Theorem 3

Suppose we have sub-gauss errors.

Suppose

$$\int_0^R H^{1/2}(u, \mathcal{G}(T), \|\cdot\|_V) du \le \psi_T(R)$$

Suppose

$$\frac{\psi_T(a+u)}{u^2}$$

are nonincreasing wrt to u for all $u > 0, a > 0$.

Then there is some constant C (only dependent on the characteristics of the sub-gassian errors) such that for all δ that satisfy

$$\sqrt{n_V}\delta^2 \ge 2C\left[\psi_T\left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2\delta\right) \vee \left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2\delta\right)\right]$$

we have with high probability that

$$\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V \le \|\hat{g}_{\tilde{\lambda}} - g^*\|_V + \delta$$

Proof for Theorem 3

The basic inequality gives us

$$\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V^2 \le \|\hat{g}_{\tilde{\lambda}}(\cdot|T) - g^*\|_V^2 + 2\langle\epsilon, \hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\rangle_V$$

Note that if $\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \le \left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V$, then

$$\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V\right)^2 \quad\le\quad \left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V^2 - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V^2$$

By a peeling argument, we have

$$\begin{aligned}
& Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \|\hat{g}_{\tilde{\lambda}} - g^*\|_V \ge \delta\right) \\
= & \sum_{s=0}^{\infty} Pr\left(2^s\delta \le \left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \le 2^{s+1}\delta\right) \\
\le & \sum_{s=0}^{\infty} Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \ge 2^s\delta \wedge \left\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right) \\
= & \sum_{s=0}^{\infty} Pr\left(\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V\right)^2 \ge 2^{2s}\delta^2 \wedge \left\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right) \\
\le & \sum_{s=0}^{\infty} Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V^2 - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V^2 \ge 2^{2s}\delta^2 \wedge \left\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right) \\
\le & \sum_{s=0}^{\infty} Pr\left(\sup_{\left\|\hat{g}_{\lambda} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta} \langle\epsilon, \hat{g}_{\tilde{\lambda}} - \hat{g}_{\lambda}\rangle_V \ge 2^{2s-1}\delta^2\right)
\end{aligned}$$

To apply the lemma based on vandegeer corollary 8.3 (see below), we must check all the conditions are satisfied.

We have chosen that δ such that

$$\begin{aligned}
\frac{\sqrt{n_V}}{8} &\ge \frac{C}{4\delta^2}\left[\psi_T\left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2\delta\right) \vee \left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2\delta\right)\right] \\
&\ge \frac{C}{2^{2s+2}\delta^2}\left[\psi_T\left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right) \vee \left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2\delta\right)\right]
\end{aligned}$$

where the second line follows from the assumption that $\psi_T(a+u)/u^2$ is nonincreasing wrt u. Hence we have satisfied the condition in corollary 8.3 that

$$\sqrt{n_V}2^{2s-1}\delta^2 \geq C\left[\psi_T\left(2\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V+2^{s+1}\delta\right)\vee\left(2\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V+2^{s+1}\delta\right)\right]$$

Hence we have

$$\begin{aligned}
& Pr\left(\left\|\hat{g}_{\hat{\lambda}}-g^*\right\|_V-\|\hat{g}_{\tilde{\lambda}}-g^*\|_V\geq\delta\wedge\|\epsilon\|_V\leq 2\sigma\wedge\|\epsilon\|_T\leq 2\sigma\right)\\
\leq\ & C\sum_{s=0}^{\infty}\exp\left(-n_V\frac{2^{4s-2}\delta^4}{4C^2\left(2\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V+2^{s+1}\delta\right)^2}\right)\\
\leq\ & C\sum_{s=0}^{\infty}\exp\left(-n_V\frac{2^{4s-2}\delta^4}{64C^2\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V^2}\right)\vee\exp\left(-n_V\frac{2^{2s}\delta^2}{196C^2}\right)\\
\leq\ & c\exp\left(-\frac{n_V\delta^4}{c^2\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V^2}\right)+c\exp\left(-\frac{n_V\delta^2}{c^2}\right)
\end{aligned}$$

for some constant c.

By Bernstein's inequality, we also know

$$Pr\left(\|\epsilon\|_T\geq 2\sigma\right)\leq\exp\left(-n_T\frac{\sigma^2}{K}\right)$$

and similarly for $Pr\left(\|\epsilon\|_V\geq 2\sigma\right)$.

