Bernstein's Inequality

Last week, we finished the analysis on the M.S.E(Mean Square Error) bias of Graph-Laplacian estimator. However, when it comes to variance, we face some technical difficulties. In the end, Prof. Wu used Bernstein's Inequality to pass over the barrier and finally got control of the variance of Graph Laplacian estimator.

Bernstein's inequality gives a powerful upper bound when given a set of identical independent distributed random variables X_i with an upper bound $|X_i| \leq a$. The inequality shows that the probability of average mean out of $E(X_i) \pm \alpha$ will decreases as the number of X_i grows.Without loss of generality, suppose the expected value $E(X_i) = 0$ and variance $Var(X_i) = \sigma^2$.The Bernstein's inequality says:

$$\Pr\{\frac{1}{n}\sum_{i=1}^{n} X_i > \alpha\} \leq exp(-\frac{n\sigma^2}{2\alpha^2 + \frac{2}{3}a\alpha})$$

Proof. On the left hand side. Since the exponential function monotonically increases, we have

$$LHS = \Pr\{exp(s\sum_{i=1}^{n} X_i) > exp(sn\alpha)\} \tag{1}$$

Using Markov's inequality,

$$LHS \leq exp[-sn\alpha)\prod_{i=1}^{n} E(exp(sX_i)] \tag{2}$$

Also,

$$\begin{aligned} E(exp(sX_i)) &= E(1 + sX_i + \sum_{m=2}^{\infty}\frac{s^m X_i^m}{m!}) \\ &\leq 1 + \sum_{m=2}^{\infty}\frac{s^m a^{m-2}\sigma^2}{m!} \\ &= 1 + \frac{\sigma^2}{a^2}\sum_{m=2}^{\infty}\frac{(sa)^m}{m!} \\ &= 1 + \frac{\sigma^2}{a^2}(e^{sa} - 1 - sa) \\ &\leq exp[\frac{\sigma^2}{a^2}(e^{sa} - 1 - sa)] \end{aligned}$$

Plug this inequality into(1)

$$LHS \leq exp[-sn\alpha + \frac{n\sigma^2}{a^2}(e^{sa} - 1 - sa)] \tag{3}$$

Let $h(s) = sn\alpha - \frac{n\sigma^2}{a^2}(e^{sa} - 1 - sa)$ we get

$$LHS \leq e^{-h(s)} \tag{4}$$

Finding where $h(s)$ has local maximum

$$\begin{aligned}
h(s) &= sn\alpha - \frac{n\sigma^2}{a^2}(e^{sa} - 1 - sa) \\
h^{'}(s) &= n\alpha - \frac{n\sigma^2}{a}e^{sa} + \frac{n\sigma^2}{a} = 0 \\
e^{sa} &= \frac{a\alpha}{\sigma^2} + 1 \\
s &= \frac{1}{a}\ln(1 + \frac{a\alpha}{\sigma^2})
\end{aligned}$$

Put s into $h(s)$ to get the local maximum,

$$\begin{aligned}
h(s) &= \frac{n\alpha}{a}\ln(1 + \frac{a\alpha}{\sigma^2}) - \frac{n\sigma^2}{a^2}[1 + \frac{a\alpha}{\sigma^2} - 1 - \ln(1 + \frac{a\alpha}{\sigma^2})] \\
&= (\frac{n\alpha}{a} + \frac{n\sigma^2}{a^2})ln(1 + \frac{a\alpha}{\sigma^2}) - \frac{n\alpha}{\sigma^2}
\end{aligned}$$

Let $t = 1 + \frac{\sigma^2}{a\alpha}$ yields,

$$\begin{aligned}
-h(s) &= \frac{n\alpha}{a}[1 - (1 + t)\ln(1 + \frac{1}{t})] \\
&\leq \frac{n\alpha}{a}\frac{-1}{2t + \frac{2}{3}} \\
&= \frac{-n\alpha^2}{2\sigma^2 + \frac{2a\alpha}{3}}
\end{aligned}$$

Finally we get,

$$\begin{aligned}
LHS &\leq e^{-h(s)} \\
&\leq exp(\frac{-n\alpha^2}{2\sigma^2 + \frac{2a\alpha}{3}})
\end{aligned}$$

□

Graph Laplacian Estimator - variance error

Now we have some basic knowledge about Bernstein's Inequality, it's time to look at the variance of Graph Laplacian estimator.

