Theorem 1

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 for all $\delta > 0$ s.t.

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

With high probability

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

Proof for Theorem 1

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

$$Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \|\hat{g}_{\tilde{\lambda}} - g^*\|_V \ge \delta\right)$$

$$= \sum_{s=0}^{\infty} Pr\left(2^s\delta \le \left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \le 2^{s+1}\delta\right)$$

$$\le \sum_{s=0}^{\infty} Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \ge 2^s\delta \wedge \left\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right)$$

$$= \sum_{s=0}^{\infty} Pr\left(\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V\right)^2 \ge 2^{2s}\delta^2 \wedge \left\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right)$$

$$\le \sum_{s=0}^{\infty} Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V^2 - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V^2 \ge 2^{2s}\delta^2 \wedge \left\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\right\|_V \le 2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right)$$

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

To apply the lemma based on vandegeer corollary 8.3 (see below), we must check that δ satisfies all the conditions. By our choice of δ and the fact that $\psi(a+u)/u$ is nonincreasing, we have that for for all s

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

1

So by the lemma below, we have

$$Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \geq \delta \wedge \|\epsilon\|_V \leq 2\sigma \wedge \|\epsilon\|_T \leq 2\sigma\right)$$

$$\leq C\sum_{s=0}^{\infty} \exp\left(-n_V \frac{2^{4s-2}\delta^4}{4C^2\left(2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V + 2^{s+1}\delta\right)^2}\right)$$

$$\leq C\sum_{s=0}^{\infty} \exp\left(-n_V \frac{2^{4s-2}\delta^4}{64C^2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V^2}\right) \vee \exp\left(-n_V \frac{2^{2s}\delta^2}{196C^2}\right)$$

$$\leq c\exp\left(-\frac{n_V\delta^4}{c^2\left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V^2}\right) + c\exp\left(-\frac{n_V\delta^2}{c^2}\right)$$

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

$$Pr\left(\left\|\hat{g}_{\hat{\lambda}} - g^*\right\|_V - \left\|\hat{g}_{\tilde{\lambda}} - g^*\right\|_V \geq \delta\right)$$

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

Corollary to Theorem 1

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,

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

Proof

By Lemma param_covering_cube, we have

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

Let $R_1 = R \wedge \sqrt{n^{t_{max}-\kappa}/C}$.

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

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

Hence it suffices to choose δ s.t.

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

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

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

So we are interested in δ s.t.

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

The quadratic formula tells us that this is satisfied when

$$
\delta \geq \frac{K + \sqrt{K^2 + 4K\omega\sqrt{n_V}}}{2\sqrt{n_V}}
$$

A more simple lower bound for δ is

$$
\delta \geq \frac{K}{\sqrt{n_V}} + \sqrt{K\|\hat{g}_{\tilde{\lambda}} - g^*\|_V\, n_V^{-1/2}}
$$

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 $\sup_{f \in \mathcal{F}_n(T)} \|f\|_{Q_m} \leq R$ and satisfies

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

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

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

we have

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

Lemma param_covering_cube

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

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

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.

If the balls have radius $\delta/4$ and are centered at C, then the balls must be disjoint and are completely contained in the box $[\lambda_{min} - \delta/4, \lambda_{max} + \delta/4]^J$.

The volume of a ball with radius ρ is $C_J\rho^J$. Therefore

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