Hence we have found for the given δ choice, we have

$$\begin{aligned}
& Pr\left(\left\|\hat{g}_{\hat{\lambda}}-g^*\right\|_V-\|\hat{g}_{\tilde{\lambda}}-g^*\|_V\geq\delta\right)\\
\leq\ & c\exp\left(-\frac{n_V\delta^4}{c^2\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V^2}\right)+c\exp\left(-\frac{n_V\delta^2}{c^2}\right)+\exp\left(-n_T\frac{\sigma^2}{K}\right)+\exp\left(-n_V\frac{\sigma^2}{K}\right)
\end{aligned}$$

Theorem 1 (Corollary to Theorem 3)

Let $\Lambda=[n^{-t_{min}},n^{t_{max}}]^J$.

Suppose that if $\|\epsilon\|_T\leq 2\sigma$, there are constants C,κ such that for any $u>0$, we have for all $\lambda\in\Lambda$

$$\|\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2\|\leq Cn^{\kappa}u\implies\|\hat{g}_{\boldsymbol{\lambda}_1}-\hat{g}_{\boldsymbol{\lambda}_2}\|_V\leq u$$

Then there are constants c,c_1,c_2 s.t. with high probability,

$$\left\|\hat{g}_{\hat{\lambda}}-g^*\right\|_V\leq\|\hat{g}_{\tilde{\lambda}}-g^*\|_V+\frac{c_1\left(J(\log n_V+c_2)\right)^{1/2}}{\sqrt{n_V}}+\sqrt{c\left(J(\log n_V+c_2)\right)^{1/2}\left\|\hat{g}_{\tilde{\lambda}}-g^*\right\|_V n_V^{-1/2}}$$

Proof

By Lemma param_covering_cube, we have

$$\begin{aligned}
N\left(u,\mathcal{G}(T),\|\cdot\|_V\right) &\leq N\left(Cn^{\kappa}u,\Lambda,\|\cdot\|_2\right)\\
&\leq \frac{1}{C_J}\left(\frac{4\left(\lambda_{max}-\lambda_{min}\right)+2Cn^{\kappa}u}{Cn^{\kappa}u}\right)^J\\
&\leq \frac{1}{C_J}\left(\frac{4n^{t_{\max}-\kappa}+2Cu}{Cu}\right)^J
\end{aligned}$$

Hence

$$H(u, \mathcal{G}(T), \|\cdot\|_V) \leq \log \left[\frac{1}{C_J} \left(\frac{4n^{t_{max}} + 2Cu}{Cu} \right)^J \right]$$

Then

$$\begin{aligned}
\int_0^R H^{1/2}(u, \mathcal{G}(T), \|\cdot\|_V) du &\leq \int_0^R \left[\log \frac{1}{C_J} + J \log \left(\frac{2n^{t_{max}-\kappa} + 2Cu}{Cu} \right) \right]^{1/2} du \\
&< \int_0^R \left[\log \frac{1}{C_J} + J \log 4 + J \log \left(\frac{4n^{t_{max}-\kappa}}{Cu} \right) \right]^{1/2} du \\
&= R \int_0^1 \left[\log \frac{1}{C_J} + J \log 4 + J \log \left(\frac{4n^{t_{max}-\kappa}}{CRv} \right) \right]^{1/2} dv \\
&\leq R \left[\int_0^1 \log \frac{1}{C_J} + J \log 4 + J \log \left(\frac{4n^{t_{max}-\kappa}}{CRv} \right) dv \right]^{1/2} \\
&= R \left[\log \frac{1}{C_J} + J(1 + \log 4) + J \log \left(\frac{4n^{t_{max}-\kappa}}{C} \right) + J \log \frac{1}{R} \right]^{1/2} \\
&\leq R \left(\left[\log \frac{1}{C_J} + J(1 + \log 4) + J \log \left(\frac{4n^{t_{max}-\kappa}}{C} \right) \right]^{1/2} + \sqrt{J \log \frac{1}{R} \vee 0} \right)
\end{aligned}$$

The second bound is crazy loose (but is okay I think). The third inequality follows from concavity of the square root.

The term $\log \frac{1}{R}$ is nasty. When choosing δ, we will replace it with $\log n_V$ since for all $R \geq \frac{1}{n_V}$, we have

$$\log \frac{1}{R} \vee 0 \leq \log n_V$$

Now apply Theorem 3. If δ is chosen such that

$$\sqrt{n_V}\delta^2 \geq 2C \left(\left\| \hat{g}_{\tilde{\lambda}} - g^* \right\|_V + \delta \right) \left(\left[\log \frac{1}{C_J} + J(1 + \log 4) + J \log \left(\frac{4n^{t_{max}-\kappa}}{C} \right) \right]^{1/2} + \sqrt{J \log n_V} \right) \quad (1)$$

then with high probability, we have

$$\left\| \hat{g}_{\hat{\lambda}} - g^* \right\|_V \leq \|\hat{g}_{\tilde{\lambda}} - g^*\|_V + \delta \quad (2)$$

