Supplement to "Local clustering of density upper level sets"

Anonymous Authors[1]

In this supplement, we present proofs for "Local Clustering of Density Upper Level Sets". We begin by providing technical lemmas, before moving on to proving the main results of the paper.

1. Technical Lemmas

For $A \subset \mathcal{X}$, let $P(A) = \mathbb{P}_{X \sim P}(X \in A)$. To simplify expressions, we will write $A_{\sigma,\sigma+r} := \{x : 0 < \rho(x, A_\sigma) \leq r\}$. We further let $\widetilde{\mathcal{E}} = |E(A_\sigma[\mathbf{X}], \mathbf{X} \setminus A_\sigma[\mathbf{X}]; G_{n,r})|$ be the number of edges between $A_\sigma[\mathbf{X}]$ and $\mathbf{X} \setminus A_\sigma[\mathbf{X}]$ in the graph $G_{n,r}$; $\widetilde{\mu} = \mathbb{E}\left[\widetilde{\mathcal{E}}\right]$ be the expected number of such edges; and $\widetilde{p} = \widetilde{\mu}/\binom{n}{2}$ the probability of any two vertices x_i and x_j having such an edge. Similarly, $\mathcal{V} = \mathrm{vol}(A_\sigma[\mathbf{X}]; G_{n,r})$ is the volume of $A_\sigma[\mathbf{X}]$; $\mu = \mathbb{E}[\mathcal{V}]$ is the expected volume; and $p = \mu/\binom{n}{2}$. Finally, we denote $rB = B(0,r)$.

1.1. Expected Values

Lemma 1. *Under the setup and conditions of Theorem 1, and for any $r < \sigma$,*

$$P(A_{\sigma,\sigma+r}) \leq 2^{d-1}\nu(A_\sigma)\frac{rd}{\sigma}\left(\tau_\sigma - \frac{r^\gamma}{\gamma+1}\right)$$

Proof. Recalling that f is the density function for P, we have

$$P(A_{\sigma,\sigma+r}) = \int_{A_{\sigma,\sigma+r}} f(x)dx \tag{1}$$

Now, for $0 = t_0 < t_1 < \ldots < t_k = 1$, we divide up $A_{\sigma,\sigma+r} = \bigcup_{i=0}^{k-1} \mathcal{T}_i$ where $\mathcal{T}_i = \{x : rt_i < \rho(x, A_\sigma) \leq rt_{i+1}\}$. We can rewrite the right hand side of (1) as

$$\int_{A_{\sigma,\sigma+r}} f(x)dx = \sum_{i=0}^{k-1}\int_{\mathcal{T}_i} f(x)dx \leq \sum_{i=0}^{k-1}\nu(\mathcal{T}_i)\max_{x\in\mathcal{T}_i} f(x).$$

By definition,

$$\nu(\mathcal{T}_i) = \nu(A_\sigma + rt_{i+1}B) - \nu(A_\sigma + rt_iB).$$

Moreover, by (A1) and (A2) we have

$$\max_{x\in\mathcal{T}_i} f(x) \leq \tau_\sigma - (rt_i)^\gamma.$$

since for all $x \in \mathcal{T}_i$, $\rho(x, A_\sigma) > rt_i$. Therefore

$$\sum_{i=0}^{k-1}\int_{\mathcal{T}_i} f(x)dx \leq \sum_{i=0}^{k-1}\Big\{\nu(A_\sigma + rt_{i+1}B) - \nu(A_\sigma + rt_iB)\Big\}\left(\tau_\sigma - (rt_i)^\gamma\right). \tag{2}$$

Now, we have that $\sigma B \subset A_\sigma$ which implies,

$$\nu(A_\sigma + rt_iB) \leq \nu(A_\sigma + \frac{rt_i}{\sigma}A_\sigma)$$

and we therefore have the upper bound

$$\begin{aligned}
&\sum_{i=0}^{k-1}\Big\{\nu(A_\sigma + rt_{i+1}B) - \nu(A_\sigma + rt_iB)\Big\}(\tau_\sigma - (rt_i)^\gamma) \\
&\leq \sum_{i=0}^{k-1}\Big\{\nu(A_\sigma + \frac{rt_{i+1}}{\sigma}A_\sigma) - \nu(A_\sigma + \frac{rt_i}{\sigma}A_\sigma)\Big\}(\tau_\sigma - (rt_i)^\gamma) \\
&= \nu(A_\sigma)\sum_{i=0}^{k-1}\Big\{(1 + \frac{rt_{i+1}}{\sigma})^d - (1 + \frac{rt_i}{\sigma})^d\Big\}(\tau_\sigma - (rt_i)^\gamma)
\end{aligned} \tag{3}$$

where the upper bound holds because $\tau_\sigma - (rt)^\gamma$ is decreasing in t.