Theorem. (variance)

$$|(\frac{L_{\epsilon,0}\vec{f}}{\epsilon})_i - (1 - T_{\epsilon,0})f(x_i)| \leq O(\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{\frac{1}{4}+\frac{d}{2}}})$$

with probability $1 - \frac{1}{n}$

Before starting the detail proof, let's recall some basic relation and properties:

$$(\frac{L_{\epsilon,0}\vec{f}}{\epsilon})_i =: \sum \frac{F_k}{G_k}$$
$$F_j = \frac{1}{\sqrt{\epsilon}} e^{-\|x_i - x_j\|^2/\epsilon}[f(x_i) - f(x_j)]$$
$$G_j = \frac{1}{\sqrt{\epsilon}} e^{-\|x_i - x_j\|^2/\epsilon}$$

The expectation value of F_j

$$\begin{aligned} E(F_j) &= \int_0^1 \frac{1}{\sqrt{\epsilon}} F_y p(y) dy \\ &= \int_0^1 \frac{1}{\sqrt{\epsilon}} e^{-\|x_i - x_y\|^2/\epsilon}[f(x_i) - f(x_y)]p(y)dy \\ &= \epsilon\{[(f(x_i) - f(y)]p(y))^{''}|_{y=x_i}\} + O(\epsilon^2) \end{aligned}$$

The expectation value of G_j

$$E(G_j) = p(x_i) + O(\epsilon)$$

The expectation value of F_j^2

$$\begin{aligned} E(F_j^2) &= \int_0^1 \frac{1}{\epsilon} e^{-\|x_i - x_y\|^2/\epsilon}[f(x_i) - f(y)]^2 p(y)dy \\ &= \frac{1}{\sqrt{\epsilon}} \int_0^1 \frac{1}{\sqrt{\epsilon}} e^{-\|x_i - x_y\|^2/\epsilon}[f(x_i) - f(y)]^2 p(y)dy \\ &= c_2\sqrt{\epsilon}[f(x_i) - f(y)]^2 p(y)^{''}|_{y=x_i} + O(\sqrt{\epsilon}) \end{aligned}$$

The variance of F_j

$$var(F_j) = d_2\sqrt{\epsilon}[(f(x_i) - f(y)]p(y))^{''}|_{y=x_i} + O(\epsilon^{\frac{3}{2}})$$

The expectation value of G_j^2

$$E(G_j^2) = \frac{1}{\sqrt{\epsilon}} P(x_i) + O(\sqrt{\epsilon})$$

The expectation of $F_j G_j$

$$E(F_j G_j) = d_2 \sqrt{\epsilon}[(f(x_i) - f(y)]p(y))^{''}|_{y=x_i} + O(\epsilon^{\frac{3}{2}})$$

We can get the covariance of F_j and G_j by the above calculations,

$$\begin{aligned} cov(F_j, G_j) &= E(F_j G_j) - E(F_j) E(G_j) \\ &= d_2 \sqrt{\epsilon}[(f(x_i) - f(y)]p(y))^{''}|_{y=x_i} + O(\epsilon^{\frac{3}{2}}) \end{aligned}$$

By scaling the covariance we get correlation,

$$corr(F_j, G_j) = \sqrt{\epsilon}$$

Go back to look at $(L_{\epsilon,0}\vec{f})_i \rightarrow \frac{E(F_j)}{E(G_j)}$

$$\begin{aligned} (L_{\epsilon,0}\vec{f})_i &= \frac{\frac{1}{n-1}\sum_{j\neq i} F_j}{\frac{1}{n-1}\sum_{j\neq i} G_j} && \text{(by Bernstein's Inequality)} \\ &\approx \frac{E(F) + \xi^F}{E(G) + \xi^G} \end{aligned}$$