We can combine the two inequalities. If we let

$$K = c \left(\left[\log \frac{1}{C_J} + J(1 + \log 4) + J \log \left(\frac{4n^{t_{max}-\kappa}}{C} \right) \right]^{1/2} + \sqrt{J \log n_V} \right)$$

and

$$\omega = \left\| \hat{g}_{\tilde{\lambda}} - g^* \right\|_V$$

Then (1) can be expressed as

$$\sqrt{n_V}\delta^2 - K\delta - K\omega \geq 0$$

We notice that (1) is precisely the quadratic inequality and is satisfied for δ such that

$$\begin{aligned}
\delta &\geq \frac{K + \sqrt{K^2 + 4K\omega\sqrt{n_V}}}{2\sqrt{n_V}} \\
&\geq \frac{K}{\sqrt{n_V}} + \sqrt{K \left\| \hat{g}_{\tilde{\lambda}} - g^* \right\|_V n_V^{-1/2}}
\end{aligned}$$

where the second inequality is provided for a more intuitive understanding. Plug this inequality back into (2) to get the final result.

Lemma (Based on Vandegeer Corollary 8.3)

Let Q_m be the empirical distributon of m observations at covariates x_i.

Suppose ϵ are m independent sub-gaussian errors. Suppose the model class $\mathcal{F}(T)$ has elements $\sup_{f\in\mathcal{F}_n(T)} \|f\|_{Q_m} \le R$ and satisfies

$$\psi_T(R) \ge \int_0^R H^{1/2}(u, \mathcal{F}(T), \|\cdot\|_{Q_m})du$$

There is C dependent only on the sub-gaussian constants such that for all $\delta > 0$ such that

$$\sqrt{m}\delta \ge C(\psi_T(R) \vee R)$$

we have

$$Pr\left(\sup_{f\in\mathcal{F}_n(T)} \left|\frac{1}{m}\sum_{i=1}^m \epsilon_i f(x_i)\right| \ge \delta \wedge \|\epsilon\|_{Q_m} \le 2\sigma\right) \le C\exp\left(-\frac{m\delta^2}{4C^2R^2}\right)$$

Lemma param_covering_cube

Suppose we have $\Lambda = [\lambda_{min}, \lambda_{max}]^J$, we have

$$N(\delta, \Lambda, \|\cdot\|_2) \le \frac{1}{C_J}\left(\frac{4(\lambda_{max} - \lambda_{min}) + 2\delta}{\delta}\right)^J$$

Proof

(Based on Lemma 2.5 in vandegeer)

Let $C = \{c_j\}_{j=1}^N \subset \Lambda$ be the largest set s.t. two distinct points c_{j_1}, c_{j_2} are at least δ apart. Then balls with radius δ centered at C cover Λ. Hence

$$N(\delta, \Lambda, \|\cdot\|_2) \le N$$

If we instead consider the balls centered at C but with radius $\delta/4$, then the balls must be disjoint and are completely contained in the box $[\lambda_{min} - \delta/4, \lambda_{max} + \delta/4]^J$. So we know the aggregate volume of these smaller balls is less than the volume of the box.

Recall the volume of a ball with radius ρ is $C_J\rho^J$ (where C_J is a constant dependent on dimension J). Hence

$$NC_J(\delta/4)^J \le (\lambda_{max} - \lambda_{min} + \delta/2)^J$$

Lemma: Lipschitz Definition Equivalence

The following two conditions are equivalent:

1. For all $u > 0$ and any $\lambda^{(1)}, \lambda^{(2)}$ that satisfy

$$\|\lambda^{(1)} - \lambda^{(2)}\| \le Cu$$

then

$$\|\hat{g}(\cdot|\lambda^{(1)}) - \hat{g}(\cdot|\lambda^{(2)})\|_D \le u$$

1. $\hat{g}(\cdot|\lambda)$ is $1/C$-Lipschitz in λ:

$$\|\hat{g}(\cdot|\lambda^{(1)}) - \hat{g}(\cdot|\lambda^{(2)})\|_D \le \frac{1}{C}\|\lambda^{(1)} - \lambda^{(2)}\|$$

Proof

It is clear that Condition 2 implies Condition 1.

To show Condition 1 implies Condition 2, suppose for any $\lambda^{(1)}, \lambda^{(2)}$, we have

$$\|\lambda^{(1)} - \lambda^{(2)}\| = d = C\frac{d}{C}$$

Then

$$\|\hat{g}(\cdot|\lambda^{(1)}) - \hat{g}(\cdot|\lambda^{(2)})\|_D \leq \frac{d}{C} = \frac{1}{C}\|\lambda^{(1)} - \lambda^{(2)}\|$$