Notation

$(\mathcal{X}, \Sigma, P)$. X is a random element valued in $\mathcal{X}$ and its law is P.

$X_1, \dots, X_n \overset{iid}{\sim} P$

$P_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$, $Pf = \int f \, dP$

Def P-Glivenkov-Cantelli Class

A class of functions on $\mathcal{X}$: $\mathcal{F}$ s.t.

$\|P_n - P\|_{\mathcal{F}} \triangleq \sup_{f\in\mathcal{F}} |P_n f - Pf| \to_{p^*} 0$

Symmetrization

$P_n' f = \frac{1}{n}\sum_{i=1}^n f(X_i')$, $X_i' =_d X_i$, $X_i' \perp X_i$

$P_n^\sigma f = \frac{1}{n}\sum_{i=1}^n \sigma_i f(X_i)$, $P(\sigma_i = 1) = P(\sigma_i = -1) = \frac{1}{2}$

$\sigma_i \perp X_i$

Lemma $\mathbb{E}\|P_n - P\|_{\mathcal{F}} \le \mathbb{E}\|P_n - P_n'\|_{\mathcal{F}} \le 2\,\mathbb{E}\|P_n^\sigma\|_{\mathcal{F}}$

Pf: $|P_n f - Pf| = |P_n f - \mathbb{E}(P_n' f \mid X)|$

$\overset{\text{Jensen}}{\le} \mathbb{E}\,|P_n f - P_n' f| \mid X$

$\Rightarrow \mathbb{E}\sup_{f\in\mathcal{F}} |P_n f - Pf| \le \mathbb{E}\sup_{f\in\mathcal{F}} \mathbb{E}|P_n f - P_n' f| \mid X$

$\le \mathbb{E}\|P_n - P_n'\|_{\mathcal{F}}$

Since $\|P_n - P_n'\|_{\mathcal{F}} =_d \|P_n^\sigma - P_n^{\sigma'}\|_{\mathcal{F}}$,

$\mathbb{E}\|P_n - P\|_{\mathcal{F}} = \mathbb{E}\|P_n^\sigma - P_n^{\sigma'}\|_{\mathcal{F}} \le 2\,\mathbb{E}\|P_n^\sigma\|_{\mathcal{F}}$

$\square$

We also mark

$$R_n(\mathcal{F}) = \mathbb{E}\,\|P_n^{\sigma}\|_{\mathcal{F}}$$

Theorem

If $\forall f \in \mathcal{F}$, $\sup_{x\in X} \sup_{f\in\mathcal{F}} |f(x)| \leq b$, then

$$\mathbb{P}(\|P-P_n\|_{\mathcal{F}} \geq 2R_n(\mathcal{F}) + \delta) \leq \exp\left(-\frac{n\delta^2}{2b^2}\right)$$

Pf: let $G(X_1, \ldots, X_n) = \|P-P_n\|_{\mathcal{F}}$,

$$= \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} \sigma_i f(x_i) - \mathbb{E}\,\sigma f(x) \right|$$

Then

$$|G(\ldots, x, \ldots) - G(\ldots, y, \ldots)| \leq \frac{2b}{n}$$

$$\Rightarrow \mathbb{P}(\|P-P_n\|_{\mathcal{F}} \geq 2R_n(\mathcal{F}) + \delta)$$

$$\leq e^{-\frac{2\delta^2}{n\left(\frac{2b}{n}\right)^2}} = e^{-\frac{n\delta^2}{2b^2}}$$

$\square$

Lemma

If $\forall f \in \mathcal{F}_1$, $\sup_{x\in X} \sup_{f\in\mathcal{F}_1} |f(x)| \le b$, and

$$|\mathcal{F}_1(X_1, \dots, X_n)| \le (n+1)^v,$$

where

$$\mathcal{F}_1(X_1, \dots, X_n) = \{(f(X_1), \dots, f(X_n)); f \in \mathcal{F}_1\}$$

then

$$R_n(\mathcal{F}_1) \le b\sqrt{\frac{v\log(n+1)}{n}}$$

Pf:

$$e^{\lambda \mathbb{E}_\sigma \sup_{f\in\mathcal{F}_1} \frac{1}{n}|\sum_{i=1}^n \sigma_i f(X_i)|}$$

$$\le \mathbb{E}_\sigma \sup_{f\in\mathcal{F}_1} e^{\lambda|\frac{1}{n}\sum_{i=1}^n \sigma_i f(X_i)|}$$

$$\le \sum_{f\in\mathcal{F}_1} \mathbb{E}_\sigma e^{\lambda|\frac{1}{n}\sum_{i=1}^n \sigma_i f(X_i)|}$$

$$\leq 2(n+1)^{v} \; e^{\left(\frac{\lambda}{n}\right)^2 / 2 \; n \, 4 b^2}$$

$$\Rightarrow R_n(\mathcal{F}) \leq \frac{v \log 2(n+1)}{\lambda} + \frac{2\lambda b^2}{n}$$

$$\Rightarrow R_n(\mathcal{F}) \leq 2b \sqrt{\frac{2v \log 2(n+1)}{n}}$$

$$\lesssim b \sqrt{v \frac{\log(n+1)}{n}}$$

Cor

$$\mathbb{P}\left(\sup_x |\hat{F}_n(x) - F(x)| \geq C \sqrt{\frac{\log(n+1)}{n}} + \delta\right)$$

$$\leq e^{-\frac{n\delta^2}{2}}$$