bump* appears.

**Identifying CCs of level sets:**
CCs of subgraphs of a $k$-NN graph [Chau., Das., Kpo., v Lux., 14]

How to identify false modes in $f_k$?
Remove all *bumps* of height $\lesssim |f_k - f| \approx 1/\sqrt{k}$. 
ALGO: As $f_k$ goes down, pick a new mode as a new bump appears.

Identifying CCs of level sets:
CCs of subgraphs of a $k$-NN graph [Chau., Das., Kpo., v Lux., 14]

How to identify false modes in $f_k$?
Remove all bumps of height $\lesssim |f_k - f| \approx 1/\sqrt{k}$. 
Identifying good modes

$x$ is $r$-salient: separated from other modes by valley of radius $r$.

**Theorem 3.** Suppose $x \in \mathcal{M}$ is $r$-salient. Let $n \geq N(x)$. W.h.p. $\exists \tilde{x} \in \mathcal{M}_n$ s.t.

$$\|\tilde{x} - x\| \lesssim k^{-1/4}, \quad \text{provided } \ln n/r^4 \lesssim k \lesssim n^{4/(4+d)}.$$  

Constants depend on $f(x)$ and $\nabla^2 f(x)$. 
Identifying good modes

$x$ is $r$-salient: separated from other modes by valley of radius $r$.

Theorem 3. Suppose $x \in \mathcal{M}$ is $r$-salient. Let $n \geq N(x)$. W.h.p. $\exists \tilde{x} \in \mathcal{M}_n$ s.t.

$$\|\tilde{x} - x\| \lesssim k^{-1/4}, \quad \text{provided } \ln n/r^4 \lesssim k \lesssim n^{4/(4+d)}.$$  

Constants depend on $f(x)$ and $\nabla^2 f(x)$. 
Identifying good modes

Suppose $x \in \mathcal{M}$ is $r$-salient. Let $n \geq N(x)$. W.h.p. $\exists \tilde{x} \in \mathcal{M}_n$ s.t.

$$\|\tilde{x} - x\| \lesssim k^{-1/4}, \quad \text{provided } \ln n/r^4 \lesssim k \lesssim n^{4/(4+d)}.$$

Constants depend on $f(x)$ and $\nabla^2 f(x)$. 

$x$ is $r$-salient: separated from other modes by valley of radius $r$. 

**Theorem 3.**
Pruning bad modes

**Theorem 4.** Suppose $f$ is Lipschitz. Assume $k \geq \ln n$. Let $\lambda_0 = \Theta(\ln n/k)$. All modes in $\mathcal{M}_n$ at $f_k$-level $\lambda > \lambda_0$ can be assigned to *distinct* modes in $\mathcal{M}$ at $f$-level $\approx \lambda_0$. 
Pruning bad modes

Theorem 4. Suppose $f$ is Lipschitz. Assume $k \geq \ln n$. Let $\lambda_0 = \Theta(\ln n/k)$. All modes in $\mathcal{M}_n$ at $f_k$-level $\lambda > \lambda_0$ can be assigned to distinct modes in $\mathcal{M}$ at $f$-level $\approx \lambda_0$.

TRUTH: 5-modes mixture $\sum_{i=1}^{5} 0.2N(2\sqrt{d}e_i, I_d)$
Merci!