Assume $f \in \mathcal{F} = \{f : \mathcal{X} \mapsto \mathbb{R}\}$ and $x_1, \ldots, x_n$ are i.i.d. Denote $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(x_i)$ and $\mathbb{P}f = \int f dP = \mathbb{E}f$.
We are interested in bounding $\frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}f$.

Worst-case scenario is the value

$$\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| .$$

The Glivenko-Cantelli property $GC(\mathcal{F}, P)$ says that

$$\mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| \to 0$$

as $n \to \infty$.

- Algorithm can output any $f \in \mathcal{F}$
- Objective is determined by $\mathbb{P}_n f$ (on the data)
- Goal is $\mathbb{P}f$
- Distribution P is unknown

The most pessimistic requirement is

$$\sup_P \mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| \to 0$$

which we denote

$$\text{uniform}GC(\mathcal{F}).$$

VC classes of sets

Let $\mathcal{C} = \{C \subseteq X\}$, $f_C(x) = I(x \in C)$. The most pessimistic value is

$$\sup_P \mathbb{E}\sup_{C\in\mathcal{C}} |\mathbb{P}_n(C) - \mathbb{P}(C)| \to 0.$$

For any sample $\{x_1, \ldots, x_n\}$, we can look at the ways that $\mathcal{C}$ intersects with the sample:

$$\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\}.$$

Let

$$\triangle_n(\mathcal{C}, x_1, \ldots, x_n) = \text{card}\ \{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\},$$

the number of different subsets picked out by $C \in \mathcal{C}$. Note that this number is at most 2^n.

Denote

$$\triangle_n(\mathcal{C}) = \sup_{\{x_1,\ldots,x_n\}} \triangle_n(\mathcal{C}, x_1, \ldots, x_n) \le 2^n.$$

We will see that for some classes, $\triangle_n(\mathcal{C}) = 2^n$ for $n \le V$ and $\triangle_n(\mathcal{C}) < 2^n$ for $n > V$ for some constant V.

What if $\triangle_n(\mathcal{C}) = 2^n$ for all $n \ge 1$? That means we can always find $\{x_1, \ldots, x_n\}$ such that $C \in \mathcal{C}$ can pick out any subset of it: "$\mathcal{C}$ **shatters** $\{x_1, \ldots, x_n\}$". In some sense, we do not learn anything.

Definition 8.1. If $V < \infty$, then $\mathcal{C}$ is called a VC class. V is called VC dimension of $\mathcal{C}$.

Sauer's lemma states the following:

Lemma 8.2.

$$\forall\{x_1,\ldots,x_n\},\qquad \triangle_n(\mathcal{C},x_1,\ldots,x_n)\le\left(\frac{en}{V}\right)^V \quad \textit{for}\ \ n\ge V.$$

Hence, $\mathcal{C}$ will pick out only very few subsets out of 2^n (because $\left(\frac{en}{V}\right)^V\sim n^V$).

Lemma 8.3. *The number $\triangle_n(\mathcal{C},x_1,\ldots,x_n)$ of subsets picked out by $\mathcal{C}$ is bounded by the number of subsets shattered by $\mathcal{C}$.*

Proof. Without loss of generality, we restrict $\mathcal{C}$ to $\mathcal{C}:=\{C\cap\{x_1,\ldots,x_n\}:C\in\mathcal{C}\}$, and we have $\text{card}(\mathcal{C})=\Delta_n(\mathcal{C},x_1,\cdots,x_n)$.
We will say that $\mathcal{C}$ is **hereditary** if and only if whenever $B\subseteq C\in\mathcal{C}$, $B\in\mathcal{C}$. If $\mathcal{C}$ is hereditary, then every $C\in\mathcal{C}$ is shattered by $\mathcal{C}$, and the lemma is obvious. Otherwise, we will transform $\mathcal{C}\to\mathcal{C}'$, hereditary, without changing the cardinality of $\mathcal{C}$ and without increasing the number of shattered subsets.
Define the operators T_i for $i=1,\cdots,n$ as the following,

$$\begin{aligned} T_i(C) &= \begin{cases} C-\{x_i\} & \text{if } C-\{x_i\} \text{ is not in } \mathcal{C}\\ C & \text{otherwise}\end{cases}\\ T_i(\mathcal{C}) &= \{T_i(C):C\in\mathcal{C}\}.\end{aligned}$$

It follows that card $T_i(\mathcal{C})=$ card $\mathcal{C}$. Moreover, every $A\subseteq\{x_1,\cdots,x_n\}$ that is shattered by $T_i(\mathcal{C})$ is also shattered by $\mathcal{C}$. If $x_i\notin A$, then $\forall C\in\mathcal{C}, A\bigcap C=A\bigcap T_i(C)$, thus $\mathcal{C}$ and $T_i(\mathcal{C})$ both or neither shatter A. On the other hand, if $x_i\in A$ and A is shattered by $T_i(\mathcal{C})$, then $\forall B\subseteq A,\exists C\in\mathcal{C}$, such that $B\bigcap\{x_i\}=A\bigcap T_i(C)$. This means that $x_i\in T_i(C)$, and that $C\backslash\{x_i\}\in\mathcal{C}$. Thus both $B\bigcup\{x_i\}$ and $B\backslash\{x_i\}$ are picked out by $\mathcal{C}$. Since either $B=B\bigcup\{x_i\}$ or $B=B\backslash\{x_i\}$, B is picked out by $\mathcal{C}$. Thus A is shattered by $\mathcal{C}$.
Apply the operator $T=T_1\circ\ldots\circ T_n$ until $T^{k+1}(\mathcal{C})=T^k(\mathcal{C})$. This will happen for at most $\sum_{C\in\mathcal{C}}\text{card}(C)$ times, since $\sum_{C\in\mathcal{C}}\text{card}(T_i(C))<\sum_{C\in\mathcal{C}}\text{card}(C)$ if $T_i(\mathcal{C})\neq\mathcal{C}$. The resulting collection $\mathcal{C}'$ is hereditary. This proves the lemma. $\square$

Sauer's lemma is proved, since for arbitrary $\{x_1,\ldots,x_n\}$,

$$\begin{aligned}\triangle_n(\mathcal{C},x_1,\ldots,x_n) &\le \text{card (shattered subsets of } \{x_1,\ldots,x_n\})\\ &\le \text{card (subsets of size } \le V)\\ &=\sum_{i=0}^{V}\binom{n}{i}\\ &\le\left(\frac{en}{V}\right)^V.\end{aligned}$$