Let $t_i = i/k$ for $i = 0, \ldots, k$. Taking the limit as $k \to \infty$, we have

$$\begin{aligned}
&\lim_{k\to\infty}\sum_{i=0}^{k-1}\left\{\left(1 + \frac{r(i+1)}{k\sigma}\right)^d - \left(1 + \frac{ri}{k\sigma}\right)^d\right\}\left(\tau_\sigma - \left(\frac{ri}{k}\right)^\gamma\right) \\
&= \int_0^1 \frac{rd}{\sigma}(1 + \frac{rt}{\sigma})^{d-1}(\tau_\sigma - (rt)^\gamma)dt \\
&\leq 2^{d-1}\frac{rd}{\sigma}\left(\tau_\sigma - \frac{r^{\gamma+1}}{\gamma+1}\right)
\end{aligned}$$

where the inequality comes from $t \leq 1$ and $r < \sigma$.

[1] Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Finally, note that (3) holds for any k and arbitrary $0 = t_0 < t_1 < \ldots < t_k = 1$. In particular, it holds for $t_i = i/k$ for $i = 0, \ldots, k$, and in the limit as $k \to \infty$. Therefore, we have

$$P(A_{\sigma,\sigma+r}) \leq \nu(A_\sigma) 2^{d-1} \frac{rd}{\sigma}\left(\tau_\sigma - \frac{r^\gamma}{\gamma+1}\right)$$

which is exactly the stated result of Lemma 1. □

Lemma 2. *Under the setup and conditions of Theorem 1, and for any $r < \sigma$,*

$$\widetilde{p} \leq \frac{2^d d}{\sigma} \nu(A_\sigma)\nu_d r^{d+1}\tau\left(\tau_\sigma - \frac{r^{\gamma+1}}{\gamma+1}\right)$$

Proof. We can write $\widetilde{\mathcal{E}}$ as the sum of indicator functions,

$$\widetilde{\mathcal{E}} = \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}(x_i \in A_{\sigma,\sigma+r})\mathbf{1}(x_j \in B(x_i, r) \cap A_\sigma) \quad (4)$$

Ignoring the cross terms (which are zero), normalizing by $1/\binom{n}{2}$, and taking expectation, we have

$$\begin{aligned}
\frac{\widetilde{\mu}}{\binom{n}{2}} &= 2\int_{A_{\sigma,\sigma+r}} f(x)\left\{\int_{B(x,r)\cap A_\sigma} f(x')dx'\right\} dx \\
&\overset{(i)}{\leq} 2\int_{A_{\sigma,\sigma+r}} f(x)\nu_d r^d \tau dx \\
&\overset{(ii)}{\leq} 2^d \nu_d r^d \tau \nu(A_\sigma)\frac{rd}{\sigma}\left(\tau_\sigma - \frac{r^{\gamma+1}}{\gamma+1}\right)
\end{aligned}$$

where (i) follows from Assumption (A3), which implies $f(x') \leq \tau$ for all $x' \in A_\sigma \setminus A$, and (ii) follows from Lemma 1. □

Lemma 3. *Under the setup and conditions of Theorem 1,*

$$p \geq 2\tau_\sigma^2 \nu(A_\sigma)\nu_d \left(\frac{r}{2}\right)^d$$

Proof. We can write $\mathcal{V}$ as the sum of indicator functions,

$$\mathcal{V} = \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}(x_i \in A_\sigma)\mathbf{1}(x_j \in B(x_i, r)) \quad (5)$$

Ignoring the cross terms (which are zero), normalizing by $1/\binom{n}{2}$, and taking expectation, we have

$$\frac{\mu}{\binom{n}{2}} = 2\int_{A_\sigma} f(x)\left\{\int_{B(x,r)} f(x')dx'\right\} dx \quad (6)$$

For $x \in A_\sigma$, take $x_0 \in A$ such that $\|x - x_0\| = \rho(x, A)$ (note that such a minimizer exists because A is closed). Then, by the triangle inequality, we have

$$B\left(\frac{x + x_0}{2}, \frac{r}{2}\right) \in A_\sigma \cap B(x, r)$$

Recall that by ((A1)), we have $f(x') \geq \tau_\sigma$ for all $x' \in A_\sigma$. We can therefore lower bound the right hand side of (6) by

$$\begin{aligned}
&2\int_{A_\sigma} f(x)\tau_\sigma \nu_d\left(\frac{r}{2}\right)^d dx \\
&\leq 2\tau_\sigma^2 \nu(A_\sigma)\nu_d\left(\frac{r}{2}\right)^d.
\end{aligned}$$

