Assume \( f \in \mathcal{F} = \{ f : \mathcal{X} \rightarrow \mathbb{R} \} \) 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 \( \mathbb{P} f = \int f d\mathbb{P} = \mathbb{E} f. \)

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 \( GC(\mathcal{F}, \mathbb{P}) \) says that

\[
\mathbb{E} \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - \mathbb{P} f| \rightarrow 0
\]
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 \( \mathbb{P} \) is unknown

The most pessimistic requirement is

\[
\sup_{\mathbb{P}} \mathbb{E} \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - \mathbb{P} f| \rightarrow 0
\]
which we denote

\[
\text{uniform} GC(\mathcal{F}).
\]

**VC classes of sets**

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

\[
\sup_{\mathbb{P}} \mathbb{E} \sup_{C \in \mathcal{C}} |\mathbb{P}_n (C) - \mathbb{P} (C)| \rightarrow 0.
\]

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

\[
\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \text{card} \{ C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C} \},
\]

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

Denote

\[
\Delta_n(\mathcal{C}) = \sup_{\{x_1, \ldots, x_n\}} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) \leq 2^n.
\]

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

What if \( \Delta_n(\mathcal{C}) = 2^n \) for all \( n \geq 1 \)? That means we can always find \( \{x_1, \ldots, x_n\} \) such that \( C \in \mathcal{C} \) can pick out any subset of it: "\( C \text{ shatters} \{x_1, \ldots, x_n\} \). 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. 
\[ \forall \{x_1, \ldots, x_n\}, \quad \Delta_n(C, x_1, \ldots, x_n) \leq \left( \frac{en}{V} \right)^V \text{ for } n \geq V. \]

Hence, \( C \) 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 \( \Delta_n(C, x_1, \ldots, x_n) \) of subsets picked out by \( C \) is bounded by the number of subsets shattered by \( C \).

Proof. Without loss of generality, we restrict \( C \) to \( C := \{ C \cap \{x_1, \ldots, x_n\} : C \in C \} \), and we have \( \text{card}(C) = \Delta_n(C, x_1, \ldots, x_n) \).

We will say that \( C \) is hereditary if and only if whenever \( B \subseteq C \in C, B \subseteq C \). If \( C \) is hereditary, then every \( C \in C \) is shattered by \( C \), and the lemma is obvious. Otherwise, we will transform \( C \to C' \), hereditary, without changing the cardinality of \( C \) and without increasing the number of shattered subsets.

Define the operators \( T_i \) for \( i = 1, \ldots, n \) as the following,

\[ T_i(C) = \begin{cases} C - \{x_i\} & \text{if } C - \{x_i\} \text{ is not in } C \\ C & \text{otherwise} \end{cases} \]

\[ T_i(C) = \{ T_i(C) : C \in C \}. \]

It follows that \( \text{card} T_i(C) = \text{card} C \). Moreover, every \( A \subseteq \{x_1, \ldots, x_n\} \) that is shattered by \( T_i(C) \) is also shattered by \( C \). If \( x_i \notin A \), then \( \forall C \in C, A \cap C = A \cap T_i(C) \), thus \( C \) and \( T_i(C) \) both or neither shatter \( A \). On the other hand, if \( x_i \in A \) and \( A \) is shattered by \( T_i(C) \), then \( \forall B \subseteq A, \exists C \in C, \text{ such that } B \cap \{x_i\} = A \cap T_i(C) \). This means that \( x_i \in T_i(C) \), and that \( C \setminus \{x_i\} \in C \). Thus both \( B \cup \{x_i\} \) and \( B \setminus \{x_i\} \) are picked out by \( C \).

Since either \( B = B \cup \{x_i\} \) or \( B = B \setminus \{x_i\} \), \( B \) is picked out by \( C \). Thus \( A \) is shattered by \( C \).

Apply the operator \( T = T_1 \circ \ldots \circ T_n \) until \( T^{k+1}(C) = T^k(C) \). This will happen for at most \( \sum_{C \in C} \text{card}(C) \) times, since \( \sum_{C \in C} \text{card}(T_i(C)) < \sum_{C \in C} \text{card}(C) \) if \( T_i(C) \neq C \). The resulting collection \( C' \) is hereditary. This proves the lemma. \( \square \)

Sauer’s lemma is proved, since for arbitrary \( \{x_1, \ldots, x_n\} \),

\[ \Delta_n(C, x_1, \ldots, x_n) \leq \text{card (shattered subsets of } \{x_1, \ldots, x_n\} \text{)} \]

\[ \leq \text{card (subsets of size } \leq V \text{)} \]

\[ = \sum_{i=0}^{V} \binom{n}{i} \]

\[ \leq \left( \frac{en}{V} \right)^V. \]