Let's check whether F_j and G_j satisfy the condition of Bernstein's Inequality. Since the only difference between them is a ratio $[f(x_i) - f(y)]$, the behaviour of F_j and G_j will be similar. First, check if F_j bounded.

$$|F_j| = |\frac{1}{\sqrt{\epsilon}} e^{-\|x_i - y\|^2/\epsilon}| \leq \frac{1}{\sqrt{\epsilon}}$$

Then calculate the variance of F_j

$$\sigma^2 = var(F_j) = d_2 \sqrt{\epsilon}[(f(x_i) - f(y)]p(y))^{''}|_{y=x_i} + O(\epsilon^{\frac{3}{2}})$$

By Bernstein's Inequality we know,

$$Pr\{\frac{1}{n-1}\sum_{j\neq i}^{n} F_j - E(F_j) > \alpha\} \leq e^{-\frac{n\alpha^2}{2\sigma^2 + \frac{2}{3}\alpha}}$$

Assume $\alpha << \epsilon$, $\alpha = \theta(\epsilon)$, estimate the factor in the exponential

$$\frac{n\alpha^2}{2\sigma^2 + \frac{2}{3}c\alpha} = \frac{n\alpha^2}{O(\sqrt{\epsilon} + \theta(\frac{1}{\sqrt{\epsilon}}\epsilon)} = \frac{n\alpha^2}{O(\sqrt{\epsilon})}$$

In order to let the above factor grows large as $n \to \infty$, which makes the exponential term become small. Take α in the following way,

$$\frac{n\alpha^2}{\sqrt{\epsilon}} = \log n$$
$$\Rightarrow \frac{n\alpha^2}{O(\sqrt{\epsilon})} = \frac{\log n}{O(\epsilon)}$$
$$\Rightarrow P(n, \alpha) \to 0$$
$$\alpha = \frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{-1/4}}$$

Put this α into Bernstein's Inequality yields,

$$|\frac{1}{n-1}\sum F_j - E(F_j)| \leq \frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{-1/4}}$$
$$w.p. \quad 1 - \frac{1}{n}$$

Using this inequality to estimate Graph Laplacian,

$$(L_{\epsilon,0}\vec{f})_i = \frac{E(F_j) + \xi^F}{E(G_j) + \xi^G} = \frac{E(F)}{E(G)} + O(\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{-1/4}})$$
$$(\frac{L_{\epsilon,0}}{\epsilon} f)_i \xrightarrow[a.s.]{n\to\infty} \Delta f(x_i) + \frac{2 \bigtriangledown f(x_i) \bigtriangledown p(x_i)}{p(x_i)}$$

Consider the M.S.E(Mean Square Error) of Graph Laplacian,

$$|(\frac{L_{\epsilon,0}}{\epsilon} f)_i - (\Delta f_i + \frac{2 \bigtriangledown f(x_i) \bigtriangledown p(x_i)}{p(x_i)}|)$$
$$= O(\epsilon) + O(\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{3/4}})$$

where $O(\epsilon)$ controls bias, $O(\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{3/4}})$ controls variance. Want the error level of variance is the same as bias,

$$\epsilon = \frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{3/4}} \Rightarrow \epsilon^{1/2+3} = \frac{\log n}{n}$$
$$\Rightarrow \epsilon = (\frac{\log n}{n})^{\frac{1}{1/2+3}}$$

To sum up, in the previous chapter we've proved that the bias in the M.S.E of Graph Laplacian is controlled by ϵ. Here we use Berstein's Inequality to help us passing $\frac{\sum F_j}{\sum E(G_j)}$ to $\frac{E(F_j)}{G_j}$ with error controlled by α which is also bounded by ϵ. Thanks to this transformation we then got an estimation on variance $O(\frac{\sqrt{\log n}}{\sqrt{n}\epsilon^{3/4}})$. Finally letting this estimation controlled by ϵ we found out that ϵ will decreases to 0 as $n \to \infty$. As a result, M.S.E of Graph Laplacian can be controlled by ϵ. Graph Laplacian is an asymptotic estimator of Laplacian operator.