□

1.2. Concentration inequalities.

Given a symmetric kernel function $k : \mathcal{X}^m \to \mathbb{R}$, and data $\{x_1, \ldots, x_n\}$, we define the *order-m U statistic* to be

$$U := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \ldots < i_m \leq n} k(x_{i_1}, \ldots, x_{i_m})$$

For both Lemmas 4 and 5, let $X_1, \ldots, X_n \in \mathcal{X}$ be independent and identically distributed. We will additionally assume the order-m kernel function k satisfies the boundedness property $\sup_{x_1,\ldots,x_m} |k(x_1, \ldots, x_m)| \leq 1$.

Lemma 4 (Hoeffding's inequality for U-statistics.)**.** *For any $t > 0$,*

$$\mathbb{P}(|U - \mathbb{E}U| \geq t) \leq 2\exp\left\{-\frac{2nt^2}{m}\right\}$$

Further, for any $\delta > 0$, we have

$$\begin{aligned}
U &\leq \mathbb{E}U + \sqrt{\frac{m\log(1/\delta)}{2n}}, \\
U &\geq \mathbb{E}U - \sqrt{\frac{m\log(1/\delta)}{2n}}
\end{aligned}$$

each with probability at least $1 - \delta$.

Lemma 5 (Bernstein's inequality for U-statistics)**.** *Additionally, assume $\sigma^2 = \mathrm{Var}(k(X_1, \ldots, X_m)) < \infty$. Then for any $\delta > 0$,*

$$\mathbb{P}(U - \mathbb{E}U \geq t) \leq \exp\left\{-\frac{n}{2m}\frac{t^2}{\sigma^2 + t/3}\right\},$$

Moreover if $\sigma^2 \leq \mu/n$,

$$\begin{aligned}
U &\leq \mathbb{E}U \cdot \left(1 + \max\left\{\sqrt{\frac{2m\log(1/\Delta)}{\mu}}, \frac{2m\log(1/\Delta)}{3\mu}\right\}\right), \\
U &\geq \mathbb{E}U \cdot \left(1 - \max\left\{\sqrt{\frac{2m\log(1/\Delta)}{\mu}}, \frac{2m\log(1/\Delta)}{3\mu}\right\}\right)
\end{aligned}$$

each with probability at least $1 - \Delta$.

2. Proof of Theorem 1

Given the previous lemmas, the proof of Theorem 1 is straightforward. We rely on Lemmas 2 and 3 to bound $\widetilde{\mu}$ and μ, respectively, and Lemma 5 to bound the deviations $\widetilde{\mathcal{E}} - \widetilde{\mu}$ and $\mathcal{V} - \mu$ with high probability.

2.1. Numerator of $\Phi_{n,r}(A_\sigma[\mathbf{X}])$.

From (4), we can see that $\widetilde{\mathcal{E}}$, properly scaled, can be expressed as an order-2 U-statistic,

$$\frac{1}{\binom{n}{2}}\widetilde{\mathcal{E}} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \widetilde{k}(x_i, x_j)$$

where

$$\begin{aligned}\tilde{k}(x_i, x_j) =& \mathbf{1}(x_i \in A_{\sigma,\sigma+r})\mathbf{1}(x_j \in B(x_i,r) \cap A_\sigma) + \\ & \mathbf{1}(x_j \in A_{\sigma,\sigma+r})\mathbf{1}(x_i \in B(x_j,r) \cap A_\sigma)\end{aligned}$$

From Lemma 4 we therefore have

$$\frac{\widetilde{\mathcal{E}}}{\binom{n}{2}} \le \widetilde{p} + \sqrt{\frac{\log(1/\delta)}{n}} \tag{7}$$

with probability at least $1 - \delta$.

