CV Proof: Mitchell's Thrm

Consider the CV problem with

$$\hat{\lambda} = \arg\min_{\lambda} \sum_{k=1}^{K} \|y - \hat{g}_{\lambda}(\cdot|D_{-k})\|_k^2$$

We consider the behavior of

$$\hat{g}_{MCV} = \sum_{k=1}^{K} \hat{g}_{\lambda}(\cdot|D_{-k})$$

and its generalization error

$$E_D\|\hat{g}_{MCV}(\cdot|D) - g^*\|^2 = E_D\left[\int \left(\hat{g}_{MCV}(x|D) - g^*(x)\right)^2 d\mu(x)\right]$$

We will assume that $\sup_{g\in\mathcal{G}} \|g\|_\infty \le G$, so Assumption A.1 and A.2 are satisfied.
Suppose the errors are bounded ($\|\epsilon\|_\infty < \infty$).
Suppose we can show

$$\|g_{\lambda_1} - g_{\lambda_2}\|_\infty \le |\lambda_1 - \lambda_2|\, n^{\kappa} C$$

So we get that

$$E\left[\|g^* - \frac{1}{k}\sum_{k=1}^{K} g_{\hat{\lambda}}(\cdot|D_{-k})\|^2\right] \le (1+a)E\left[\|g^* - \frac{1}{k}\sum_{k=1}^{K} g_{\tilde{\lambda}}(\cdot|D_{-k})\|^2\right] + c_a J^2 \max_{k=1:K} \frac{(\log n)^2}{n_k}$$

Proof

First note that the assumption A.2 is satisfied:

$$\begin{aligned}\int \left(g^*(x) - g_\lambda(x|D_0)\right)^4 d\mu(x) &\le \|\left(g^* - g_\lambda\right)^2\|_1 \|\left(g^* - g_\lambda\right)^2\|_\infty \\ &\le 4G^2\|g^* - g_\lambda\|_2^2\end{aligned}$$

Let

$$\mathcal{Q}(T) = \left\{q_\lambda(x,y) = (g^* - \hat{g}_\lambda)^2 : \lambda \in \Lambda\right\}$$

To compute the upper bound, we need the $\|\cdot\|_{\psi_1}$ and $\|\cdot\|_{L_2}$ entropy of the function class

$$\mathcal{Q}_d^{L_2}(T) = \left\{q \in \mathcal{Q}(T) : \|Q(Z)\|_2 \le \sqrt{d}\right\}$$

From our assumptions, we note that

$$\begin{aligned}\|q_{\lambda_1} - q_{\lambda_2}\|_{\psi_1} &= \|(g^* - g_{\lambda_1})^2 - (g^* - g_{\lambda_2})^2\|_{\psi_1} \\ &= \|(g_{\lambda_1} - g_{\lambda_2})(2g^* - g_{\lambda_1} - g_{\lambda_2})\|_{\psi_1} \\ &\le \|g_{\lambda_1} - g_{\lambda_2}\|_\infty \|2g^* - g_{\lambda_1} - g_{\lambda_2}\|_\infty \\ &\le |\lambda_1 - \lambda_2|\, 4Cn^\kappa G\end{aligned}$$

and similarly for $\|\cdot\|_{L_2}$ (we're using a loose bound for $\|\cdot\|_{L_2}$ since calculating its inverse/convex conjugate is difficult).

Hence (for a different constant κ)

$$H(u, \mathcal{Q}_d^{L_2}(T), \|\cdot\|_{\psi_1}) \le J\left(\log\frac{1}{u} + \log(4Cn^\kappa G)\right)$$

and

$$H(u, \mathcal{Q}_d^{L_2}(T), \|\cdot\|_{L_2}) \leq J\left(\log\frac{1}{u} + \log(4Cn^\kappa G)\right)$$

We calculate each component of the complexity term $J(d)$:

$$\begin{aligned}
\gamma_1(\mathcal{Q}_d^{L_2}(T), \|\cdot\|_{\psi_1}) &= \int_0^G H(u, \mathcal{Q}_d^{L_2}(T), \|\cdot\|_{\psi_1})du \\
&= JG\left(1 + \log(4Cn^\kappa G)\right)
\end{aligned}$$

and

$$\begin{aligned}
\gamma_2(\mathcal{Q}_d^{L_2}(T), \|\cdot\|_{L_2}) &= \int_0^{\sqrt{d}} \left[H(u, \mathcal{Q}_d^{L_2}(T), \|\cdot\|_{L_2})\right]^{1/2} du \\
&= \sqrt{d}\int_0^1 \left[J\left(\log\frac{1}{u} + \log(4Cn^\kappa G)\right)\right]^{1/2} du \\
&\leq \sqrt{d}\left[\int_0^1 J\left(\log\frac{1}{u} + \log(4Cn^\kappa G)\right) du\right]^{1/2} \\
&= \sqrt{d}\left[J\left(1 + \log(4Cn^\kappa G)\right)\right]^{1/2}
\end{aligned}$$

So we can define

$$\begin{aligned}
J(d) &\equiv J\left(1 + \log(4Cn^\kappa G)\right)\left[\sqrt{d} + \left(\max_{k=1:K}\frac{\log n_k}{\sqrt{n_k}}\right)\right] \\
&\geq \sqrt{d}\left[J\left(1 + \log(4Cn^\kappa G)\right)\right]^{1/2} + \left(\max_{k=1:K}\frac{\log n_k}{\sqrt{n_k}}\right) JG\left(1 + \log(4Cn^\kappa G)\right)
\end{aligned}$$

Then$J^{-1}(b)$ is

$$J^{-1}(b) = \left(\frac{b}{J\left(1 + \log(4Cn^\kappa G)\right)} - \left(\max_{k=1:K}\frac{\log n_k}{\sqrt{n_k}}\right)\right)^2$$

The convex conjugate of $J^{-1}(b)$ is

$$\psi(v) = \frac{1}{2}\left(vJ\left(1 + \log(4Cn^\kappa G)\right)\right)^2 + J\left(1 + \log(4Cn^\kappa G)\right)\left(\max_{k=1:K}\frac{\log n_k}{\sqrt{n_k}}\right)v$$

Therefore

$$\begin{aligned}
\epsilon_q(1/q) &= \psi\left(\max_{k=1:K}\frac{2q(1+a)}{a\sqrt{n_k}}\right) \\
&= \frac{1}{2}\left(\frac{2q(1+a)}{a\sqrt{n_k}}J\left(1 + \log(4Cn^\kappa G)\right)\right)^2 + J\left(1 + \log(4Cn^\kappa G)\right)\left(\max_{k=1:K}\frac{\log n_k}{\sqrt{n_k}}\right)\frac{2q(1+a)}{a\sqrt{n_k}} \\
&\leq \max_{k=1:K}\left(\frac{2q(1+a)}{a\sqrt{n_k}}J\left(1 + \log(4Cn^\kappa G)\right)\right)^2
\end{aligned}$$

Finally, we get

$$E\left[\|g^* - \frac{1}{k}\sum_{k=1}^K g_{\hat{\lambda}}(\cdot|D_{-k})\|^2\right] \leq (1+a)E\left[\|g^* - \frac{1}{k}\sum_{k=1}^K g_{\tilde{\lambda}}(\cdot|D_{-k})\|^2\right] + \frac{c(1+a)^2}{a}\frac{J^2\left(1 + \log(4Cn^\kappa G)\right)^2}{\min_{k=1:K} n_k}$$