Multiplicative bound: As $\tilde{k}(x_1, x_2)$ is the sum of two Bernoulli random variables with negative covariance (since $\mathbf{1}(x_i \in A_{\sigma,\sigma+r})\mathbf{1}(x_j \in B(x_i,r) \cap A_\sigma) = 1$ implies $\mathbf{1}(x_j \in A_{\sigma,\sigma+r})\mathbf{1}(x_i \in B(x_j,r) \cap A_\sigma) = 0$ and vice versa), we can upper bound Var $\left(\tilde{k}(x_1,x_2)\right) \le \widetilde{p}$, where we recall

$$\widetilde{p} = 2 \cdot \mathbb{P}\left(\mathbf{1}(x_1 \in A_{\sigma,\sigma+r})\mathbf{1}(x_2 \in B(x_1,r) \cap A_\sigma)\right)$$

From Lemma 5, we therefore have

$$\frac{\widetilde{\mathcal{E}}}{\binom{n}{2}} \le \widetilde{p} + \max\left\{\sqrt{\frac{4\log(1/\Delta)\widetilde{p}}{n}}, \frac{4\log(1/\Delta)}{3n}\right\}$$

with probability at least $1 - \Delta$.

Denominator of $\Phi_{n,r}(A_\sigma[\mathbf{X}])$. We follow a very similar set of steps as above.

By (4), we see that $\mathcal{V}$ can also be expressed as an order-2 U-statistic,

$$\frac{\mathcal{V}}{\binom{n}{2}} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} k'(x_i, x_j)$$

with

$$\begin{aligned}k'(x_i, x_j) =& \mathbf{1}(x_i \in A_\sigma)\mathbf{1}(x_j \in B(x_i,r)) + \\ & \mathbf{1}(x_j \in A_\sigma)\mathbf{1}(x_i \in B(x_j,r))\end{aligned}$$

From Lemma 4 we therefore have

$$\frac{\mathcal{V}}{\binom{n}{2}} \ge p - \sqrt{\frac{\log(1/\delta)}{n}} \tag{8}$$

with probability at least $1 - \delta$.

Multiplicative bound: The two terms on the right hand side are both distributed $\text{Bernoulli}(p/2)$. Moreover, since $\mathbf{1}(x_i \in A_\sigma) = 1$ implies $\mathbf{1}(x_j \in A_\sigma) = 0$, they have negative covariance. We can therefore upper bound $\text{Var}(k'(x_i,x_j)) \le p$, and so from Lemma 5, we have

$$\frac{\mathcal{V}}{\binom{n}{2}} \ge p - \max\left\{\sqrt{\frac{4\log(1/\Delta)p}{n}}, \frac{4\log(1/\Delta)}{3n}\right\}$$

with probability at least $1 - \Delta$.

Proof of the additive error bound. Noting that $\Phi_{n,r}(A_\sigma[\mathbf{X}]) = \widetilde{\mathcal{E}}/\mathcal{V}$, and multiplying and dividing by $\binom{n}{2}$, we have

$$\Phi_{n,r}(A_\sigma[\mathbf{X}]) = \frac{\widetilde{p} + \left(\frac{\widetilde{\mathcal{E}}}{\binom{n}{2}} - \widetilde{p}\right)}{p + \left(\frac{\widetilde{\mathcal{V}}}{\binom{n}{2}} - p\right)} \tag{9}$$

We assume (7) and (8) hold, keeping in mind that this will happen with probability at least $1 - 2\delta$. Along with (9) this means

$$\Phi_{n,r}(A_\sigma[\mathbf{X}]) \le \frac{\widetilde{p} + \text{Err}_n}{p - \text{Err}_n}$$

for $\text{Err}_n = \sqrt{\frac{\log(1/\delta)}{n}}$. Now, some straightforward algebraic manipulations yield

$$\begin{aligned}\frac{\widetilde{p} + \text{Err}_n}{p - \text{Err}_n} &= \frac{\widetilde{p}}{p} + \left(\frac{\widetilde{p}}{p - \text{Err}_n} - \frac{\widetilde{p}}{p}\right) + \frac{\text{Err}_n}{p - \text{Err}_n} \\ &= \frac{\widetilde{p}}{p} + \frac{\text{Err}_n}{p - \text{Err}_n}\left(\frac{\widetilde{p}}{p} + 1\right) \\ &\le \frac{\widetilde{p}}{p} + 2\frac{\text{Err}_n}{p - \text{Err}_n}.\end{aligned}$$

Finally, combining the upper bound given by Lemma 3 with the lower bound on n specified in the statement of Theorem 1, we have

$$2\frac{\text{Err}_n}{p - \text{Err}_n} \le \epsilon$$

By Lemmas 2 and Lemma 3, we have

$$\frac{\widetilde{p}}{p} \le C_\sigma \frac{\tau}{\tau_\sigma} \frac{\left(\tau_\sigma - \frac{r^{\gamma+1}}{\gamma+1}\right)}{\tau_\sigma}$$

and thus we have shown (4) occurs with probability at least $1 - 2\delta$. Plugging in $\delta' = \delta/2$ gives the exact statement in Theorem 1.